



Samuel Heffer

**Sensor Approaches for the Non-Destructive Assessment of
Food Safety and Authenticity**

School of Water, Energy and the Environment

PhD

Academic Year: 2021 - 2025

Supervisor: Dr Fady Mohareb

Associate Supervisor: Dr Maria Anastasiadi

March 2025



School of Water, Energy and the Environment

PhD

Academic Year 2021 - 2025

Samuel Heffer

“Sensor Approaches for the Non-Destructive Assessment of Food
Safety and Authenticity”

Supervisor: Dr Fady Mohareb
Associate Supervisor: Dr Maria Anastasiadi
March 2025

This thesis is submitted in partial fulfilment of the requirements for
the degree of PhD

© Cranfield University 2024. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

Academic integrity declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

Ensuring access to high-quality, safe food is typically the responsibility of national governing bodies overseeing food production and distribution, and is usually secured by enforcing compliance with national food standards at the producer level. However, recent and historical food safety transgressions have illustrated that conventional food standards frameworks have fallen short in some respects: conventional batch-based compliance testing has been shown to result in delays between identification of problems and corrective action, resulting in product recalls, batch destruction, and reduced consumer trust. In severe cases, delayed responses have exposed consumers to unsafe products before action was taken, leading to harm.

Modern, high throughput sensor technologies, coupled with appropriate informatics support have the potential to improve the responsiveness of current food standards frameworks, offering a means to monitor the food production and distribution chain in higher resolution, while allowing for close-to-real-time detection of emergent issues, before exposure to the public and before incurring costs for the food business operator.

This work shows a two-part appraisal of this approach. First, mathematical models were developed to assess the accuracy and reliability of candidate sensors for prediction of food safety and quality parameters in meat products, both in isolation and in a sensor fusion context. Second, an online platform was constructed to support development and deployment of predictive models built using multiple channels of high-throughput sensor data in a manner enabling performance evaluation and real-time decision support capabilities.

The methods and platform developed in this work demonstrate that employment of high-throughput, minimally invasive sensor technologies may supplement existing monitoring practices when coupled with suitable machine learning pipelines, and such pipelines, including those integrating multiple streams of multi-modal sensor data, can be deployed efficiently in a decision support capacity.

Keywords:

Chemometrics, Bioinformatics, Sensors, Food Analysis, Food Science, Spoilage, Authentication

Acknowledgements

I would like to thank Prof. Fady Mohareb and Dr. Maria Anastasiadi for their immense support throughout this period of work - without them I probably would have floundered.

Additionally, I would like to express my sincere gratitude towards Prof. George-John Nychas, who ultimately coordinated the EU project DiTECT (grant agreement No 861915) which provided this PhD opportunity.

I would also like to thank the members of the Bioinformatics Group at Cranfield, who have offered valuable feedback and insights throughout my research during meetings and presentations.

Finally, I would like to thank my parents, who throughout this period of study, as always, have encouraged and supported me enormously. I am immensely grateful.

Table of Contents

Academic integrity declaration.....	i
Abstract	ii
Acknowledgements	iv
List of Figures.....	viii
List of Tables	xvi
List of Equations.....	xx
List of Abbreviations	xxi
1 Introduction and Literature Review.....	22
1.1 Background.....	22
1.2 Food Supply Issues in the UK and Internationally	23
1.3 Examples of Food Control Incidents	26
1.3.1 Chinese Milk Scandal.....	27
1.3.2 Horse Meat Scandal.....	30
1.3.3 Common Themes in Food Standards Incidents	33
1.4 Process Analytical Technology and Digital Twin.....	35
1.5 Maintaining Traceability in the Food Supply Chain	41
1.6 Common Analytical/Machine Learning Methodologies	44
1.6.1 Common Machine Learning Methods Applied In Sensor Analysis ...	45
1.6.2 Preparation of Sensor Data for Food Analysis	56
1.7 Sensor/Data Fusion	60
1.8 Modern Sensor Technologies	75
1.8.1 Overview of Chromatography Techniques	75
1.8.2 Overview of Spectroscopic Sensor Technologies	79
1.8.3 Chemical Probes – Electronic Nose and Tongue.....	96
1.9 Significance of the Research Problem.....	98
1.10 Hypothesis Statements	99
1.11 Objectives	100
References	101
2 Application of Spectroscopic Sensor Technologies for Meat Spoilage Prediction	121
2.1 Abstract.....	121
2.2 Introduction	122
2.3 Materials and Methods.....	126
2.3.1 Experimental Design, Sample Preparation and Microbiological analysis	126
2.3.2 Multi-Sensor Fusion Strategies	135
2.4 Results.....	138
2.4.1 Microbial assessment of chicken thigh spoilage.....	138
2.4.2 Microbial Assessment of Beef Spoilage	140

2.4.3 Fourier-Transform Infrared and Multispectral Imaging Measurements of Chicken Thigh and Beef Mince	142
2.4.4 Estimation of microbial activity in chicken thigh using machine learning analysis	147
2.4.5 Estimation of microbial activity in beef mince using machine learning analysis	158
2.4.6 Discussion and Conclusions	167
References	171
3 SorfML2: A New Web Platform for Analysis of High Throughput Sensor Data in Food Spoilage and Authentication Tasks.....	175
3.1 Introduction	175
3.1.1 Existing Platforms	177
3.1.2 Determination of Desired Features for a Candidate Food Analysis Platform.....	185
3.1.3 Platform Architecture Design Considerations.....	193
3.2 Specific Aims and Objectives of this Development Project.....	199
3.3 Technical Specifications	200
3.3.2 System Requirements	203
3.3.3 Data Handling and Storage	205
3.3.4 User Interfaces	207
3.4 Implementation Details	208
3.4.1 Platform-wide Authentication and Authorisation	208
3.4.2 Modules and Components	209
3.4.3 Final User Interface Design.....	217
3.4.4 Semantic Data Querying	226
3.4.5 Final Service Architecture	227
3.5 Platform Evaluation.....	229
3.5.1 Materials and Methods.....	229
3.5.2 Case Study 1 – Chicken Spoilage Model Benchmarking and Ranking.....	236
3.5.3 Case Study 2 – Chicken Spoilage Model Development and Prediction	255
3.5.4 Case Study 3 – Evaluation of Bayesian Framework and Support Vector Machine Models for Chicken Spoilage Prediction.....	263
3.6 Discussion	274
3.6.1 Comparison With Existing Platforms	274
3.6.2 Limitations of Current Implementation.....	278
3.6.3 Future Work	279
3.7 Conclusion	281
References	282
4 Discussion and Conclusions.....	288
4.1 Meat Spoilage Detection with Minimally Invasive Sensors	289

4.1.1 Interpretation of Results	289
4.1.2 Practical Considerations for Sensor Fusion Deployment	293
4.2 Development of the SorfML2 Platform	296
4.3 Recommendations for Future Work	298
5 Conclusion.....	300
References	301
Appendices.....	304
Ethical approval letter	304
Appendix A:	305
SPARQL Functionalities	305

List of Figures

Figure 1-1: A Conceptual Overview of a typical food production pipeline (A), versus a PAT-augmented production pipeline (B). The greater frequency of non/minimally-invasive testing throughout production is intended to maintain consistent output quality.....	39
Figure 1-2: Generalised topology of an Early/Data-level fusion strategy, typically resulting in the concatenation of data from multiple sources (and modalities) .	63
Figure 1-3: Generalised topology of a Mid/Feature-level fusion strategy, typically resulting in the concatenation of features from multiple sources (and modalities)	69
Figure 1-4: Generalised topology of a Late/Decision-level fusion strategy, typically resulting in the concatenation of decisions from multiple sources (and modalities).....	74
Figure 2-1: General schematic for a single fold of Nested Cross Validation. This process is repeated for k folds. In the case of repetitions, the data is shuffled randomly before the same process occurs again for a new selection of k outer folds.....	132
Figure 2-2: Graphical representation of early, feature (mid-) and decision (late) fusion approaches implemented using multispectral imaging (MSI), multispectral imaging fluorescent (MSIF), and Fourier transform infrared spectroscopy (FTIR) for total viable count (TVC) prediction. The upper examples represent decision fusion, wherein trained first-layer model predictions (3.) are passed to a linear model for combination of their predictions. The feature fusion approach uses first-layer partial least squares (PLS) models to generate projections for passing into a second-layer PLS model (2.), the number of latent variables u is tuned on these outputs. The lower approach represents early fusion, wherein the unprocessed data matrices are concatenated into a single matrix and standardised (1.) prior to passing to a subsequent PLS model. The number of latent variables retained was tuned in a manner similar to the feature fusion approach.....	137

Figure 2-3: Mean and Standard Deviation of Total Viable Count (TVC) recorded in chicken thigh for aerobic (batch 1: blue, batch 2: dark blue) and vacuum (batch 1: blue, batch 2: dark blue) packaging conditions at 0°C, 5°C and 10°C.....	139
Figure 2-4: Mean and Standard Deviation of Total Viable Count (TVC) recorded in beef mince for aerobic (batch 1: red, batch 2: dark red) and MAP (batch 1: blue, batch 2: dark blue) packaging conditions at 0°C, 5°C and 10°C.	141
Figure 2-5: Plot showing mean spectral values of Fourier transform infrared spectroscopy (FTIR) data across the aerobically and vacuum-stored chicken thigh samples (solid black line). The standard deviation across the dataset is indicated by grey shading. Mean values are indicated for 0 °C/0 h (dotted blue line aerobic n = 10, vacuum n = 8) and 0 °C/216 h (dashed orange line aerobic n = 21, vacuum n = 8).	143
Figure 2-6: Plot showing mean spectral values of Fourier transform infrared spectroscopy (FTIR) data across the aerobically and modified atmosphere packaged (MAP) beef mince samples (solid black line). The standard deviation across the dataset is indicated by grey shading. Mean values are indicated for 0 °C/0 h (dotted blue line aerobic n = 4, MAP n = 4) and 0 °C/336 h (dashed orange line aerobic n = 4, MAP n = 4).	144
Figure 2-7: Plot showing mean and standard deviation of spectral values of MSI data across the aerobically and vacuum-stored chicken thigh samples (solid black line). Mean and standard deviation values are also indicated for 0°C/0h (blue line aerobic n=10, vacuum n=8) and 0°C/216h (orange line aerobic n=21, vacuum n=8).	145
Figure 2-8: Plot showing mean and standard deviation of spectral values of MSI data across the aerobically and modified atmosphere-stored minced beef samples (solid black line). Mean and standard deviation values are also indicated for 0°C/0h (blue line aerobic n=4, MAP n=4) and 0°C/336h (orange line aerobic n=4, MAP n=4).	146
Figure 2-9: The root mean squared error of prediction of the repeated nested cross validation process for chicken thigh data.	153

Figure 2-10: Performance plots of testing for single-sensor, single-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets.	154
Figure 2-11: Performance plots of testing for two/three-sensor, single-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.....	155
Figure 2-12: Performance plots of testing for single-sensor, two-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets.	156
Figure 2-13: Performance plots of testing for two/three-sensor, two-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.....	157
Figure 2-14: The root mean squared error of prediction of the repeated nested cross validation process for beef mince data.....	162
Figure 2-15: Performance plots of testing for single-sensor, single-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets.....	163
Figure 2-16: Performance plots of testing for two/three-sensor, single-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.	164
Figure 2-17: Performance plots of testing for single-sensor, two-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets.....	165
Figure 2-18: Performance plots of testing for two/three-sensor, two-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.	166

Figure 3-1: An example of a SPARQL query representing the request: “Find all machine learning models trained with data collected using Sensor_A at temperatures between 0°C and 25°C”	198
Figure 3-2: A generalised view of sensor data inputs into the new platform. The primary upload process happens in a similar fashion to SorfML, however a secondary FAIRification step is initiated by the API service asynchronously.	206
Figure 3-3: The initial intended topology of the “SorfML2 Platform” application. This highlights the interrelationship of the services expected to be necessary for the complete functionality of the platform. The dashed lines output from the “Command Queue” service are to indicate the asynchronous nature of their communication.	207
Figure 3-4: Entity Relationships in the SorfML platform. Simplistic data concepts limited the extensibility of the platform beyond the representation of singular models.....	210
Figure 3-5: Entity relationship diagram of the new platform. This design makes use of new concepts, particularly “Sensor”, “Pipeline”, “Nextflow Process” and “Nextflow Log” to model many more data processing scenarios than were available in SorfML.....	213
Figure 3-6: The output hierarchy, determined by the type of analysis run. This allows the platform to display information interactively in the browser about the pipeline run outputs, using a Data API-mediated approach that doesn't expose the complete directory structure of the host.	216
Figure 3-7: The Sensor Configuration Page in the new platform. Users select a representative Dataset tabular file (1), and use it as a reference to annotate the header concepts found within it. Any subsequent uploads of datasets matching this Sensor configuration will be automatically processed with the ontological concepts annotated in this menu (2 and 3).....	217
Figure 3-8: The Dataset upload Page in the new platform. The user selects a dataset to upload (1), and matches it to a stored Sensor Configuration (2). The	

Dataset input is sanity checked with the stored Sensor Configuration, by matching the stored Sensor's header hash with the hash of the Dataset. 218

Figure 3-9: The pipeline configuration page within the web application. The user configures “channels” of sensors and associated datasets, and subsequent data analysis or machine learning pipelines. Valid configurations are also given visual feedback, in this case indicated by a *solid* green border around the process form – invalid configurations will result in a *dashed* red border. 220

Figure 3-10: Once datasets are assigned to a sensor channel, the user can conduct a variety of exploratory data analysis steps, to view the structure of the data and annotate outliers for exclusion from subsequent data processing... 220

Figure 3-11: For complex spectral data, the user is able to manually direct the process of feature selection, through selection of included regions on a representative data trace (1) - informed by prior research or publications, the feature ranges (2) will be used exclusively in any following pipeline steps or inferencing tasks. Alternatively, comma separated value (CSV) files of spectral ranges from prior work can be supplied to automate the process of range selection (3). 221

Figure 3-12: A demonstration of the configuration of an analysis pipeline. In this example, the user has selected four modular processes (1), and is contextually able to modify each of their parameters, if supported (2). 222

Figure 3-13: Entity diagram for a Model Endpoint. This supplies pertinent information to the platform about the specification of models saved by a user. ModelV1InferenceRequest is used in the GUI to request inference from multiple model configurations as specified in the “models” field, and data is supplied as an array of strings, with each entry in the array representing a single sensor's input (if the model expects more than one channel), in the order specified in the “sensors” field. ModelV1DirectQuery is used to make requests directly via the REST API. Only singular models can be specified in this approach, and data order must exactly match that expected by the requested model. 224

Figure 3-14: The ways in which the new platform's model inference functionality can be actuated. All requests made through the GUI must be authenticated regardless of the requested model's public visibility, whereas requests made via the API can either be public or authenticated, depending on the requested model. 225

Figure 3-15: The process for making a graphical SPARQL query using the Web Application. In this example the user is requesting all Datasets, matching the temperature regular expression “10” and the packageType regular expression “map” 227

Figure 3-16: The final architecture of the platform, divided with dashed lines into services for primary data handling of uploaded and processed data, semantic functionalities related to annotation of data (FAIRification) and task handling, which manages the Nextflow execution environment. 228

Figure 3-17: Configuration of the MSI Sensor Entity used for the MSI data in this study. The representative file is supplied only as a reference for the interactive header annotation wizard, and is not stored. Header ranges are selected either graphically via the tabular view or via the form underneath. Previously selected ranges are highlighted in the tabular view, and are listed and removable in the bottom section of the wizard. The “Type” field is populated with one of a variety of concepts pre-loaded onto the platform, in this case the significance of the numerical values in columns 1-18 (mean values) and 19-36 (standard deviations from the segmentation process used in conversion of Videometer data from a graphical to tabular data representation). 234

Figure 3-18: Upload procedure for Videometer (MSI) aerobic training data. Individual tabular files are selected and annotated with basic metadata. Following this, the appropriate sensor configuration is selected, and sample/observation labelling takes place, using ontological concepts from the associated ontology. Following upload to the file API, the ETL service will split and store the observation identifier concepts using this schema. 235

Figure 3-19: As an optional step, users can further annotate the uploaded dataset with geographical information detailing the location of origin of the dataset. Users can select this location directly from the map view, or via the location search or latitude/longitude inputs. In this instance and as an example, the MSI data was annotated with the details of the Agricultural University of Athens, the originating institution of the experimental data used in this work. 236

Figure 3-20: Configuration of a Nested Cross Validation Benchmarking Run. The user first selects the Sensor Channels, assigns data, and then selects the SKLEARN_Regression_Model_CV process, configuring the number of K-Fold splits and the preprocessing steps to compare. 242

Figure 3-21: Pipeline Run Log page showing the completed run log, with the auto-generated name "compassionate_jones", which contains the benchmarking performance reports for the selected data..... 243

Figure 3-22: Interacting with the report generated by the analysis pipeline: in the first panel, the work flow output directory is explorable from the browser, the view panel showing the interactive HTML report from the SKLEARN_Regression_Nested_CV process. Other representative plots are included in the report, including an RMSE performance heat map, which is shown below..... 244

Figure 3-23: Cross Validation RMSE performance Heat Map for the combined MSI Aerobic and Vacuum condition data. In this example, the best mean performing model trends across preprocessing strategies are clearly visible as ARD Regression, Bayesian Ridge and PLS Regression..... 245

Figure 3-24: Configuration of a hyperparameter search in the GUI. The user first selects the Sensor Channels, assigns data similarly to the nested cross-validation procedure, and selects the SKLEARN_Regression_Random_Grid_Search_CV process, configuring the number of k-fold splits, repetitions and the preprocessing steps to compare..... 249

Figure 3-25: The pipeline configuration page, showing the MSI and FTIR sensor configurations. Visual feedback notifies the user to misconfigured aspects of the

pipeline, in this case the lack of specified polynomial order in the Savinsky-Golay Preprocessing step.....	256
Figure 3-26: Selection of a serialised model object in the File Explorer UI. This allows for easy configuration of a machine learning endpoint on the platform.	257
Figure 3-27: The inference GUI. The left pane allows users to input tabular inference data, and match it with sensor configurations found on the platform. The middle pane is updated in real time with model endpoints on the platform that support those sensor "channels". The final pane shows metadata details about the model selected in the middle pane.	259
Figure 3-28: The results page, after clicking the “Predict” button in the inference GUI, containing a per-model result for every row supplied in the inference data.	260
Figure 3-29: With observed values supplied in the interface (top right) the platform acts in a model “auditing” capacity, allowing users to compare the performance of many models simultaneously.	260
Figure 3-30: With observed values supplied in the interface (top right), selecting the “plot metrics” tab allows for interactive graphical comparison of model performance, including bar plots and observed/predicted plots	261
Figure 3-31: The model configuration page, after selecting the three model types used in this work: Automatic Relevance Detection, Bayesian Ridge and Support Vector Regression. These models are added to the pipeline, and the training process is executed sequentially by the workflow execution service.....	265

List of Tables

Table 1-1: Examples of SVM kernel configurations in agri-food analyses.....	51
Table 1-2: Examples of artificial neural network configurations in agri-food analyses	55
Table 1-3: Table of references describing early fusion methods used in agri-food spoilage or authenticity analyses.....	64
Table 1-4 Table showing a non-exhaustive list of characteristic functional groups, their general expected wavebands and the expected intensity of response to excitation, demonstrating the utility for food analytical purposes. Table based on information from online resources (Mohrig et al., 2024; Reusch, 2013).	85
Table 2-1: Table of calibration/cross validation (CV) and test set sizes for chicken and beef spoilage scenarios using batch validation approach	131
Table 2-2: Grid Search Space for the Repeated Nested Cross Validation Approach. In the case of Random Search for MSI-FTIR Mid Fusion and MSI-MSIF-FTIR Mid/Late Fusion, 1500 (~37% coverage), 2500 (~12% coverage) and 400 (~30% coverage) iterations were performed per inner grid search respectively, with each parameter sampled from a uniform distribution. All other model types used exhaustive grid search.	134
Table 2-3: Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in chicken thigh samples using batch-on-batch validation (Aerobic n=70, Vacuum n=65, Combined n=135).....	151
Table 2-4: Generalised Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in chicken thigh samples using repeated nested cross validation (10 x 5 folds).....	152
Table 2-5: Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in beef mince samples	

using batch-on-batch validation (Aerobic n=46, Vacuum n=57, Combined n=103)	160
Table 2-6: Generalised Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in beef mince samples using repeated nested cross validation (50 folds).....	161
Table 3-1: This table represents an acceptable format for tabular data input into SorfML. The Sample ID column is correctly indicated as the first column. The input (independent) variable columns can be given any name, except those that indicate the Sample ID.	181
Table 3-2: Tabular summary of the major technological frameworks and libraries that enable the functionalities of the SorfML2 platform.....	204
Table 3-3: Example of a coded Sample ID, and the concepts string describing the ontological concepts represented within it.	215
Table 3-4: Table of the training and testing partition sizes for the three scenarios explored in this study.....	232
Table 3-5: Table of hyperparameter spaces explored in the nested cross-validation process for all MSI and FTIR training data. This process was conducted for the three preprocessing methods selected (standard scaling (Z-score normalisation), mean centring and raw/no preprocessing).....	239
Table 3-6: Table of hyperparameter spaces explored in the hyperparameter grid search operation for all MSI and FTIR training data. This process was conducted for the three preprocessing methods selected (standard scaling (Z-score normalisation), mean centring and raw/no preprocessing).....	248
Table 3-7: The twenty best performing algorithms from Nested Cross Validation using the VideometerLab (MSI) combined aerobic and vacuum training partition, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 5 outer and 5 inner folds. Results were generated with the <i>SKLEARN_Regression_Nested_CV</i> process. Note that the inclusion of a mean centring step in the Scikit-Learn implementations of Bayesian Ridge regression	

and PLS regression leads to consistent results regardless of user-specified transformation.	251
Table 3-8: The twenty best performing algorithms from the Nested Cross Validation using the FTIR combined aerobic and vacuum training partition, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 5 outer and 5 inner folds. Results were generated with the SKLEARN_Regression_Nested_CV process. Note that the inclusion of a mean centring step in the Scikit-Learn implementations of Bayesian Ridge regression and PLS regression leads to consistent results regardless of user-specified transformation.	252
Table 3-9: Best performing configurations of ARD, Bayesian Ridge and SVM Regression following hyperparameter tuning using the VideometerLab (MSI) aerobic, vacuum and combined condition training partitions, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 3 repeats of 5 fold cross-validation.....	253
Table 3-10: Best performing configurations of ARD, Bayesian Ridge and SVM Regression following hyperparameter tuning using the FTIR aerobic, vacuum and combined condition training partitions, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 3 repeats of 5 fold cross-validation.	254
Table 3-11: Results from the employment of four different PLS regression derived machine learning approaches in the new platform, for prediction of Total Viable Count values using aerobic or vacuum packaged chicken spoilage data from either MSI (VideoMeter) and/or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72)	262
Table 3-12: Results from the employment of four different PLS regression derived machine learning approaches in the new platform, for prediction of Total Viable Count values using the combined aerobic and vacuum packaged chicken spoilage training data from either MSI (VideoMeter) and/or FTIR sensors. (Combined aerobic and vacuum test batch size: 138).....	262

Table 3-13: Results from the employment of Bayesian Ridge, ARD and SVM Regression with Z-score normalisation, for prediction of Total Viable Count (log CFU/g) values, using aerobic or vacuum packaged chicken thigh datasets from MSI (VideoMeter) or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72).....	266
Table 3-14: Results from the employment of Bayesian Ridge, ARD and SVM Regression with mean centring, for prediction of Total Viable Count (log CFU/g) values, using aerobic or vacuum packaged chicken thigh datasets from MSI (VideoMeter) or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72).....	267
Table 3-15: Results from the employment of Bayesian Ridge, ARD and SVM Regression with Z-score normalisation, for prediction of Total Viable Count (log CFU/g) values, using combined aerobic and vacuum packaged chicken thigh datasets from either MSI (VideoMeter) or FTIR sensors. (Combined aerobic and vacuum test batch size: 138).....	268
Table 3-16: Results from the employment of Bayesian Ridge, ARD and SVM Regression with mean centring, for prediction of Total Viable Count (log CFU/g) values, using combined aerobic and vacuum packaged chicken thigh datasets from either MSI (VideoMeter) or FTIR sensors. (Combined aerobic and vacuum test batch size: 138)	268

List of Equations

Equation 1 General equation for the Mahalanobis Distance metric.....	57
Equation 2 Generalised Formula for Simple Moving Average.....	59
Equation 3 Generalised Formula For SNV	59
Equation 4 Calculation of residual matrix $E1$, where $X1$ is the inverse transform (reconstruction) of $X1$	67
Equation 5 Calculation of the standard deviation of the fused predictions in Robert <i>et al.</i> 's late fusion approach, where $\sigma RaIRRaIR$ represents the respective standard deviations of the fused data, Raman and FTIR sensors.....	71
Equation 6 Calculation of fused prediction in Robert <i>et al.</i> 's late fusion approach, where $XRaIR$ represents the fused average, XRa and XIR represent the Raman and FTIR sensor averages, and $\sigma RaIRRaIR$ represents their respective standard deviations.	71
Equation 7 The Beer-Lambert Law.....	80

List of Abbreviations

CFU	Colony Forming Units
FBO	Food Business Operator
FTIR	Fourier-transform Infra-red
LAB	Lactic Acid Bacteria
LDA	Linear Discriminant Analysis
MSI(F)	Multi-spectral Imaging (Fluorescence)
OWL	Web Ontology Language
PCA	Principal Component Analysis
PLS(R)	Partial Least Squares (Regression)
RMSE (C/P/CV)	Root Mean Squared Error (of Calibration/Prediction/Cross-Validation)
SNV	Standard Normal Variate
SVM	Support Vector Machine
TVC	Total Viable Counts

1 Introduction and Literature Review

1.1 Background

In a similar manner to the wider domain of chemometric and other data-driven study, the advancement of the disciplines of food safety, authenticity and quality analysis is driven in part by the constantly evolving landscape of information technology, equipping the analyst with a growing arsenal of devices and techniques to harvest actionable sensor data that facilitates their work (Nychas et al., 2021).

As the prevalence of these data-driven practices increases due to sensor miniaturisation and increasing cost-effectiveness, food business operators (FBOs), in addition to food standards authorities, are exploring the feasibility of incorporating this type of technology throughout the food production and distribution pipeline, to supplement their existing testing methods, which are often batch-based, slow and result in food destruction and waste (Lazou et al., 2021; Yousefi et al., 2019).

As a consequence of this, the established methodologies applied throughout collection, transformation and analysis of food safety, authenticity and quality data are in a constant state of iteration, to capitalise on the potential offered by these advancements and improve on the performance of existing measures, for, among others, reasons of public health concern, or reinforcement of trust in the food production and distribution network.

As the breadth of available sensor coverage options for food analysis increases, a case can be made for assessing the utility and benefits of multi-modal monitoring for assessing product state throughout its production and distribution lifetime (Huang et al., 2024).

Sensor/Data fusion of a union of non-overlapping technology coverage may record a more informative, holistic view of the state of a sample over a wider spectrum of variables, whilst the intersection of coverage of similar technologies may increase the “depth” of analysis, and offer a degree of redundancy and confidence when establishing the relationship between the recorded state and

the underlying biological and chemical processes taking place, a desirable trait when making decisions that have potential for damage to the health of consumers and the reputation of FBOs and regulatory bodies (Calvini and Pigani, 2022).

This desire to form robust pipelines for food analysis decision-making is complicated by multiple issues related to the data itself. The multi-modality of the data, i.e. the variety of different sensor technologies the raw data emerges from, in addition to the representational structures and concepts encountered, make it inappropriate to adopt a “one-size-fits-all” approach to ingestion into any analytical operation (Nychas et al., 2021).

In addition, the sheer volume of data produced by these information-era sensors in a real-world scenario precludes the manual intervention of operatives in processing this data, a scenario lending itself to automated analysis, which is complicated by the aforementioned multi-modal nature of high throughput sensor data.

1.2 Food Supply Issues in the UK and Internationally

The issues facing food supply infrastructure at the national and international level can be placed into several distinct categories, such as those involving international trade and economics, supply chain and logistics disruptions caused by climate change, natural disasters or geopolitical issues, or skilled labour shortages (Rojas-Reyes, Rivera-Cadavid and Peña-Orozco, 2024), however the issues that fall within the actionable realm of food analysis and associated informatics approaches can generally be categorised into those related to food safety, quality control and food fraud, supply chain transparency and waste optimisation.

The safety of food is likely the most important deciding factor in a consumer's dietary choices. Yearly, an estimated 600 million people (roughly 1 in 10) fall ill after consuming contaminated food and of those, roughly 420,000 die. Further to this, an estimated global cost of \$110 billion is attributable to lost productivity and associated medical expenses in middle and low-income countries (WHO, 2024). In the UK and other higher-income countries, the general impact on consumer

health caused by food safety issues is likely lessened due to enforcement of well-developed safety protocols, however recent investigations have suggested a far higher incidence of food-borne illness in the UK than previously thought – a 2022 Food Standards Agency (FSA) report estimated around 2.4 million cases yearly, an increase from the 2009 estimate of 1 million (O'Brien et al., 2020). Furthermore, public trust in governing bodies to enforce food standards is fragile: the FSA's "Food and You 2" survey of a stratified sample of 4000 households in the UK indicated that between 2020 and 2023, public trust in the FSA fluctuated between 69% to 78% in each two/three-month sampling wave (DEFRA, 2024).

Food quality is another factor in consumer choice – as populations become increasingly cognisant of how high-quality, nutritious food forms a critical component of a healthy lifestyle, suppliers must ensure that food reaching shop shelves is sufficiently attractive to warrant purchase (Nychas et al., 2021). The parameters that define a food product's quality are varied; many are related to a product's perceived "freshness": these might be considered in a continuum of spoilage, where microbial action or endogenous enzyme activity causes changes to the food morphology and chemistry, leading to unpleasant textures, off-flavours and smells (Sahar and Dufour, 2014; Snyder, Martin and Wiedmann, 2024); in some cases, severe microbial contamination leads to the aforementioned issue of food-borne illness, due to the presence of pathogens and pathogenic metabolites, both bacterial and fungal, which cause acute illness; common genera include bacteria such as *Campylobacter*, *Escherichia*, *Listeria*, *Salmonella* and fungi such as *Fusarium* (Swarnakar, Chowhan and Chakraborty, 2024). Some biological contamination causes both acute and long-term issues, such as the carcinogenic aflatoxins produced by fungi in *Aspergillus* (WHO, 2024).

Other factors are often perceived as markers of food quality: these include single-provenance foods, such as those with protected designation of origin (PDO) status in the EU (European Commission, 2023), or food derivatives with ingredients themselves perceived to be of higher quality than alternatives, such as extra-virgin olive oils (Guzmán et al., 2015) or monofloral honey (Gan et al.,

2016). High regard for “high quality” food products or their ingredients incurs an associated premium, with such products costing more than lower-quality alternatives. This may represent financial motivation for bad actors to conduct fraudulent activities in order to exploit consumers. Brooks *et al.* classified types of food fraud transgressions into three categories, based on their impact on the consumer, based on the work of Spink *et al.* (Brooks *et al.*, 2021; Spink *et al.*, 2019):

Direct impact represents immediate risk to the consumer i.e. the inclusion of substances that are toxic or lethal. This may include inorganic substances, such as lead chromate found in turmeric to improve yellow colouring (Cowell *et al.*, 2017) or organic substances such as melamine in milk products to feign protein content (Xiu and Klein, 2010). This categorisation also includes substitution of foodstuffs with allergens with inadequate notification, such as peanut derivatives (oils or flour) (Luparelli *et al.*, 2022) or dairy proteins in non-dairy products (Yang *et al.*, 2022).

Indirect impact is fraud caused by longer-term actions or inactions, such as low dose exposure to chronically toxic chemical agents, including banned herbicides or pesticides, or inorganic bulking agents that may not be acutely toxic, such as talc or brick powder seen in herbs and spices (Cottenet, Andrey and Dubascoux, 2024). Chronic omission of beneficial constituents such as preservatives or nutrition (vitamins, protein) is also included in this category: in 2004, the sale of infant milk powder of little nutritional value led to the deaths of more than 60 babies through malnutrition (BBC, 2004).

Technical fraud refers to incidences of fraud that are non-material in nature or are related to provenance claims and result in unfair advantage for the transgressor. This includes forgery of provenance information or other misrepresentation food products, such as mislabelling battery farmed or conventional meat as free-range, or the complete or partial substitution/adulteration of the aforementioned examples of olive oil or honey with inferior replacements, such as other vegetable oils or non-virgin olive oils in the case of the former (Poulli, Mousdis and

Georgiou, 2006; Sayago, Morales and Aparicio, 2004), and cheaper honeys or sugar syrups in the case of the latter (Dodd et al., 2025; Shehata et al., 2024).

The degree of economically motivated adulteration (EMA) occurring globally is challenging to estimate due to different methods of calculation, however cost estimates range from \$15 billion to as much as \$40 billion yearly (FDA, 2024). Some critical assessments of the global supply chain estimate as much as 10% of foods contain a fraudulent element, many of which evade detection due to the lack of a globally accepted, standardised fraud prevention program and increasingly complex international food distribution chains (Barrere et al., 2020).

The length of supply chains and poor oversight of weak points may allow bad actors the freedom to conduct their fraudulent activities, and conceal themselves amongst the volume of concurrent legitimate business, or allow non-fraud food control failures to progress (Stadler et al., 2016). Poor communication among the links in the regulatory chain may impede information flow and prevent findings in one jurisdiction from adequately progressing to the next, limiting the ability to form a complete view of the progression of a food crime or emergent incident, and pinpoint the origin of the event(s) (Soon, Manning and Smith, 2019).

1.3 Examples of Food Control Incidents

The importance of actioning a robust standards compliance testing strategy prior to food retail and consumption is paramount: without such, there is the potential for severe public health concern and damage to the reputations of involved Food Business Operators (FBOs) and governmental food standards authorities, both short and long term, due to the ramifications of the original incident and the public's lingering distrust in implemented testing measures (Li et al., 2021b). Food control incidents are reasonably common, and many are harmless and do not capture the public's attention – in the UK, a total of 69 Product Recall Information Notices (PRINs) were issued in the 2023-2024 period by the Food Standards Agency (FSA) and Food Standards Scotland (FSS), which are typically issued at or after food product recall due to concerns caused by contamination risk, mis-packing or mislabelling (DEFRA, 2024). However, recall events can be complex and costly (Whipple, Voss and Closs, 2009), damage an FBO's reputation, and

public perception of the category of products (Marucheck et al., 2011; Trienekens and Zuurbier, 2008). Some food control incidents and the subsequent mitigation efforts have become high-profile due to their perceived or realised impact on consumer bases, and led to changes in national and international food testing strategies.

1.3.1 Chinese Milk Scandal

Fraudulent adulteration of food to comply with quality-control/standards-compliance testing, or to evade it altogether, can pose serious health concerns. In the 2007-2008 period, a significant food safety incident emerged due to the deliberate augmentation of powdered baby milk products produced in China through the addition of the organic compound melamine.

The ostensible intention behind the producer's deception was to artificially increase the powder's nitrogen content to give the illusion that it contained a higher proportion of protein, in order to pass quality control testing. As a result of the adulteration, approximately 300,000 babies and young children suffered the effects of melamine poisoning, of which roughly 50,000 were admitted to hospital with kidney stones or complete renal failure – the Chinese health ministry officially confirmed that six infant deaths were directly attributable to the melamine contamination (Lancet, 2009).

As a result of the poisoning, a range of measures were taken by importer nations, from the complete ban of dairy and dairy-containing products from China (Xiu and Klein, 2010), to testing of all Chinese food imports, or only those implicated (Bhalla et al., 2009).

A noteworthy aspect of the incident's first emergence was the lag period between the first complaints made by parents of children affected by the chemical in December 2007, the main producer of the contaminated products first detecting the presence of melamine in milk products in June 2008, and the delay in reporting the findings to local authorities in August of the same year, and the eventual dissemination to the World Health Organisation's INFOSAN in late

September 2008. In that time, the sale and export of products containing the contaminant had continued (Gossner et al., 2009).

An effort from the World Health Organisation at the investigative conclusion of the Chinese milk product scandal in 2008 was to draw attention to the danger that the rapidly modernising international food trade had outpaced the legislation and monitoring infrastructure implemented at that time, and the ease with which bad actors could exploit that legislation for profit. Of paramount importance in the protection of consumer health when encountering suspected contamination issues in food production is the expedient transfer of findings to the relevant regulatory body, to have a greater chance of limiting the spread of affected produce and thus limit the size of any potential public health emergency.

It is reasonable to consider some possible reasons for similar delays in reporting of instances of food fraud or other food control violations:

- Cover-ups: Whilst the dissemination of information with food authorities is paramount for preservation of public health, there may be illicit considerations taken into account by the food producers, i.e. “damage control” – attempts to destroy evidence, clean up production and avoid repercussions before resorting to sharing of information with others (Zeng et al., 2018)
- Normalisation of deviance: Left unchecked, behaviour or practice that strays from the accepted standard becomes the norm and accepted by the workforce (Banja, 2010). This behaviour is likely compounded by poor oversight and the resultant sense of security for transgressors (a low chance of being caught) (Bogard et al., 2015; Etienne et al., 2020), leading to conditioning of workers and detachment from the effects of their actions (Lord et al., 2017; Lotta and Bogue, 2015)
- Internal affairs: Internal inefficiencies or miscommunication in the food industry between producer, distributor, and retailer may lead to certain events not being prioritised correctly (Soon, Manning and Smith, 2019)

- Industry Structure: Cultural practices that may discourage reporting issues externally, for fear of consequences for workers, and few protections are guaranteed for those who do report externally after reasoning that no resolution would be achieved from an internal report (Etienne et al., 2020)

A 2009 report on the extent of the health problems caused by the contamination made note of the fact that the Chinese authorities acknowledged the events when the WHO requested information following initial investigations, rather than the information being freely volunteered to the WHO upon first discovery (Gossner et al., 2009). Whilst noting the delays in the initial reporting, the review highlights the successful networking of scientific knowledge through the WHO International Food Safety Authorities Network (INFOSAN). This network facilitated the rapid dissemination of intelligence with all member parties, and Chinese-led expert meetings aided in exchange of current scientific knowledge and assessing the best course of action for member states to take to minimise the public health impact of the incident.

A consideration to be made during international public health incidents is the likelihood of public health guidance and legislation regarding hazard scenarios differing between different member states. For example, whilst food testing programmes for melamine were under way following the 2008 incident, some countries took the policy of accepting goods that displayed a concentration below a maximum threshold for contamination, whilst some took a no-tolerance policy, and rejected most imports of dairy products from China as a result (Bhalla et al., 2009). This difference in approach was likely caused by the lack of consensus on some domain concepts, particularly the lack of establishment of the threshold value of detectable melamine in foodstuffs at which action should be taken. Following the expert-led meeting in 2008, the Chinese authorities were commended for their rapid supply of useful information pertinent to the incident to the other INFOSAN partner nations (Gossner et al., 2009). Member feedback suggests that standardised international food information networks such as INFOSAN are useful for overcoming some of the complexities of knowledge exchange between different food authorities and acting as a “community of

practice”, facilitating knowledge exchange between domain experts and collaborative meta-analyses to assess different regional approaches for dealing with issues. (Savelli and Mateus, 2020). Systematic reporting following INFOSAN incidents has revealed that a minority of member states contribute the majority of information exchanged through the network (Savelli et al., 2019). More than half of respondents to a 2019 user survey of 239 INFOSAN members in 137 countries have suggested the existence of a number of barriers to participation, particularly related to the need for a standardised way to share information at a national level, difficulties conducting national-level food safety risk assessments, limited funding for food monitoring/response infrastructure, and limited realised monitoring/response infrastructure (Savelli and Mateus, 2021).

1.3.2 Horse Meat Scandal

In 2013, the so-called “horse meat scandal” affected the European food sector in such a way that the effectiveness of quality, authenticity and safety controls were brought into disrepute.

Food safety inspectors from the Food Safety Authority of Ireland (FSAI) discovered varying proportions of horse meat in frozen products labelled as other meats (DAFM, 2013), a discovery that sparked similar investigations in other European nations. Across the EU, 27 member states conducted 4144 tests for the presence of horse meat in similarly packaged and distributed products, and of those 193 showed a positive result (4.66%) (European Commission, 2014a).

Methods were used that had not been routinely applied to agri-food assessment – initial detection of the adulteration was made after employment of advanced DNA testing methods which had not been required under food regulations, including real-time polymerase chain reaction (RT-PCR/qPCR) (DAFM, 2013; Salihah et al., 2016). Other methods, such as the enzyme-linked immunosorbent assay (ELISA), which uses enzymes to detect the quantitative or qualitative presence of “fingerprint” antigens in proteins (Nešić, Stojanović and Baltić, 2017) were used, however subsequent study in beef and pork has suggested ELISA requires a higher threshold of meat adulteration for reliable detection (Perestam et al., 2017).

Whilst clearly an issue of food fraud, the presence of horse meat raised accompanying concerns about the suitability of the meat for human consumption. Whilst an industry for horse meat does exist in parts of the EU, there was no guarantee that the meat entering the food chain fraudulently was sourced or prepared suitably for human consumption. As such, testing for trace quantities of veterinary drugs took place; of particular concern was the potential of phenylbutazone or “bute” entering the food chain, a drug that presents a potential health risk to humans.

Of the 3115 tests for bute, 16 proved to be positive (0.51%). A further 7951 follow-up tests were undertaken by food business operators (FBOs) for the presence of the drug, of which 110 samples indicated a positive result (1.38%) (European Commission, 2014b).

The resulting investigation undertaken by the European Commission established that the risk to public health was low, but the entry of illegitimate produce into the food chain damaged consumer trust and raised questions about the functioning of food testing infrastructure, both at a national and international level. International consignments not only experience the potential “weak-link” in storage and monitoring whilst moving across large geographical distances, but also the issues that may arise when crossing national borders and thus moving into different food standards jurisdictions (Soon, Manning and Smith, 2019).

As a result of the original findings, the subsequent investigation and the resultant public scandal, the European Commission authored a five point plan of areas of potential changes that the food standards infrastructure across the EU should move to enact (European Commission, 2014c).

Several of the proposed actions were specific to the control of processing and distribution of horse meat through the enforcement of rules surrounding horse passports:

- Distinct identification to be assigned to horses and thus horse products entering the human food chain, and members states to report transparently on the measures they take to enforce the rulings on such, particularly on how the passport is filled following veterinary drug administration, and to perform regular controls to ensure compliance with legislation.
- To make mandatory the utilisation of a national database to consolidate the recording of issuance of horse passports.

Following the conclusion of the investigation's testing efforts, other more generic/wider reaching suggestions were made for the purposes of strengthening food assurance infrastructure – in summary:

- Better coordination between EU member states regarding food safety/quality/authenticity monitoring and standards compliance, in collaboration with enforcement agencies such as Europol.
 - Dissuasive random spot checks to ensure compliance with legislation and to investigate suspected instances of food fraud
 - Implementation of newer legislation to ensure member states' transparency about internal food standards affairs
- Development of a monitoring platform, similar to, or integrating with, the EU's RASFF (Rapid Alert System for Food and Feed - As RASFF was primarily concerned with risks to health from food and feed, it was not initially triggered after discovery of adulteration)

The overall determination of the European commission's investigation attributed the scandal to fraudsters exploiting weaknesses in the food production and distribution network (European Commission, 2014c), and emphasised the need for improved national and international food control networking. For example, the EU made recommendations for central databases for horse passports, offering greater efficiency for authorities making enquiries on the status of animals prior to sale or export, and which would simplify assurance that all necessary data is associated with individual registrants, such as their drug status and thus suitability for consumption (European Commission, 2014c); information systems for livestock tracing and certification issuance had already been in established use in the EU by 2013, an example being the TRAdE Control and Expert System (TRACES) under the competence of the Directorate-General for Health and Food Safety (DG SANTE) (European Commission, 2003). Subsequent web service plug-ins have been developed to complement the functionality in a modular fashion to fulfil the requirements of additional EU directives: Bovine Exchange (BOVEX) is designed for the automatic harmonisation of bovine animal identification data across the national databases connected to it.

Following the events of the scandal, the United Kingdom's government commissioned the Elliott review in 2014, a critical evaluation of the state of the food industry. One of the main outcomes of the review was the foundation of the National Food Crime Unit, a law enforcement arm of the Food Standards Agency, with the remit to discover and pursue consequence for fraudsters (Elliott, 2014).

1.3.3 Common Themes in Food Standards Incidents

A common theme between these events is a seeming reliance on the corporate responsibility of food suppliers to report honestly to food authorities about the safety and authenticity of produce in their purview, in a world where increasing pressures are being placed on food authorities due to the demands from increasing populations, financial stresses and an increasingly global food industry.

The aforementioned incidents demonstrate in some capacity that such reliance, when coupled with legislation to dissuade fraudulent reports of food authenticity or safety, is not necessarily enough to deter bad actors from exploiting their customers for profit, when presented with the right conditions; as a result of the melamine scandal in China, known for having robust food adulteration legislation, 21 people were sentenced with varying severity (Gossner et al., 2009).

Incidents that occur due to negligence or other non-intentional factors are also non-optimally addressed with a food testing infrastructure that relies on batch-oriented compliance testing. Such testing gives a “top-down”, retrospective determination of the authenticity or safety status of a sample, thus possesses little real-time power to determine the weak point in the food chain where the issues originate (Nychas et al., 2021). Such testing may be time-, cost- or location- prohibitive, requiring specialist equipment and expertise that cannot be distributed efficiently throughout the food chain, or as a necessity incorporates a time delay, such as microbial culture/colony counting for the assessment of biological contamination.

Outsourcing of testing duties to external specialist laboratories may result in additional delays to reach decisions on the fate of food consignments and likely cannot offer sufficient capacity or coverage of the whole food chain in a high enough resolution to identify points of failure. To use the horsemeat scandal as an example, the date of the first positive test for equine DNA was 10th December 2012. Following this, the Food Safety Authority of Ireland (FSAI) formally requested that the Irish Department of Agriculture, Food and the Marine obtain further samples, sent for lab study on 4th January 2013, with stores starting to issue product withdrawals in mid-January, roughly 6 weeks later. During this period however, adulterated products were still freely sold to consumers in the EU (European Commission, 2013).

A means to augment the currently established food testing frameworks of the EU and China is desirable, working in concert with the empirical laboratory based food testing strategies currently employed, to provide coverage in real- or close-to-real-time of parts of the food chain that are currently poorly monitored, and

facilitating the rapid networked exchange of data regarding a specific food production and distribution pipeline across commercial and international boundaries (Nychas et al., 2021).

Real- or close-to-real- time monitoring of food consignments moving throughout the food chain would offer a preventative methodology for identifying and rectifying production failures, rather than relying entirely on batch-based testing or spot testing conducted randomly or as the result of an incident report, in a concept known as Process Analytical Technology (PAT) (van den Berg et al., 2013).

A modernised system working with oversight from relevant food authorities may also allow for identification of suspected instances of food fraud, as the ability to track a specific consignment would generate a production and distribution process status trail, so deviations from expected operating norms are thus more easily identified and traceable (Brooks et al., 2021).

1.4 Process Analytical Technology and Digital Twin

The desire to incorporate real- or near-real- time, non-invasive sensor technology into the production or distribution line of food products can be succinctly explained as falling under the umbrella term of Process Analytical Technology, itself an element of modern manufacturing practice/initiatives, including concepts such as “quality by design”, “real-time release” and “continuous manufacturing” (Clegg, 2020).

The formalisation of the term emerged from the US’ Food And Drug Administration (FDA) assessment of the performance metrics of drug manufacturing and guidance for enabling recommended changes; stated in its original documentation is the definition that PAT is *“a system for designing, analysing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality.”* (FDA, 2020)

The general intention behind the concept of PAT was to facilitate consistency in the output of drug production pipelines, through the integration of monitoring in-, at- or on-line, to offer granular insights into the state of a sample or consignment at any stage in the pipeline, and the potential of greater consistency of output due to finer control over operating parameters. This is achieved through the non-destructive monitoring of Critical Process Parameters (CPPs) and thereby maintenance of Critical Quality Attributes (CQAs).

Van Den Berg *et al.* described the potential shape and monitoring arrangement of a PAT framework when applied to food production control, using the existing concepts of process dynamics as a basis of understanding. A candidate *process analyser* (for example, the sensor component) must accurately reconstruct the true (but unknown) signal of interest in a pipeline for use as an input for subsequent adjustment, for example, the temperature of a sample, or the correct proportions of ingredients in a mixture. Candidate analysers were considered unsuitable for this task if they possessed one or more negative traits (Van den Berg, Hoefsloot and Smilde, 2002):

1. Large uncertainty of measurement: All sensing devices possess some degree of error, however sensors that exhibit unpredictable traits in real-world working conditions are unsuitable
2. Low frequency of sampling (and thus high sampling *period*, the reciprocal of frequency): crucial for real-time sensor analysis, the frequency of sampling determines the resolution of the reconstructed signal
3. (Related to point 2) A high sampling *period*: results in a smoothing process where variation in the true signal will be captured and averaged during one cycle of the sensor device
4. Carry-over effects: Sensors may possess a *memory* effect, where the results from a previous cycle of the device are inadequately purged before the recording of the next cycle, resulting in a smoothed (convoluted) reconstructed signal as above, resulting in deviation from the true signal

5. High processing (dead time) delay: periods where the sensor or associated computational device is processing a cycle's results, and thus not progressing with recording of the next. This lag may offset results by minutes, hours or days depending on the device or method used

In their assessment of how these concepts aligned with the real-time objectives of PAT, the utilisation of spectroscopic sensors, such as ultraviolet–visual absorption (UV-VIS), near infrared (NIR) and infrared (IR) absorption, Raman scatter and excitation emission fluorescence spectroscopy (EEM), were suggested to be favourable for integration into a *during-process* monitoring scheme. Their reasoning was due to their responsiveness through their high measurement frequency (and thus low sampling *period*, the reciprocal of frequency), and their non-invasive nature limiting carry-over effects, when compared to other sensor device types (van den Berg et al., 2013).

Figure 1-1 shows a conceptual overview of a Process Analytical Technology implementation. In the conventional testing approach in Figure 1-1A, the consignment moves through the pipeline unhindered, and is subject to compliance testing at the end-stage. One of the main drawbacks of this approach is that any deviation from acceptable sampling parameters, be it due to chemical/biological contamination, incorrect manufacturing protocol or otherwise, is recorded at a point where very little can be extracted from the collected data to isolate the specific chronology of the deviation. In the eventuality that a deviation is detected, the lag time between its emergence and its recording, and the potential inability to isolate the specific point of failure may lead to uncertainty in the state of other samples in the current operating run, leading to batch release delays as lab-centralised diagnostic testing is completed and any actionable rectification steps are taken, or in severe cases, complete batch destruction.

Figure 1-1B shows a pipeline for a similar task, with an attached PAT-like monitoring strategy. By integrating sensor technologies throughout, individual points where deviation from operating norms are likely to occur can be monitored in close-to or in real time, and thus more reactive measures can be taken to resolve emergent issues and return the line to optimal conditions. This

“checkpointed” approach is intended to give the operator(s) stronger confidence in the consistency of output, as not only does the additional data grant greater resolving power in overview of the system, but chronological thresholds can be drawn that can isolate consignments affected by the specific deviation, thus minimising wastage due to whole-batch destruction, and costly withdrawals from distribution and sale.

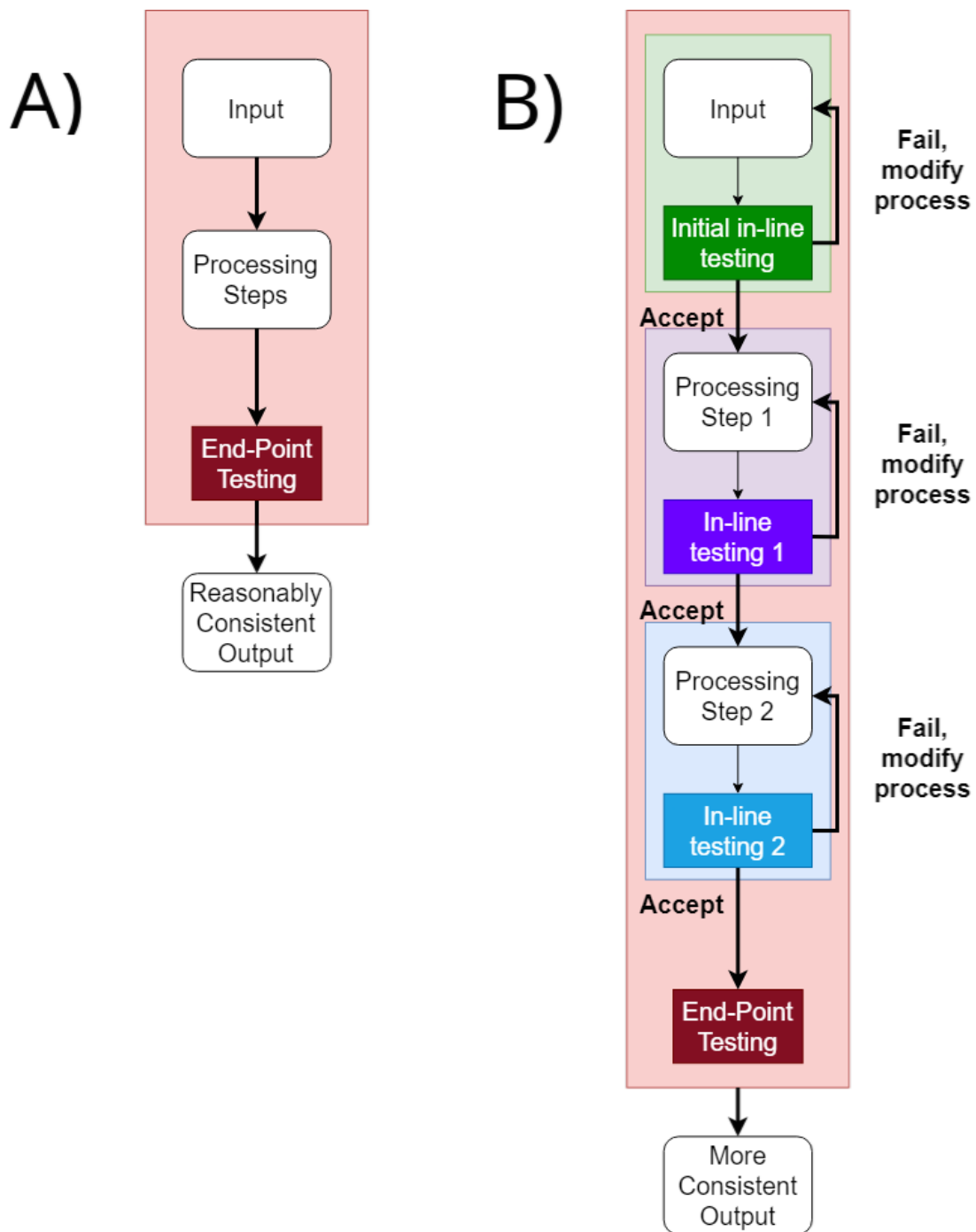


Figure 1-1: A Conceptual Overview of a typical food production pipeline (A), versus a PAT-augmented production pipeline (B). The greater frequency of non/minimally-invasive testing throughout production is intended to maintain consistent output quality.

In the context of the food sector, reviewers have suggested the use of PAT in concert with other data-driven and Internet of Things (IoT) practices (Nychas et al., 2021).

Digital Twin (DT) is a concept emerging from industry and now viewed by ISO as an international standard (Grieves, 2015; ISO, 2021), in which digital facsimiles of physical objects or processes are created with the intention of fusing real-time data with mathematical models built on prior experience, gathered both experimentally or in otherwise controlled conditions. This synergistic convergence between the real and virtual twins allows for predictions to be made about the state of the physical entity being modelled, which shows utility in the context of real-world manufacturing, where outcomes can be forecast for said entities in a variety of simulated conditions without risk of economic or consumer harm. Additionally, the difference between the predicted and observed values (residuals) can be fed back into the digital model, for further validation of the built models, their predictive power potentially improving over time, depending on the architecture of the built solution.

When integrated into a production or distribution pipeline as reflected above, the type of data representative of networked, high performance sensor technologies may allow for real-time, low latency forecasting of quality, safety and authentication metrics in a wide variety of simulated conditions.

An example in the context of food safety may be the incorporation of the results of prior bacterial growth and contamination experimentation into such a system, enabling the operator to forecast bacterial growth in their manufacturing processes, allowing for best- and worst-case scenarios to be accounted for, when factoring in the current configuration of the tooling, and their sum uncertainties/precisions.

This would give a confidence interval for production chain outcomes and inform the correct setting of parameters for succeeding entities passing through the pipeline, in a continuously updating negative feedback loop based on the criteria of minimising this interval and decreasing the absolute observed value of contamination.

Whilst the general conceptualization of DT typically encompasses long-lived products rather than non-durables like food (Srai et al., 2019), actualised examples exist; digital twinning has been accomplished in fruit, with the digital twin models recapitulating the mechanisms for mango spoilage based on physical fruit spoilage data and calibration using “physical twins”, physical facsimiles of real fruit with connected sensor payloads, configured to validate statistical models based on the relationship between the internal conditions of the real fruit and the external storage conditions (Defraeye et al., 2019).

The commercial implementation of DT has been explored in domain-specific contexts: Siemens developed a data management solution that brought together data from across the lifespan of food entities, and coupled it with their proprietary IoT platform, “MindSphere”, now known as “Insights Hub”, a System-as-a-Service (SaaS) solution supported by proprietary networking hardware and open Application Programming Interfaces (APIs) (Siemens, 2017). These are a crucial component of modern interoperable networked solutions, as they allow for users/operators of the service to “build to” the specifications of the platform, thus developing their own domain specific iteration of the platform.

1.5 Maintaining Traceability in the Food Supply Chain

In a modernised monitoring system, food control data would need to be networked in such a way to accompany the food product throughout its commercial lifespan in a de-fragmented fashion, potentially across national boundaries, whilst maintaining a degree of incontrovertibility. While Hollands *et al.* stress that such data is no substitute for inspection, audit and analysis in investigation of authenticity and provenance, blockchain technologies may be suitable for providing a write-only chronological ledger of timestamped assessments of food consignments, useful for reasons of efficiency in issuing product recalls, accountability for producers and distributors, and in stock keeping (Hollands et al., 2018).

Inheritance of properties through a chronological chain may be watched algorithmically for the detection of product deviation/control heuristics, such as spikes in refrigeration temperatures for cold-chain products, or discrepancies in

retail performance across food batches, potentially indicative of cross-batch variation or quality issues. Increased traceability and perceived trustworthiness of food origins may also be appreciated by the consumer, who would be able to view the progression of food through the food supply chain from “farm to fork” (Galvez, Mejuto and Simal-Gandara, 2018). The increased accountability and traceability in production and distribution may prove to be a deterrent for bad actors, as stored records are impossible to edit retroactively, thus granting investigators an additional data stream with which to examine the progression and evolution of issues (Behnke and Janssen, 2020; Brooks et al., 2021).

A blockchain is in essence a Distributed Ledger Technology (DLT) that records “transactions” in a secure, decentralised, immutable, and transparent manner, originally forming a component of the Bitcoin protocol, intended for use in the financial sector. Each “transaction” or block in the chain is linked cryptographically to the preceding one, forming a continuous, verifiable chain of events. (Nakamoto, 2008). The peer-to-peer nature of the protocol ensures that all have access to the same version of events, potentially enhancing traceability and trust in food chains (Marchese and Tomarchio, 2022; Tian, 2016).

Perceptions of the protocol have gradually acquiesced to the concept of blockchain detached from the Bitcoin/financial connotations, from an initial focus on privacy and security perspectives with little focus on concrete application and evaluation in non-Bitcoin applications, such as smart contracts and licensing (Yli-Huumo et al., 2016).

The utility of blockchain in agri-food has been explored more recently in the research and commercial sectors, in the aforementioned facilitatory capacity for food business operators, but also for regulatory compliance / fraud prevention purposes. In 2018, the UK’s Food Standards Agency (FSA) successfully trialled the use of a blockchain ledger in a cattle slaughterhouse, for the recording of animal welfare and post-mortem conditions, the first application of the technology in a regulatory capacity (Brewer et al., 2021; FSA, 2019).

Walmart, in partnership with IBM, have demonstrated that the application of blockchain can expedite many of the logistical and customer-facing tasks that

face an extremely large food retailer. In feasibility studies of blockchain using a platform based on the Hyperledger Open-Source project (Kamath, 2018), mangoes were used as the pilot foodstuff to assess the acceleration of product traceability querying – in a comparative assessment where it took a team nearly a week to trace the origins of consignments of sliced mango, equivalent blockchain-enabled querying took 2.2 seconds (Stristy, 2021). A subsequent pilot in collaboration with China's Tsinghua University implemented a system for pork traceability, fostering Chinese academic interest in agri-food blockchain systems (Kamath, 2018).

Further commercial interest in IBM's Food Trust platform has resulted in the French retailer Carrefour forming a similar commercial partnership, using the technology to boost customer-facing traceability and trust in food provenance and quality claims (Reuters, 2019). More recently, the Middle Eastern, African and Asian franchises of the European retailer joined the project, the initiative covering the store's own-brand chicken and hydroponic-grown micro-greens, to grant customers *"immediate access to actionable food supply chain data, from farm to store shelf"* (IBM, 2021).

Despite its benefits, the open, distributed nature of blockchain may be a double-edged sword: while the enhancement of transparency, trust, and traceability in food supply chains is likely a benefit for consumers, potential risks are also present for food business operators (FBOs). The immutability of blockchain records, though a strength in preventing fraud, can become a liability if inaccurate or misleading information is entered, as corrections may require complex additional transactions rather than simple modifications, which coupled with computational complexity and scale, may become costly to fix (Treiblmaier, 2018). Moreover, interoperability at the national and international level remains a major hurdle in agri-food data exchange, an issue likely compounded by utilisation of proprietary platforms such as IBM's Food Trust without adequate employment of open, established data standards, largely limiting DLTs, like blockchain, to internal use (Casino, Dasaklis and Patsakis, 2019; Pearson et al., 2019).

1.6 Common Analytical/Machine Learning Methodologies

The diverse modalities of high throughput sensor technologies used in food analysis often necessitate a variety of processing steps before the data can be presented intelligibly. The usage of machine learning techniques is commonplace, in both an unsupervised fashion, for the purposes of data exploration and dimensionality reduction/segmentation tasks for subsequent analysis, but also in supervised applications, for the construction of models for the prediction of external and internal properties of food samples related to authenticity, quality and safety analysis, based on their observable state – using data from either optical/visual analysis using various modalities of spectroscopic sensor, or from chemical analysis, be it exploration of the products of biological processes occurring in the sample, or recording of tell-tale fingerprint traces of non-biological contamination, both conducted with a variety of sensors, including electronic noses and various chromatographic and spectrographic approaches (see Section 1.8: Modern Sensor Technologies).

In order to construct predictive supervised models in such a fashion, accompanying laboratory work must take place to provide the “target” information - the “real” values for whatever property the researcher has an interest in predicting, for example, the size of bacterial populations existing on the surface of samples at stored for different lengths of time or at different temperatures (a typical regression task), or the spoilage classification of a sample, based on its volatile profile.

The capturing of sample state information with applicable sensors must occur in concert with empirical laboratory work, to minimise the possibility of unaccounted factors influencing the construction of the predictive model, for example, the progression of biological processes in the time between laboratory recording and sensor capture. Further to this, other extraneous factors must be considered in a large scale or long-running data capture project. For example, the configuration of sensor devices must be kept consistent; certain modalities of sensor may be susceptible to extraneous phenomena such as drift, and may require regular calibration (Bosch et al., 2022).

1.6.1 Common Machine Learning Methods Applied In Sensor Analysis

Machine Learning (ML) is a subset of Artificial Intelligence (AI) concerned with developing algorithms that “learn” from data and can be used to make predictions on new, unseen data without being explicitly programmed (Hastie, Tibshirani and Friedman, 2009).

Supervised learning forms the foundation of predictive modelling in this work, where the goal is to build models that map input variables to target outcomes. Unsupervised learning, which derives structure and patterns from unlabelled data, is useful for clustering and feature extraction applications, and will be described in the following sections where appropriate. Reinforcement learning, which learns optimal decision-making strategies through trial and error using reward-based feedback, has seen significant advancements in fields such as robotics and sequential decision-making (Sutton and Barto, 2018). However, given the non-sequential nature of food analysis expected in this work, this category of ML was not explored.

The selection of a suitable supervised ML algorithm for a particular task depends on the fundamental nature of the problem, for example, whether the dependent variables (target variables) are numerical (indicative of a regression problem) or categorical (indicative of a classification problem).

A further consideration is the characteristics of the data upon which the ML model is to be “trained”. An aphorism commonly, although dubiously attributed to statistician George E.P Box is that “*All models are wrong, but some are useful*”. In a 1976 paper published in the Journal of the American Statistical Association, he described aspects of the scientific method that should be applied to development of statistical methods for effective results. As a completely “correct” model is impossible to obtain, the more elegant solution that solves a problem to a sufficient margin of error – an “*economical description of natural phenomena*”, fits with the law of parsimony in that hypotheses with fewer assumptions should be preferred (Box, 1976).

However, as the inevitable drive towards more accurate results in Machine Learning continues, a compromise must be set between the accuracy and the interpretability of models; it may be hard to justify more simple, interpretable models over more complex, and in some cases functionally opaque “black box” models, if the results are suitably validated as being more accurate (Kuhn and Johnson, 2013). Deep Learning (DL), a field of research associated with the construction of neural networks, can be considered in the realm of “black box” algorithm design due to the heavy parameterisation of resultant models that makes interpretation a complex task.

Chemometrics, the application of mathematical and statistical methods to chemical data, has increasingly leveraged ML algorithms for spectral data analysis to enhance predictive accuracy and interpretability. Chemometric datasets, especially in spectroscopic and analytical chemistry applications, tend to have high collinearity and relatively small sample sizes compared to the number of variables (Bro and Smilde, 2014). Deep learning models, while often powerful, typically require large datasets upon which to train (Goodfellow, Bengio and Courville, 2016) in the absence of more complex approaches for dealing with this problem, such as careful pre-training strategies or transfer learning techniques (Min, Lee and Yoon, 2017), which are likely precluded due to the context-dependent nature of chemometrics, where spectral characteristics of a training dataset are highly dependent on sample composition, instrument settings, and preprocessing techniques.

Conventional algorithmic choices have been proven to display accurate performance for the types of highly dimensional data explored in this work. These include Principal Component Analysis and Principal Component Regression (PCA/PCR) and Partial Least Squares Regression (PLSR) (Martens and Naes, 1992). Such methods have also been evaluated favourably in the context of Process Analytical Technology (PAT) applied to the food industry (van den Berg et al., 2013).

These commonly-applied algorithms will be explained in more detail in the following sections. Following this, Section 1.8: Modern Sensor Technologies will

examine the use of modern sensor technologies in concert with ML techniques as may be found in a prototypical sensor-enabled food monitoring pipeline. Section 1.7: Sensor/Data Fusion will examine avenues for potential improvement of Machine Learning performance in food analysis through the fusion of multi-modal monitoring signals with the goal of providing greater analytical coverage of samples and enhancing reliability of predictions.

1.6.1.1 Least Squares Techniques

Partial Least Squares (PLS) methods are a family of algorithms that find consistent use in the chemometrics and Bioinformatics fields, due to their aptitude for spectral data analysis tasks (Wold, Sjöström and Eriksson, 2001). PLS excels in handling the challenges posed by highly dimensional data typically expected from spectral data.

PLS methods are an extension of the earlier Multiple Linear Regression (MLR) methods (Wold, Sjöström and Eriksson, 2001), and are similar in some respects to Principal Component Regression methods. PLSR shows clear benefits over MLR for the analysis of spectral data. MLR is suited to the analysis of data wherein the predictor variables are relatively few, and display linear independence. This is not representative of many modern sensor technologies' output, such as that the spectroscopic methods mentioned previously (and described in greater detail later in this work), where many variables may exhibit a response to certain phenomena existing in the sample. A domain specific demonstration of this could be within an IR spectrogram, wherein the presence of alcohol groups in spoiled food samples may cause response across multiple wavenumbers in concert ($3400\text{-}3700\text{cm}^{-1}$) (Mohrig et al., 2024; Reusch, 2013).

Furthermore, the PLS algorithm is suited to the commonly-encountered circumstance in chemometrics where the number of observations (samples) is less than the number of predictors used. This condition, sometimes known as the “small n , large p ” problem, typically results in the predictor ($\mathbf{X}$) matrix being low rank due to collinearity, which can complicate some of the steps of executing some machine learning algorithms (Abdi, 2010).

PLSR methods handle these type of collinear predictor variables efficiently; loosely speaking, they examine the relationship between the $\mathbf{X}$ (predictor) and $\mathbf{Y}$ (response) aspects of the training data used to build the model, with the aim of capturing the multidimensional direction in the $\mathbf{X}$ space that explains the most variance in the $\mathbf{Y}$ space. This is in contrast to Principal Component Regression (PCR), another method intended to handle data displaying collinearity. PCR performs the Principal Component Analysis (PCA) process on the $\mathbf{X}$ matrix, using the generated principal components as the input features for a multiple linear regression model for prediction of $\mathbf{Y}$ (Abdi, 2010). The principal components generated in the PCA process offer an orthogonal vector-based explanation of the variance in the dataset, avoiding the problems of collinear data in the resultant matrix. However, the PCR method does not explain any of the variance observed in $\mathbf{Y}$. In this regard, there is no guarantee that any of the principal components generated will correspond to the variance observed in $\mathbf{Y}$ in any meaningful way.

On the other hand, the objectives of the dimensionality reduction steps in PLS are to **a)** explain the variance in matrix $\mathbf{X}$, **b)** explain the variance in matrix $\mathbf{Y}$, whilst **c)** maintaining the best explanation of the covariance between $\mathbf{X}$ and $\mathbf{Y}$ in combination.

Diagnostic interpretation of a built model can occur in a relatively simple fashion, due to the presence of multiple metrics describing the training process. The scores generated from the transformation of the $\mathbf{X}$ and $\mathbf{Y}$ matrix, $\mathbf{t}$ and $\mathbf{u}$, can be used for interpretation of the model, as along with their accompanying weights they can be used to determine the importance of certain predictors in the training data and their effect on variance in the response variable(s) (indicated by numerically large weight values), or indicate redundant variables that may provide the same information (suggested by similar weight profiles)(Wold, Sjöström and Eriksson, 2001). Furthermore, the training process for PLSR generates residuals for both $\mathbf{X}$ and $\mathbf{Y}$ data. Residuals represent the parts of the data not explained by the model in $\mathbf{Y}$, and can indicate poor training performance in the case of consistent large values, or outlying values in the relationship between $\mathbf{T}$, the transformed feature space, and $\mathbf{Y}$. The residuals generated for $\mathbf{X}$

indicates the part of the data not used for training, and can be used for outlier detection within the predictor space (Wold, Sjöström and Eriksson, 2001).

The partial least squares model, whilst originally intended for the prediction of continuous variables (regression), has been modified for classification/discrimination tasks (Barker and Rayens, 2003; Ståhle and Wold, 1987). Named partial least squares-discriminant analysis (PLS-DA), the objective of the algorithm is to maximise the separation between predefined classes in a response variable, analogous to the objectives of PLS regression of maximising the covariance between $\mathbf{X}$ and $\mathbf{Y}$.

1.6.1.2 Support Vector Machine Methods

Support Vector Machines (SVM) are a popular alternative to PLS and PCA derived methods in sensor-enabled food analyses. They function differently to the aforementioned methods, exploiting linear algebraic theory to find separation between individuals in a dataset. They do this by attempting to define hyperplanes in n -dimensional space, for the purposes of maximising separation between correctly labelled groups. While a geometrically complex procedure to visualise, especially when considering highly dimensional data with more than three dimensions, SVM's kernel machine design is an attempt to simplify the problem (Hearst et al., 1998). In essence, the algorithm attempts to project the input space, $\mathbf{X}$, non-linearly into a feature space, $\mathbf{F}$, for the purposes of finding the hyperplane, yielding the decision boundary, which is transferable to feature space. The concept emerged as a result of a concept in 1963, originally implemented as a linear classifier (Vapnik and Lerner, 1963). Through the use of a kernel function and the associated "kernel trick", there is no need to explicitly map the input data into the feature space, enabling the discovery of this boundary without the excessive computational steps involved in mapping highly dimensional matrices. The kernel or Gram matrix represents a measure of similarity within the dataset, and is calculated for all pairs of observations (samples, or rows in the dataset). The selection of the kernel function to use is dependent on the task, and while there is typically no substitute for cross-

validation in determination of suitability (Li et al., 2021a), Table 1-1 is a general summary of kernel functions and their areas of use in agri-food analyses.

Similar to principal-component- or partial-least-squares- regression, SVM also shares some strengths when dealing with highly dimensional data that exhibits more features than observations – in typical linear methods, increasing the dimensionality of the feature space will naturally increase the sparsity of the data, i.e. the distance between data points, making it difficult for algorithms to find patterns in the data. This so-called “curse of dimensionality” generally means that for an increase in the number of features desired in a machine learning model, a corresponding increase in the number of observations is also necessary (Keogh and Mueen, 2017). SVM methods typically display robustness in the face of this problem; the hyperplane can still be generated in sparser, higher dimensional space, due to the determination of decision boundaries (hyperplanes) through the location of so-called “support vectors”, the data points that most determine the geometry of said boundaries, in both classification and regression tasks.

Additionally, the application of the “kernel trick” allows for the implicit mapping of input features into higher dimensional spaces, without explicit computation of the transformation. When coupled with the right kernel function, the computation of hyperplanes remains efficient. The regularisation parameter, C , can be tuned to provide a suitable compromise between overfitting (a common failure of models built on sparse data) and the “smoothness” of decision boundaries. Additional preliminary dimensionality reduction techniques, such as PCA or PLS can be employed to further reduce the sparsity of the input space; Kaddour *et al.* demonstrate the effectiveness in employing PCA and PLS dimensionality reduction steps before the implementation of an SVM classifier for beef authentication using fluorescence spectroscopy data (Ait-Kaddour et al., 2018). This remains a decision dependent on the task at hand, as application of these techniques may compress and reduce the “resolution” of the data through disposal of features, or may affect the generalisability of subsequent models.

As support vector classifiers/regressors use distance-based metrics to examine the relationships between variables, such models may respond well to

standardisation of the data (fitting to a standard normal distribution), if the disparity between feature scales is dramatic (Luor, 2015).

Table 1-1: Examples of SVM kernel configurations in agri-food analyses

Kernel Name	Kernel Equation	
Linear Kernel	$k(x, y) = x^T y, x, y \in \mathbb{R}^d$	(Argyri et al., 2013) – Showed marginally inferior performance than RBF and polynomial kernels for microbial count prediction and freshness classification using Fourier-Transform Infrared (FTIR) and Raman spectroscopic data in minced beef samples stored under different temperature conditions.
Gaussian Kernel (Radial Basis Function - RBF)	$k(x, y) = \exp\left(\frac{-\ x - y\ ^2}{2\sigma^2}\right), x, y \in \mathbb{R}^d, \sigma > 0$	(Argyri et al., 2013) – Gave performance on par with polynomial kernel for microbial count prediction and freshness classification using FTIR and Raman spectroscopic data in minced beef samples stored under different temperature conditions. (Wu et al., 2008) – RBF with LS-SVM shows better performance for tea classification tasks than Linear LS-SVM, partial least squares regression, and RBF with Neural Networks. (Yin and Zhao, 2013) – SVMs, tuned with genetic algorithm (GA), showed the best performance with RBF for the classification of tomato plants showing early blight symptoms.
Sigmoid Kernel	$k(x, y) = \tanh(\gamma \cdot x^T y + r)$	(Argyri et al., 2013) – Gave poor performance for microbial count prediction and freshness classification using FTIR and Raman spectroscopic data in minced beef samples stored under different temperature conditions – results from validation phase were not carried forward to testing.
Polynomial Kernel	$k(x, y) = (x^T y + r)^n, x, y \in \mathbb{R}^d, r \geq 0, n \geq 1$	(Argyri et al., 2013) – Gave performance on par with RBF for microbial count prediction and freshness classification using FTIR and Raman spectroscopic data in minced beef samples stored under different temperature conditions. (Du and Sun, 2005) – When coupled with HSV colour space transformation,

computer vision-based food topping classification showed good performance versus linear and RBF kernels (Ren et al., 2021) – Black tea classification of quality using leaf shape performed using LS-SVM with polynomial kernel shows 12% higher correct discrimination rate (CDR) than polynomial SVM method (100% based on six shape features).

1.6.1.3 Deep Learning methods / Artificial Neural Networks

The development of neural network-derived methodologies and algorithms form a large portion of activity within the field of artificial intelligence. A wide variety of algorithms and model topologies constitute deep learning – as a consequence, only those methods that have been documented in the realm of food science will be discussed.

All neural networks share some conceptual similarities: each network is composed of many individual nodes, called “neurons”, which reflects their analogous functionality to their biological counterpart. In an abstract sense, a neuron will take an input, perform an operation with a weight and activation (transfer) function, and produce a transformed output (Huang, Kangas and Rasco, 2007). The effect of the weight on the input data dictates either a strong or weak overall response from the neural unit. Through the modulation of weights in the collection of neurons in the model, the overall response to input data can be configured. Neurons are arranged in layers, and different configurations of layers will achieve different objectives in the built model. Layers sandwiched between the input and output layers are known as “hidden” layers, as their internal functioning is obscured externally.

Common “hyperparameters”, or tunable parameters for neural networks are the activation function used, the number of neurons that make up each layer, and the “dropout” parameter, which will negate the contribution of varying numbers of neurons to the succeeding layer. This can help introduce sparsity into the models' representation of the problem space, reduce co-adaptation (the tendency for one

hidden unit to correlate with another, and thus not contribute independently), and may help improve the efficiency of the model, as the computational costs are reduced (Gerum et al., 2020; Srivastava et al., 2014). During the training process in the simplest neural network type, the multi-layer perceptron (MLP, consisting of an input layer, hidden layer and output layer), the weights can be progressively updated as data travels throughout the model, and passed backwards to update preceding layers (backpropagation), after each training epoch.

Some of the most popular and well-known families of deep learning neural network models are Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). CNNs are some of the most well-known examples of artificial neural network architecture, due to their suitability for semi-supervised image processing. They are able to undertake feature selection and engineering within the input data without any human intervention (Gu et al., 2018). CNNs notably differ from MLPs in that neurons share weights, whereas in MLPs, each neuron has an associated weight vector for the preceding neuronal inputs. This removes the need for separate, per-connection weight calculations, significantly reducing the training time and cost through the introduction of sparsity (Alzubaidi et al., 2021).

Recurrent Neural Networks (RNNs) are so called because of their ability to retain a “memory” of sorts, due to repetition of certain actions at the neuron level in a feedback loop. This is in contrast to the other generalised category of artificial neural networks, known as “feedforward” neural networks, so-called due to the flow of information throughout the model (forwards only). Their ability to repeat certain actions with inputs, and feed the resulting outputs into the weighted calculations for subsequent inputs lends the model type utility in the analysis of sequential data, in forecasting tasks for example, wherein multiple observations are passed into the model, and based on the trends in the input data, a predictive output is given for the next item in the progression (Schäfer and Zimmermann, 2006).

Table 1-2 summarises some applications of artificial neural networks applied to tasks in the agri-food domain. Common themes include the application of the

aforementioned CNN architecture to tasks involving image recognition. In this regard, recent literature has shown the efficacy of so-called “transfer learning”, wherein pre-trained Deep Learning (DL) architectures are applied to new research problems. The “depth” of DL refers to the high number of network layers forming the model, which typically require large quantities of high-quality training data to correctly assign model weights. As such, transfer learning may be useful in situations where training data is scarce, or computational training costs are high (Iman, Arabnia and Rasheed, 2023).

Table 1-2: Examples of artificial neural network configurations in agri-food analyses

Model Type	Configuration information	Task
Multilayer Perceptron	Provided by the R nnet package (Venables and Ripley, 2002).	(Spyrelli et al., 2021) - Application of multiple machine learning techniques using the SorfML Platform (Estelles-Lopez et al., 2017), for the prediction of microbial counts in chicken samples. Proved to be on par with LARS (Least Angle Regression) techniques.
Multilayer Perceptron	1,315 food items with 60 parameters from United States Food and Nutrient Database for Dietary Studies (FNDDS) used as input.	(Gorbachev et al., 2022) - Predicting the antiradical potential (ARP) of food ingredients - results revealed a positive correlation ($r = 0.539$) between dietary fibre and ARP and a weaker negative correlation ($r = -0.432$) between calorific content and ARP.
Convolutional Neural Network (Transfer Learning)	Various applications of pre-trained CNN; best results achieved with DenseNet201 architecture. 19,526 images of 6 fruit types with three quality grades ("good", "bad", "mixed") from FruitNet database (Meshram and Patil, 2022).	(Morshed et al., 2022) - Automated visual quality classification of fruits – results showed good classification performance with 99.67% accuracy.

1.6.2 Preparation of Sensor Data for Food Analysis

A common task in the interpretation of data emerging from usage of information-age sensor technologies, such as multi-spectral imaging and other spectroscopic analyses, is the simplification of its depth and breadth. This process is typically highly context-dependent, and thus an innumerable list exists of combinations of possible outlier detection, dimensionality reduction, transformation, and feature extraction/engineering steps that could be taken before the processed data is passed into the predictor model construction and subsequent utilisation steps.

The following subsections are intended to give a working understanding of some of these regularly encountered techniques, and some examples of their application in the food analysis domain.

1.6.2.1 Outlier Detection

A common necessity in the processing of sensor data generated in bioimaging and chemometric research is the detection of outliers; that is, sample data that differs from the general trends in the data by such a degree that may introduce unintended tendencies in any predictive modelling conducted.

As above, the specifics of selection of outlier detection algorithms is dependent on the task at hand, however the detection of outliers from the types of sensors described in this document is likely complicated by the complexity and dimensionality of the data, for example, spectral data emerging from a spectrometer may represent thousands of wavelengths, and thus the trends of each feature of the dataset in isolation may not represent the trends of the observation as a whole, meaning that rather than dropping samples that exhibit isolated outlying values across a spectrum, analysis of a dataset may be better served by treating features with a high number of outliers or otherwise erroneous data differently, either through subsetting of columns or imputation, the technique of inserting synthetic values based on values held by similar samples in the dataset.

A commonly used property in the detection of outliers within spectral data is that of Mahalanobis distances (Mahalanobis, 1936). Calculation of the Mahalanobis

Distance of a sample within a dataset generates a metric to assist in determination of distance from a distribution for data displaying covariance, i.e. variance between multiple variables (features) in a dataset. The general formula for the calculation of the Mahalanobis distance for a multidimensional dataset is given by Equation 1,

$$D^2 = \sqrt{((x - m)^T \cdot C^{-1} \cdot (x - m))} \quad \text{Equation 1 General equation for the Mahalanobis Distance metric}$$

where **D** is the Mahalanobis Distance, **x** is the vector of the observation (the sample row), **m** is the mean values of each variable (mean of each column in the dataset), and **C**⁻¹ is the inverse of the covariance matrix calculated for the matrix of independent variables (each column of the dataset).

In some cases, the calculated distances are compared with a Chi-squared distribution, with the number of degrees of freedom set to the number of columns of data, and a significance of 0.95. Any observation (row) that fails to meet the threshold is then considered an outlier (Hershberger et al., 2021).

Component Analytical methods are often employed for exploratory data analysis (EDA) and outlier detection, the most widely used likely being PCA (Principal Component Analysis). PCA works by transforming observations into a new set of uncorrelated variables (principal components). Outliers are then more easily detectable by projecting the observations into the transformed space, the axes representative of decreasing degrees of variance in the data. Clustered information can be indicative of commonalities in the data, whereas isolated values may indicate outlying data, depending on the context.

A variety of non-parametric methods exist for outlier detection; non-parametric methods differ from parametric methods in that the former do not make assumptions about the underlying distribution of the data, whereas the latter do.

Isolation forests are sometimes employed for the detection of outliers in spectral data, and in novelty detection (detection of outlying data in new data, based on existing model experience). They are a specialised implementation of the Decision Tree concept, and do not make use of target data to calculate leaf node

statistics for branching, instead relying solely on path-length approximation (Liu, Ting and Zhou, 2012), making it an unsupervised machine-learning method. The same paper demonstrates the method's efficiency in dealing with high dimensional data, due to its linear time complexity characteristics.

1.6.2.2 Correction/Smoothing Techniques

Another step in the pre-processing pipeline that may prove useful for highly dimensional sensor data is accounting for other irregularities often caused by noise in the data capture process, an inevitability in many experimental situations due to the manner in which many sensors function. Such technologies are often optimised for the maximisation of the signal-to-noise ratio (SNR), in essence the amount of meaningful data captured by a sensor versus the amount of random noise generated in the recording process.

Techniques for “smoothing” the data, with the intention of reducing the effects that the noise may have on the general trends of the data are numerous. The most mathematically simple is the moving average. In this case, a window is considered, moving across an individual observation (row) within the dataset – a simple average for each value x with window size n is taken at each point within this window, with the intention of smoothing out any irregularities within that area.

$$SMA_t = \frac{x_{(t-n)} + \dots + x_t + \dots + x_{(t+n)}}{2n + 1}$$

Equation 2 Generalised Formula for Simple Moving Average

However, a problem with this method is its sensitivity to outliers; that is, all values within the moving window are given equal weight, regardless of their concordance with general trends in the spectrum.

A commonly utilised alternative in the field of signals processing and chemometric study is Savitzky-Golay Filtering (Savitzky and Golay, 1964). First described in the 1960s, the method works by fitting polynomial functions to local areas within the spectrum via the use of linear least squares. This is a comparatively complex method, overcoming the issues described with moving averages. The polynomial-fitting approach is more likely to capture more of the general trends of the spectrum, especially when dealing with dramatic changes in shape, and less likely to flatten out important peaks and valleys in the data (Isnanto, 2011).

Another approach sometimes explored in the processing of spectral data is Standard Normal Variate (SNV) filtering. The motivations for this method within spectral analysis are typically to account for multiplicative effects occurring between spectra and baseline shifts, typically as the result of technical variability in the conduction of the experimentation (Barnes, Dhanoa and Lister, 1989). In Barnes *et al.*'s original documentation, the motivation for developing the method was to reduce the multiplicative effects of scattering and particle size in the analysis of hay samples in diffuse reflectance NIR spectroscopy, and reduce differences in the global intensities of the signals (Zeaiter and Rutledge, 2009). Each spectrum is centred and then scaled by dividing by its standard deviation. This approach is applied row-wise across each spectrum, where $\mathbf{X}$ is the original spectrum, $\hat{\mathbf{X}}$ is the mean of the spectrum, and $\mathbf{s}$ is the standard deviation of the spectrum:

$$X_{SNV} = \frac{X - \hat{X}}{s}$$

Equation 3 Generalised Formula For SNV

1.7 Sensor/Data Fusion

Data fusion is an umbrella term for a variety of data-integrative approaches for combining sources of information, for the purposes of improvement of explanatory power. This process can take place at multiple levels: consolidation of the raw data directly from the source, use of features selected or engineered from raw data, or fusion of constructs generated from the data, such as machine learning decisions. A widely accepted definition of data fusion was provided by the Joint Directors of Laboratories/Data Fusion Information Group (JDL/DFIG), intended to formalise some of the order of operations for conducting a data fusion process (White, 1987). However, this offers limited practical information on how to design solutions for domain-specific scenarios that involve integrating data from multiple types of minimally invasive sensors.

Hall and Llinas (Hall and Llinas, 1997) give the description of sensor-based data fusion as *“data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences than could be achieved by the use of a single sensor alone.”*

The particulars of sensor fusion are highly dependent on the particulars of the devices used in the sample analysis, as a great deal of heterogeneity exists in the space, due to a wealth of sensor types, models, and configurations that complicate the process of facilitating the union of useful data. For example, the analysis of data emerging from the sensor types described in Section 1.8 typically require an intermediary step of preprocessing before becoming useful in any subsequent machine learning analysis pipeline, and configuration of these steps are likely specific to individual sensor modalities.

Most sources are in agreement into the general categorisation of sensor fusion approaches, based on the stage at which the integration occurs.

1.7.1.1 Early/Data-level fusion

Low Level (Early) Fusion: Typically resulting from the fusion of sensor data at the dataset level, i.e. with minimal or no preprocessing undertaken with the data. This type of fusion approach is often considered to be the easiest to implement, as it

is a case of concatenating data into one data matrix representative of all modalities - Figure 1-2 shows a general topology of the early fusion approach. Smolinska *et al.* (Smolinska et al., 2019) describe the fact that the observations are in the shared mode, and describe how common unsupervised exploratory data analytical methods may not function well when data matrices are representative of heterogenous data – for example, a Mass-Spectrometry dataset may contain 3000 variables or features, and a prepared Multispectral imaging (MSI) dataset may only contain 15-30. In this extreme case, many methods, such as the typical implementations of Principal Component Analysis, will not take into account that the data originated from different sources, and thus will apply an even weighting in the calculation of the eigenvectors of the matrix. As a consequence, data integrative modifications of the PCA algorithm, such as SUM-PCA, described by Smilde *et al.* (Van Deun et al., 2009), can be used for the explorations of two or more coupled data matrices. This method aims to find linear combinations of variables in the concatenated matrix that describe the largest amount of variance between the observations (Smolinska et al., 2019; Van Deun et al., 2009).

However, methods like SUM-PCA are not explicitly aware of the fact that the matrices originate from different sources and are potentially equally impaired experimentally. Such methods may even disregard the contributions of the subsequent matrices if the covariance in the first matrix is far larger than the others. A block-level scaling factor can be applied, before the matrices are concatenated, that seeks to account for these differences. How best to scale the matrices is still an open problem (Smolinska et al., 2019), however several common methods are used; in some situations, each matrix is scaled by its Frobenius norm, meaning that the total sum of squares equals one (Smilde et al., 2005), prior to concatenation. Alternatively, a maximum likelihood approach has been suggested, for determination of the best scaling factors (Wilderjans et al., 2011).

Smolinska *et al.* emphasise that often the aim of unsupervised analysis in data fusion tasks is not to assess the maximum variance in the combined dataset, but

to express the relationship between the constituent matrices in terms of shared and unique information (Acar et al., 2014; Smilde et al., 2017; Smolinska et al., 2019). Methods such as canonical correlation analysis have shown utility in this regard, wherein the algorithm is used to identify linear combinations of the variables of each matrix that are strongly correlated to the other matrix (or matrices) (Glahn, 1968; Tenenhaus and Tenenhaus, 2014). This strong correlation indicates commonalities in the data that may be indicative of redundancies or collinearity for later analyses, if not sufficiently handled.

Within the context of food safety, quality and authenticity analysis, a number of examples of early fusion can be found within published research. A selection of published examples with a well-described methodology for conducting the fusion process is given in Table 1-3.

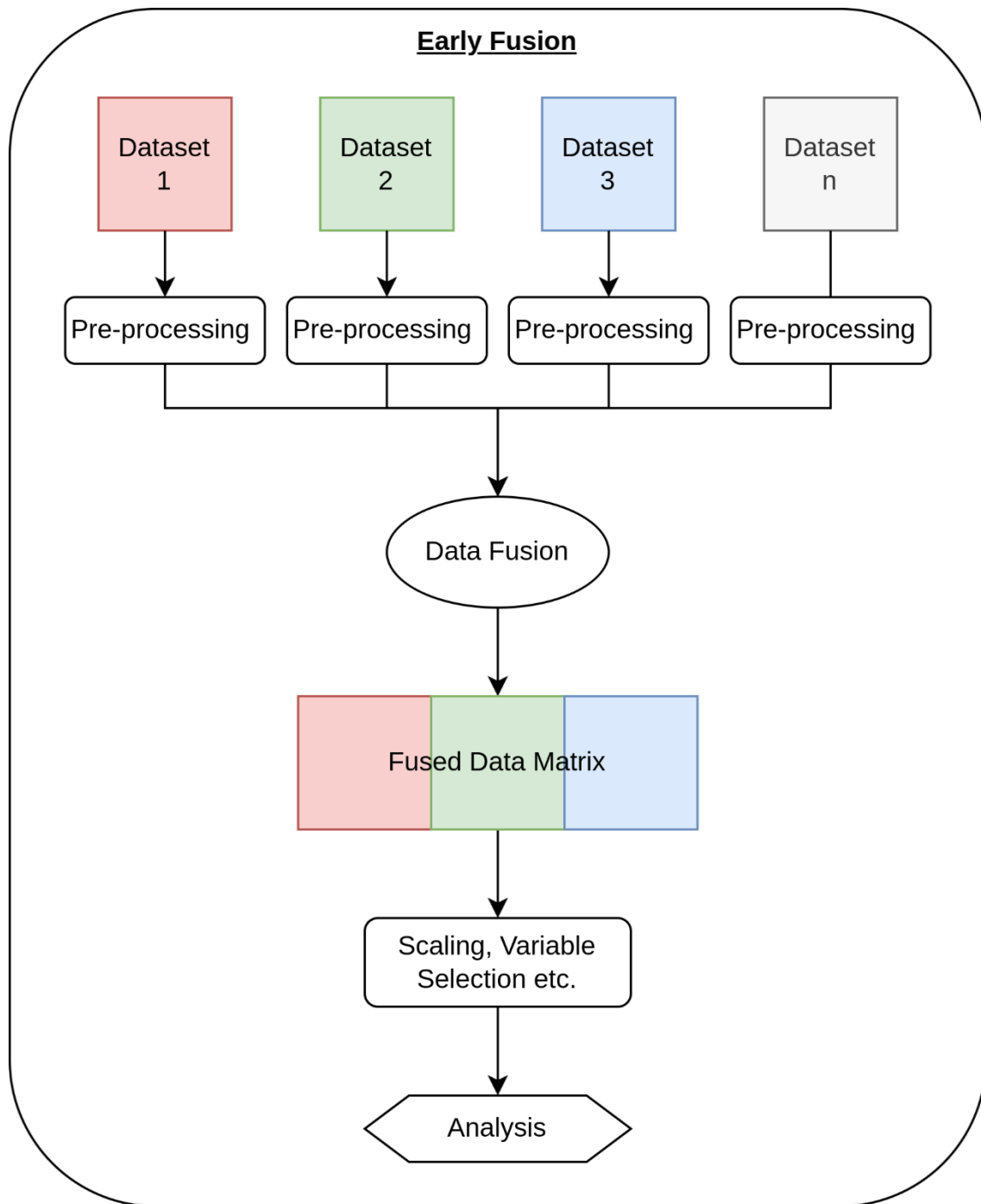


Figure 1-2: Generalised topology of an Early/Data-level fusion strategy, typically resulting in the concatenation of data from multiple sources (and modalities)

Table 1-3: Table of references describing early fusion methods used in agri-food spoilage or authenticity analyses.

Approach	Application	Outcome
Fusion of near Infra-red (NIR), mid-range Infra-red (MIR) and Raman Spectroscopy features as input into PLS-DA model.	Species-specific discrimination of meat and bone meal (MBM).	Significant improvement of discrimination of origin species from bone fragment and meat fragment components of MBM (Gao et al., 2024).
Fusion of near Infra-red (NIR) and electronic nose (E-nose) data as input into support vector machine (SVM) and backpropagation neural network (BPNN) models	Detection of adulterated lamb	Significant decrease in accuracy by attempting an early fusion approach (less than 70% accuracy for both SVM and BPNN models, versus more than 90% accuracy for single-sensor model types) (Jia, Qin and Zhao, 2024).
Fusion of multispectral imaging (MSI), Fourier-transform Infra-red (FTIR) and Visible light Spectroscopy	Total viable count (TVC) estimation in pork mince	Fusion model offered marginally improved test performance over FTIR, the best performing single sensor, with an R2 of 0.862 versus 0.834, and an RMSE of 0.836 log CFU/g versus 0.915 log CFU/g (Fengou et al., 2019).

1.7.1.2 Mid/Feature-level Fusion

Mid- or Feature-level fusion is conducted through the concatenation of data matrices that have been subject to some type of preparatory feature engineering or dimensionality reduction steps. Due to the large array of algorithms that may be employed in this regard prior to the fusion step, there are a variety of implementations of this strategy that have been described in scientific literature.

The general motivation for employing mid-level fusion is to address the problems during data integration that are described in Section 1.7.1.1: Early/Data-level fusion, namely regarding the nature of the data types being explored. Many of the sensor types employed in modern chemometric research are capable of outputting data in a highly collinear fashion, which can impact the effectiveness of some machine learning algorithms and the ability to combine the information from multiple sensor sources, if there is a great disparity in dimensionality between the respective data matrices.

The dimensionality reduction algorithms applied to each data matrix need not match; common examples include PCA and PLS. After the application of the selected algorithm(s), a number of components of the transformed data are retained and fused in a similar manner to early fusion. Smolinska *et al.* note that component selection after the dimensionality reduction step typically results in a similar number contributed by each data matrix, thus avoiding the incidence of block scaling issues (Smolinska et al., 2019). Figure 1-3 shows the general topology of this fusion approach.

There are a number of published applications of mid-level sensor data fusion within the agri-food sector focused on safety, quality and authenticity analysis. Much *et al.* compared mid-level and low-level UV-Vis and NIR spectroscopy fusion approaches for the detection of adulteration of “high-quality” vegetable oils (olive, rapeseed, flaxseed and chia oils) with either sunflower, corn or soya bean oil (Much et al., 2024). Their early-fusion approach produced better results than the mid-level approach; a decrease in RMSE of between 0.6-0.1 % v/v contamination in all five adulteration scenarios was observed. The authors discuss this performance disparity, a possible reason being that the unsupervised

nature of the PCA dimensionality reduction step taken on the sensor data may have caused loss of data otherwise relevant to the subsequent construction of the PLS regression model on the fused dataset. The use of PLS for dimensionality reduction may have retained more of this relevant information, as the algorithm attempts to maximise the variance with regard to the predictors and the target variable.

On the other hand, Cevoli *et al.* recorded mid-level fusion of Fourier-Transform Infrared (FTIR) spectroscopy and Vis/NIR hyperspectral imaging as showing superior performance over early fusion in quality parameter prediction of yellow-flesh kiwifruit (Cevoli *et al.*, 2024). The authors employed both PCA and PLS dimensionality reduction steps in two different mid-level approaches, prior to PLS regression in both. The PLS-PLS regression approach outperformed the PCA-PLS approach, showing a decrease in RMSE test performance of between 20-40% in prediction of soluble solid content, flesh firmness and hue, and dry matter metrics.

In the specific field of meat spoilage and quality assessment, Robert *et al.* combined Raman and Infrared spectroscopy for the prediction of red meat (beef, lamb, and venison) quality, using pH and the percentage of intra-muscular fat (% IMF) as metrics (Robert *et al.*, 2021). The authors compared the performance of early-, mid- and late-level data fusion strategies. Their mid-level approaches employed either multivariate curve resolution (MCR) or PCA in the feature extraction/dimensionality reduction roles, and the resultant selected principal components were fused into a matrix for the construction of PLS regression models. Both mid-level approaches were substantially outperformed by either the early or late fusion methods, with the early fusion approach showing performance parity with the best performing single sensor model for the prediction of percentage of intra-muscular fat (a PLS model built with Raman data), with RMSEP percentages matching at 8.47%. The late fusion approach showed the best results for prediction of pH values and exceeding the single best performing sensor (Raman) with an RMSEP of 12.9% versus 15.9%. Similarly to Cevoli *et al.* (Cevoli *et al.*, 2024), the authors suggest that the under-performance of the

mid-level approach may have been attributable to the dimensionality reduction process, specifically information loss in the selection of features to concatenate. The authors note that simplification of the mid-level fusion models by reducing the number of included features did not yield improved results.

The common theme in the discussions of both of the poorly performing mid-level fusion applications is the nature of the dimensionality reduction step and the possible loss of relevant statistical information. This lost information may otherwise be useful when combined with the other dimensionally-reduced matrices. Geurts *et al.* acknowledged this possibility and accounted for it with the concept of “sustainable mid-level fusion” (Geurts et al., 2016). In their fusion approach, dimensionality reduction is conducted on the data matrices as per the typical method, however an additional step is coordinated with the intention of “recycling” the lost information when in the context presented by the other dimensionally-reduced data matrices.

As an example, two sensor data matrices, X_1 and X_2 , are dimensionally reduced, producing T_1 and T_2 .

Residual matrices are then calculated for each, leaving E_1 and E_2 , representing the information within X_1 and X_2 that was not captured in T_1 and T_2 , using the general form in Equation 4.

$$E_1 = X_1 - \hat{X}_1$$

Equation 4 Calculation of residual matrix E_1 , where $\hat{X}_1$ is the inverse transform (reconstruction) of X_1

Then, PLS regression models are built:

- 1) with E_1 as the training data, and the columns of T_2 as the target.
(number of PLS models equal to number of features in T_2)
- 2) with E_2 as the training data, and the columns of T_1 as the target.
(number of PLS models equal to number of features in T_1)

Predictions are made by these models for every column (feature) in T_1 and T_2 , and concatenated into matrices A_1 and A_2 , regressing the residual matrices E_2 and E_1 respectively. The “A” matrices are designed to capture how the “discarded” information present in the residuals relates to the other datasets’ contributed features. T_1 and T_2 are concatenated with A_1 and A_2 , and these are used as features, as before; in the authors’ example scenarios, PLS regression and discriminant analysis (classification) were used. Using this “sustainable fusion” approach, the authors were able to demonstrate an improvement of ~3% for RMSEP performance in their beer bitterness scenario, which the authors determined as statistically significant in that circumstance. This methodology shows merit, due to its compromise between early and mid- level fusion. Whilst early/low level fusion is able to produce models that have full awareness of the so-called “multivariate” advantage (Olivieri, 2008), they often suffer with poor performance when not accounting for differences in the dimensionality of input data. For the reasons stated previously, mid-level fusion addresses these issues, with a scaling and dimensionality-reducing action, but trades away potentially useful information in this step. The “recycling” action described here is potentially able to recover this information, whilst retaining the benefits of early fusion approaches.

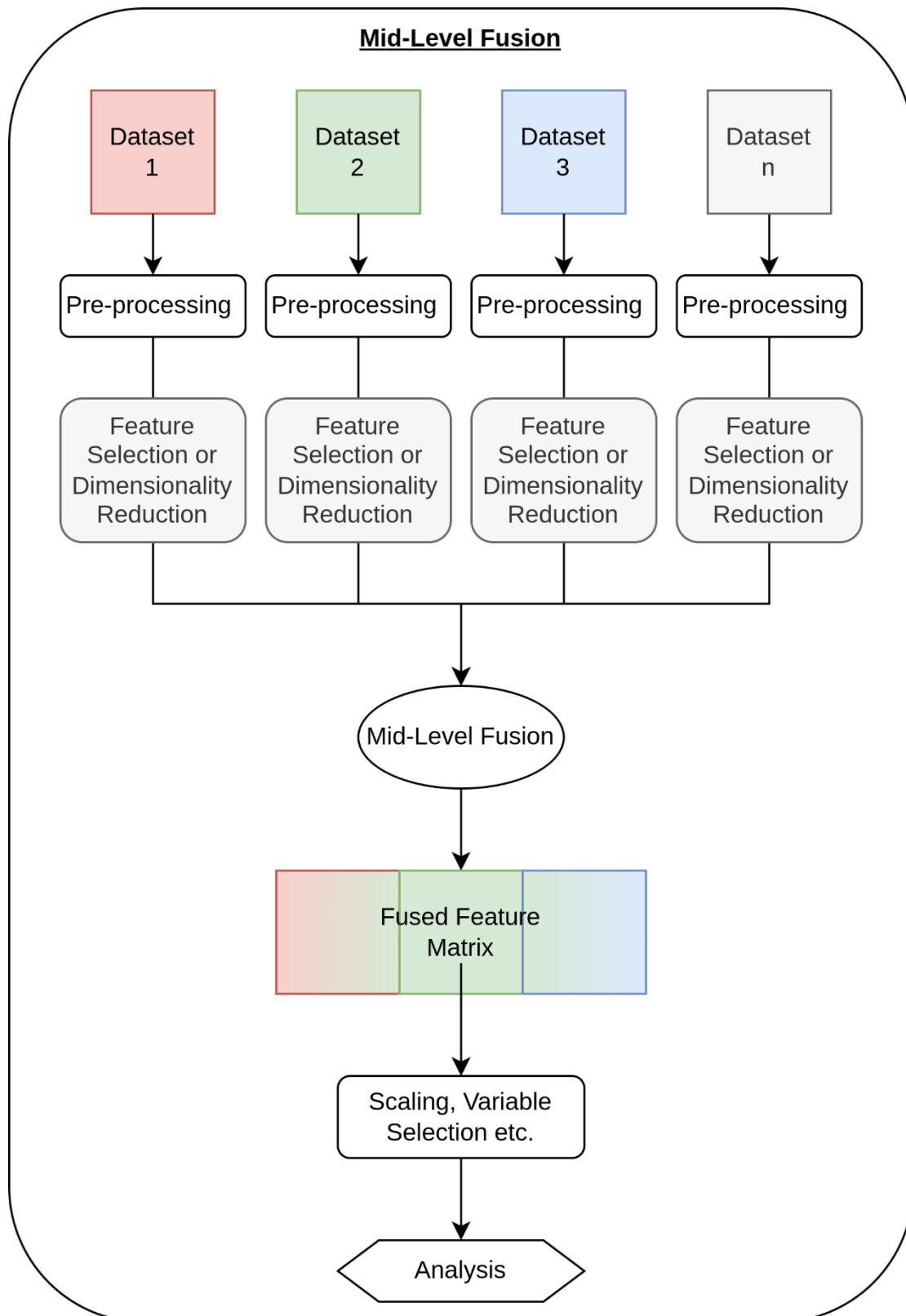


Figure 1-3: Generalised topology of a Mid/Feature-level fusion strategy, typically resulting in the concatenation of features from multiple sources (and modalities)

1.7.1.3 Late/Decision-level Fusion

As the name suggests, decision fusion occurs through concatenation of statistical outcomes formed about the specific input matrices. These decisions are often predictions made by individual machine learning models that independently take a single data matrix as input. These concatenated decisions are then acted upon in some way; the specific methodology for the formulation of overall decisions is context dependent, with various approaches finding more application to either classification or regression problems. The general topology of this approach is described in Figure 1-4.

Late fusion is well equipped to handle the dimensionality and scaling issues presented by the fusion of chemometric data, through the independent handling of datasets by models configured for those types. Multiple models can be fitted for the same sensor types, and decisions can be made about that sensor's data in an ensemble fashion, from for example, a selection of linear and non-linear methods, such as PLS, decision trees, random forests or neural networks. Late-fusion architectures can be considered to trade a good deal of biological interpretability and explainability for the pursuit of accuracy however, as no observations about the latent structure of the data can be made at the fusion level (Smolinska et al., 2019).

The simplest strategy with which to fuse the predictions of the models is through averaging of the results, however the use of this approach assumes that all of the constituent predictions are equally important. This approach can be modified to account for this by weighting the predictions of certain models in the final calculation; these weights would require determination through a cross-validation procedure of some kind. For classification tasks for two or more member models, a majority voting approach can be used, to determine the most “popular” decision for a particular observation (Breiman, 2001).

Robert *et al.*, as described in Section 1.7.1.2: Mid/Feature-level Fusion assessed the performance of implementations of the three fusion-level archetypes for the quality analysis of beef (Robert et al., 2021). For their late fusion approach, they fused the results from PLS regression models built with Raman and FTIR

spectroscopy through a central limit theory (CLT) influenced strategy (Fourati, H., 2017; Salina and Malathi, 2014), which takes the standard deviation of the predictions by the individual sensors as the weighting for the fusion process.

$$\sigma_{RaIR}^2 = [\sigma_{Ra}^{-2} + \sigma_{IR}^{-2}]^{-1}$$

Equation 5 Calculation of the standard deviation of the fused predictions in Robert *et al.*'s late fusion approach, where $\sigma_{RaIR/Ra/IR}$ represents the respective standard deviations of the fused data, Raman and FTIR sensors.

$$X_{RaIR} = \sigma_{RaIR}^2 [\bar{X}_{Ra} \sigma_{Ra}^{-2} + \bar{X}_{IR} \sigma_{IR}^{-2}]$$

Equation 6 Calculation of fused prediction in Robert *et al.*'s late fusion approach, where $\bar{X}_{RaIR}$ represents the fused average, $\bar{X}_{Ra}$ and $\bar{X}_{IR}$ represent the Raman and FTIR sensor averages, and $\sigma_{RaIR/Ra/IR}$ represents their respective standard deviations.

The authors demonstrated the strategy as showing improvements in performance over the single sensor models and the other fusion approaches for the prediction of sample pH, an important indicator of meat quality, typically declining *post mortem* due to glycolytic activity and an accumulation of lactate (Ferguson and Gerrard, 2014).

Other strategies more complex than averaging have been explored and published, including those based on Bayesian inference (Roussel *et al.*, 2003), while more contemporary approaches may draw on ensemble learning techniques. Stacked generalisation, or “stacking” is a machine learning ensemble strategy that utilises a second “meta-learning” model layer to form predictions based on the pooled decision matrix from the first layer “expert models”, regressing it against the target variable. The intention behind the construction of this second layer is for it to learn the performance relationship between the model

types and the target variable. Selection of a suitable second layer algorithm may allow for prioritisation of better performing models and lessen lower-performing model's contributions towards the overall output of the ensemble.

The vocabulary used to describe specific variations of the practice is inconsistent: the general theory has been discussed in the literature over a period longer than two decades (Breiman, 1996), however the modern literature typically refers to the “Super Learner” approach described by Van der Laan *et al.* in their 2007 paper (Laan, Polley and Hubbard, 2007). This method introduces a cross-validation step in the construction of the meta-learning layer, as it is not advisable to train the meta-learning layer on the same data partition as that used to train the expert models (first layer). This is due to the likelihood of these first-layer models producing unrealistically accurate predictions from the same data upon which they were trained, thus causing the meta-learner to learn an inflated view of their performance.

In order to prevent overfitting in the Super Learner approach, out-of-fold predictions are made by training and making predictions with the first-layer model configurations using a cross-validation approach, such as k-fold or leave-one-out, the predictions forming the training set used by the second layer. This in effect synthesises a second training set, and allows the meta-learner to train on the performance trends of the first-layer models and account for discrepancies in their performance, in addition to learning the best performing combination of first-layer models in the regression/classification task.

An alternative to the cross-validation approach used in super-learning is that used in “blending”, wherein a small validation set is set apart from the training data, and used to train the meta-learner in a hold-out fashion. This method is more appropriate when large quantities of training data are available, to allow the meta-learner to generalise well.

Although the authors of the super learner specification demonstrated its applicability to regression and classification tasks within their originating publication (Laan, Polley and Hubbard, 2007), stacking has not seen a large degree of utilisation in the chemometric food analysis domain. Cheng *et al.*

demonstrated a variant of the stacking approach for fusion of hyperspectral imaging and electronic nose data for the prediction of pork moisture content, their decision fusion approach achieving an R^2 in prediction of 0.9492 versus their best performing imaging feature model that achieved an R^2 of 0.9016 (Cheng et al., 2022).

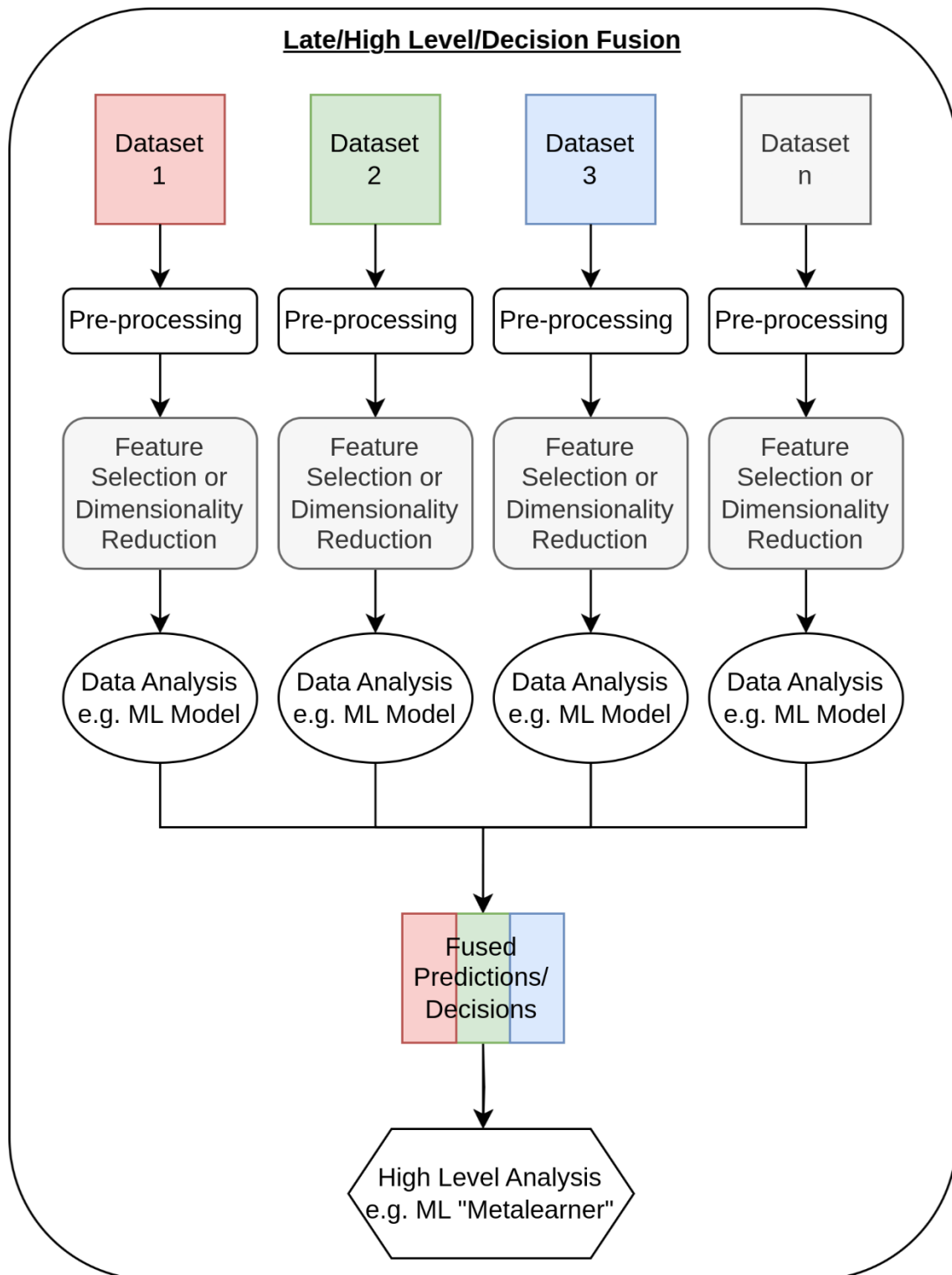


Figure 1-4: Generalised topology of a Late/Decision-level fusion strategy, typically resulting in the concatenation of decisions from multiple sources (and modalities)

1.8 Modern Sensor Technologies

Sensor-enabled analysis already forms the basis for many supply chain compliance protocols in the agri-food sector - a wide variety of in-laboratory sensor types have found applications and niches supporting food safety and authentication tasks, in many task-dependent forms.

1.8.1 Overview of Chromatography Techniques

A variety of chromatographic techniques exist, such as gas chromatography (GC) (Lehotay and Hajšlová, 2002; Tiedink et al., 1990) and high-performance liquid chromatography (HPLC) (Salghi, Armbruster and Schwack, 2014; Tsuzuki and Ushida, 2009; Velkoska-Markovska, Petanovska-Ilievska and Markovski, 2018). These differ in the nature of their mobile and stationary phases – a high-level overview follows: in the case of GC use of an inert gas as the mobile phase (e.g. helium) and volatile or otherwise vaporised samples are passed through the column containing a typically solid/packed stationary phase, whereas HPLC makes use of an appropriate solvent for the mobile phase and a variety of gels or other bonded/solid-supported phases for the stationary phase. The intention is that fractionation of the sample occurs due to varying affinities for the stationary phase, and as such each constituent component of the sample passes through and exits the machine in sequence. The exit of constituents is recorded by a sensor appropriate for the specific task, non-exhaustively including detectors for flame ionization, thermal conductivity or electron capture for GC and UV-Visible Light/Fluorescence detectors for HPLC.

The processed data is in most cases representative of detected signal intensity over retention time, and as such the area under the curve is proportional to the concentration of the corresponding constituents recorded in the sample. Detectors can be grouped into the modalities of *Ionisation*, the excitation of compounds to produce ions, *bulk physical property*, or *optical*, measuring the eluted compound's interaction with light (Marriott, 2004).

In the case of gas chromatography ionisation detection, a common method of analyte detection and discrimination is through the use of flame ionisation

detectors (FID) (Stez, 2015). This is a destructive method; upon elution, compounds are burnt within a hydrogen flame and a potential is applied across it. A current is produced, which is proportional to the change in organic ionic composition of the combustion products (Hage, 2018). The detector's destructive nature limits the possibility of further analytical steps, however the wide linear analytical response, high sensitivity and specificity for detection of organic compounds ensures its commonplace status in food analysis. Recent noteworthy applications of the detector include in food nutrition, for the quantification of galactose content in a spread of Korean agricultural products as a reference for patients with galactosemia, where previous HPLC analyses had shown low sensitivity in this regard (Jeong et al., 2023).

A detector with a similar mode of operation is the Nitrogen-Phosphorus Detector (NPD), a name indicative of the detector's specificity, and is also known as a Thermionic Selective Detector (TSD). The nitrogen-phosphorus specificity lends utility in detection of biological components, such as nucleic acids and bases (Hage, 2018; Hinshaw, 1992). The NPD device has been used for the quantification of the carcinogen acrylamide in fried potato using ethyl acetate as a solvent; acrylamide was quantifiable in concentrations between 0.5-100 ppm and a recovery rate of $106 \pm 8\%$ was achieved; Kim *et al.* suggested that such a method may have utility in the food industry as a screening tool in the food industry prior to the use of more advanced/expensive methods (Kim, Hwang and Lee, 2011).

Electron Capture Detection (ECD) is sensitive to compounds containing electronegative groups, such as halogens or nitro-groups. It detects analytes by measuring the capture of electrons emitted from a radioactive source. Although highly selective, limitations such as incompatibility with oxygen and water, as well as the radioactive element involved, may restrict its widespread application in food analysis. However, gas chromatography with ECD has been applied to tasks involving organochlorine pesticide residue detection in water (Chowdhury et al., 2024).

The Thermal Conductivity Detector (TCD) detects compounds based on their thermal conductivity, measuring resistance changes in a heated filament when analytes pass through (Marriott, 2004). It is versatile and non-destructive, detecting a wide range of substances including non-ionic compounds and water. However, it has relatively low sensitivity and specificity, potentially complicating data analysis due to background noise (Hage, 2018). Another non-destructive sensor is the Photo-ionization Detector (PID), which uses ultraviolet light to ionise analytes, especially effective for aromatic and alkene compounds due to their π -electron structures. Its non-destructive nature allows integration upstream of other detectors like the FID, providing additional analytical insights without loss of sample sensitivity resulting from splitting the gas stream between devices (Marriott, 2004).

Liquid chromatography typically employs optical detection methods, predominantly absorbance and fluorescence (Smith, 2004). Absorbance detectors measure how compounds absorb UV or visible light, widely used due to their versatility in detecting aromatic and unsaturated organic compounds, however a common necessity with such detectors is the usage of solvents, buffers and other spectator substances that do not exhibit properties that interfere with the measurements taking place (Hage, 2018). HPLC with ultraviolet absorption detection (HPLC-UV) has been used to quantify adulteration in edible oil samples, the results proving highly satisfactory for detection in monovarietal extra-virgin olive oils, showing an overall error of 2.9% with adulteration proportions as low as 2.5% when coupled with Partial Least Squares (PLS) modelling (Carranco et al., 2018).

Fluorescence detectors offer enhanced sensitivity for compounds exhibiting or derivatised to exhibit fluorescent properties (Lunn and Hellwig, 1998). Such apparatus has been used for the detection and quantification of Nitrofurans, an antibacterial feed additive with carcinogenic and teratogenic risks, in meat products including fish, chicken, and pork (Sheng et al., 2013; Wang et al., 2022).

Less sensitive methods such as refractive index detection, used mainly for lipids and carbohydrates, have largely been replaced by Light Scattering Detection

(LSD) (Smith, 2004), which identifies non-volatile components through scattered light. Applications within food analysis include the detection of lipids, sugars and other compounds that are not readily detected with the preceding methods. Examples include rapid quantification of sugar types in fruit juice (Benvenuti and Burgess, 2012), and the determination/quantification of polysorbate 80 content in therapeutic protein formulations – In the latter example, a strong linear relationship was observed between log-transformed detector response and log-transformed polysorbate 80 concentration (Nayak et al., 2012).

In some cases, gas or liquid chromatography's discriminatory facilities can be supplemented with the addition of spectrometric techniques. After leaving the chromatograph's column, the analyte can be passed into a mass spectrometer, which can be used to identify the species within a particular sample with greater resolving power than the spectrograph in isolation, in a process abbreviated to GC-MS or LC-MS, depending on the mobile phase used. Due to their versatility, these combinations of devices have been used in a variety of agri-food applications. LC-MS has been used for pollutant risk assessment studies in fish in South African seas, for the detection of pharmaceuticals, perfluoroalkyl and endocrine disrupting compounds; a high health and environmental risk was determined, necessitating intervention (Ojemaye and Petrik, 2019). Previous work by Norli, Christiansen and Deribe had employed GC-MS for the quantification of organic pollutants in fish, including organochlorine pesticides, with the QuEChERS (Quick, Easy, Cheap, Effective, Rugged and Safe) method, originally developed by Anastassiades *et al.* for the extraction and quantification of pesticides in fruit and vegetables, also using GC-MS (Anastassiades et al., 2003; Norli, Christiansen and Deribe, 2011).

Whilst these types of sensors are typically associated with end-stage/destructive laboratory compliance testing of samples due to their typically high cost, low portability, and complex sample preparatory steps, they have found certain niche uses in non-destructive testing of products throughout the food production and distribution chain. For example, vaporous components ("offgassing") produced during fermentation processes can be monitored using "headspace sampling", in

which the sensor does not require direct interface with the sample. Further usages include authenticity testing of meat products via analysis of volatile marker compounds expected from slaughtered animals fed on approved diets (Segura-Borrego et al., 2022). Headspace solid-phase microextraction combined with gas chromatography/mass spectrometry (HS/SPME–GC/MS) has also been used to monitor spoilage in minced beef stored in different packaging (air and modified atmosphere packaging, MAP) and temperature conditions (0, 5, 10, and 15 °C). Chemometric techniques correlated the results with microbial counts and sensory scores, identifying specific compounds (such as 2-pentanone, ethanol, diacetyl, and ethyl esters) as potential indicators of spoilage. PLS models built with the data were able to accurately estimate microbial populations and sensory qualities in both conditions, with 91.8% of predictions within $\pm 20\%$ error for total viable counts (TVC) (Argyri et al., 2015).

1.8.2 Overview of Spectroscopic Sensor Technologies

Spectroscopic sensor technologies are useful components in the modern food analysis suite. In an abstracted sense, spectroscopic techniques record the interaction between matter i.e. the analyte and absorbed, transmitted and/or radiated photonic energy. These sensor types have shown great utility in analysing quality, authenticity, and safety parameters of food over several decades (Al-Jowder, Kemsley and Wilson, 2002; Guzmán et al., 2015; Kyriakidis and Skarkalis, 2000).

A feature that distinguishes some, but not all members of the spectroscopic family of sensors from some other chemical analytical techniques is the non-destructive nature of the study, meaning that samples are not destroyed through the operation of the device. This offers flexibility to the observer, as analysis could feasibly be conducted either with a single spectroscopic sensor type, or with multiple sensor modalities working in an ensemble fashion, without great concern about small quantities of analyte to sample from. Furthermore, modern portable adaptations of the technology could also be employed in-, on- or at-line in a food production and distribution chain, and thus sampled individuals could be returned to the food chain intact and untainted (Dietrich, 2018).

Following is a short summary of some of the most prevalent categories of spectroscopic sensor currently exploited in the food analysis sector.

1.8.2.1 Ultraviolet – Visible Light Spectroscopy (UV-Vis)

Ultraviolet – Visible Light spectroscopy has found use in a variety of food authenticity tasks. It functions according the principle of the Beer-Lambert Law:

$$A = \log(I_0/I) = \epsilon cl \quad \text{Equation 7 The Beer-Lambert Law}$$

Where **A** is the measured absorbance of the sample, equal to the product of ϵ , the molar concentration coefficient at a specific wavelength of light, **c** is the concentration of the sample, and **l** is the length of the path through which the light travels in the sample. This is equivalent to the log value of the ratio of the incident light intensity I_0 , to the intensity of light transmitted through the sample **I**.

Noteworthy use cases of such sensors are those involved in the analysis of food quality and authenticity. Often, this revolves around the quantitative measurement of the “colour” of the sample. For example, in the authenticity analysis of food oils, an important factor of interest is the presence of anthocyanins and carotenoids in the sample (Giusti and Wrolstad, 2001). Carotenoids are a large family of molecules, containing plant pigments, amongst others. These plant pigments, are found in all photosynthetic species, and alongside chlorophyll are typically expected to be present in varying quantities according to a variety of different parameters, such as the type of the oil being analysed, its origin, or in the case of olive oil its “virgin” status, descriptive of the way in which the oil is extracted and thus its purported flavour profile (Giusti and Wrolstad, 2001). Excess chlorophyll has been shown to be indicative of oil originating from younger olives, and thus could be associated with oils that contain a larger concentration of antioxidants and a “greener flavour” (Blatchly, Delen and O’Hara, 2014), whilst the authenticity of oils, from either bulking of higher quality oils with lower quality oils or complete substitution, can be characterised with the technology (Popa et al., 2020; Salghi, Armbruster and Schwack, 2014).

UV-Visible Spectroscopy has also seen applications in the assessment of food quality, through the quantification of sample absorbency correlated with the rancidification of oils and edible fats. The *p-anisidine* value calculated from spectral readings is used as a measure of the aldehyde compounds present in the sample, which are formed as secondary oxidation products of the lipids in the oil (Tompkins and Perkins, 1999).

These aldehyde compounds have not necessarily been associated with the alteration of flavour profiles through spoilage, but lower scores may represent a marker for minimal oil degradation.

1.8.2.2 Fluorescent Spectroscopy

Fluorescent Spectroscopy is a methodology that allows for rapid, non-destructive assessment of sample state. It exploits the phenomenon of fluorescence, in which incident light is absorbed by a molecule, called a fluorophore, and later emitted as light of a different wavelength. This is due to the excitation of a fluorophore's electrons to a higher energy state by incident electrons and subsequently (around 10^{-8} seconds later) the relaxation of electrons back to their ground state.

The wavelength of the emitted light is proportional to the difference between energy levels of the excited and ground state of the electron - an importance note is the distinction between fluorescence and simple reflectance, as the promotion of an electron to a discrete higher energy level (orbital) results in the emission of light at a different wavelength, whereas reflection is the result of excitation of electrons to a non-discrete energy, which is almost immediately lost and re-emitted at the same wavelength.

This exploitation of the relationship between monochromatic incident light and the wavelength of the radiated fluorescence lends the technology its utility – the high specificity of organic fluorophores for specific excitatory wavelengths allows the researcher to search for the presence of candidate compounds that are known to fit the excitatory and emission wavelength envelopes under investigation, whilst offering a good degree of sensitivity, as the monochromatic nature of the incident light means that other wavelengths can be eliminated from the test cell, reducing

the likelihood of interference patterns emerging from sample interactions with other bands of light.

Fluorescent Spectroscopy has an important application in food testing due to its aptitude for the detection of very low concentrations of toxins and microbes, in addition to non-organic contamination with food additives. A natural property of many bacterial species is that of fluorescence, either of constituents of the microbial cell, or the cell's metabolites.

Fluorescent spectroscopy has been used for the detection of lactic acid bacteria (LAB) in dry sausage samples, not only showing good results in identification of the presence of bacteria in general, but also showing good discrimination in 28 of 29 wild strains of bacteria at the genus and species level, using the fluorescent qualities of aromatic amino acids and nucleic acids (AAA+NA) (Ammor et al., 2004). Other examples include the assessment of *E. coli* contamination on the surface of meat (beef) samples following intentional inoculation – the results after 5 hours of contamination showed a distinct modification of values in a fingerprint wavelength region of 500-560 nm (Mekhrengin et al., 2018). The authors state that whilst the results show the ability to detect high quantities of bacterial contamination using a semi-portable fibre-optic setup and thus serve the food industry as a non-invasive means of conducting first-stage quality assurance, the results may be improvable through the application of fluorescent dyes to the sample or employing time-resolved spectroscopy to increase the sensitivity of detection.

Microbial toxin detection exploits the natural fluorescence of metabolic compounds, typically associated with fungal metabolism (mycotoxins). Aflatoxins are typically associated with the fungi *Aspergillus parasiticus* and *Aspergillus flavus*, via the polyketide pathway (Alcaide-Molina et al., 2009; Ali et al., 2005), and are regarded as being some of the most carcinogenic biologically-derived compounds. As such, there is great interest in development of methods for their identification and minimisation in foodstuffs fit for human and livestock consumption, for the interests of preserving long-term consumer health.

The four important aflatoxin groupings are Aflatoxins B1, B2, G1 and G2. These are typically differentiated in the experimental setting through their fluorescent properties: AFB1 and AFB2 typically fluoresce in the UV/green to blue wavebands, whilst G1 and G2 typically fluoresce in yellow-green wavebands (Singh and Mehta, 2020). Further classes of mycotoxins detectable with fluorescence sensing equipment include Zearalenone, Citrinin and Ochratoxin, each displaying different toxic modalities (Creppy, 1999; Mishra and Das, 2003; Bhat, Rai and Karim, 2010). The fluorescent distinction allows for accurate discrimination of each in a variety of food safety oriented analyses (Huang and Elmashni, 2007; Wu, Xu and Xu, 2019; Smeesters et al., 2023).

Further uses of fluorescent spectroscopy include the quality and compositional analysis of liquids, in particular food grade oils. In a similar fashion to UV-Vis Spectroscopy, fluorescent spectroscopy can be used for the detection of food fraud attempted through the adulteration of higher-quality virgin olive oils with other inferior-quality oils, such as that from refined olives, seed and nut oils. Sayago *et al.* used a fluorospectroscopic method for the classification of olive oils with varying contamination with hazelnut oils, the resultant pipeline when coupled with stepwise linear discriminant analysis (LDA) showing high efficacy in classification (100%) (Sayago, Morales and Aparicio, 2004). Poulli *et al.* demonstrated a method for the detection of adulteration of olive oil with common sunflower oil, and noted that the latter possessed a fluorescence peak in the 325-385nm region, whereas the former did not. The developed partial least squares regression (PLSR) model was found to be suitable for detection of quantities of sunflower oil adulteration down to 3.4% (weight per volume or w/v), and rapid, producing results in 150 seconds (Poulli, Mousdis and Georgiou, 2006) further evidencing the technology's utility for rapid food control tasks.

1.8.2.2.1 Portable and Non-Invasive Implementations of Fluorescent Spectroscopy

The operating characteristics of fluorescent microscopy may make it more suited to miniaturisation over some other members of the spectroscopic family of sensors. The conditions required for a fluorochemometric sensor's monitoring of sample emissions as the result of monochromatic stimulation of may be simpler to produce than those required for other sensors, such as those relying on the principles of UV-Visible light spectroscopy, which assess the transmittance of incident light throughout a sample, typically a liquid, as this may necessitate preparation of the sampling environment, i.e. filling of a cuvette, calibration, and sensor cleaning with control compounds etc. before analysis can be conducted.

The Freshdetect platform is an attempt to miniaturise a fluorescent sensor platform into a hand-held device, for the purposes of facilitating on-line testing of meat produce throughout the food chain. The claims, made in the EU funding agreement state that the device can accurately assess the Total Microbial Count (TVC) of meat products in a fraction of the time of laboratory-based assays (5-10s versus up to 72h for lab-based assessment) (Dietrich, 2018). This performance acceleration may represent a candidate device for the food operator to conduct real-time process controls on produce.

Smeesters *et al.* demonstrated the utility of a hand-held sensor device in the analysis of aflatoxin contamination in maize, analysing individual kernels by classification of their *a priori* unknown aflatoxin contamination into categories (Smeesters et al., 2023). The authors of the paper make note of the reduction in size of the portable "IndiGo" device versus the conventional laboratory setup, whilst maintaining the high sensitivity and predictive accuracy of the latter, and incorporating the desirable quality of being non-destructive/invasive.

Mekhrengin *et al.*'s assessment of the *E. coli* contamination encountered on beef samples made use of a fibre-optic probe device attached to a desktop sensor, allowing for non-invasive assessment of meat substrates (Mekhrengin et al., 2018).

1.8.2.3 Infrared Spectroscopy

Infrared spectroscopy represents a large family of absorption spectroscopy devices, due to the range of wavelengths that the technology represents, from approximately 78nm to 1mm. IR spectroscopy is typically useful for analysis of the chemical structure of compounds, as excitation with infrared and near-microwave radiation will result in the absorbance of specific wavelengths, associated with the mechanical excitation of atomic bonds that occur in the sample species, at resonant frequencies universally characteristic to specific bond strengths and atomic masses, therefore representative of specific functional groups.

Table 1-4 Table showing a non-exhaustive list of characteristic functional groups, their general expected wavebands and the expected intensity of response to excitation, demonstrating the utility for food analytical purposes. Table based on information from online resources (Mohrig et al., 2024; Reusch, 2013).

Functional Group	Absorption Locations (cm ⁻¹)	Absorption Intensity
Alkane (C-H)	2850 – 2975	Medium/Strong
Alcohol (O-H)	3400 – 3700	Strong, but Broad
Alkene (C=C)	1640 – 1680	Weak/Medium
(C=C-H)	3020 – 3100	Medium
Alkyne (C≡C)	2100 – 2250	Medium
(C≡C-H)	3300	Strong
Nitrile (C≡N)	2200 – 2250	Medium
Aromatics	1650 – 2000	Weak, Broad
Amines (N-H)	3300 – 3350	Medium
	3500	
Carbonyls (C=O)	1700 – 1750	Strong
Aldehyde (CHO)		
Ketone (RCOR)		
Ester (RCOOR)		
Carboxyl (RCOOH)		

The differentiation between near (14,000–4,000 cm⁻¹), medium (4,000–400 cm⁻¹) and far (400–10 cm⁻¹) dictates the type of vibrational, vibrational-rotational, or rotational excitation expected of a molecule. An apparent majority of IR spectroscopy used in food analysis occurs in the near- and medium infrared categories (Rohman et al., 2020).

A commonly used implementation of near- and mid-IR spectroscopy is Fourier-transform Infrared (FTIR). The spectrometer collects data over a wide array of wavelengths, and uses a mathematical transform process to convert measurements into a readable spectrum.

Instead of using a monochromatic beam of light to assess the sample and recording the resultant absorbency (in the case of UV-Vis) or fluorescence, the FTIR device makes use of a range of different wavelengths emitted together. Subsequent readings are achieved with differing ranges of wavelengths emitted by the sensor. The device uses an interferometer to modify the spectrum due to wave interference, and the resultant absorption spectra (interferograms) are passed into an analytical element to compute the results by applying the Fourier transform, inverting the domain of the collected data/function (light absorption versus mirror displacement in centimetres into light absorption versus wavenumber, measured in cm^{-1}).

This methodology confers some advantages over more conventional “dispersive” spectroscopy techniques that function through “scanning” emission of monochromatic light in sequence, such as some UV-vis technologies, that generally result in a higher signal-to-noise ratio (SNR) (Sellar and Boreman, 2005). This advancement in technological implementation could indicate that methods that make use of monochromators are now obsolete, however the usage of such technologies is still maintained in many circumstances due to other operational factors, such as cost and simplicity of operation.

Applications of FTIR are far ranging, due to the large envelope of spectral wavelengths/wavenumbers covered by spectrometers of the type, and the precision/accuracy of the sensor type. Pertinent to this document is its use for the purposes of food authenticity, quality and safety testing.

An important development to accompany the use of IR spectroscopy in the food sector was that of Attenuated Total Reflectance (ATR), expanding the usage envelope of IR/FTIR spectroscopy outside of applications related to transmission, reducing the need for adulteration of the sample through sectioning, reconstitution or otherwise. Furthermore, use of an ATR accessory minimises the

penetrance of the incident light, avoiding issues related to attenuation throughout highly absorbent media like aqueous solutions, and extending the range of suitable sample media to undiluted liquids and thicker solids. This offers further benefits to operation, such as increasing of sample-to-sample reproducibility of experimentation, minimising and accelerating necessary maintenance of the sensor (only requiring cleaning of interfacing sensor head) and offering high throughput relative to the available energy supplied to the sample compartment, resulting in high quality, high SNR results, that allow for more precise analyses (Blum and John, 2012).

Such ATR probes require physical contact with a sample to allow for modulation of a standing/evanescent wave generated as a total internal reflection in a crystal of a higher refractive index than that of the sample. As such, the technique may be considered as “minimally invasive” versus “non-invasive”. The throughput increases due to the minimisation of experimental setup and reset time and analytical optimisation for solids and liquids in an unmodified state make the combination of FTIR and ATR an accessible tool for food control applications (Rodriguez-Saona and Allendorf, 2011). Tests are rapid, taking roughly 60-120 seconds, require very little if any sample preparation and generate minimal waste. The resultant spectra can be examined for multiple classes of chemical contaminants simultaneously, due to fingerprint capture of multiple spectral elements.

In the case of food control, the mid-infrared (MIR) band is often chosen, due to spectral absorbencies often being correlated with fundamental vibrations of functional groups of interest (Guillén and Cabo, 1997), containing bands rich in structural information associated with proteins, lipids, polysaccharides and carotenoids. An important element of this type of assessment is the quantitative dimension, as the intensities of the bands recorded is proportional to the concentration of the representative functional group (Paré and Bélanger, 1997).

Following the Chinese melamine scandal of 2008-9, tests of NIR and MIR techniques coupled with multivariate analysis/machine learning showed the ability to distinguish between adulterated (1 part per million, ppm) and

unadulterated infant formula samples (Mauer et al., 2009). Whilst satisfactorily detecting the threshold value set at 1ppm for infant foods by the FDA, performance still fell below the 250ppb limit of detection (LOD) achievable with the prescribed Liquid Chromatography/Mass Spectrometry method, however, the higher performance of the LC-MS pipeline is offset by its time-consuming and labour-intensive nature making it unsuitable for screening large quantities of samples, which may be encountered in a food production control environment.

Partial Least Squares Regression (PLSR) models built using FTNIR data in the 3,330–2,993 cm^{-1} and 1,321–983 cm^{-1} regions, corresponding to the stretching vibration of amino groups and the fingerprint region, showed good results, with an R^2 of 0.95 and RMSEcv of 1 ppm, whilst results with NIR data in the 12,497–6,098 cm^{-1} and 5,450–4,248 cm^{-1} region showed marginally better results, with R^2 of 0.999 and RMSEcv of 0.62 (Mauer et al., 2009).

Subsequent analysis has confirmed the efficacy of such a method for the rapid quantification of melamine within powdered infant milk products; when coupled with Partial Least Squares regression, the analysis of FTNIR data at similar wavelengths to the previous FTNIR study showed a coefficient of prediction (R^2) of >0.99, and an LOD of > 0.5 $\mu\text{g/L}$ (García-Miguel et al., 2018).

Honey authentication is another noteworthy example of food authentication using FTIR. Honey is a high value item for producers and consumers, thus a motivation exists for some farmers and producers to conduct fraudulent activities to maximise profits. Detection of adulteration of honey products with inferior bulking agents, such as poorer grades of honey or sugar syrup is of paramount importance for the preservation of consumer trust in purchased goods, and maintenance of a competitive environment for producers with a reduced risk of fair business practices being undercut. Additionally, sugar content analysis in honeys is importance for quality control purposes –according to Bogdanov, honey is typically 95% carbohydrate, the remainder being water and other trace constituents (Bogdanov et al., 2008). The carbohydrate composition is partly dependent on the nectar collected by the bees and thus varies with the climatic conditions of the region and the local flora (Ouchemoukh et al., 2010). Anjos et

al. demonstrate a methodology for the identification and quantification of individual sugar components in 63 honey samples, using mid-range FTIR couple with an ATR accessory; the 1500-750 cm^{-1} wave number range allowed for the construction the best performing model in the paper – a PLS regression model demonstrating an R^2 of 86.60 and 86.01 for identification of fructose and glucose respectively, with a relative percentage difference (RPD) of 2.6% and 2.55% respectively (Anjos et al., 2015).

IR Spectroscopy and its subcategories have found usage in authentication and spoilage analysis of meat produce in addition to plant-based or plant-derived produce. FTIR has been utilised in a variety of applications, coupled with multivariate analysis; often dimensionality-reducing algorithms are employed, including partial least squares and principal component techniques, whilst Support Vector Machines (SVMs) have demonstrated their utility in finding boundary separation in multidimensional data.

Mabood *et al.*'s 2020 paper illustrates the utilisation of FTIR, coupled with Partial Least Squares – Discriminant Analysis (PLS-DA), PLS-R and Principal Component Analysis (PCA) for effective classification and regression of meat samples containing varying degrees of pork, emulating the circumstances of several previous food authentication failures in which pork had erroneously entered the food chain, e.g. the horse meat scandal in 2013. In their study, the researchers sourced 39 different types of meat, including beef, mutton, lamb, pork, camel, chicken and veal from various regions and sourced from various Omani markets (Rizvi et al., 2018).

All meats were minced, and were mixed with 8 degrees of pork (i.e. 0%, 1%, 5%, 10%, 30%, 50%, 70%, and 90% pork). All samples were scanned using an NIR spectrophotometer in the range 10000 to 4000 cm^{-1} in reflectance. Upon construction of PLS models using the spectral data, the researchers established that the majority of contribution to the models was made by peaks in the spectral range of 7000 to 4730 cm^{-1} , due to higher concentrations of polyunsaturated fatty acids in muscle and adipose tissues than other meat types tested, the most stark difference being the percentage linoleic acid composition of meat muscle and

adipose tissue, found to be 14.3%, 1.3%, 1.1% for pork, sheep, and cattle adipose tissue respectively, and 14.2%, 2.7%, 2.4% for pork, sheep and cattle muscle tissue. The optimal PLS-R model, built using three factors, was found to explain the dependent variability significantly, with $R^2 = 0.9774$ and good cross validation performance, with RMSECV = 1.08%. PLS-DA conducted on the different proportional groups of meat adulteration performed well, classifying with 100% accuracy, sensitivity, and selectivity displayed for the 10% adulterated samples, but with decreasing performance for the 5% and 1% classes (93%,92%,92% and 85%,88%,87% accuracy, sensitivity and selectivity for 5% and 1% classes respectively). This analysis provided a good degree of supporting evidence for the utilisation of FTIR for robust adulteration analysis in differing meat types, due to the variety of meats used, but also due to the quantity of individuals generated by the project showing good generalised performance (~6000 samples over 3 biological replicates of each condition).

Further applications of IR spectroscopy for meat analysis includes spoilage quantification; Spyrelli *et al.*'s 2021 paper describes the utilisation of FTIR for the recording of the surface state of chicken fillets and assessment of microbial metabolic activity, followed by multivariate analysis with a variety of different algorithms to assess their suitability for the task (Spyrelli *et al.*, 2021). Meat was acquired directly from Greek suppliers, and was transferred to aerobic storage in four isothermal conditions (0, 5, 10, 15°C), for up to 480h, depending on the storage condition, and at regular intervals, readings were taken using FTIR and MSI (Multispectral Imaging) sensors.

Assessment of algorithms was conducted with Unscrambler® software and the SorfML web application (Estelles-Lopez *et al.*, 2017). The researchers sought to compare and contrast the performance of the two sensor types for the purposes of quantification of bacterial load in the meat samples (Total Viable Count (TVC) and *Pseudomonas*), calibrated using bacterial culture data collected at the same time intervals.

Unscrambler was used to construct PLS-R models using each sensor data type independently. Cross validation performance of the models was mediocre,

showing R^2 values of 0.609 and 0.493 for Leave One Out Validation (LOOCV) and validation sets respectively for TVC, and 0.549 and 0.481 for the *Pseudomonas* assessment's LOOCV and validation set.

Accuracy was a metric devised to give a perspective on useful performance: it represented the proportion of predicted bacterial values that fell within ± 1 log CFU/cm². Both batches had a training partition of 115 spectra and a test partition of 99 spectra.

The authors determined that the FTIR sensor was more suitable for assessment of chicken spoilage and prediction of microbial loads than MSI, and identified regions that fell in agreement with existing literature on informative wavebands for spoilage in meat – 1426 - 1369cm⁻¹ and 1567 - 1464 cm⁻¹, correlating with absorption bands for amides and free amines (Böcker et al., 2007), and falling in agreement with observations linking absorption bands 1650, 1550, and 1400-1200 cm⁻¹ to amide I, II and III and the proteolytic activity of *Pseudomonas spp.* on meat (Nychas and Tassou, 1997).

1.8.2.4 Raman Spectroscopy

Raman spectroscopy is similar in some respects to Infrared Spectroscopy, in that both are vibrational microscopy techniques, as they are used to examine molecular vibrations in sample compounds; some research has reported on usage and performance of Raman spectroscopy in a complementary fashion to the aforementioned.

Raman and IR spectroscopy differ in the way in which they interact with the sample molecules and the information yielded from such interactions. In summary, IR spectroscopy measures absorption of IR radiation at different wavelengths, due to molecules, bonds and functional groups undergoing vibrational transitions, e.g. rotation, stretching, and bending. These types of interaction occur when the frequency of infrared matches the natural fundamental vibrational frequency of species in the compound, or at overtones – multiples of the fundamental frequency.

Raman Spectroscopy measures the inelastic scattering of monochromatic light, by molecules and functional groups. When molecules are irradiated with monochromatic laser light, the majority of the scattered light is of the same frequency as the incident light, a phenomenon known as Rayleigh Scattering. A small proportion of the light is returned at a different energy level (i.e. frequency) due to molecular vibrations – the difference between the energy of the incident photon and that of the scattered photon is known as the Raman Shift, and is related to the type of bond being excited, and the intensity of the returned signal is proportional to the number of the bonds and/or the analyte concentration (Farber et al., 2019).

Raman spectroscopy is often used in situations where other techniques may be unsuitable due to compositional incompatibilities of the analyte. For example, IR spectroscopy excels for the characterisation of polar functional groups, however performs generally poorly for symmetric or weakly dipolar bonds. Raman spectroscopy displays opposing tendencies, generally performing better with bonds displaying weak polarity or none at all (Farber et al., 2019). Additionally, Raman spectroscopy typically performs better with aqueous or hydroxyl-bond rich solutions, an issue that can affect non-ATR equipped IR spectrophotometers.

A key benefit of Raman technology is that it can be deployed in a non-destructive fashion for analysis of a variety of sample forms (e.g., liquids, solutions, gels, solids). However, a key weakness of Raman implementation is the generally poor response to samples that exhibit fluorescence, as the radiated light from such phenomena can overwhelm the magnitudes-smaller Raman shift response. This may present a problem in the biological domain, as many organic molecules including microbial metabolic products are fluorescent to some degree, and steps must be taken to calibrate the sensor, or otherwise account for this property.

As mentioned previously, Raman and IR spectroscopy have often found use in concert in the food analysis sector, due in part to their complementary strengths offering different views of the same analytical target. Some research has sought to further exploit the complementarity of the two modalities of sensor by integrating them into a sensor fusion pipeline. Márquez *et al.* employed several

sensor fusion strategies for multivariate qualitative analysis of adulteration of hazelnut pastes with almond pastes (Márquez et al., 2016). Fraudulent adulteration of hazelnut products is made economical for bad actors due to the demand for such substances as an ingredient for other foodstuffs, like confectionery. Padding of nut pastes with a lower value alternative, such as almond, or in extreme cases, other plant derivatives such as chickpea paste, can increase the profit margins for producers.

The paper's authors demonstrate two methodologies for achieving sensor fusion for classification into either non-adulterated or adulterated classes using both a mid- and late fusion approach for combining the transformed features and decisions made with the IR and Raman data respectively.

The fused pipelines both showed better results than either of the two sensor types in isolation in unadulterated cases, and roughly equal performance for adulterated cases: sensitivity and specificity were both 100% for the mid-level fusion pipeline for unadulterated cases, with no inconclusive results. FT-Raman gave 75%, 100%, 93% for sensitivity, almond specificity, and chickpea specificity, with 25 inconclusive results, whilst NIR showed 75 %, 96%, 100% for sensitivity and specificity for almond and chickpea cases respectively, with 21 inconclusive results.

1.8.2.5 Imaging Technologies

Other light-based technologies have uses in remote sensing tasks across the data-science and process monitoring domain. Multi- and Hyper-spectral imaging could be considered as special cases of spectroscopy (the nomenclature deriving from NASA's Airborne Imaging Spectrometer (AIS) and AVIRIS projects in the mid-1980s) and are the product of the information age – such sensors are designed to capture spectral information about a sample from across a wider range of the electromagnetic spectrum than spectroscopic techniques in isolation. As imaging technologies, they are often employed to capture more of the spacial information presented in a sample that spectroscopy can offer.

Multi- and Hyper-spectral imaging technologies typically capture information about a specific waveband of light, both visible and invisible, with less resolution than a dedicated spectrophotometer, however the large range and quantity of data that is captured is often well suited for the physical assessment of sample state in a standardised fashion, when compared to other more qualitative measures such as domain expert opinion.

Multi-spectral imaging (MSI) is a less resource intensive form of imaging technology, when compared with hyper-spectral imaging (HSI). MSI is configured to collect spectral and spacial information subsetting or subsampled from regions of interest, whilst HSI is often configured to capture information from a wide spectrum of EM radiation, and as a consequence is a far more resource intensive sensor device. HSI measures continuous spectral bands, when compared with spectroscopic techniques, which measure discrete and spaced spectral bands (Hagen and Kudenov, 2013). The inclusion of the additional spectral dimension into the original spatial representation of a typical two-dimensional image results in the formation of a so-called “datacube”, in which layers of additional spectral information are stacked upon the spatial information.

1.8.2.5.1 Portable and Non-Invasive Implementations of Multispectral Imaging

The near-real-time imaging nature of some implementations of MSI makes it a desirable technology for development into an in-/at-line food assessment strategy, for simultaneous assessment of spectral and spatial state of samples in various stages of production and distribution.

Sensor implementations of note include Videometer (VideometerLab) (Videometer, Denmark, 2021), a family of portable sensors designed for at-line, near real time process control applications. A specialised implementation of MSI, the platform is equipped with a machine-learning for users to gather pertinent information for assessment of the physical and chemical state of a sample in a non-destructive fashion.

The VideoMeterLab sensor records measurements at 18 wavelengths between 405-970nm. Samples are successively strobed with light from light emitting

diodes (LEDs) within the sensor in varying combinations, which generate monochromatic light with a high degree of accuracy and of a diffuse/homogeneous intensity, due to the Ulbricht sphere component. This spectral scanning process results in a datacube with the dimensions $m \times n \times 18$, where m and n represent the height and width of the image in pixels.

Videometer has been widely cited in the food analysis domain, in a variety of contexts, including quality/safety control and food fraud detection. Pertinent applications to the current document include usage in the detection of meat spoilage and quality analysis (Tsakanikas et al., 2016). The paper describes the utilisation of a Videometer sensor for the classification of beef samples into one of two qualitatively assigned classes based on the plate count interval ($\text{TVC} < 2\log_{10} \text{CFU/g}$ and $\text{TVC} > 2\log_{10} \text{CFU/g}$), and regression for prediction of the real-time TVC present in samples based on multispectral imaging data.

The sensor was used to collect images at different time intervals, and they were processed using an in-house novel unsupervised segmentation pipeline to eliminate extraneous information from the images, (employing standard normal variate, K-means clustering and Gaussian Mixture Models (McLachlan, Lee and Rathnayake, 2019)) resulting in condensed imaging data, which was then fed into an SVM regressor, which showed high prediction performance, with an RMSEP = $0.14 \log_{10}\text{CFU/g}$ and $R^2=0.98$.

Estimation of pork microbiological spoilage was also examined using the Videometer technology, alongside visible and IR spectroscopy (Fengou et al., 2019). In a similar strategy to the previous example, spoilage data for pork samples stored in different isothermal conditions was collected at regular time intervals using the plate counting method and sensor reading taken simultaneously. Partial least squares regression models were constructed using the individual sensor data, with the FTIR, VIS, and MSI models displaying RMSEP values of 0.915, 1.173, and $1.034 \log_{10}\text{CFU/g}$, and R^2 values of 0.834, 0.727, and 0.788 respectively. The conclusions of the authors were that the sensor types exhibited potential for spoilage prediction. The authors also attempted a naive fusion approach for the spectral data, but achieved negligible performance

differences to single sensor models. Of note was the difference in processing of the Videometer data: rather than employing a novel segmentation strategy and processing the image data as-is, image segmentation was performed with the supplied VideometerLab software, through contrast maximisation and canonical discriminant analysis. After this process, the mean reflectance spectra in the informative area, i.e. the mean intensity of pixels was recorded.

1.8.3 Chemical Probes – Electronic Nose and Tongue

The human sensory experience, like that of many other animals, has evolved in part for sensing the state of potential foodstuffs e.g., the visual impression of the item, its smell and its taste. This sensory input is vital for deciding whether the item is fit for ingestion. A large human-sense detectable aspect of food spoilage is its change in odour, caused by the metabolic breakdown of plant and animal tissue by a range of microbe types.

This property has warranted the use of “sensory panels” in the food production domain for many years, and is likely one of the oldest forms of quality control in food manufacture: in such a system, individuals, either trained internal personnel or third parties, form a panel in which participants are required to assess their human sensory experience of a particular sample of food or beverage.

In order to attempt standardisation of such testing regimes in the food industry, the International Organization for Standardisation (ISO) publishes and updates approved testing guidance, methodologies and controlled vocabularies for respondents (e.g. ISO 67.240 - Sensory analysis), but this form of testing is still subjective to the individuals taking part, so the use of large panels of individuals, generally consisting of more than 8 and less than 20 individuals (Djekic et al., 2021) helps to average out any outliers from the testing data. The aggregated feedback from their individual experiences is then fed back to analysts, who make use of the data in a qualitative sense, either in a descriptive or discriminative capacity i.e., for use in quality control testing.

Whilst the sensory panel plays an important role in qualitative analysis in food science, they are cumbersome to implement. Sourcing of members is the first

hurdle, and it can be challenging for small food business operators to implement such a testing scheme. Additionally, there is likely a money and time cost for training of a food testing panel, to maintain abilities pursuant to ISO standards. Further time delays are likely incurred upon collation of sensory panel results from a specific test. In general, they do not offer the affordability, flexibility, or responsiveness of some data-driven sensor analysis techniques, which can offer the additional benefit of being functionally impartial, serving as a simple probe of the objective chemical make-up of a sample of food or beverage.

Electronic nose and tongue are a closely related technology: in essence, they are arrays of sensors, calibrated to detect compounds or groups thereof present in a food sample. The modes of operation of sensors mimicking human gustatory and olfactory receptors are varied, however many common instances in the food industry are of the metal–oxide–semiconductor (MOS) type, in which presence of a target gas changes the resistance value of the metal oxide layer in the transistor, and allowing a current to flow, generating a response/signal that can be detected. Such a response is thus proportional to the concentration of the analyte, and requires a recovery period and possible purging to return to the neutral response after being exposed to a sample.

E-nose arrays used in food assessment include probes that are pertinent to the spoilage processes expected, detecting Volatile Organic Compounds (VOCs), however are generally non-specific (Roy and Yadav, 2022; Yan et al., 2015). Instead, target analytes will likely elicit a response from multiple sensors to varying degrees, generating a characteristic “fingerprint” response.

Literature has described the use of electronic nose sensors in the authentication/safety and quality analysis of food and beverage products, including detection of both spoilage processes due to microbial activity and human-induced issues such as deliberate contamination or adulteration for fraudulent purposes, in food oils (Bougrini et al., 2014; Mildner-Szkudlarz and Jeleń, 2008), milk (Tohidi et al., 2018) meat products (Tian et al., 2019; Wang et al., 2019), and honey (Gan et al., 2016), amongst others.

1.9 Significance of the Research Problem

Modern high throughput, data-driven sensor devices have shown promise for facilitating food control objectives, including safety, quality, and authenticity analysis for many different food types, including fruits and vegetables, cereals and meat.

Recent publications have shown that data derived from such devices can be coupled with modern machine learning techniques and produce results that can offer a worthwhile supplement to existing laboratory-based techniques, that are often batch-based and thus lacking in high-resolution coverage of the entire food production and distribution pipeline, slow and unable to react to emergent issues in a preventative manner i.e. before sale, and result in sample or whole-batch destruction (Nychas et al., 2021).

However, the successful implementation of these sensor types in a “production” food monitoring environment, meaning deployment in-, on- or at- line, is contingent on the simultaneous development of infrastructure to support the ingestion, storage, and application of resultant data. Such a system must be capable of accounting for different configurations of device with varying output dimensionalities and accompanying meta-data describing the primary data’s scientific provenance, enabling application of gathered knowledge in the construction of valid and reusable decision support pipelines. With such a system in place, complex integration of sensor data can take place in a semi-automated fashion with minimal technical input from the operator.

An application of food monitoring devices that could be facilitated by such a system is to be found in sensor fusion, the combination of information gathered from multiple modalities of sensor data with the potential to increase accuracy of predictions through a more complete view of sample state, and to grant greater robustness and reliability to predictions through sensor redundancy. Sensor fusion has seen comparatively little study within the sphere of food spoilage, quality and authenticity analysis, and research in this area may serve to further improve upon the documented performance of predictive models built with non- or minimally-invasive sensor technologies.

1.10 Hypothesis Statements

1. High throughput sensor devices can offer a valid alternative or supplement to existing approaches for assessing the spoilage and quality metrics of meat products through the development of appropriate statistical and machine learning pipelines.
2. Development of multi-modal fusion approaches for meat spoilage and quality assessment that incorporate spectroscopic and multi-spectral imaging device data will improve the accuracy of machine learning outputs when compared to single-modality approaches.
3. Development of a web platform for the upload, storage, and utilisation of food analytical sensor data will allow for successful reproduction of the machine learning pipelines implemented manually in points one and two in an accessible manner suitable for reapplication to new data and sensor technologies.

1.11 Objectives

This research aims to simultaneously develop new, accurate machine learning approaches for meat spoilage and authenticity prediction tasks, and develop an online platform to facilitate the ingestion, storage, and use of food analysis sensor data for the construction of machine learning models, for deployment in a similarly online fashion.

The main objectives are as follows:

- 1 Develop machine learning regression pipelines incorporating spectroscopic and multispectral imaging data suitable for prediction of microbial load in chicken and beef samples
 - 1.1 Establish the suitability of individual sensor types for this objective, via the assessment of performance of machine learning models built with representative data
 - 1.2 Evaluate the performance of different data fusion approaches for integrating information from multiple sensors, with the intention of increasing performance over single-sensor models
- 2 Develop a modular online food analysis platform and machine learning framework to allow for ingestion, storage, and utilisation of sensor data for construction and deployment of reusable predictive pipelines
 - 2.1 Demonstrate and validate the functionality of the platform through reproduction of the existing machine learning research described in the prior objectives, previously conducted in a manual fashion

References

- Abdi, H. (2010) 'Partial least squares regression and projection on latent structure regression (PLS Regression)', *WIREs Computational Statistics*, 2(1), pp. 97–106.
- Acar, E., Papalexakis, E.E., Gürdeniz, G., Rasmussen, M.A., Lawaetz, A.J., Nilsson, M. and Bro, R. (2014) 'Structure-revealing data fusion', *BMC Bioinformatics*, 15(1), p. 239.
- Ait-Kaddour, A., Loudiyi, M., Ferlay, A. and Gruffat, D. (2018) 'Performance of fluorescence spectroscopy for beef meat authentication: Effect of excitation mode and discriminant algorithms', *Meat Science*, 137 Oxford: Elsevier Sci Ltd, pp. 58–66.
- Alcaide-Molina, M., Ruiz-Jiménez, J., Mata-Granados, J.M. and Luque de Castro, M.D. (2009) 'High through-put aflatoxin determination in plant material by automated solid-phase extraction on-line coupled to laser-induced fluorescence screening and determination by liquid chromatography-triple quadrupole mass spectrometry', *Journal of Chromatography. A*, 1216(7), pp. 1115–1125.
- Ali, N., Hashim, N.H., Saad, B., Safan, K., Nakajima, M. and Yoshizawa, T. (2005) 'Evaluation of a method to determine the natural occurrence of aflatoxins in commercial traditional herbal medicines from Malaysia and Indonesia', *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association*, 43(12), pp. 1763–1772.
- Al-Jowder, O., Kemsley, E.K. and Wilson, R.H. (2002) 'Detection of adulteration in cooked meat products by mid-infrared spectroscopy', *Journal of Agricultural and Food Chemistry*, 50(6), pp. 1325–1329.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L. (2021) 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', *Journal of Big Data*, 8(1), p. 53.
- Ammor, S., Yaakoubi, K., Chevallier, I. and Dufour, E. (2004) 'Identification by fluorescence spectroscopy of lactic acid bacteria isolated from a small-scale facility producing traditional dry sausages', *Journal of Microbiological Methods*, 59(2), pp. 271–281.
- Anastassiades, M., Lehotay, S.J., Štajnbaher, D. and Schenck, F.J. (2003) 'Fast and Easy Multiresidue Method Employing Acetonitrile Extraction/Partitioning and "Dispersive Solid-Phase Extraction" for the Determination of Pesticide Residues in Produce', *Journal of AOAC International*, 86(2), pp. 412–431.

- Anjos, O., Campos, M.G., Ruiz, P.C. and Antunes, P. (2015) 'Application of FTIR-ATR spectroscopy to the quantification of sugar in honey', *Food Chemistry*, 169, pp. 218–223.
- Argyri, A.A., Jarvis, R.M., Wedge, D., Xu, Y., Panagou, E.Z., Goodacre, R. and Nychas, G.-J.E. (2013) 'A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage', *Food Control*, 29(2), pp. 461–470.
- Argyri, A.A., Mallouchos, A., Panagou, E.Z. and Nychas, G.-J.E. (2015) 'The dynamics of the HS/SPME–GC/MS as a tool to assess the spoilage of minced beef stored under different packaging and temperature conditions', *International Journal of Food Microbiology*, 193, pp. 51–58.
- Banja, J. (2010) 'The normalization of deviance in healthcare delivery', *Business Horizons*, 53(2), p. 139.
- Barker, M. and Rayens, W. (2003) 'Partial least squares for discrimination', *Journal of Chemometrics*, 17(3), pp. 166–173.
- Barnes, R.J., Dhanoa, M.S. and Lister, S.J. (1989) 'Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra', *Applied Spectroscopy*, 43(5), pp. 772–777.
- Barrere, V., Everstine, K., Théolier, J. and Godefroy, S. (2020) 'Food fraud vulnerability assessment: Towards a global consensus on procedures to manage and mitigate food fraud', *Trends in Food Science & Technology*, 100, pp. 131–137.
- BBC (2004) *China 'fake milk' scandal deepens*, 22 April. Available at: <http://news.bbc.co.uk/1/hi/world/asia-pacific/3648583.stm> (Accessed: 12 March 2025).
- Behnke, K. and Janssen, M.F.W.H.A. (2020) 'Boundary conditions for traceability in food supply chains using blockchain technology', *International Journal of Information Management*, 52, p. 101969.
- Benvenuti, M.E. and Burgess, J.A. (2012) *Determination of Food Sugars in Fruit Juice Using Evaporative Light Scattering Detection and BEH Amide Column Chemistry*. Available at: <https://www.waters.com/nextgen/gb/en/library/application-notes/2012/determination-of-food-sugars-in-fruit-juice-using-evaporative-light-scattering-detection.html> (Accessed: 30 March 2025).
- van den Berg, F., Lyndgaard, C.B., Sørensen, K.M. and Engelsens, S.B. (2013) 'Process Analytical Technology in the food industry', *Trends in Food Science & Technology*, 31(1), pp. 27–35.
- Bhalla, V., Grimm, P.C., Chertow, G.M. and Pao, A.C. (2009) 'Melamine nephrotoxicity: an emerging epidemic in an era of globalization', *Kidney International*, 75(8), pp. 774–779.

Bhat, R., Rai, R.V. and Karim, A. a. (2010) 'Mycotoxins in Food and Feed: Present Status and Future Concerns', *Comprehensive Reviews in Food Science and Food Safety*, 9(1), pp. 57–81.

Blatchly, R.A., Delen, Z. and O'Hara, P.B. (2014) 'Making Sense of Olive Oil: Simple Experiments To Connect Sensory Observations with the Underlying Chemistry', *Journal of Chemical Education*, 91(10) American Chemical Society, pp. 1623–1630.

Blum, M.-M. and John, H. (2012) 'Historical perspective and modern applications of Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy (ATR-FTIR)', *Drug Testing and Analysis*, 4(3–4), pp. 298–302.

Böcker, U., Ofstad, R., Wu, Z., Bertram, H.C., Sockalingum, G.D., Manfait, M., Egelanddal, B. and Kohler, A. (2007) 'Revealing covariance structures in fourier transform infrared and Raman microspectroscopy spectra: a study on pork muscle fiber tissue subjected to different processing parameters', *Applied Spectroscopy*, 61(10), pp. 1032–1039.

Bogard, K., Ludwig, T.D., Staats, C. and Kretschmer, D. (2015) 'An Industry's Call to Understand the Contingencies Involved in Process Safety: Normalization of Deviance', *Journal of Organizational Behavior Management*, 35(1–2) Routledge, pp. 70–80.

Bogdanov, S., Jurendic, T., Sieber, R. and Gallmann, P. (2008) 'Honey for Nutrition and Health: A Review', *Journal of the American College of Nutrition*, 27(6) Taylor & Francis, pp. 677–689.

Bosch, S., de Menezes, R.X., Pees, S., Wintjens, D.J., Seinen, M., Bouma, G., Kuyvenhoven, J., Stokkers, P.C.F., de Meij, T.G.J. and de Boer, N.K.H. (2022) 'Electronic Nose Sensor Drift Affects Diagnostic Reliability and Accuracy of Disease-Specific Algorithms', *Sensors (Basel, Switzerland)*, 22(23), p. 9246.

Bougrini, M., Tahri, K., Haddi, Z., Saidi, T., El Bari, N. and Bouchikhi, B. (2014) 'Detection of Adulteration in Argan Oil by Using an Electronic Nose and a Voltammetric Electronic Tongue', *Journal of Sensors*, 2014 Hindawi, p. e245831.

Box, G.E.P. (1976) 'Science and Statistics', *Journal of the American Statistical Association*, 71(356), pp. 791–799.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32.

Breiman, L. (1996) 'Stacked regressions', *Machine Learning*, 24(1), pp. 49–64.

Brewer, S., Pearson, S., Godsiff, P. and Maull, R. (2021) *Food Data Trust: A framework for information sharing*. Zenodo. Available at: 10.5281/zenodo.4575565 (Accessed: 14 February 2024).

Bro, R. and Smilde, A.K. (2014) 'Principal component analysis', *Analytical Methods*, 6(9), pp. 2812–2831.

Brooks, C., Parr, L., Smith, J.M., Buchanan, D., Snioch, D. and Hebishy, E. (2021) 'A review of food fraud and food authenticity across the food supply chain, with an examination of the impact of the COVID-19 pandemic and Brexit on food industry', *Food Control*, 130, p. 108171.

Calvini, R. and Pigani, L. (2022) 'Toward the Development of Combined Artificial Sensing Systems for Food Quality Evaluation: A Review on the Application of Data Fusion of Electronic Noses, Electronic Tongues and Electronic Eyes', *Sensors*, 22(2) Multidisciplinary Digital Publishing Institute, p. 577.

Carranco, N., Farrés-Cebrián, M., Saurina, J. and Núñez, O. (2018) 'Authentication and Quantitation of Fraud in Extra Virgin Olive Oils Based on HPLC-UV Fingerprinting and Multivariate Calibration', *Foods*, 7(4) Multidisciplinary Digital Publishing Institute, p. 44.

Casino, F., Dasaklis, T.K. and Patsakis, C. (2019) 'A systematic literature review of blockchain-based applications: Current status, classification and open issues', *Telematics and Informatics*, 36, pp. 55–81.

Cevoli, C., Iaccheri, E., Fabbri, A. and Ragni, L. (2024) 'Data fusion of FT-NIR spectroscopy and Vis/NIR hyperspectral imaging to predict quality parameters of yellow flesh "Jintao" kiwifruit', *Biosystems Engineering*, 237, pp. 157–169.

Cheng, J., Sun, J., Yao, K., Xu, M., Tian, Y. and Dai, C. (2022) 'A decision fusion method based on hyperspectral imaging and electronic nose techniques for moisture content prediction in frozen-thawed pork', *LWT*, 165, p. 113778.

Chowdhury, N.J., Akbor, Md.A., Nahar, A. and Ali Shaikh, Md.A. (2024) 'Techniques for quantification of organochlorine pesticides from a validated method by using gas chromatography-electron capture detector', *Heliyon*, 10(14), p. e34548.

Clegg, I. (2020) 'Chapter 7 - Process analytical technology', in Riley, C. M., Rosanske, T. W. and Reid, G. (eds.) *Specification of Drug Substances and Products (Second Edition)*. Elsevier, pp. 149–173. Available at: 10.1016/B978-0-08-102824-7.00007-5 (Accessed: 21 June 2023).

Cottenet, G., Andrey, D. and Dubascoux, S. (2024) 'Evaluation of ED-XRF for the detection of inorganic adulterants in turmeric, paprika and oregano.', *Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment*, 41(5) England, pp. 461–467.

Cowell, W., Ireland, T., Vorhees, D. and Heiger-Bernays, W. (2017) 'Ground Turmeric as a Source of Lead Exposure in the United States', *Public Health Reports*, 132(3), pp. 289–293.

Creppy, E.E. (1999) 'Human Ochratoxycosis', *Journal of Toxicology: Toxin Reviews*, 18(3–4) Taylor & Francis, pp. 277–293.

DAFM (2013) *Equine DNA & Mislabelling of Processed Beef Investigation - Department of Agriculture, Food and the Marine*. Department of Agriculture, Food and the Marine. Available at: <https://library2.nics.gov.uk/pdf/dard/2014/EEXA.pdf> (Accessed: 16 March 2024).

DEFRA (2024) *United Kingdom Food Security Report 2024: Theme 5: Food Safety and Consumer Confidence - Department for Environment, Food & Rural Affairs* (DEFRA)., GOV.UK Available at: <https://www.gov.uk/government/statistics/united-kingdom-food-security-report-2024/united-kingdom-food-security-report-2024-theme-5-food-safety-and-consumer-confidence> (Accessed: 12 January 2025).

Defraeye, T., Tagliavini, G., Wu, W., Prawiranto, K., Schudel, S., Assefa Kerisima, M., Verboven, P. and Bühlmann, A. (2019) 'Digital twins probe into food cooling and biochemical quality changes for reducing losses in refrigerated supply chains', *Resources, Conservation and Recycling*, 149, pp. 778–794.

Dietrich, O. (2018) *FRESHDETECT – food safety – fast and reliable | FRESHDETECT Project | Fact Sheet | H2020., CORDIS | European Commission* Available at: <https://cordis.europa.eu/project/id/822732> (Accessed: 11 December 2023).

Djekic, I., Lorenzo, J.M., Munekata, P.E.S., Gagaoua, M. and Tomasevic, I. (2021) 'Review on characteristics of trained sensory panels in food science', *Journal of Texture Studies*, 52(4), pp. 501–509.

Dodd, S., Kevei, Z., Karimi, Z., Parmar, B., Franklin, D., Koidis, A. and Anastasiadi, M. (2025) 'Detection of sugar syrup adulteration in UK honey using DNA barcoding', *Food Control*, 167, p. 110772.

Du, C.J. and Sun, D.W. (2005) 'Comparison of three methods for classification of pizza topping using different colour space transformations', *Journal of Food Engineering* 68(3) Oxford: Elsevier Sci Ltd, pp. 277–287.

Elliott, C. (2014) '*Elliott Review into the Integrity and Assurance of Food Supply Networks – Final Report*', , p. 146.

Estelles-Lopez, L., Ropodi, A., Pavlidis, D., Fotopoulou, J., Gkousari, C., Peyrodie, A., Panagou, E., Nychas, G.-J. and Mohareb, F. (2017) 'An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling', *Food Research International*, 99, pp. 206–215.

Etienne, J., Campbell, L., Jager, E., Verstraeten, Y., Rigoni, M. and McEntaggart, K. (2020) *Barriers to and Enablers of the Reporting of Intelligence Regarding Food Crime*. ICF Consulting Services Ltd., UK Food Standards Agency.

European Commission (2023) *Geographical indications and quality schemes explained* - European Commission. Available at:

https://agriculture.ec.europa.eu/farming/geographical-indications-and-quality-schemes/geographical-indications-and-quality-schemes-explained_en (Accessed: 12 March 2025).

European Commission (2014a) '*Fighting food fraud*'

European Commission (2014b) *Test results - European Commission*. Available at: https://food.ec.europa.eu/food-safety/eu-agri-food-fraud-network/eu-coordinated-actions/horse-meat-2013-14/test-results_en (Accessed: 13 March 2025).

European Commission (2014c) *Horsemeat: one year after., European Commission* Available at: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_14_113 (Accessed: 5 December 2023).

European Commission (2003) *2003/623/EC: Commission Decision of 19 August 2003 concerning the development of an integrated computerised veterinary system known as Traces (notified under document number C(2003) 2983)*. OJ L. Available at: <http://data.europa.eu/eli/dec/2003/623/oj/eng> (Accessed: 13 March 2025).

European Commission (2013) *Questions and answers surrounding the horse meat scandal*. Available at: https://food.ec.europa.eu/safety/eu-agri-food-fraud-network/eu-coordinated-actions/horse-meat-2013-14/q-and_en (Accessed: 8 September 2024).

Farber, C., Li, J., Hager, E., Chemelewski, R., Mullet, J., Rogachev, A.Yu. and Kurouski, D. (2019) 'Complementarity of Raman and Infrared Spectroscopy for Structural Characterization of Plant Epicuticular Waxes', *ACS Omega*, 4(2) American Chemical Society, pp. 3700–3707.

FDA (2024) *Economically Motivated Adulteration (Food Fraud)*. Available at: <https://www.fda.gov/food/compliance-enforcement-food/economically-motivated-adulteration-food-fraud> (Accessed: 12 January 2025).

FDA (2020) *PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance.*, U.S. Food and Drug Administration - Center for Drug Evaluation and Research Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance> (Accessed: 3 September 2022).

Fengou, L.-C., Spyrelli, E., Lianou, A., Tsakanikas, P., Panagou, E.Z. and Nychas, G.-J.E. (2019) 'Estimation of Minced Pork Microbiological Spoilage through Fourier Transform Infrared and Visible Spectroscopy and Multispectral Vision Technology', *Foods*, 8(7), p. 238.

Ferguson, D.M. and Gerrard, D.E. (2014) 'Regulation of post-mortem glycolysis in ruminant muscle', *Animal Production Science*, 54(4), p. 464.

Fourati, H. (ed.) (2017) *Multisensor Data Fusion: From Algorithms and Architectural Design to Applications*. Boca Raton: CRC Press.

FSA (2019) *Insights and learnings from the Food Standards Agency (FSA) exploring the use of Blockchains - Food Standards Agency*. Available at: <https://www.food.gov.uk/sites/default/files/media/document/fsa-blockchain-first-steps-to-next-steps.pdf> (Accessed: 14 February 2024).

Galvez, J.F., Mejuto, J.C. and Simal-Gandara, J. (2018) 'Future challenges on the use of blockchain for food traceability analysis', *TrAC Trends in Analytical Chemistry*, 107, pp. 222–232.

Gan, Z., Yang, Y., Li, J., Wen, X., Zhu, M., Jiang, Y. and Ni, Y. (2016) 'Using sensor and spectral analysis to classify botanical origin and determine adulteration of raw honey', *Journal of Food Engineering*, 178, pp. 151–158.

Gao, B., Qin, Q., Xu, X., Han, L. and Liu, X. (2024) 'Enhancement of species-specific analysis for meat and bone meal by matrix fragments-related spectral fusion', *Vibrational Spectroscopy*, 130, p. 103644.

García-Miguel, E., Meza-Márquez, O.G., Osorio-Revilla, G., Téllez-Medina, D.I., Jiménez-Martínez, C., Cornejo-Mazón, M., Hernández-Martínez, D.M. and Gallardo-Velazquez, T. (2018) 'Detection of Cyanuric Acid and Melamine in Infant Formula Powders by Mid-FTIR Spectroscopy and Multivariate Analysis', *Journal of Food Quality*, 2018 Hindawi, p. e7926768.

Gerum, R.C., Erpenbeck, A., Krauss, P. and Schilling, A. (2020) 'Sparsity through evolutionary pruning prevents neuronal networks from overfitting', *Neural Networks*, 128, pp. 305–312.

Geurts, B.P., Engel, J., Rafii, B., Blanchet, L., Suppers, A., Szymańska, E., Jansen, J.J. and Buydens, L.M.C. (2016) 'Improving high-dimensional data fusion by exploiting the multivariate advantage', *Chemometrics and Intelligent Laboratory Systems*, 156, pp. 231–240.

Giusti, M.M. and Wrolstad, R.E. (2001) 'Characterization and Measurement of Anthocyanins by UV-Visible Spectroscopy', *Current Protocols in Food Analytical Chemistry*, 00(1), p. F1.2.1-F1.2.13.

Glahn, H.R. (1968) 'Canonical Correlation and Its Relationship to Discriminant Analysis and Multiple Regression' Available at: https://journals.ametsoc.org/view/journals/atasc/25/1/1520-0469_1968_025_0023_ccairt_2_0_co_2.xml (Accessed: 9 March 2025).

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. MIT Press.

- Gorbachev, V., Nikitina, M., Velina, D., Mutallibzoda, S., Nosov, V., Korneva, G., Terekhova, A., Artemova, E., Khashir, B., Sokolov, I., Dimitrieva, S. and Nikitin, I. (2022) 'Artificial Neural Networks for Predicting Food Antiradical Potential', *Applied Sciences*, 12(12) Multidisciplinary Digital Publishing Institute, p. 6290.
- Gossner, C.M.-E., Schlundt, J., Ben Embarek, P., Hird, S., Lo-Fo-Wong, D., Beltran, J.J.O., Teoh, K.N. and Tritscher, A. (2009) 'The Melamine Incident: Implications for International Food and Feed Safety', *Environmental Health Perspectives*, 117(12), pp. 1803–1808.
- Grieves, M. (2015) '*Digital Twin: Manufacturing Excellence through Virtual Factory Replication*'
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T. (2018) 'Recent advances in convolutional neural networks', *Pattern Recognition*, 77 Elsevier, pp. 354–377.
- Guillén, M.D. and Cabo, N. (1997) 'Infrared spectroscopy in the study of edible oils and fats', *Journal of the Science of Food and Agriculture*, 75(1), pp. 1–11.
- Guzmán, E., Baeten, V., Pierna, J.A.F. and García-Mesa, J.A. (2015) 'Evaluation of the overall quality of olive oil using fluorescence spectroscopy', *Food Chemistry*, 173, pp. 927–934.
- Hage, D.S. (2018) 'Chromatography', in *Principles and Applications of Clinical Mass Spectrometry*. Elsevier, pp. 1–32. Available at: 10.1016/B978-0-12-816063-3.00001-3 (Accessed: 28 December 2023).
- Hagen, N.A. and Kudenov, M.W. (2013) 'Review of snapshot spectral imaging technologies', *Optical Engineering*, 52(9) SPIE, p. 090901.
- Hall, D.L. and Llinas, J. (1997) 'An introduction to multisensor data fusion', *Proceedings of the IEEE*, 85(1), pp. 6–23.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. and Scholkopf, B. (1998) 'Support vector machines', *IEEE Intelligent Systems and their Applications*, 13(4), pp. 18–28.
- Hershberger, J., Morales, N., Simoes, C.C., Ellerbrock, B., Bauchet, G., Mueller, L.A. and Gore, M.A. (2021) 'Making waves in Breedbase: An integrated spectral data storage and analysis pipeline for plant breeding programs', *The Plant Phenome Journal*, 4(1), p. e20012.
- Hinshaw, J. (1992) 'A Compendium of Gc Terms and Techniques', *LC GC-MAGAZINE OF SEPARATION SCIENCE*, 10(7) Duluth: Advanstar Communications, pp. 516–22.

Hollands, T., Martindale, W., Swainson, M. and Keogh, J.G. (2018) 'Blockchain or bust for the food industry?', *Food Science and Technology*, 32(4), pp. 40–45.

Huang, J. and Elmasni, D. (2007) '*Analysis of Aflatoxins Using Fluorescence Detection*'

Huang, W., Yin, M., Xia, J. and Zhang, X. (2024) 'A review of cross-scale and cross-modal intelligent sensing and detection technology for food quality: Mechanism analysis, decoupling strategy and integrated applications', *Trends in Food Science & Technology*, 151, p. 104646.

Huang, Y., Kangas, L.J. and Rasco, B.A. (2007) 'Applications of artificial neural networks (ANNs) in food science', *Critical Reviews in Food Science and Nutrition*, 47(2), pp. 113–126.

IBM (2021) *Majid Al Futtaim Taps IBM Food Trust to Deliver Food Traceability Across Carrefour Stores Using Blockchain Technology.*, IBM Newsroom Available at: <https://newsroom.ibm.com/2021-02-22-Majid-Al-Futtaim-Taps-IBM-Food-Trust-to-Deliver-Food-Traceability-Across-Carrefour-Stores-Using-Blockchain-Technology> (Accessed: 14 February 2024).

Iman, M., Arabnia, H.R. and Rasheed, K. (2023) 'A Review of Deep Transfer Learning and Recent Advancements', *Technologies*, 11(2) Multidisciplinary Digital Publishing Institute, p. 40.

Isnanto, R.R. (2011) 'Comparison on Several Smoothing Methods in Nonparametric Regression', *Jurnal Sistem Komputer*, Vol.1(No.1) Department of Computer Engineering, pp. 41–47.

ISO (2021) *ISO 23247-1:2021(en), Automation systems and integration — Digital twin framework for manufacturing — Part 1: Overview and general principles.* Available at: <https://www.iso.org/obp/ui/en/#iso:std:iso:23247:-1:ed-1:v1:en> (Accessed: 14 November 2023).

Jeong, H.-N., Kwon, R.H., Kim, Y., Yoo, S.-H., Yoo, S.M. and Wee, C.-D. (2023) 'Gas Chromatography with Flame-Ionization Detection-Based Analysis of Sugar Contents in Korean Agricultural Products for Patients with Galactosemia', *Foods (Basel, Switzerland)*, 12(5), p. 1104.

Jia, W., Qin, Y. and Zhao, C. (2024) 'Rapid detection of adulterated lamb meat using near infrared and electronic nose: A F1-score-MRE data fusion approach', *Food Chemistry*, 439, p. 138123.

Kamath, R. (2018) 'Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM', *The Journal of The British Blockchain Association*, 1(1) The British Blockchain Association Available at: [10.31585/jbba-1-1-\(10\)2018](https://doi.org/10.31585/jbba-1-1-(10)2018) (Accessed: 14 March 2025).

Keogh, E. and Mueen, A. (2017) 'Curse of Dimensionality', in Sammut, C. and Webb, G. I. (eds.) *Encyclopedia of Machine Learning and Data Mining*. Boston,

MA: Springer US, pp. 314–315. Available at: 10.1007/978-1-4899-7687-1_192 (Accessed: 3 January 2024).

Kim, S.H., Hwang, J.-H. and Lee, K.-G. (2011) 'Analysis of acrylamide using gas chromatography-nitrogen phosphorus detector (GC-NPD)', *Food Science and Biotechnology*, 20(3), pp. 835–839.

Kuhn, M. and Johnson, K. (2013) *Applied Predictive Modeling*. Springer.

Kyriakidis, N.B. and Skarkalis, P. (2000) 'Fluorescence spectra measurement of olive oil and other vegetable oils', *Journal of AOAC International*, 83(6), pp. 1435–1439.

Laan, M.J. van der, Polley, E.C. and Hubbard, A.E. (2007) 'Super Learner', *Statistical Applications in Genetics and Molecular Biology*, 6(1) De Gruyter Available at: 10.2202/1544-6115.1309 (Accessed: 7 September 2024).

Lancet, T. (2009) 'Melamine and food safety in China', *The Lancet*, 373(9661) Elsevier, p. 353.

Lazou, T.P., Gelasakis, A.I., Chaintoutis, S.C., Iossifidou, E.G. and Dovas, C.I. (2021) 'Method-Dependent Implications in Foodborne Pathogen Quantification: The Case of *Campylobacter coli* Survival on Meat as Comparatively Assessed by Colony Count and Viability PCR', *Frontiers in Microbiology*, 12 Frontiers Available at: 10.3389/fmicb.2021.604933 (Accessed: 12 March 2025).

Lehotay, S.J. and Hajšlová, J. (2002) 'Application of gas chromatography in food analysis', *TrAC Trends in Analytical Chemistry*, 21(9), pp. 686–697.

Li, J., Zuo, L., Su, T. and Guo, Z. (2021a) 'The Choice of Kernel Function for One-Class Support Vector Machine', *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, pp. 195–198. Available at: 10.1109/ICBASE53849.2021.00044 (Accessed: 2 January 2024).

Li, S., Wang, Y., Tacken, G.M.L., Liu, Y. and Sijtsema, S.J. (2021b) 'Consumer trust in the dairy value chain in China: The role of trustworthiness, the melamine scandal, and the media', *Journal of Dairy Science*, 104(8), pp. 8554–8567.

Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2012) 'Isolation-Based Anomaly Detection', *ACM Transactions on Knowledge Discovery from Data*, 6(1), p. 3:1-3:39.

Lord, N., Spencer, J., Albanese, J. and Flores Elizondo, C. (2017) 'In pursuit of food system integrity: the situational prevention of food fraud enterprise', *European Journal on Criminal Policy and Research*, 23(4), pp. 483–501.

Lotta, F. and Bogue, J. (2015) 'Defining Food Fraud in the Modern Supply Chain', *European Food and Feed Law Review*, 10(2) Lexxion Verlagsgesellschaft mbH, pp. 114–122.

Lunn, G. and Hellwig, L.C. (1998) *Wiley-VCH - Handbook of Derivatization Reactions for HPLC*. Available at: <https://www.wiley-vch.de/en/areas-interest/natural-sciences/handbook-of-derivatization-reactions-for-hplc-978-0-471-23889-8> (Accessed: 29 December 2023).

Luor, D.-C. (2015) 'A comparative assessment of data standardization on support vector machine for classification problems', *Intelligent Data Analysis*, 19(3) IOS Press, pp. 529–546.

Luparelli, A., Losito, I., De Angelis, E., Pilolli, R., Lambertini, F. and Monaci, L. (2022) 'Tree Nuts and Peanuts as a Source of Beneficial Compounds and a Threat for Allergic Consumers: Overview on Methods for Their Detection in Complex Food Products', *Foods*, 11(5), p. 728.

Mahalanobis, P.C. (1936) 'Reprint of: Mahalanobis, P.C. (1936) "On the Generalised Distance in Statistics."', 80(1), pp. 1–7.

Marchese, A. and Tomarchio, O. (2022) 'A Blockchain-Based System for Agri-Food Supply Chain Traceability Management', *SN Computer Science*, 3(4), p. 279.

Márquez, C., López, M.I., Ruisánchez, I. and Callao, M.P. (2016) 'FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud', *Talanta*, 161, pp. 80–86.

Marriott, P.J. (2004) 'Chapter 8 Gas chromatography', in *Journal of Chromatography Library*. Elsevier, pp. 319–368. Available at: 10.1016/S0301-4770(04)80014-9 (Accessed: 28 December 2023).

Martens, H. and Naes, T. (1992) *Multivariate calibration*. John Wiley & Sons.

Maruchek, A., Greis, N., Mena, C. and Cai, L. (2011) 'Product safety and security in the global supply chain: Issues, challenges and research opportunities', *Journal of Operations Management*, 29(7), pp. 707–720.

Mauer, L.J., Chernyshova, A.A., Hiatt, A., Deering, A. and Davis, R. (2009) 'Melamine Detection in Infant Formula Powder Using Near- and Mid-Infrared Spectroscopy', *Journal of Agricultural and Food Chemistry*, 57(10) American Chemical Society, pp. 3974–3980.

McLachlan, G.J., Lee, S.X. and Rathnayake, S.I. (2019) 'Finite Mixture Models', *Annual Review of Statistics and Its Application*, 6(1), pp. 355–378.

Mekhrengin, M.V., Meshkovskii, I.K., Kaftyreva, L.A., Guryev, V.I. and Pogorelaya, D.A. (2018) 'Fluorescent method of bacterial contamination control on meat surface', *Journal of Physics: Conference Series*, 1124, p. 031012.

Meshram, V. and Patil, K. (2022) 'FruitNet: Indian fruits image dataset with quality for machine learning applications', *Data in Brief*, 40, p. 107686.

- Mildner-Szkudlarz, S. and Jeleń, H.H. (2008) 'The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil', *Food Chemistry*, 110(3), pp. 751–761.
- Min, S., Lee, B. and Yoon, S. (2017) 'Deep learning in bioinformatics', *Briefings in Bioinformatics*, 18(5), pp. 851–869.
- Mishra, H.N. and Das, C. (2003) 'A review on biological control and metabolism of aflatoxin', *Critical Reviews in Food Science and Nutrition*, 43(3), pp. 245–264.
- Mohrig, J.R., Alberg, D., Holifmeister, G., Schatz, P.F. and Hammond, C.N. (2024) *Laboratory techniques in organic chemistry*. 4th edn. W.H. Freeman.
- Morshed, Md.S., Ahmed, S., Ahmed, T., Islam, M.U. and Ashikur Rahman, A.B.M. (2022) 'Fruit Quality Assessment with Densely Connected Convolutional Neural Network', *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, pp. 1–4. Available at: 10.1109/ICECE57408.2022.10088873 (Accessed: 30 March 2025).
- Much, D.G., Alcaraz, M.R., Camiña, J.M., Goicoechea, H.C. and Azcarate, S.M. (2024) 'Leveraging the performance of conventional spectroscopic techniques through data fusion approaches in high-quality edible oil adulteration analyses', *Talanta Open*, 9, p. 100313.
- Nakamoto, S. (2008) *Bitcoin: A peer-to-peer electronic cash system*.
- Nayak, V.S., Tan, Z., Ihnat, P.M., Russell, R.J. and Grace, M.J. (2012) 'Evaporative Light Scattering Detection Based HPLC Method for the Determination of Polysorbate 80 in Therapeutic Protein Formulations', *Journal of chromatographic science*, 50(1), pp. 21–25.
- Nešić, K., Stojanović, D. and Baltić, Ž.M. (2017) 'Authentication of meat and meat products vs. detection of animal species in feed – what is the difference?', *IOP Conference Series: Earth and Environmental Science*, 85(1) IOP Publishing, p. 012043.
- Norli, H.R., Christiansen, A. and Deribe, E. (2011) 'Application of QuEChERS method for extraction of selected persistent organic pollutants in fish tissue and analysis by gas chromatography mass spectrometry', *Journal of Chromatography A*, 1218(41), pp. 7234–7241.
- Nychas, G.-J., Sims, E., Tsakanikas, P. and Mohareb, F. (2021) 'Data Science in the Food Industry', *Annual Review of Biomedical Data Science*, 4(1), pp. 341–367.
- Nychas, G.-J.E. and Tassou, C.C. (1997) 'Spoilage Processes and Proteolysis in Chicken as Detected by HPLC', *Journal of the Science of Food and Agriculture*, 74(2), pp. 199–208.

O'Brien, S.J., Iturriza-Gomara, M., Knight, A., Lees, D., Lowther, J., Cook, N., D'Agostino, M., Allen, D.J., Elviss, N., Fox, A., Hunter, P., Maas, J., Rushton, S. and Sanderson, R. (2020) *FS101040: Assessing the Contribution Made by the Food Chain to the Burden of UK-Acquired Norovirus Infection*. University of Liverpool, Leatherhead Food Research, Cefas, Fera, Public Health England, University of East Anglia, Newcastle University.

Ojemaye, C.Y. and Petrik, L. (2019) 'Occurrences, levels and risk assessment studies of emerging pollutants (pharmaceuticals, perfluoroalkyl and endocrine disrupting compounds) in fish samples from Kalk Bay harbour, South Africa', *Environmental Pollution*, 252, pp. 562–572.

Olivieri, A.C. (2008) 'Analytical Advantages of Multivariate Data Processing. One, Two, Three, Infinity?', *Analytical Chemistry*, 80(15) American Chemical Society, pp. 5713–5720.

Ouchemoukh, S., Schweitzer, P., Bachir Bey, M., Djoudad-Kadji, H. and Louaileche, H. (2010) 'HPLC sugar profiles of Algerian honeys', *Food Chemistry*, 121(2), pp. 561–568.

Paré, J.R.J. and Bélanger, J.M.R. (1997) *Instrumental Methods in Food Analysis*. Elsevier.

Pearson, S., May, D., Leontidis, G., Swainson, M., Brewer, S., Bidaut, L., Frey, J.G., Parr, G., Maull, R. and Zisman, A. (2019) 'Are Distributed Ledger Technologies the panacea for food traceability?', *Global Food Security*, 20, pp. 145–149.

Perestam, A.T., Fujisaki, K.K., Nava, O. and Hellberg, R.S. (2017) 'Comparison of real-time PCR and ELISA-based methods for the detection of beef and pork in processed meat products', *Food Control*, 71, pp. 346–352.

Popa, S., Milea, M.S., Boran, S., Nițu, S.V., Moșoarcă, G.E., Vancea, C. and Lazău, R.I. (2020) 'Rapid adulteration detection of cold pressed oils with their refined versions by UV–Vis spectroscopy', *Scientific Reports*, 10(1) Nature Publishing Group, p. 16100.

Poulli, K.I., Mousdis, G.A. and Georgiou, C.A. (2006) 'Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower oil', *Analytical and Bioanalytical Chemistry*, 386(5), pp. 1571–1575.

Ren, G., Gan, N., Song, Y., Ning, J. and Zhang, Z. (2021) 'Evaluating Congou black tea quality using a lab-made computer vision system coupled with morphological features and chemometrics', *Microchemical Journal*, 160, p. 105600.

Reusch, W. (2013) *Infrared Spectroscopy*. Available at: <https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/infrared/infrared.htm> (Accessed: 8 January 2024).

Reuters (2019) Carrefour says blockchain tracking boosting sales of some products, *Reuters*, 3 June. Available at: <https://www.reuters.com/article/business/carrefour-says-blockchain-tracking-boosting-sales-of-some-products-idUSKCN1T42C1/> (Accessed: 14 February 2024).

Rizvi, T.S., Mabood, F., Ali, L., Al-Broumi, M., Al Rabani, H.K.M., Hussain, J., Jabeen, F., Manzoor, S. and Al-Harrasi, A. (2018) 'Application of NIR Spectroscopy Coupled with PLS Regression for Quantification of Total Polyphenol Contents from the Fruit and Aerial Parts of *Citrullus colocynthis*', *Phytochemical Analysis*, 29(1), pp. 16–22.

Robert, C., Jessep, W., Sutton, J.J., Hicks, T.M., Loeffen, M., Farouk, M., Ward, J.F., Bain, W.E., Craigie, C.R., Fraser-Miller, S.J. and Gordon, K.C. (2021) 'Evaluating low- mid- and high-level fusion strategies for combining Raman and infrared spectroscopy for quality assessment of red meat', *Food Chemistry*, 361, p. 130154.

Rodriguez-Saona, L.E. and Allendorf, M.E. (2011) 'Use of FTIR for Rapid Authentication and Detection of Adulteration of Food', *Annual Review of Food Science and Technology*, 2(1), pp. 467–483.

Rohman, A., Ghazali, M.A.B., Windarsih, A., Irnawati, Riyanto, S., Yusof, F.M. and Mustafa, S. (2020) 'Comprehensive Review on Application of FTIR Spectroscopy Coupled with Chemometrics for Authentication Analysis of Fats and Oils in the Food Products', *Molecules*, 25(22) Multidisciplinary Digital Publishing Institute, p. 5485.

Rojas-Reyes, J.J., Rivera-Cadavid, L. and Peña-Orozco, D.L. (2024) 'Disruptions in the food supply chain: A literature review', *Heliyon*, 10(14) Elsevier Available at: [10.1016/j.heliyon.2024.e34730](https://doi.org/10.1016/j.heliyon.2024.e34730) (Accessed: 12 March 2025).

Roussel, S., Bellon-Maurel, V., Roger, J.-M. and Grenier, P. (2003) 'Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties', *Chemometrics and Intelligent Laboratory Systems*, 65(2), pp. 209–219.

Roy, M. and Yadav, B.K. (2022) 'Electronic nose for detection of food adulteration: a review', *Journal of Food Science and Technology*, 59(3), pp. 846–858.

Sahar, A. and Dufour, É. (2014) 'Use of Fourier transform-infrared spectroscopy to predict spoilage bacteria on aerobically stored chicken breast fillets', *LWT - Food Science and Technology*, 56(2), pp. 315–320.

Salghi, R., Armbruster, W. and Schwack, W. (2014) 'Detection of argan oil adulteration with vegetable oils by high-performance liquid chromatography–evaporative light scattering detection', *Food Chemistry*, 153, pp. 387–392.

Salihah, N.T., Hossain, M.M., Lubis, H. and Ahmed, M.U. (2016) 'Trends and advances in food analysis by real-time polymerase chain reaction', *Journal of Food Science and Technology*, 53(5), pp. 2196–2209.

Salina, B. and Malathi, P. (2014) 'FPGA implementation of data fusion algorithm for object localization', *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*., pp. 1–5. Available at: 10.1109/ICRAIE.2014.6909200 (Accessed: 7 September 2024).

Savelli, C.J., Bradshaw, A., Ben Embarek, P. and Mateus, C. (2019) 'The FAO/WHO International Food Safety Authorities Network in Review, 2004–2018: Learning from the Past and Looking to the Future', *Foodborne Pathogens and Disease*, 16(7) Mary Ann Liebert, Inc., publishers, pp. 480–488.

Savelli, C.J. and Mateus, C. (2020) 'Looking Inside the International Food Safety Authorities Network Community Website', *Journal of Food Protection*, 83(11), pp. 1889–1899.

Savelli, C.J. and Mateus, C. (2021) 'Exploring the International Food Safety Authorities Network as a Community of Practice: Results from a Global Survey of Network Members', *Journal of Food Protection*, 84(2), pp. 262–274.

Savitzky, Abraham. and Golay, M.J.E. (1964) 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures.', *Analytical Chemistry*, 36(8), pp. 1627–1639.

Sayago, A., Morales, M.T. and Aparicio, R. (2004) 'Detection of hazelnut oil in virgin olive oil by a spectrofluorimetric method', *European Food Research and Technology*, 218(5), pp. 480–483.

Schäfer, A.M. and Zimmermann, H.G. (2006) 'Recurrent Neural Networks Are Universal Approximators', Kollias, S. D., Stafylopatis, A., Duch, W. and Oja, E. (eds.) *Artificial Neural Networks – ICANN 2006*. Berlin, Heidelberg: Springer. Lecture Notes in Computer Science, pp. 632–640.

Segura-Borrego, M.P., Martín-Gómez, A., Ríos-Reina, R., Cardador, M.J., Morales, M.L., Arce, L. and Callejón, R.M. (2022) 'A non-destructive sampling method for food authentication using gas chromatography coupled to mass spectrometry or ion mobility spectrometry', *Food Chemistry*, 373, p. 131540.

Sellar, R.G. and Boreman, G.D. (2005) 'Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer', *Applied Optics*, 44(9) Optica Publishing Group, pp. 1614–1624.

Shehata, M., Dodd, S., Mosca, S., Matousek, P., Parmar, B., Kevei, Z. and Anastasiadi, M. (2024) 'Application of Spatial Offset Raman Spectroscopy (SORS) and Machine Learning for Sugar Syrup Adulteration Detection in UK Honey', *Foods*, 13(15)

Sheng, L.-Q., Chen, M.-M., Chen, S.-S., Du, N.-N., Liu, Z.-D., Song, C.-F. and Qiao, R. (2013) 'High-performance liquid chromatography with fluorescence detection for the determination of nitrofuran metabolites in pork muscle', *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment*, 30(12), pp. 2114–2122.

Siemens (2017) MindSphere Whitepaper. Available at: https://www.plm.automation.siemens.com/media/global/en/Siemens_MindSphere_Whitepaper_tcm27-8639.pdf (Accessed: 10 October 2024).

Singh, J. and Mehta, A. (2020) 'Rapid and sensitive detection of mycotoxins by advanced and emerging analytical methods: A review', *Food Science & Nutrition*, 8(5), pp. 2183–2204.

Smeesters, L., Kuntzel, T., Thienpont, H. and Guilbert, L. (2023) 'Handheld Fluorescence Spectrometer Enabling Sensitive Aflatoxin Detection in Maize', *Toxins*, 15(6) Multidisciplinary Digital Publishing Institute, p. 361.

Smilde, A.K., Måge, I., Næs, T., Hankemeier, T., Lips, M.A., Kiers, H.A.L., Acar, E. and Bro, R. (2017) 'Common and distinct components in data fusion', *Journal of Chemometrics*, 31(7), p. e2900.

Smilde, A.K., van der Werf, M.J., Bijlsma, S., van der Werff-van der Vat, B.J.C. and Jellema, R.H. (2005) 'Fusion of mass spectrometry-based metabolomics data', *Analytical Chemistry*, 77(20), pp. 6729–6736.

Smith, R.M. (2004) 'Chapter 2 Column liquid chromatography', in *Journal of Chromatography Library*. Elsevier, pp. 95–138. Available at: 10.1016/S0301-4770(04)80008-3 (Accessed: 28 December 2023).

Smolinska, A., Engel, J., Szymanska, E., Buydens, L. and Blanchet, L. (2019) 'General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences', in *Data Handling in Science and Technology*. Elsevier, pp. 51–79. Available at: 10.1016/B978-0-444-63984-4.00003-X (Accessed: 26 March 2023).

Snyder, A.B., Martin, N. and Wiedmann, M. (2024) 'Microbial food spoilage: impact, causative agents and control strategies', *Nature Reviews Microbiology*, 22(9) Nature Publishing Group, pp. 528–542.

Soon, J.M., Manning, L. and Smith, R. (2019) 'Advancing understanding of pinch-points and crime prevention in the food supply chain', *Crime Prevention and Community Safety*, 21(1), pp. 42–60.

Spink, J., Chen, W., Zhang, G. and Speier-Pero, C. (2019) 'Introducing the Food Fraud Prevention Cycle (FFPC): A dynamic information management and strategic roadmap', *Food Control*, 105, pp. 233–241.

Spyrelli, E.D., Ozcan, O., Mohareb, F., Panagou, E.Z. and Nychas, G.-J.E. (2021) 'Spoilage assessment of chicken breast fillets by means of fourier transform

infrared spectroscopy and multispectral image analysis', *Current Research in Food Science*, 4, pp. 121–131.

Srai, J.S., Settanni, E., Tsolakis, N. and Aulakh, P.K. (2019) 'Supply Chain Digital Twins: Opportunities and Challenges Beyond the Hype', , p. 6.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting'

Stadler, R.H., Tran, L.-A., Cavin, C., Zbinden, P. and Konings, E.J.M. (2016) 'Analytical Approaches to Verify Food Integrity: Needs and Challenges', *Journal of AOAC INTERNATIONAL*, 99(5), pp. 1135–1144.

Ståhle, L. and Wold, S. (1987) 'Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study', *Journal of Chemometrics*, 1(3), pp. 185–196.

Stez, G. (2015) *Flame Ionization vs. Thermal Conductivity.*, *AnalyteGuru* Available at: <https://www.thermofisher.com/blog/analyteguru/detector-detective-part-i-flame-ionization-vs-thermal/> (Accessed: 27 December 2023).

Stristy, A. (2021) *Blockchain in the food supply chain - What does the future look like?*. Available at: https://tech.walmart.com/content/walmart-global-tech/en_us/blog/post/blockchain-in-the-food-supply-chain.html (Accessed: 4 February 2025).

Sutton, R.S. and Barto, A.G. (2018) *Reinforcement Learning: An Introduction*. Bradford Books.

Swarnakar, S., Chowhan, P. and Chakraborty, A.P. (2024) 'An Overview of Food-Borne Pathogenic Microbes: Classification, Mode of Actions, and Recent Advances in Strategies for Management of Food-Borne Pathogens', in Chakraborty, R., Mathur, P. and Roy, S. (eds.) *Food Production, Diversity, and Safety Under Climate Change*. Cham: Springer Nature Switzerland, pp. 307–318. Available at: 10.1007/978-3-031-51647-4_25 (Accessed: 12 January 2025).

Tenenhaus, A. and Tenenhaus, M. (2014) 'Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis', *European Journal of Operational Research*, 238(2), pp. 391–403.

Tian, F. (2016) 'An agri-food supply chain traceability system for China based on RFID & blockchain technology', *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*. Kunming, China: IEEE, pp. 1–6. Available at: 10.1109/ICSSSM.2016.7538424 (Accessed: 14 March 2025).

Tian, X., Wang, J., Ma, Z., Li, M. and Wei, Z. (2019) 'Combination of an E-Nose and an E-Tongue for Adulteration Detection of Minced Mutton Mixed with Pork', *Journal of Food Quality*, 2019 Hindawi, p. e4342509.

Tiedink, H.G.M., Hissink, A.M., Lodema, S.M., van Broekhoven, L.W. and Jongen, W.M.F. (1990) 'Several known indole compounds are not important precursors of direct mutagenic N-nitroso compounds in green cabbage', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 232(2), pp. 199–207.

Tohidi, M., Ghasemi-Varnamkhasti, M., Ghafarinia, V., Bonyadian, M. and Mohtasebi, S.S. (2018) 'Development of a metal oxide semiconductor-based artificial nose as a fast, reliable and non-expensive analytical technique for aroma profiling of milk adulteration', *International Dairy Journal*, 77, pp. 38–46.

Tompkins, C. and Perkins, E.G. (1999) 'The evaluation of frying oils with the p-Anisidine value', *Journal of the American Oil Chemists' Society*, 76(8), pp. 945–947.

Treiblmaier, H. (2018) 'The impact of the blockchain on the supply chain: a theory-based research framework and a call for action', *Supply Chain Management: An International Journal*, 23(6), pp. 545–559.

Trienekens, J. and Zuurbier, P. (2008) 'Quality and safety standards in the food industry, developments and challenges', *International Journal of Production Economics*, 113(1), pp. 107–122.

Tsakanikas, P., Pavlidis, D., Panagou, E. and Nychas, G.-J. (2016) 'Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples', *Talanta*, 161, pp. 606–614.

Tsuzuki, W. and Ushida, K. (2009) 'Preparative separation of cis- and trans-isomers of unsaturated fatty acid methyl esters contained in edible oils by reversed-phase high-performance liquid chromatography', *Lipids*, 44(4), pp. 373–379.

Van den Berg, F.W., Hoefsloot, H.C. and Smilde, A.K. (2002) 'Selection of optimal process analyzers for plant-wide monitoring', *Analytical chemistry*, 74(13) ACS Publications, pp. 3105–3111.

Van Deun, K., Smilde, A.K., van der Werf, M.J., Kiers, H.A. and Van Mechelen, I. (2009) 'A structured overview of simultaneous component based data integration', *BMC Bioinformatics*, 10(1), p. 246.

Vapnik, V. and Lerner, A.Ya. (1963) 'Pattern Recognition Using Generalized Portrait Method.', *Automation and Remote Control*, 24(6), pp. 774–780.

Velkoska-Markovska, L., Petanovska-Ilievska, B. and Markovski, A. (2018) 'Application of High Performance Liquid Chromatography to the Analysis of Pesticide Residues in Apple Juice', *Contemporary Agriculture*, 67, pp. 93–102.

Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S*. Fourth. New York: Springer.

Videometer, Denmark (2021) *VideometerLab - multispectral imaging technology*. Available at: <https://videometer.com/videometerlab/> (Accessed: 18 March 2024).

Wang, K., Wang, M., Kou, Y., Guo, C., Ma, X., Tian, B., Qi, Y., Wang, W. and Wang, J. (2022) 'Determination of nitrofuran metabolites in meat products by UHPLC-fluorescence with ultrasonic-assisted derivatization', *Journal of Food Composition and Analysis*, 107, p. 104375.

Wang, Q., Li, L., Ding, W., Zhang, D., Wang, J., Reed, K. and Zhang, B. (2019) 'Adulterant identification in mutton by electronic nose and gas chromatography-mass spectrometer', *Food Control*, 98, pp. 431–438.

Whipple, J.M., Voss, M.D. and Closs, D.J. (2009) 'Supply chain security practices in the food industry: Do firms operating globally and domestically differ?', *International Journal of Physical Distribution & Logistics Management*, 39(7), pp. 574–594.

White, F.E. (1987) *Data Fusion Lexicon, Technical Panel For C3*.

WHO (2024) Food safety., Food Safety Fact Sheet Available at: <https://www.who.int/news-room/fact-sheets/detail/food-safety> (Accessed: 12 October 2024).

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I. and van den Berg, R.A. (2011) 'Simultaneous analysis of coupled data matrices subject to different amounts of noise', *British Journal of Mathematical and Statistical Psychology*, 64(2) United Kingdom: Wiley-Blackwell Publishing Ltd., pp. 277–290.

Wold, S., Sjöström, M. and Eriksson, L. (2001) 'PLS-regression: a basic tool of chemometrics', *Chemometrics and Intelligent Laboratory Systems*, 58(2), pp. 109–130.

Wu, D., Yang, H., Chen, X., He, Y. and Li, X. (2008) 'Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine', *Journal of Food Engineering*, 88(4), pp. 474–483.

Wu, Q., Xu, J. and Xu, H. (2019) 'Discrimination of aflatoxin B1 contaminated pistachio kernels using laser induced fluorescence spectroscopy', *Biosystems Engineering*, 179, pp. 22–34.

Xiu, C. and Klein, K.K. (2010) 'Melamine in milk products in China: Examining the factors that led to deliberate use of the contaminant', *Food Policy*, 35(5), pp. 463–470.

Yan, J., Guo, X., Duan, S., Jia, P., Wang, L., Peng, C. and Zhang, S. (2015) 'Electronic Nose Feature Extraction Methods: A Review', *Sensors*, 15(11) Multidisciplinary Digital Publishing Institute, pp. 27804–27831.

Yang, L.G.L., Brewster, R.K., Donnell, M.T. and Hirani, R.N. (2022) 'Risk characterisation of milk protein contamination in milk-alternative ice cream

products sold as frozen desserts in the United States', *Food Additives and Contaminant Part A-Chemistry Analysis Control Exposure & Risk Assessment*, 39(1) Abingdon: Taylor & Francis Ltd, pp. 1–13.

Yin, X. and Zhao, S. (2013) 'Hyperspectral Recognition of Processing Tomato Early Blight Based on GA and SVM', Wang, Y., Tan, L. and Zhou, J. (eds.) *Fifth International Conference On Machine Vision (ICMV 2012): Computer Vision, Image Analysis And Processing*. Bellingham: Spie-Int Soc Optical Engineering. Proceedings of SPIE, Vol.8783, p. 87831D. Available at: 10.1117/12.2014142 (Accessed: 2 January 2024).

Yli-Huomo, J., Ko, D., Choi, S., Park, S. and Smolander, K. (2016) 'Where is current research on blockchain technology?—A systematic review', *PLoS ONE*, 11(10)

Yousefi, H., Su, H.-M., Imani, S.M., Alkhaldi, K., M. Filipe, C.D. and Didar, T.F. (2019) 'Intelligent Food Packaging: A Review of Smart Sensing Technologies for Monitoring Food Quality', *ACS Sensors*, 4(4) American Chemical Society, pp. 808–821.

Zeaiter, M. and Rutledge, D. (2009) '3.04 - Preprocessing Methods', in Brown, S. D., Tauler, R. and Walczak, B. (eds.) *Comprehensive Chemometrics*. Oxford: Elsevier, pp. 121–231. Available at: 10.1016/B978-044452701-1.00074-0 (Accessed: 31 December 2023).

Zeng, L., Zhou, L., Pan, P.-L. and Fowler, G. (2018) 'Coping with the milk scandal: A staged approach to crisis communication strategies during China's largest food safety crisis', *Journal of Communication Management*, 22(4), pp. 432–450.

2 Application of Spectroscopic Sensor Technologies for Meat Spoilage Prediction

Samuel Heffer¹, Fady Mohareb¹, Maria Anastasiadi¹, George-John Nychas²

¹ Bioinformatics Group, Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, Bedfordshire MK43 0AL, United Kingdom

² Agricultural University of Athens, 11855 Athens, Greece

³ International Joint Research Lab (China and Greece) of Digital Transformation as an Enabler for Food Safety and Sustainability, Tai'an 271018, China

2.1 Abstract

High-throughput and portable sensor technologies are increasingly used in food production/distribution tasks as rapid and non-invasive platforms offering real-time or near real-time monitoring of quality and safety. These are often coupled with analytical techniques including Machine Learning, for estimation of sample state through monitoring of key physical attributes. However, predictive models built with these data have shown varying degrees of accuracy, depending on food type, storage conditions, sensor platform, and sample sizes. This work explores various fusion approaches for potential predictive enhancement, through summation of information gathered from different observational spaces: infrared spectroscopy is supplemented with multi-spectral imaging for prediction of chicken and beef spoilage through estimation of bacterial counts in differing environmental conditions. For most circumstances, at least one of the fusion methodologies outperformed single-sensor models in prediction accuracy. Improvement in aerobic, vacuum, and mixed aerobic/vacuum chicken spoilage scenarios was observed, with RMSE performance enhanced by up to 15%. The improved cross-batch performance of these models suggests enhanced model robustness, suggesting further evaluation in meat quality analysis is necessary. The batch-based results were corroborated with a repeated nested-cross validation approach, to give an out-of-sample generalised view of model performance across the whole dataset.

2.2 Introduction

Meat is an important staple in diets worldwide, containing a large spread of macro- and micro-nutrients. Poultry and beef are important livestock for global markets, accounting for ~33% and ~24% of consumed meat, respectively, according to the United Nations Food and Agriculture Organisation (USDA, 2024).

Fresh meat products have a limited sale and consumption window, due to their perishable nature as a result of the activity of endogenous enzymes and microbes. Concerning the latter, their activity and potential pathogenicity is linked to many of the undesirable outcomes that characterise poor quality meat such as unpleasant flavours, malodour or changes to visual or textural appearance, and public health concerns.

The negative impact of spoilage-causing microbial activity necessitates steps to control or even inhibit population growth. Amongst others, refrigeration and the use of modified atmosphere packaging (MAP), which may include the application of vacuum conditions or ratios of gases incompatible with targeted microbial growth for that specific food type, including nitrogen, carbon dioxide or oxygen, have been applied (Kandeepan and Tahseen, 2022; Kuzelov et al., 2015).

Improving the accuracy and responsiveness of food quality and safety assessment has been a topic of active research for many years, partly driven by growing interest in improving the state of food supply infrastructure (Nychas et al., 2021). Furthermore, clear monetary motivations exist for producers and suppliers to optimise their operations through minimisation of wastage – globally, an estimated 13% of food is wasted or lost between harvest/production and retail (United Nations, 2023).

Food testing methodologies have been standardised within legislation at a national and international level (EFSA 2073/2005). Current strategies feature laboratory-based methods for quantifying sizes and types of microbial populations within food samples (Nychas, Panagou and Mohareb, 2016), in addition to organoleptic and sensory analyses (Rodrigues, Dias and Teixeira,

2024; Stelmach and Cwiek-Ludwicka, 2006). Over the last 10-15 years, data-driven approaches for employing sensor technologies to digitise the physicochemical properties of food samples in a non-/minimally-invasive, high throughput manner have been employed (Cozzolino, 2022; Nychas et al., 2021; Yousefi et al., 2019). Application of sensor technologies may serve to increase the responsiveness and coverage of testing within the food production and distribution pipeline and complement the capabilities of the aforementioned methodologies, whilst not suffering the shortcomings of conventional techniques, i.e. their time-consuming, technically involved nature requiring specialist personnel (Nychas et al., 2021).

The motivations for deployment of these sensor types throughout the production and distribution pipeline are summarised in the concept of process analytical technology (PAT) (Nychas, Panagou and Mohareb, 2016).

This operating methodology originated in the pharmacological industry to monitor sample critical control attributes through the determination of critical control points (CCPs) in a responsive manner and at a higher resolution than that achievable by relying solely on end-stage compliance testing, allowing for faster identification and rectification of points of failure within the pipeline (Clegg, 2020).

A plurality of sensor modalities exists, including biomimetic sensors that are designed to record a sample's chemometric properties in a manner that emulates the human sensory experience (e.g. electronic nose and electronic tongue), and spectroscopic sensors that record the optical properties of a sample (Pampoukis et al., 2022).

Techniques such as Fourier-transform infrared (FTIR) spectroscopy, Raman spectroscopy and multispectral imaging (MSI) have been investigated in meat quality and spoilage analyses (Böcker et al., 2007; Pampoukis et al., 2022; Rahman et al., 2018; Spyrelli et al., 2021). A large number of machine learning algorithms, such as partial least squares discriminant analysis (PLS-DA) and regression (PLS-R) models have been used to analyse such data (Nychas et al., 2021).

Spyrelli et al. compared microbial count regression performance to models built on sample images captured with the Videometer proprietary MSI platform and confirmed the utility of FTIR spectroscopy to detect spoilage in chicken (Spyrelli et al., 2021). The researchers found that while FTIR-trained models outperformed MSI models, the difference may have been due to intra-batch variation rather than Videometer data having less predictive power. The authors also use MSI/Videometer sensor data to predict chicken slaughter time (Spyrelli et al., 2020). In their machine learning objectives, the resultant models used data from one sensor modality as predictor variables. Sensor/data fusion may improve machine learning model prediction if different sensor modalities operate in different observational spaces and record different sample states.

Data fusion is a wide topic within the information-technology domain, and covers many approaches for combining useful information originating from different sources. Within the bioinformatics sphere, the main approaches for reaching these objectives can be subdivided into categories, based on the stage of the information processing workflow at which the fusion operation takes place (see Figure 2-2):

Early fusion – Typically resulting from the concatenation of data matrices from different modalities, before predictions are made on the subsequent matrix

Mid-level/Feature fusion – Individual data matrices undergo feature extraction; the resultant features are concatenated and used to make predictions (Smolinska et al., 2019; Yang et al., 2020).

Late/Decision fusion – Individual machine learning models are trained on the data from each modality, and the decisions from each model are combined in some way, either through some form of averaging, or through the construction of 2nd layer (meta-learning) models, which are built upon the library of 1st layer decision vectors, in a form similar to a *stacked generalisation* ensemble, which would typically exploit some degree of orthogonality in 1st level predictions, typically generated from a heterogeneous library of algorithm types and/or with varied data/feature engineering steps (Naimi and Balzer, 2018; Wolpert, 1992).

A variety of data fusion architectures have been explored in the biosciences. Attempts to fuse data emerging from food sensor technologies may hamper implementation of early- and mid-level architectures – differing data dimensionalities or scaling may limit the effectiveness of models trained on concatenated raw data or features without some intermediary scaling, as one data matrix may influence the model unfairly (Smolinska et al., 2019). Borràs *et al.* note that while early and mid-level data fusion are often recorded in food analysis literature as producing better results than high-level approaches, a strength of the latter approach is the ease with which new modalities can be added to the overall model (Borràs et al., 2015).

Smolinska *et al.* compare representative implementations of mid- and late-fusion approaches for the prediction of gas composition, bitterness and foam stability in beer using sensor data from gas and liquid chromatography coupled with mass spectrometry; their determination was that all approaches improved prediction accuracy over single-modality modelling and that late fusion performed consistently well in terms of accuracy, whilst the mid-level fusion approaches eased biological interpretation when fusing sensor data (Smolinska et al., 2019).

The aim of this research was to assess the comparative efficacy of individual high-throughput, minimally-invasive sensor devices and subsequent fusion of their outputs for prediction of total viable counts of bacteria found in fresh poultry (chicken) and beef mince samples through the use of machine learning models trained on data recorded in aerobic and modified atmosphere or vacuum conditions. Data fusion workflows were implemented for combining the outputs of both sensor modalities (Fourier Transform Infra-Red and Multi-Spectral Imaging), for the investigation of predictive performance versus single-sensor models.

2.3 Materials and Methods

2.3.1 Experimental Design, Sample Preparation and Microbiological analysis

2.3.1.1 Chicken Thigh Samples

Fresh chicken thigh and minced beef samples were sourced from Greek meat distributors rapidly after preparation. Chicken thigh fillets were provided by the food company KOTINO, Athens, Greece. Sample storage and preparation is fully described elsewhere (Fengou et al., 2024).

Chicken samples were prepared and stored in either vacuum-packaged or an aerobic state. The samples were stored at one of three isothermic temperatures for both packaging conditions, 0°C, 5°C (safe), or 10°C (marginally abusive), and monitored at 24–48h intervals, until visual deterioration of the samples became apparent.

Data for each storage condition were recorded in two batches, corresponding with meat sourcing on separate days, with a variable number of replicates (2–10) for each sampling condition in each batch.

Maximum storage time varied depending on the storage temperature and packaging condition: for the vacuum-packaged samples, those stored at 0°C and 5°C were monitored for a maximum of 240h after collection, while those stored at 10°C were monitored for a maximum of 120h. For the aerobically packaged samples, those stored at 0°C were monitored for a maximum of 216h, those stored at 5°C were monitored for a maximum of 192h, and those at 10°C were monitored for a maximum of 96h.

2.3.1.2 Beef Mince Samples

The beef mince samples were acquired specifically for the purpose of this study, sourced from a central butcher shop in Athens and transported under refrigeration to the laboratory within 30 min. Approximately 75–80g portions were placed on Styrofoam trays and stored either in air or under modified atmosphere packaging (MAP) conditions (20% CO₂/80% O₂). For aerobic storage, the samples were covered with plastic food wrap typically used for domestic purposes. For MAP

storage, the samples were sealed in plastic pouches with gas permeability rates of approximately 25cm³/m² per day for CO₂ and 90 cm³/m² per day for O₂ at 20°C and 50% relative humidity, using a Henkovac 1900 machine (Henkovac, 's-Hertogenbosch, Netherlands). The samples were stored at one of three isothermic temperatures for both packaging conditions (as above), and monitored at 24–48h intervals, until visual deterioration of the samples became apparent. Data for each storage state were recorded in two batches, corresponding to meat sourcing on separate days, with a variable number of replicates (2–11) for each sampling condition in each batch.

Maximum storage time varied depending on the storage temperature and packaging condition: for the modified atmosphere meat samples, those stored at 0°C and 5°C were monitored with all sensor configurations for a maximum of 336 and 312h after collection, respectively, and for 144h for the samples stored at 10°C. For the aerobically packaged samples, those stored at 0°C were monitored for a maximum of 336h, those stored at 5°C were monitored for a maximum of 144h, and those at 10°C were monitored for a maximum of 120h.

2.3.1.3 Microbiological Analysis

From each meat sample, 25 g portions were aseptically weighed into 400 mL sterile Stomacher bags (Seward Medical, London, UK) containing 225 mL of sterile quarter-strength Ringer's solution (LabM Limited, Lancashire, UK) and homogenised for 60 s using a Lab Blender 400 (Seward Medical). Appropriate serial dilutions were then prepared using the same Ringer's solution. From these dilutions, duplicate samples of 0.1 or 1 mL were either spread or mixed onto the following media: plate count agar (PCA, Biolife 4,021,452, Milan, Italy) for total viable counts (TVC), incubated at 30°C for 48–72 h. All plates were visually examined for typical colony types and morphological characteristics that were associated with the growth medium.

2.3.1.4 Spectral Acquisition – Multi-Spectral Imaging

MSI spectra for both the chicken and beef spoilage scenarios were captured using the VideometerLab platform (Videometer, Denmark) (Videometer, Denmark, 2021), which captures information in a spread of 18 monochromatic

wavelengths ranging from 405 nm to 970 nm, at intervals between 15 and 150 nm. The imaging data produced from the sensor is representative of two spatial dimensions and one wavelength dimension in a layered *width×height×18* data cube, with each layer representing the reflected response from illumination with each wavelength. To convert the data into a tabular form that can facilitate analysis with conventional data processing methods, a segmentation and averaging process is undertaken, wherein Canonical Discriminant Analysis (CDA) is performed, using the accompanying VideoMeter software version 2.12.39. In an averaging process described in Fengou *et al.* (Fengou et al., 2019b), the mean and standard deviation of intensity of reflectance across the target area was recorded. MSIF (Multispectral Imaging Fluorescent) spectra are captured at 8 bands in the UV/blue light spectrum, ranging from 270 to 405nm. Fluorescence bands were measured using a filter for combined reflectance/fluorescence.

2.3.1.5 Spectral Acquisition – Fourier-Transform Infrared Spectroscopy

FTIR spectra were recorded between 4000 and 400 cm^{-1} using a zinc selenide (ZnSe) HATR (Horizontal Attenuated Total Reflectance) crystal and an FTIR-6200 spectrometer (JASCO, Tokyo, Japan) equipped with a standard sample chamber, a triglycine sulphate (TGS) detector, and a Ge/KBr beamsplitter. The HATR crystal had a refractive index of 2.4 and a penetration depth of 2 μm at 1000 cm^{-1} . Using the Spectra Manager™ Code of Federal Regulations (CFR) software version 2, spectra were collected by accumulating 100 scans with a resolution of 4 cm^{-1} and a total integration time of 2 min.

2.3.1.6 Data Partitioning, Preprocessing, and Analysis

The overall aim of the data analysis segment of the research was to generate machine learning regression models for the prediction of TVC, using the sensor data collected under each experimental condition. For each meat type, single-sensor models were constructed, along with early-, feature-, and decision- fusion approaches (Figure 2-2) for combining the outputs of the individual sensor devices with the intention of improving the overall predictive accuracy.

Machine learning pipelines were constructed in the Python programming language (Version 3.11.7), using the Scikit-Learn framework (Version 1.3.0).

Two data partitioning methods were utilised for data analysis; the first was a batch-on-batch validation approach, in which models were trained and optimised with the first batch and tested with the second. Pre-processing steps were undertaken on each training and testing set separately, to avoid data leakage. Observations were matched across datasets from different sensor modalities by using the observation naming scheme in the dataset. A further calibration/test split was synthesised, by combining the aerobic/vacuum chicken and aerobic/modified atmosphere packaged beef calibration sets, and their respective testing sets, to assess the performance of models constructed with data from all respective packaging conditions. The resultant sample sizes for each storage condition are described in Table 2-1.

For the MSI/MSIF datasets, Standard Normal Variate (SNV) filtering was undertaken to minimise baseline shifts in the dataset, similar to prior research (Barnes, Dhanoa and Lister, 1989; Spyrelli et al., 2021). SNV was also applied to the chicken thigh FTIR data for both storage conditions, to remove baseline shifts in the dataset. For machine learning model construction, all mean and standard deviation information was used, as per prior research (Fengou et al., 2019b; Ropodi et al., 2015). Savinsky-Golay filtering was applied to FTIR data for both scenarios, with a first-order polynomial and a window size of 11 proving sufficient for smoothing without loss of useful information. A preliminary feature selection step was undertaken by restricting the FTIR spectral data to the wavenumber range 2000–900 cm^{-1} , as guided by the existing literature (Argyri et al., 2013; Spyrelli et al., 2021), and known as the ‘spoilage fingerprint region’ in many food group analyses, due to its correspondence to the excitation of polysaccharides, peptide bonds, and phosphate groups (Sahar and Dufour, 2014).

An alternative approach was implemented to provide the algorithms with larger calibration sets to assess the impact of the smaller training sample sizes used in the previous method, and to generalise the predictive out-of-sample performance achievable with the different strategies.

Ten repeats of five-fold nested cross validation were undertaken with all model/fusion strategies under investigation, and all combinations of data – aerobic, vacuum/modified atmosphere, and combined conditions as before. This was intended to provide a more reliable estimate of model performance across the whole dataset. Each nested cross validation round consisted of the random shuffling and splitting of the data into five folds, wherein five combinations of four folds were used for the nested hyperparameter tuning process with an *inner* five-fold cross validation procedure. The best candidate's parameters in this nested validation were used to retrain on the whole *inner* dataset, and the model was tested with the respective held-out *outer* fifth fold. The process was repeated a further nine times, for a total of fifty assessments of performance, each made with random subsections of training and testing data. The utility of this process in minimising optimistic bias/overfitting in model assessment without split-to-split information leakage when compared to more conventional k-fold cross validation has been acknowledged by Cawley and Talbot (Cawley and Talbot, 2010). A schematic of this process is shown at Figure 2-1. Crucially, employment of the *same* random seed allows for all folds in the respective assessments using this technique to use the same series of random data splits. This allows for true comparison of model performance on a fold-to-fold basis, removing the opportunity for one algorithm to be assessed on an advantageous series of folds, while disadvantaging other algorithms with alternative splits. A similar approach has been employed by Borràs *et al.* for their prediction of olive oil sensory descriptors, which coincidentally also explored the efficacy of early- and feature-level fusion with PLS modelling (Borràs *et al.*, 2016).

Table 2-1: Table of calibration/cross validation (CV) and test set sizes for chicken and beef spoilage scenarios using batch validation approach

Partition	Aerobic condition	Vacuum condition	Combined conditions
Chicken Spoilage			
Calibration	70	65	135
CV	70	65	135
Prediction	66	72	138
Repeated Nested CV	136	137	273
Beef Spoilage			
Calibration	46	57	103
CV	46	57	103
Prediction	46	56	102
Repeated Nested CV	92	113	205

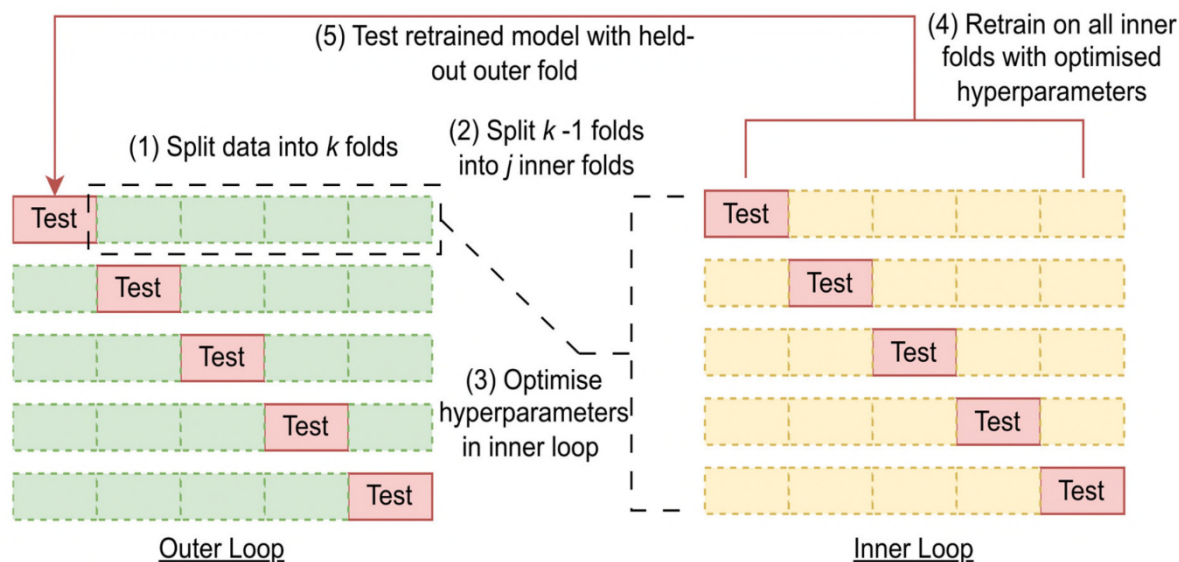


Figure 2-1: General schematic for a single fold of Nested Cross Validation. This process is repeated for k folds. In the case of repetitions, the data is shuffled randomly before the same process occurs again for a new selection of k outer folds.

2.3.1.7 Machine Learning Methods

Partial least squares (PLS) methods were selected as the algorithm for the machine learning tasks, widely utilised in chemometric analysis due to their aptitude for dealing with the multicollinearity of data expected from spectroscopic sensor technologies, where the number of linearly independent columns of the resultant data matrix is far fewer than the total number of columns (low rank).

Furthermore, the family of algorithms are often considered in situations where the number of predictor variables (independent variables) far exceed the number of observations, a situation known as the “small n , large p ” problem, which can complicate some other types of machine learning methodologies which cannot account for predictor ($\mathbf{X}$) matrices having low rank (Abdi, 2010). Furthermore, previous research has shown the algorithm as effective with similar meat spoilage scenarios (Fengou et al., 2019a; Rahman et al., 2018; Sahar and Dufour, 2014).

A series of regression models were built, using the sensor-associated total viable counts (TVC) as the dependent variable.

For the batch-based performance assessment, the optimisation of machine learning models was made through the use of repeated k-fold cross validation, in which the generalised performance of the selected machine learning model was assessed through 3 repetitions of dividing the dataset into 10 partitions; each fold is iteratively held out for evaluation, with remaining data used for model training. This methodology often provides more stable and robust performance estimates than Leave-One-Out Cross-Validation (LOOCV), which can yield higher variance due to the high correlation between its nearly identical training sets (James et al., 2021).

Optimisation of the PLS models took place in a grid search, wherein models built using varying numbers of latent variables (LVs) between 2 and 20 were assessed. Selection of optimal numbers of PLS LVs was based on minimisation of root mean square of error in cross-validation (RMSE_{cv}).

The exploration of the grid search space used for optimisation of hyperparameters was performed through either randomized (Bergstra and Bengio, 2012) or exhaustive search depending on the complexity of the model and search space; information on hyperparameters is given in Table 2-2.

Individual PLS regression models were trained on each data type and storage condition, for a total of 6 single-storage-condition models for both the chicken and beef spoilage datasets. Additionally, single-sensor-modality models were trained on the data from both storage conditions, by combining the training partitions for each sensor type, resulting in 6 further models. Assessment of the complete models was achieved through use of the test set: root mean square of error in prediction (RMSE_p) (log CFU/g) was recorded as the main performance metric for the test results, along with the coefficient of determination (R^2) value. This metric signifies the proportion of variance in the dependent (target) variable explained by the model's independent (predictor) variables (Draper and Smith, 1998). A domain-specific metric was also recorded as an effort to categorise performance in otherwise regression-based analyses: “accuracy” is defined as the proportion of TVC predictions that fall within ± 1 log CFU/g of observed values. This threshold of acceptability is based on prior microbiological growth research

(Fengou et al., 2019b; Spyrelli et al., 2020, 2021). To explore the robustness of built models when encountering data from unseen conditions, the test sets from both storage conditions were used to assess model performance independently. Improvement of cross-condition performance was seen as a potential benefit for models incorporating data from both storage conditions.

Table 2-2: Grid Search Space for the Repeated Nested Cross Validation Approach. In the case of Random Search for MSI-FTIR Mid Fusion and MSI-MSIF-FTIR Mid/Late Fusion, 1500 (~37% coverage), 2500 (~12% coverage) and 400 (~30% coverage) iterations were performed per inner grid search respectively, with each parameter sampled from a uniform distribution. All other model types used exhaustive grid search.

Model	Range
<i>Single Sensor Models and Mid/Late Fusion Base Models</i>	
MSI - pls__n_components	2-17
FTIR - pls__n_components	2-20
MSIF - pls__n_components	2-7
<i>Early Fusion Models</i>	
pls__n_components	2-20
<i>Mid Fusion Models</i>	
pls__n_components	2-17

2.3.2 Multi-Sensor Fusion Strategies

Pipelines characterising the three main archetypes of data fusion (early, mid/feature and late/decision) were developed to compare their microbial count predictive performance for both meat types and their respective storage conditions, in addition to the combination of conditions as previously described.

A naive early fusion approach for all conditions was examined, using the same procedure as described by Fengou *et al.* (Fengou *et al.*, 2019b). The heterogeneity of the data emerging from the project's MSI, MSIF, and FTIR sensors necessitated preprocessing steps before combining their respective information. From each modality, the data were standardised (z-score normalised) before being concatenated into a single predictor matrix, then used as input for a single PLS regression model. To implement the mid-level approach, the previously developed single-sensor PLS models were used to transform the raw sensor data into the feature space of the respective models. These transformed features were horizontally concatenated into new matrices and used for the construction of a single new PLS model. As before, the repeated k-fold cross-validation scheme was used to determine the optimum number of latent variables. As described earlier, differences in dimensionality or scaling in early and mid-level fusion can lead to concatenated data matrices being dominated by the contributions of one modality (Smolinska *et al.*, 2019). A late (decision) fusion approach was developed alongside the former, which is by nature less prone to domination with information from a single modality.

The TVC prediction output from the single-sensor models for all experimental conditions (aerobic storage, vacuum storage and combined) were concatenated into respective results matrices and served as predictor variables in a further regression layer used as a meta-learning step, with the observed TVC results serving as target variables, forming the final regression result. All three fusion methodologies are described in Figure 2-2.

The final approach employed the stacked generalisation methodology of training the meta-learner on the out-of-fold (OOF) predictions made by the first-layer models, described by Wolpert (Wolpert, 1992) and later by Laan *et al.* in their

published strategy (Laan, Polley and Hubbard, 2007). The training of the meta-learning layer on the OOF predictions of the first layer models gives a representation of how the model may perform on unseen data. The meta-learner is thus trained on the predictions made by the first layer of the model, for all observations in the original training data, with the intention of reducing the risk of overfitting.

For this process, 20-fold cross validation was used in the construction of the meta-learner training set, Linear regression was used as the meta learning algorithm., which was also used in Laan *et al.*'s simulation.

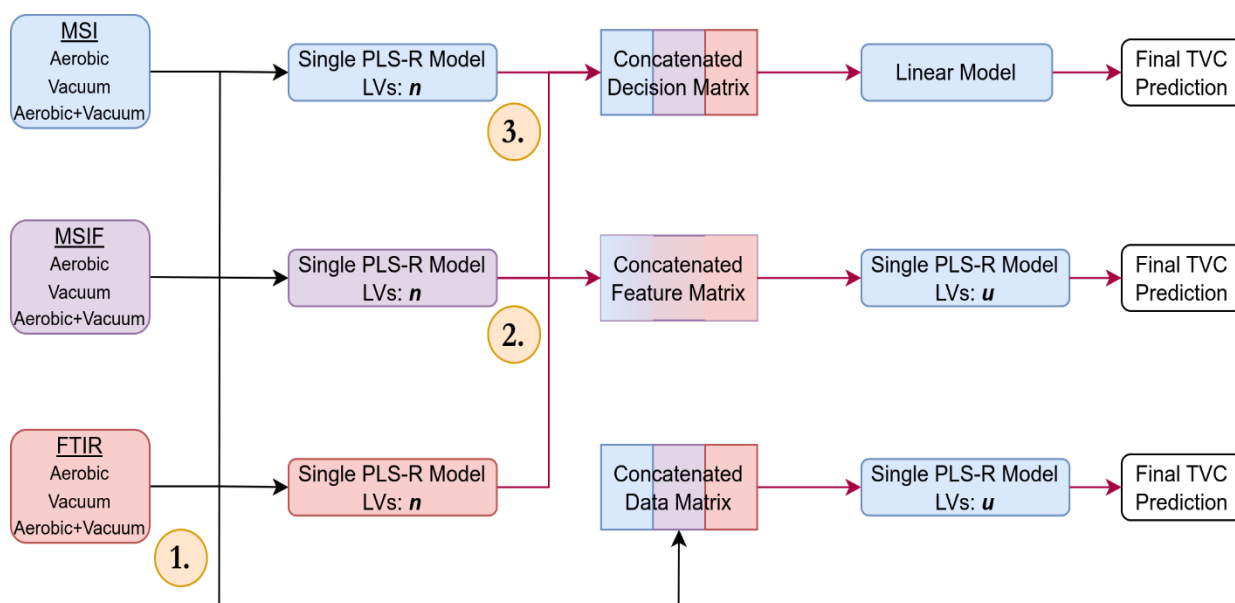


Figure 2-2: Graphical representation of early, feature (mid-) and decision (late) fusion approaches implemented using multispectral imaging (MSI), multispectral imaging fluorescent (MSIF), and Fourier transform infrared spectroscopy (FTIR) for total viable count (TVC) prediction. The upper examples represent decision fusion, wherein trained first-layer model predictions (3.) are passed to a linear model for combination of their predictions. The feature fusion approach uses first-layer partial least squares (PLS) models to generate projections for passing into a second-layer PLS model (2.), the number of latent variables u is tuned on these outputs. The lower approach represents early fusion, wherein the unprocessed data matrices are concatenated into a single matrix and standardised (1.) prior to passing to a subsequent PLS model. The number of latent variables retained was tuned in a manner similar to the feature fusion approach.

2.4 Results

2.4.1 Microbial assessment of chicken thigh spoilage

The microbial (TVC) populations for chicken spoilage are recorded in Figure 2-3. The initial TVC load in the sampled chicken thigh was shown to fall within the range of 4.5-6 log CFU/g. Storage temperature had an apparent large impact on the kinetics throughout storage. Initial disparity between batches for both atmospheric conditions appeared minimal, barring the mean initial measurements for batch 1 of the vacuum-packaged meat samples (5.72 versus 5.22 log CFU/g). Across all conditions, initial load appeared high – previous research conducted in similar conditions showed lower values of between 2 and 3.5 log CFU/g (Sahar and Dufour, 2014; Spyrelli et al., 2021).

Storage in the safe conditions (0°C and 5°C) shows slowing of bacterial growth in vacuum packaged meat when compared to aerobic storage – beyond 48h at 5°C and 72h at 0°C, batches of the former condition exhibit a widening TVC difference against the aerobically stored meat until the end of sampling.

Mean TVC loads at 0°C and 216h were 7.39 and 7.29 log CFU/g for the vacuum condition and 8.04 and 9.11 log CFU/g for the aerobic condition. At 5°C and 216h, TVC was recorded at 8.75 and 8.63 log CFU/g for the vacuum condition. For the aerobic condition 5°C and 168h, TVC was recorded at 9.01 and 9.38 log CFU/g.

The marginally abusive condition (10°C) exhibited the most severe microbial growth: aerobic counts at the end of sampling were 9.50 and 9.27 log CFU/g at 96h, whereas the vacuum-packaged meat peaked at 8.25 and 8.53 log CFU/g at 96h and 216h for batch 1 and 2 respectively.

Monitoring for all conditions was terminated at points past the 7-8 log CFU/g range previously reported as representative of spoilage in chicken (Rouger, Tresse and Zagorec, 2017).

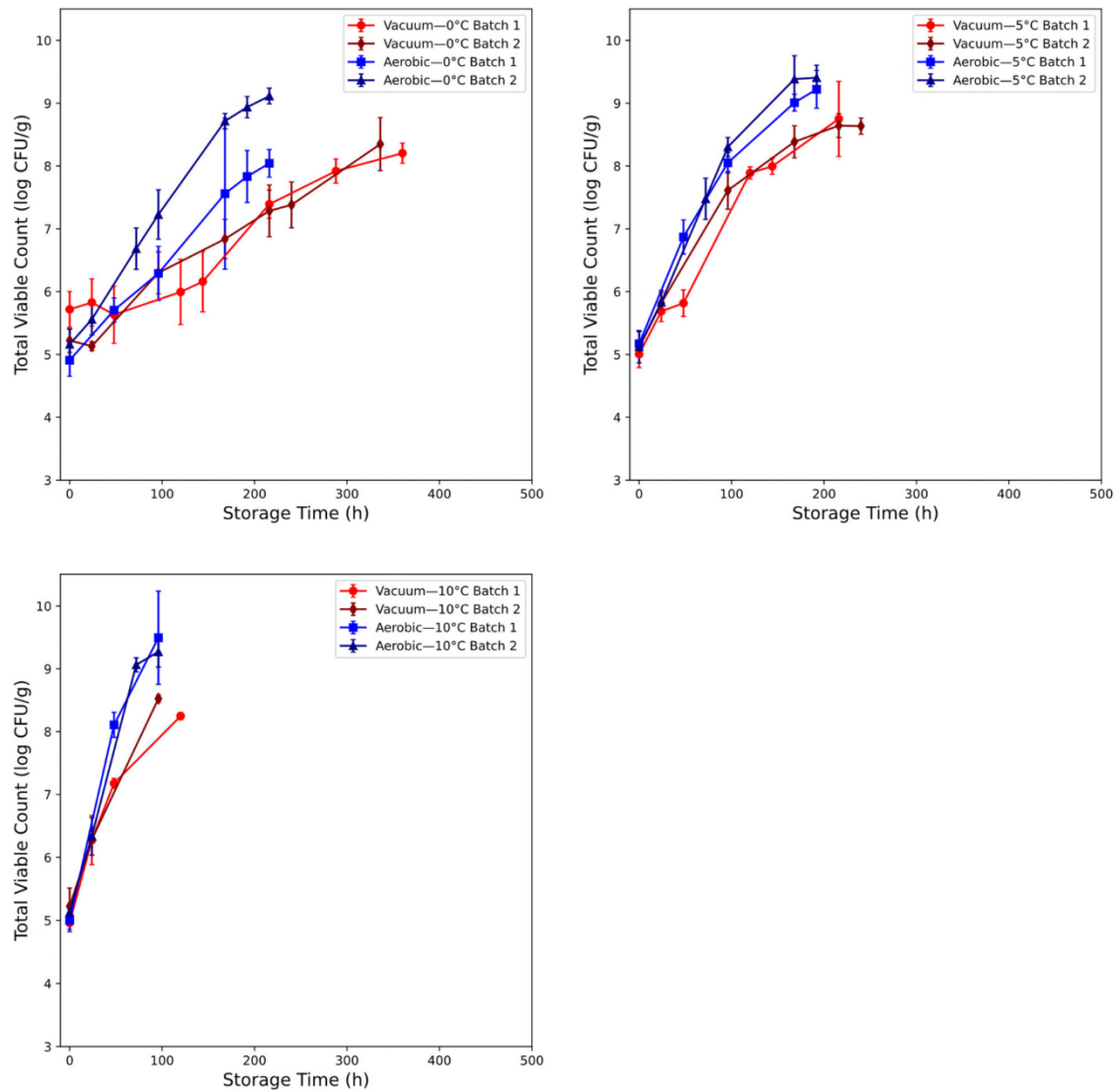


Figure 2-3: Mean and Standard Deviation of Total Viable Count (TVC) recorded in chicken thigh for aerobic (batch 1: blue, batch 2: dark blue) and vacuum (batch 1: blue, batch 2: dark blue) packaging conditions at 0°C, 5°C and 10°C.

2.4.2 Microbial Assessment of Beef Spoilage

The microbial (TVC) populations for beef mince spoilage are recorded in Figure 2-4. Initial TVC load for both packaging conditions show separation by batch, of between 0.5 to 1 log CFU/g. Batch 1 of the aerobic and modified atmosphere samples showed mean initial counts of 4.95, 4.72 and 5.04 log CFU/g at 0°C, 5°C and 10°C respectively, whereas batch 2 showed counts of 4.07, 3.91 and 4.02 log CFU/g.

This ~ 0.5 - 1 log CFU/g difference in initial TVC values seen in both aerobic and modified atmosphere packaging batches likely reflects the common origin of the batches and their simultaneous preparatory steps. During the preparation of patties for storage, the mincing process is likely to distribute any surface contamination present in certain areas throughout the substrate and inoculate areas that were previously functionally sterile (US Institute of Medicine, 2002), in addition to decreasing the consistent baseline morphology of the underlying meat tissue, which may affect microbial growth in unpredictable ways.

Despite the variability apparent in both the aerobic and modified atmospheric packaging (MAP) initial conditions, the aerobic batches tend towards a closer final count in both the 0°C and 5°C conditions.

The MAP batches stored at 0°C show a reversal of highest counts at 216h, possibly indicating a change in growth conditions, as batch 2 shows a declination of counts at 288h, followed by a more severe, apparently exponential growth pattern onwards, mean counts in batch 2 of the MAP condition jump significantly from 7.21 log CFU/g at 408h to 9.77 log CFU/g at 456h.

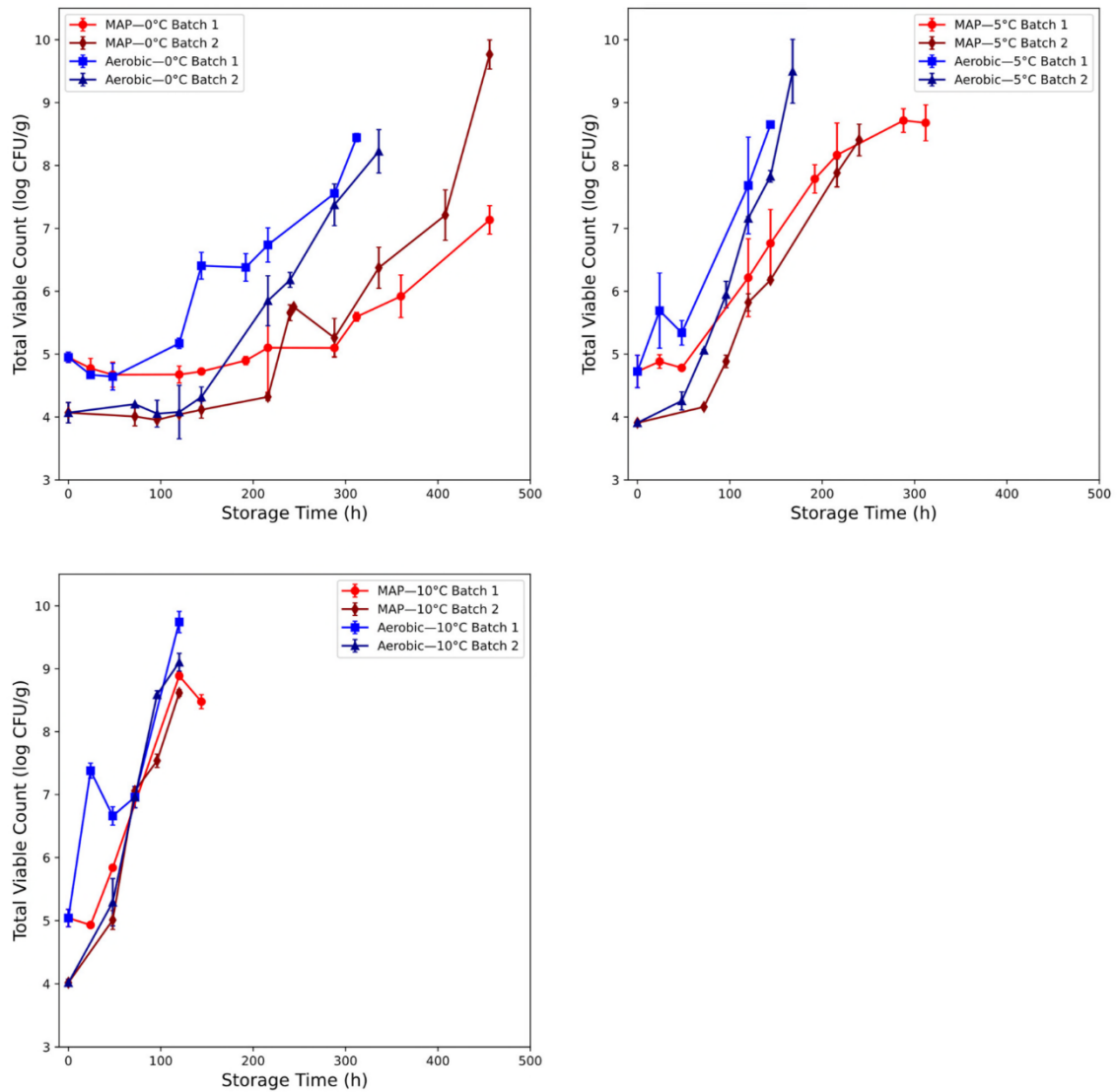


Figure 2-4: Mean and Standard Deviation of Total Viable Count (TVC) recorded in beef mince for aerobic (batch 1: red, batch 2: dark red) and MAP (batch 1: blue, batch 2: dark blue) packaging conditions at 0°C, 5°C and 10°C.

2.4.3 Fourier-Transform Infrared and Multispectral Imaging Measurements of Chicken Thigh and Beef Mince

FTIR Spectral plots corresponding to the mean and standard deviation recorded within the processed calibration sets of the chicken thigh and beef mince datasets are found in Figure 2-5 and Figure 2-6. Previous domain specific research has highlighted the contribution of the $1800\text{-}1400\text{cm}^{-1}$ band in spoilage analysis (Spyrelli et al., 2021); the spectra gathered through this research corroborate these contributions: large degrees of variance exist in the $1750\text{-}1400\text{cm}^{-1}$ region indicated by the wide standard deviation margin in both aerobic and vacuum-stored samples, although this is partially obscured by the gradient of the peaks.

The spectral regions peaking at $\sim 1550\text{cm}^{-1}$ and 1640cm^{-1} have been associated with amide II vibrations (Ellis et al., 2002), suggesting the presence of proteins and potential free amino acids, and is concordant with microbial spoilage-related proteolytic changes (Alexandrakis, Downey and Scannell, 2012).

Figure 2-7 shows a large degree of variation in the aerobic packaged chicken thigh MSI 450-700nm region for the 0 to 216h window, and more limited variation in the 459-570nm region for the vacuum packaged chicken. Similar observations have been made for published chicken-based food products (Spyrelli et al., 2020), with reflectance in this region attributed to pigments including myoglobin, oxymyoglobin and metmyoglobin (570/590/630nm) (Pu, Kamruzzaman and Sun, 2015). The more limited variation in the longer-wavelength region may be indicative of the vacuum storage condition's inhibitory effect on the metabolic processes of spoilage-causing microbiota and meat oxidation.

Across the aerobic and modified atmosphere beef measurements in Figure 2-8, the difference in reflectance between the 0h and 336h samples is more uniform across the wavelengths, except for a region in the aerobic data between 590-700nm, in which the overall trend is inverted, with the 336h samples showing a lesser degree of reflectance. This regional reduction in reflectance has been attributed to an effect similar to that observed with the chicken samples, namely colour changes (browning) due to breakdown of meat pigment proteins (Liu et al., 2003; Panagou et al., 2014).

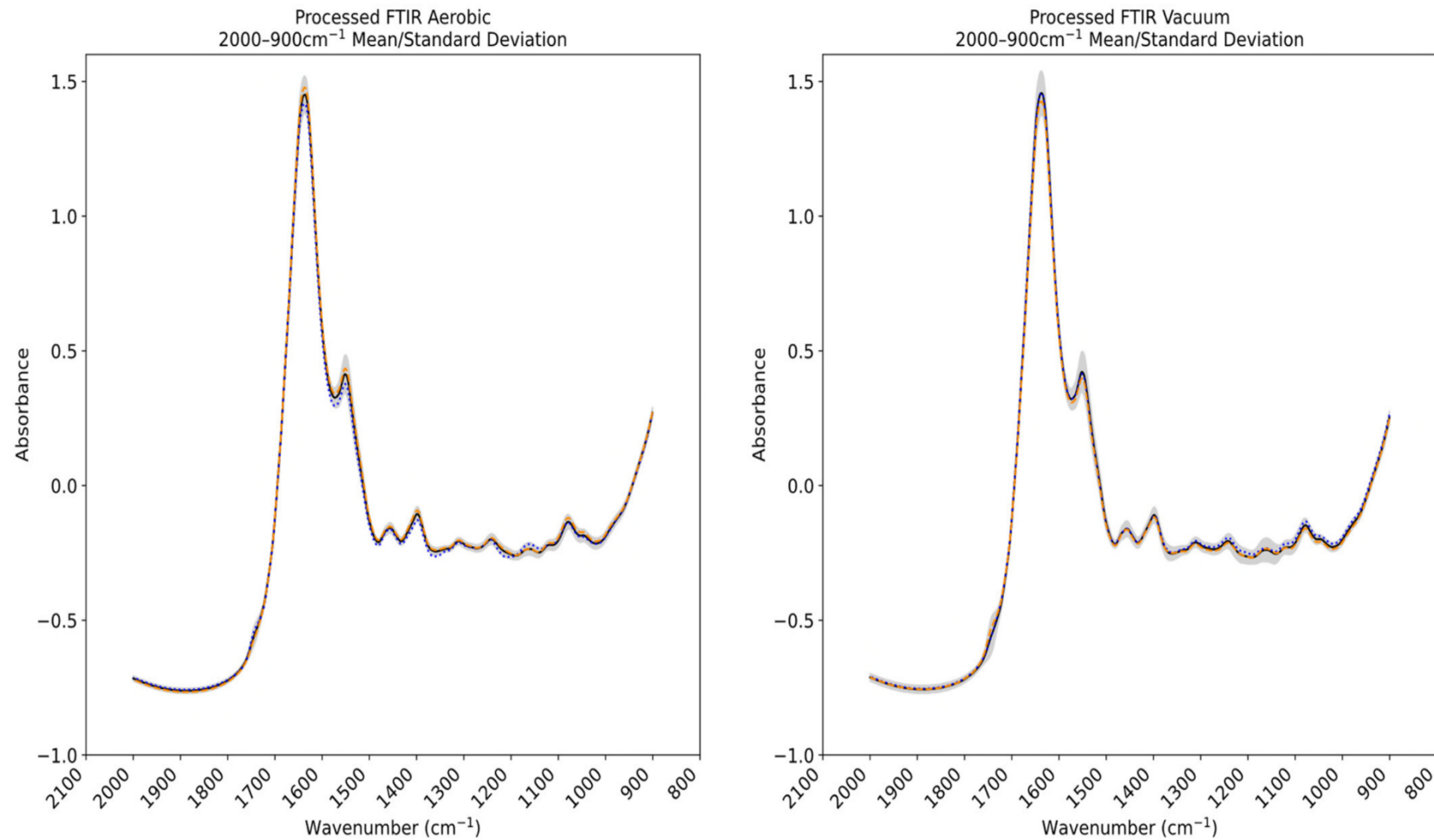


Figure 2-5: Plot showing mean spectral values of Fourier transform infrared spectroscopy (FTIR) data across the aerobically and vacuum-stored chicken thigh samples (solid black line). The standard deviation across the dataset is indicated by grey shading. Mean values are indicated for 0 °C/0 h (dotted blue line aerobic n = 10, vacuum n = 8) and 0 °C/216 h (dashed orange line aerobic n = 21, vacuum n = 8).

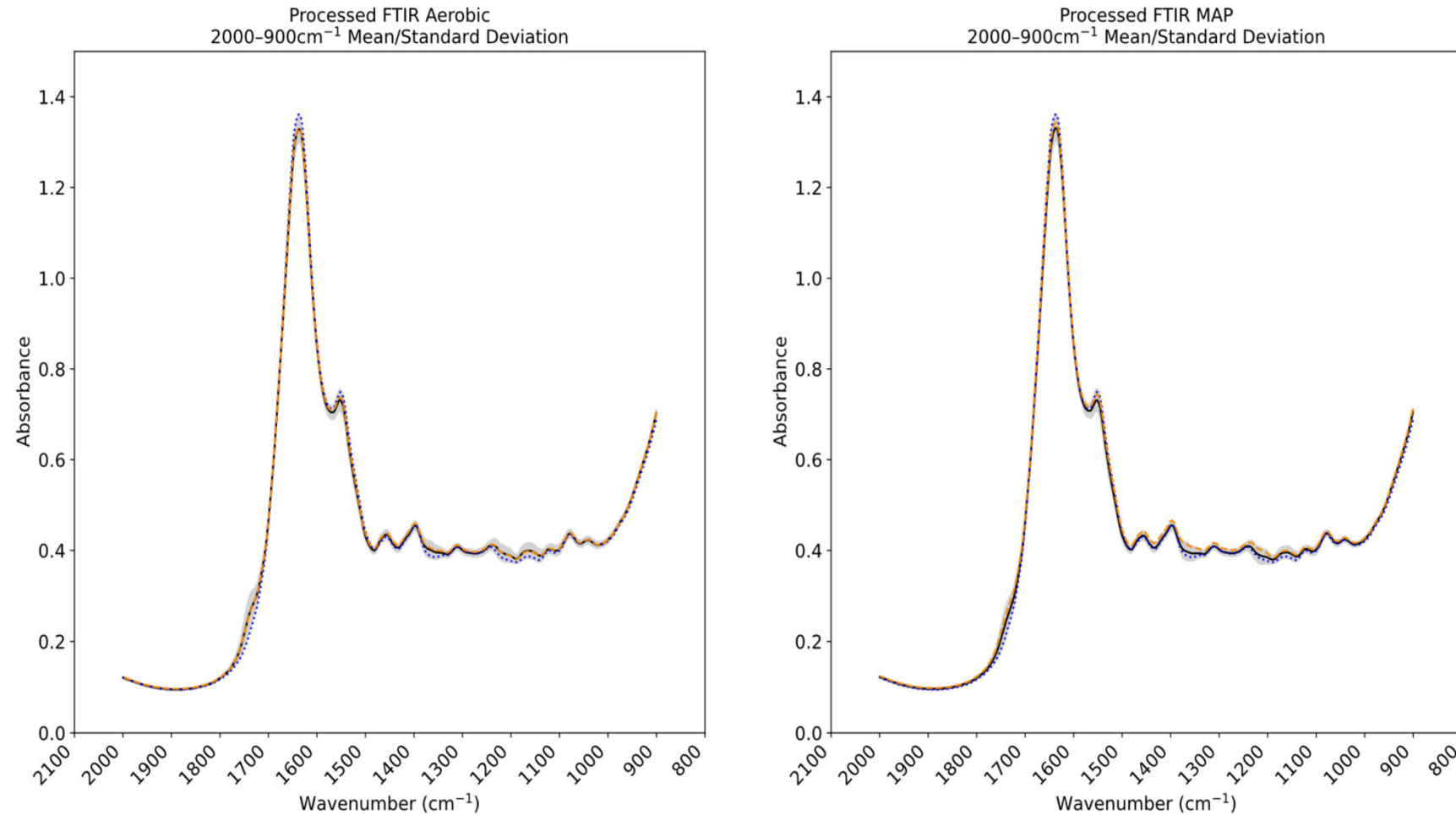


Figure 2-6: Plot showing mean spectral values of Fourier transform infrared spectroscopy (FTIR) data across the aerobically and modified atmosphere packaged (MAP) beef mince samples (solid black line). The standard deviation across the dataset is indicated by grey shading. Mean values are indicated for 0 °C/0 h (dotted blue line aerobic n = 4, MAP n = 4) and 0 °C/336 h (dashed orange line aerobic n = 4, MAP n = 4).

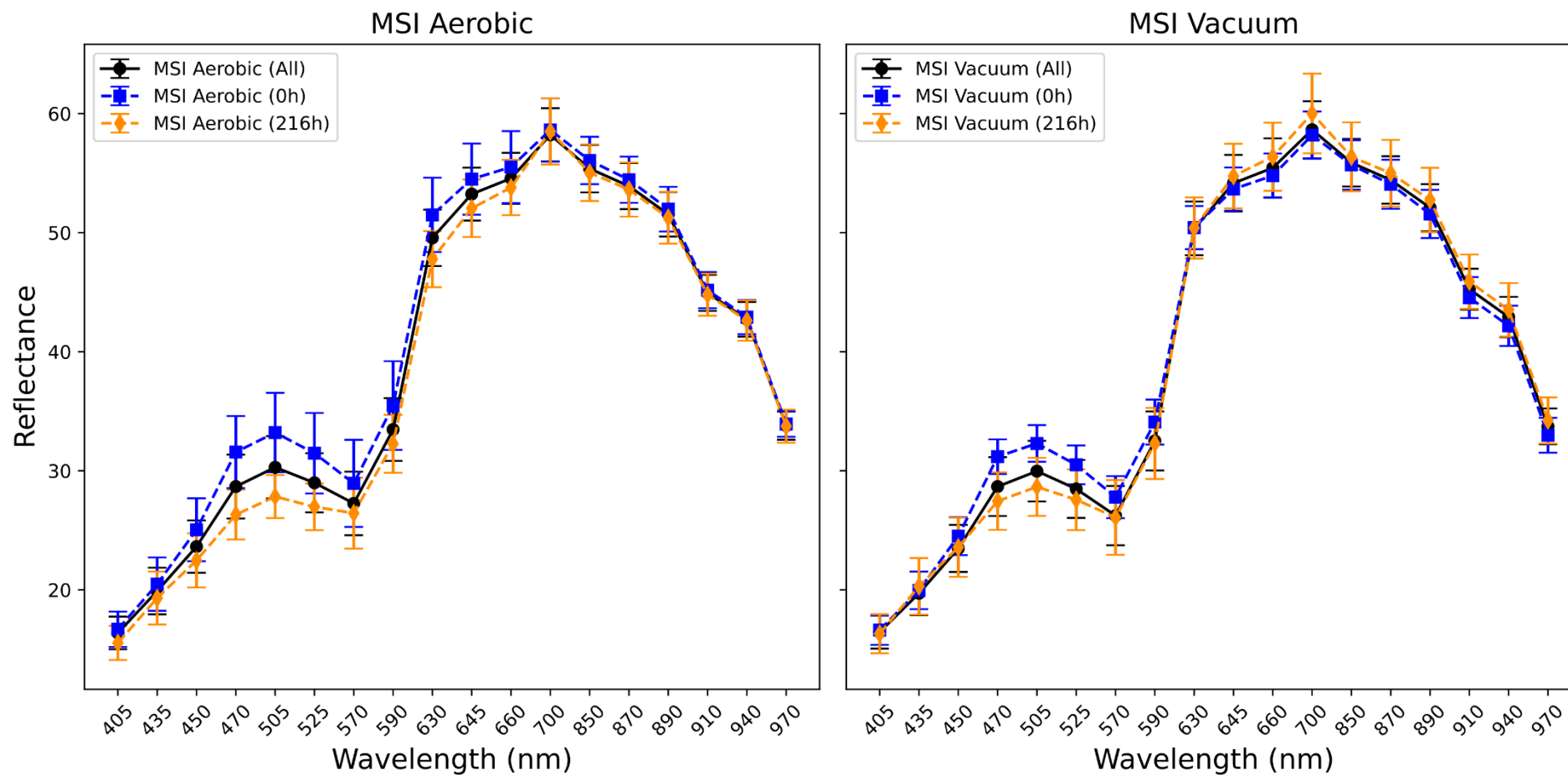


Figure 2-7: Plot showing mean and standard deviation of spectral values of MSI data across the aerobically and vacuum-stored chicken thigh samples (solid black line). Mean and standard deviation values are also indicated for 0°C/0h (blue line aerobic n=10, vacuum n=8) and 0°C/216h (orange line aerobic n=21, vacuum n=8).

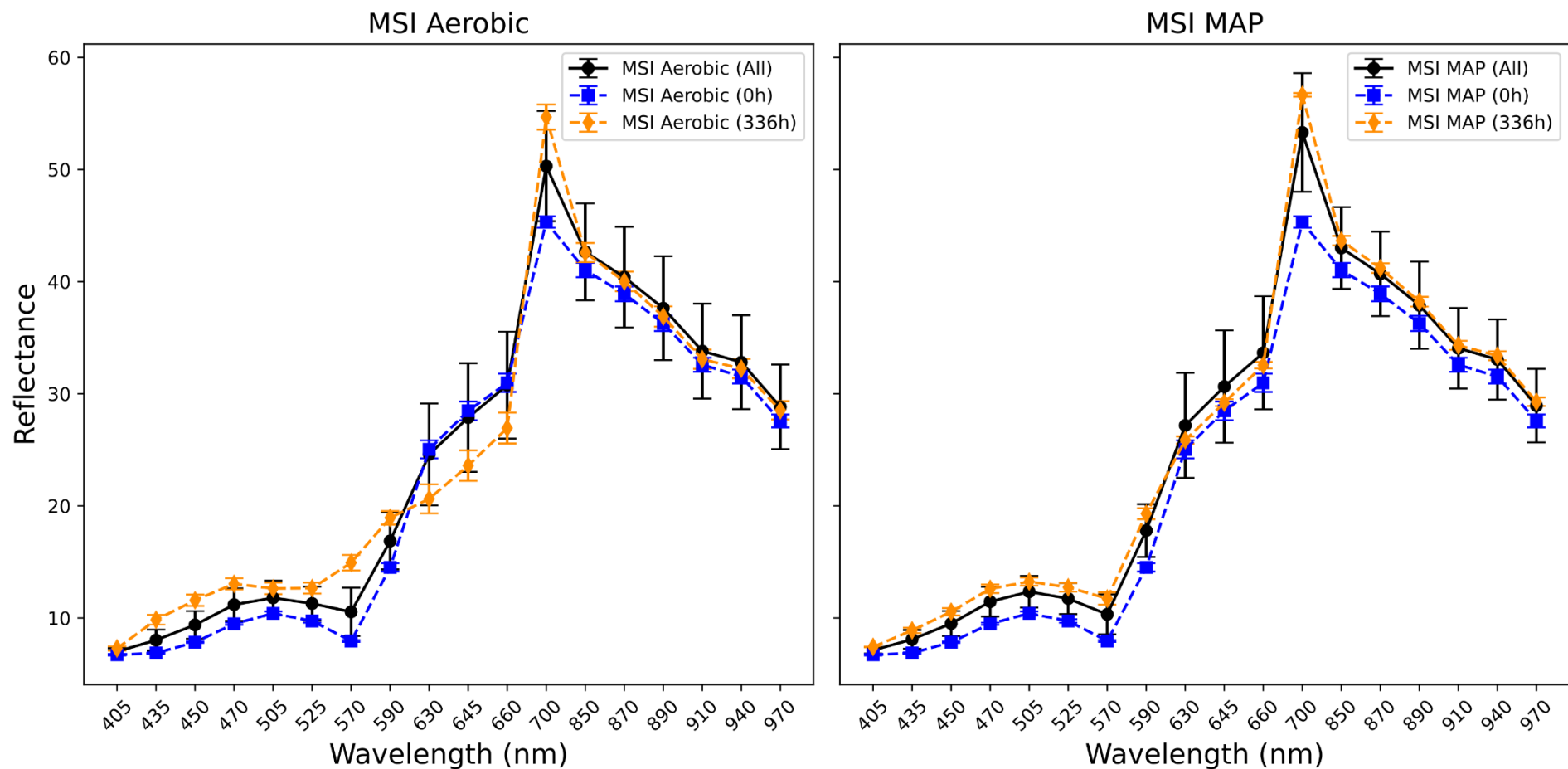


Figure 2-8: Plot showing mean and standard deviation of spectral values of MSI data across the aerobically and modified atmosphere-stored minced beef samples (solid black line). Mean and standard deviation values are also indicated for 0°C/0h (blue line aerobic n=4, MAP n=4) and 0°C/336h (orange line aerobic n=4, MAP n=4).

2.4.4 Estimation of microbial activity in chicken thigh using machine learning analysis

Performance metrics for the two data partitioning methodologies for chicken thigh samples are shown in Table 2-3 and Figure 2-9/Table 2-4.

Table 2-3 describes the performance of the complete model against the test set for the relevant modality and the “opposing” condition, for example, the recording of performance of a PLS model trained on MSI data from the aerobic packaging condition when tested with the test set from the vacuum-packaged meat. For the models trained with the concatenated training sets from both aerobic and vacuum conditions, performance was assessed with the respective test sets, giving performance metrics for each condition individually. Performance of the complete models on the test partition that fell within the cross-validation standard deviation envelope was considered to be indicative of a better-generalised model.

Of the single sensor, single packaging condition models developed and validated in the batch-on-batch fashion, only the MSI models were found to perform within this envelope and achieve sub-1 log CFU/g performance; RMSEp values of 0.615 log CFU/g for the aerobic condition and 0.793 log CFU/g for the vacuum condition were recorded. These results can be seen in Figure 2-11. For the MSI model built with the combined aerobic and vacuum training data, performance was found to be a compromise between the predictive accuracies of the single-condition models when tested with the appropriate test set as seen in Figure 2-13.

Models built with either the MSIF or FTIR data underperformed, with all single-condition models failing to achieve $R^2 \geq 0.5$. The two-condition FTIR and MSIF models narrowly reached this threshold, with FTIR achieving RMSEp performance of 1.131 log CFU/g for the aerobic condition and 1.030 log CFU/g for the vacuum condition. The two-modality (FTIR and MSI) decision (late) and feature (mid-level) approaches for all three conditions appeared to offer a greater degree of overall predictive utility than the single-modality models and those additionally incorporating the MSIF data. Improved test results were observed when comparing with the best-performing MSI models, with the late-fusion MSI-FTIR aerobic model test set validation RMSE decreasing from 0.615 to 0.547 log

CFU/g, and the corresponding vacuum model RMSE decreasing from 0.793 to 0.697 log CFU/g. These results are shown in Figure 2-12.

The two-condition MSI-FTIR feature fusion models offered improvements in performance over the respective MSI model, with RMSE values improving from 0.731/0.809 log CFU/g for aerobic and vacuum data, respectively, to 0.678/0.752 log CFU/g and accuracy improving from 81.8%/72.2% to 89.4%/87.5%. Overall, addition of the MSIF data did not offer sizeable performance benefits for the early and mid-/feature fusion models. Potential explanations could derive from the similarity in observational space captured by both the MSI and MSIF modalities, as both emerge from different configurations of the same sensor device.

Low-level fusion batch-on-batch validation results were unremarkable and were not suggestive of an alternative to the former approaches. As described in the prior literature (Fengou et al., 2019b; Smolinska et al., 2019), attempts to utilise sensor modalities with greatly differing dimensionalities may result in the larger “dominating” the useful effects of the smaller in resultant predictions. The naïve approach described here of simply concatenating the scaled dataset matrices does not offer any counter to the occurrence of these effects.

Previous food spoilage research has highlighted the importance of informative datasets that describe the plurality of food states in experimental scenarios when constructing robust models that generalise well to new data (Fengou et al., 2019a; Spyrelli et al., 2021). In addition to adopting a batch-derived partitioning approach as described here for the quantification of farmed sea bream microbial activity, Fengou *et al.* devised a random sampling approach for partitioning data into training and testing sets (Fengou et al., 2019a). A more robust implementation of this approach is used here to generalise performance of the algorithms for out-of-sample data. Through the use of repeated cross-validation for fifty 80%/20% splits, average performance of the single-modality and fusion architectures is explored, in essence automating the training and testing process examined in the referenced work. This gives an indication of the relative differences in performance between models in mean RMSE, Accuracy, and R^2 , but also an indication of the generalisation and stability of performance across

folds through the recording of standard deviation of these metrics. With limited training and testing data, evidence of model stability moving across folds is considered an important characteristic in food monitoring applications. Results for this process are shown graphically in Figure 2-9 and tabularly in Table 2-4.

Across all conditions, it was found that on average, MSI outperformed the other single-modality models, however the larger standard deviation in RMSE performance in the single-condition scenarios when compared to the equivalent sensor fusion models may be indicative of more unstable performance. The high mean and small standard deviation in the accuracy score for the MSI aerobic condition model suggests consistent placement of predictions in the ± 1 log CFU/g window, regardless of RMSE performance, which the other two modalities are not able to achieve. The appropriate feature (mid-) and decision (late) fusion approaches were able to marginally exceed mean performance for aerobic MSI. For the vacuum condition models, only late fusion was able to significantly outperform the single-sensor modalities through increased performance across all three recorded metrics, while reducing the standard deviation recorded in each. The most significant results were obtained with the combined aerobic and vacuum data due to the larger sample: MSI/FTIR/MSIF feature (mid-) fusion was shown to offer the greatest degree of performance improvement over the single-sensor approaches in all metrics, showing an RMSE_{cv} decrease of ~ 0.15 log CFU/g and a $>7\%$ increase in prediction accuracy over the best-performing single-sensor model (MSI), while substantially reducing standard deviation in both metrics, a result suggestive of a model with more predictive stability when encountering mixed condition data.

Whilst not entirely representative of the experimental conditions encountered in this research, Spyrelli *et al.*'s examination of the performance of PLS regression models using both FTIR and MSI for the prediction of spoilage in chicken achieved similar levels of performance – their conclusions were that the efficacy of their models were impacted by the inter-batch variability observed due to separation of meat sampling by four months (Spyrelli *et al.*, 2021). Regardless, in their case FTIR proved the most appropriate sensor type for informative model

development. The poorer efficacy of FTIR models built in this research may be attributable to minimal preprocessing of the spectra, beyond restriction of the datasets to the 2000-900 cm^{-1} region as described, leading to models fitting on noisy data with a potential degree of confounding information.

Table 2-3: Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in chicken thigh samples using batch-on-batch validation (Aerobic n=70, Vacuum n=65, Combined n=135)

Model	CV Performance			Aerobic Performance			Vacuum Performance		
		RMSEcv (SD)	LVs (Meta LVs)	RMSEp	Acc %	R ²	RMSEp	Acc %	R ²
MSI	Aerobic	0.6779 (0.1664)	9	0.6147	86.3636	0.8604	1.4145	56.9444	-0.1725
	Vacuum	0.5199 (0.1702)	4	0.8956	75.7576	0.7037	0.7933	75.0000	0.6312
	A+V	0.6928 (0.1311)	6	0.7311	81.8182	0.8025	0.8090	72.2222	0.6165
MSIF	Aerobic	0.8329 (0.2677)	3	2.2473	30.3030	-0.8655	1.7561	41.6667	-0.8073
	Vacuum	0.7304 (0.3058)	5	1.6805	42.4242	-0.0432	1.0976	63.8889	0.2941
	A+V	0.8333 (0.1779)	6	1.5469	42.4242	0.1161	0.9222	70.8333	0.5016
FTIR	Aerobic	0.8936 (0.2631)	6	1.6040	53.0303	0.0496	1.7679	50.0000	-0.8315
	Vacuum	0.6321 (0.1211)	6	1.4657	56.0606	0.2064	1.2924	62.5000	0.0212
	A+V	0.8764 (0.1789)	7	1.1310	63.6364	0.5275	1.0304	68.0556	0.3779
<i>Late (Decision) Fusion</i>									
	Aerobic	0.6421 (0.2086)	9,6	0.5466	93.9394	0.8896	1.3863	52.7778	-0.1263
	Vacuum	0.4829 (0.1616)	4,6	0.9821	63.6364	0.6437	0.6971	81.9444	0.7152
	A+V	0.6314 (0.1126)	6,7	0.6779	84.8485	0.8302	0.6959	84.7222	0.7162
<i>Mid (Feature) Fusion</i>									
MSI +FTIR	Aerobic	0.6483 (0.1857)	9,6 (6)	0.5607	93.9394	0.8838	1.3778	56.9444	-0.1124
	Vacuum	0.4753 (0.1293)	4,7 (5)	0.9415	66.6667	0.6725	0.6638	84.7222	0.7418
	A+V	0.6049 (0.1400)	6,7 (6)	0.6782	89.3939	0.8301	0.7520	87.5000	0.6686
<i>Early Fusion</i>									
	Aerobic	0.8334 (0.2347)	6	1.2906	72.7273	0.3847	1.7791	44.4444	-0.8548
	Vacuum	0.5429 (0.1905)	7	1.3672	53.0303	0.3094	1.2745	69.4444	0.0481
	A+V	0.7672 (0.1736)	6	1.1962	66.6667	0.4713	1.2714	58.3333	0.0527
<i>Late (Decision) Fusion</i>									
	Aerobic	0.6399 (0.1930)	9,6,3	0.6533	87.8788	0.8424	0.9940	69.4444	0.4210
	Vacuum	0.4997 (0.1603)	4,6,5	0.9978	63.6364	0.6323	0.6918	83.3333	0.7195
	A+V	0.5943 (0.1210)	6,7,6	0.7611	80.3030	0.7860	0.6520	84.7222	0.7509
<i>Mid (Feature) Fusion</i>									
MSI +FTIR +MSIF	Aerobic	0.6269 (0.1938)	9,6,3 (6)	0.5907	90.9091	0.8711	1.0228	65.2778	0.3870
	Vacuum	0.5026 (0.1956)	4,6,5 (5)	1.0467	62.1212	0.5952	0.7010	83.3333	0.7120
	A+V	0.5748 (0.1300)	6,7,6 (5)	0.7818	75.7576	0.7742	0.7544	80.5556	0.6665
<i>Early Fusion</i>									
	Aerobic	0.7731 (0.2164)	6	1.0802	72.7273	0.5690	1.6575	44.4444	-0.6099
	Vacuum	0.5472 (0.2002)	7	1.3565	45.4545	0.3203	1.2171	70.8333	0.1320
	A+V	0.6884 (0.1774)	6	1.1749	63.6364	0.4901	1.0544	63.8889	0.3486

Table 2-4: Generalised Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in chicken thigh samples using repeated nested cross validation (10 x 5 folds)

Model	RMSE	RMSE	Acc %	Acc %	R2	R2
	Mean	SD	Mean	SD	Mean	SD
<i>Aerobic Models</i>						
PLS-Regression-MSI	0.6270	0.0766	89.4921	5.5495	0.8313	0.0426
PLS-Regression-FTIR	0.8853	0.1355	75.4497	8.1551	0.6548	0.1428
PLS-Regression-MSIF	1.0940	0.1226	61.1323	8.1930	0.4876	0.1170
PLS-MSI-FTIR-EarlyFusion	0.6731	0.1165	86.3942	6.6007	0.8013	0.0757
PLS-MSI-FTIR-MSIF-EarlyFusion	0.6328	0.1007	88.5344	6.4148	0.8251	0.0590
PLS-MSI-FTIR-MidFusion	0.6010	0.0956	90.4233	6.0550	0.8432	0.0538
PLS-MSI-FTIR-MSIF-MidFusion	0.5899	0.0930	90.0026	5.7044	0.8494	0.0504
PLS-MSI-FTIR-LateFusion	0.5923	0.0875	91.6270	5.3435	0.8485	0.0479
PLS-MSI-FTIR-MSIF-LateFusion	0.5785	0.0851	91.7619	4.7053	0.8554	0.0471
<i>Vacuum Models</i>						
PLS-Regression-MSI	0.6276	0.1076	89.1217	6.4258	0.7432	0.1014
PLS-Regression-FTIR	0.7385	0.1087	84.9841	5.0404	0.6475	0.1137
PLS-Regression-MSIF	0.8266	0.1254	78.0291	6.6680	0.5587	0.1415
PLS-MSI-FTIR-EarlyFusion	0.5988	0.0829	90.3651	5.3492	0.7688	0.0691
PLS-MSI-FTIR-MSIF-EarlyFusion	0.6176	0.0813	89.3439	5.0660	0.7537	0.0755
PLS-MSI-FTIR-MidFusion	0.5928	0.0893	91.1164	5.4086	0.7717	0.0801
PLS-MSI-FTIR-MSIF-MidFusion	0.5912	0.0927	90.1746	6.0481	0.7727	0.0804
PLS-MSI-FTIR-LateFusion	0.5586	0.0782	91.3201	5.5276	0.7977	0.0698
PLS-MSI-FTIR-MSIF-LateFusion	0.5691	0.0941	91.0291	5.5210	0.7891	0.0802
<i>Aerobic+Vacuum Models</i>						
PLS-Regression-MSI	0.7358	0.1018	84.2222	4.4987	0.7344	0.1007
PLS-Regression-FTIR	0.7641	0.0727	82.1293	4.5691	0.7218	0.0599
PLS-Regression-MSIF	0.9682	0.0771	68.9360	6.2140	0.5536	0.0817
PLS-MSI-FTIR-EarlyFusion	0.6203	0.0650	89.7461	4.2534	0.8159	0.0431
PLS-MSI-FTIR-MSIF-EarlyFusion	0.6116	0.0524	90.0094	3.5852	0.8221	0.0324
PLS-MSI-FTIR-MidFusion	0.6141	0.0534	90.4431	3.2687	0.8198	0.0393
PLS-MSI-FTIR-MSIF-MidFusion	0.5892	0.0494	91.5091	3.2946	0.8346	0.0323
PLS-MSI-FTIR-LateFusion	0.6392	0.0693	88.9044	3.3687	0.8029	0.0561
PLS-MSI-FTIR-MSIF-LateFusion	0.6163	0.0603	89.8949	3.7594	0.8180	0.0440

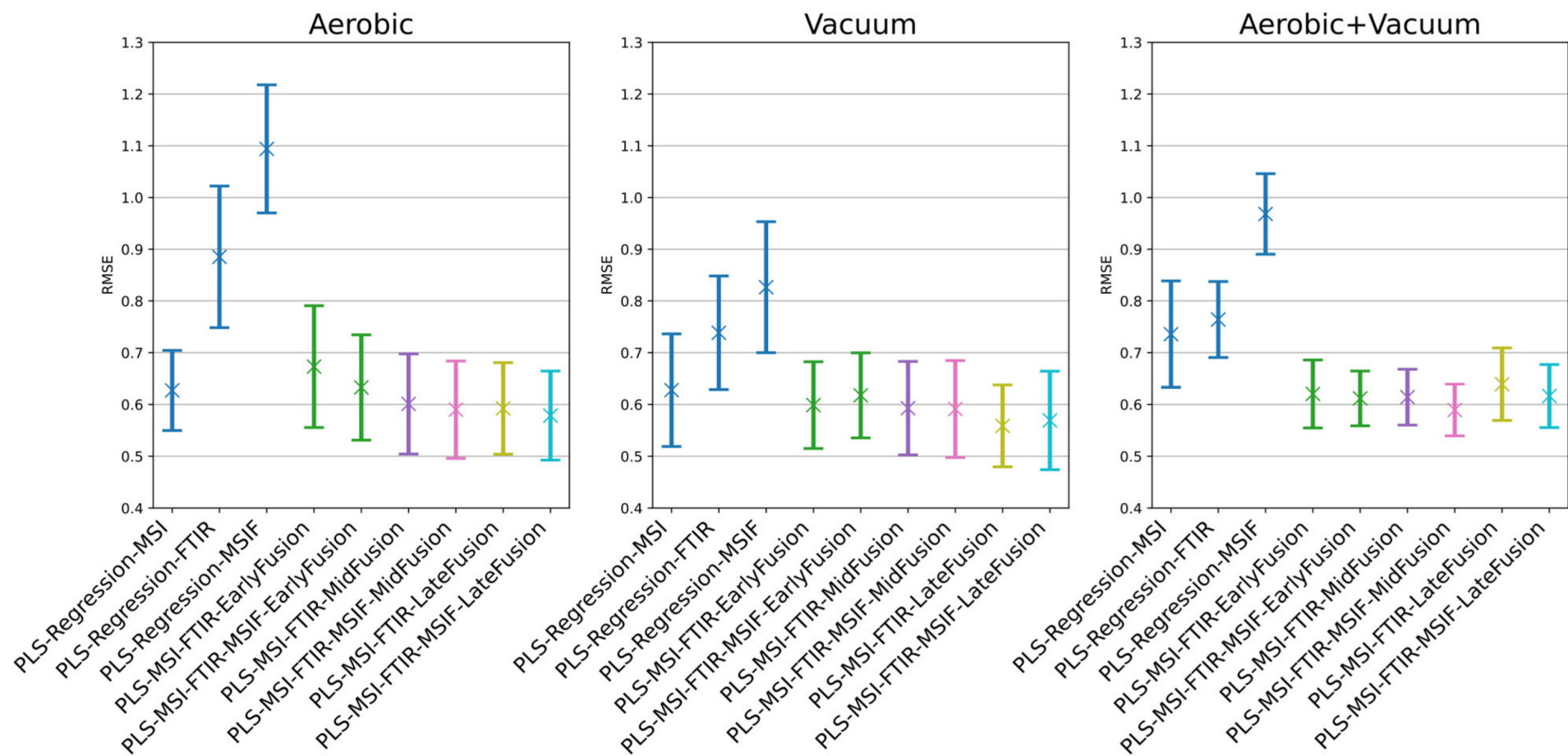


Figure 2-9: The root mean squared error of prediction of the repeated nested cross validation process for chicken thigh data.

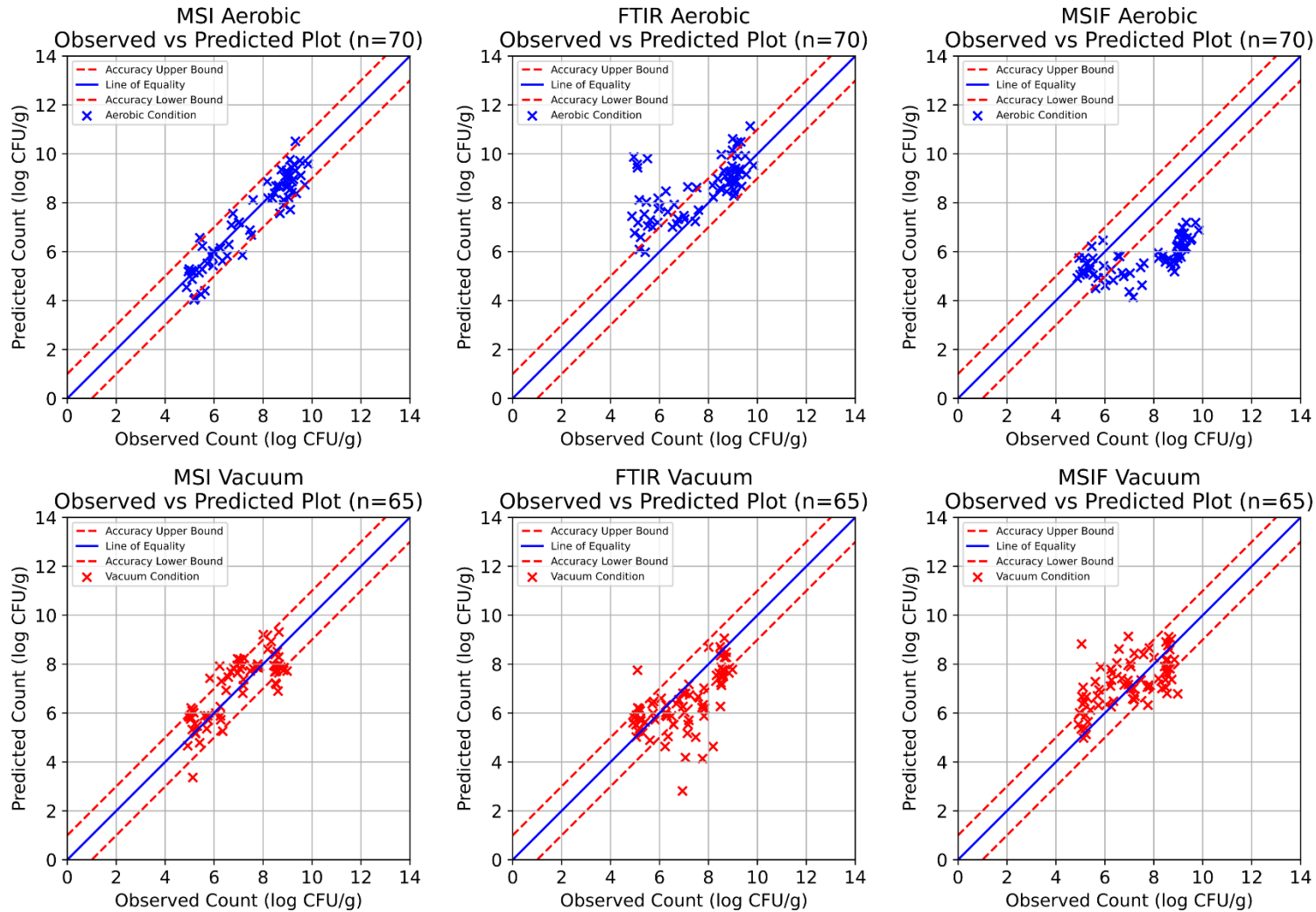


Figure 2-10: Performance plots of testing for single-sensor, single-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets.

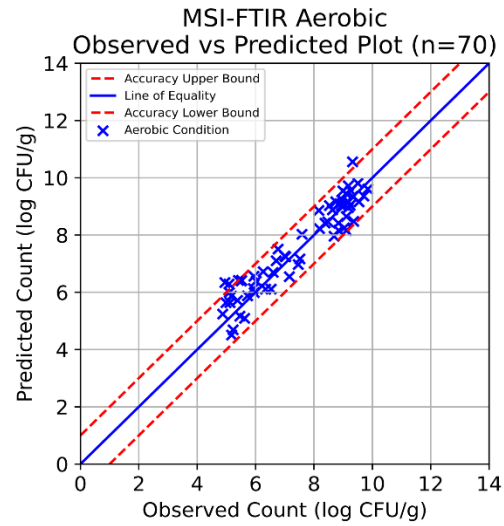
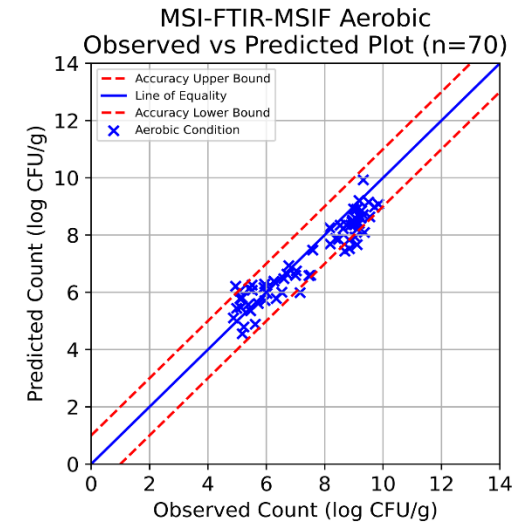
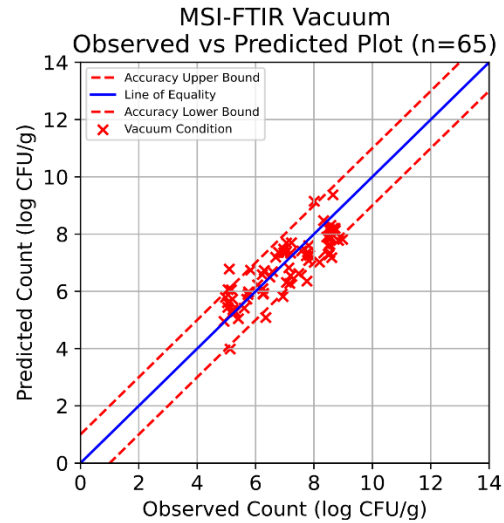
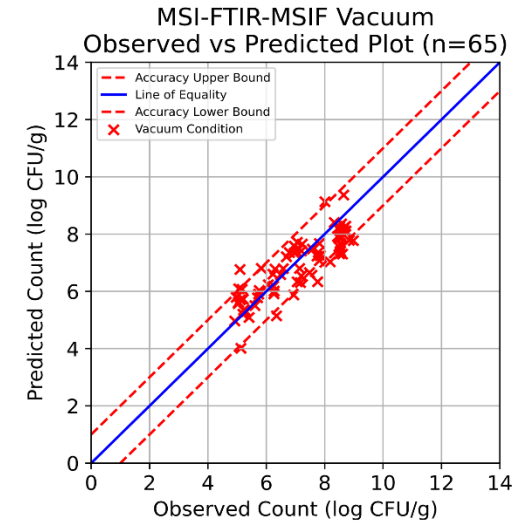


Figure 2-11: Performance plots of testing for two/three-sensor, single-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.

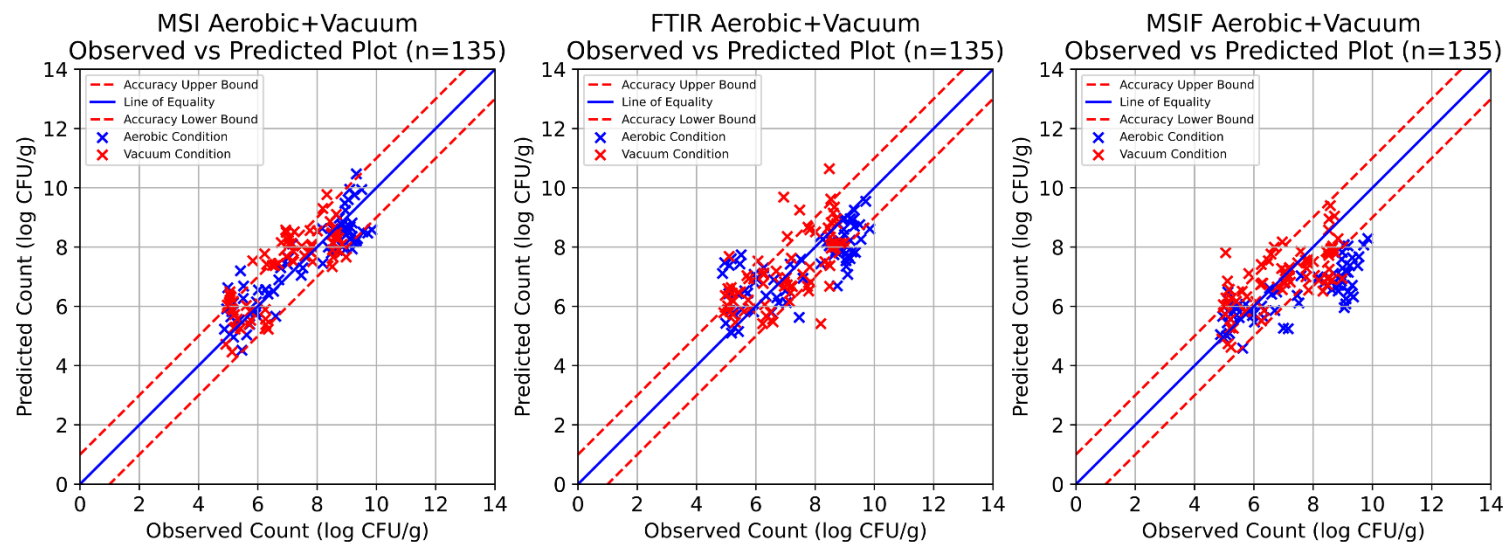


Figure 2-12: Performance plots of testing for single-sensor, two-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets.

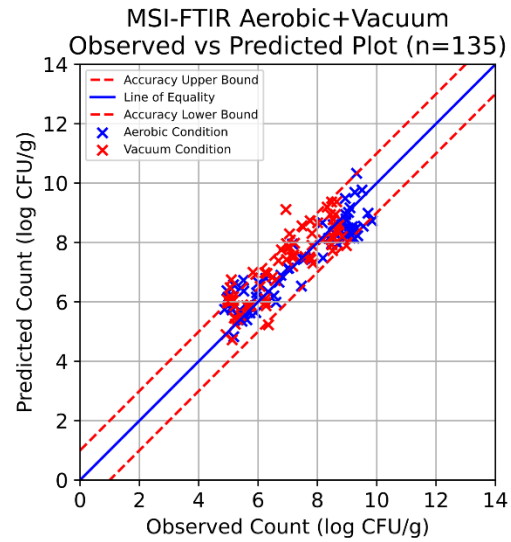
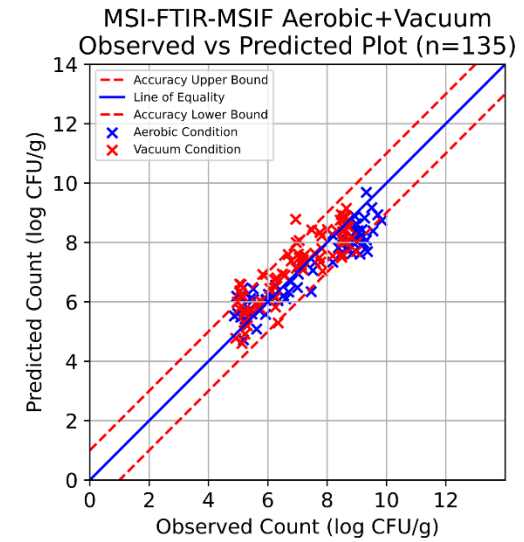


Figure 2-13: Performance plots of testing for two/three-sensor, two-packaging-condition chicken thigh spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.

2.4.5 Estimation of microbial activity in beef mince using machine learning analysis

Performance metrics for beef mince data are shown in Table 2-5 for the batch-on-batch approach, and Figure 2-15/Table 2-6 for the alternative approach.

Results for the batch-on-batch validation scheme appeared affected by the inter-batch variation visible in Figure 2-4. This is suggested by the strong intra-batch cross validation performance exhibited by most single sensor models.

Batch-on-batch cross-validation RMSE performance across the late fusion approaches is varied: in both the aerobic and combined aerobic and MAP conditions, the associated two- and three-sensor decision fusion models exceeded the cross-validation performance of all single-sensor models.

Predictions for the aerobic condition were improved with the late fusion MSI/FTIR model, with RMSE values decreasing from 1.306 to 1.167 log CFU/g and an accuracy improving from 45.7% to 56.5%. This aerobic performance was mirrored in the feature (mid-) fusion approach. The poor test set validation of the models is likely attributable to a combination of small training sets, along with the variation between the batches. None of the explored approaches were able to achieve RMSEp results lower than 1 log CFU/g. Only the combined aerobic and MAP MSIF model shows performance approaching this threshold, with an RMSEp of 1.08 log CFU/g, however, the configuration appears unsuitable for prediction of the aerobic test TVC.

The results of this data partitioning approach highlight the high degree of variability between the batches used for training and testing the respective model types, stressing the importance of supplementation of the training data with further batches to construct more representative models that are applicable to “production” use cases.

Figure 2-17 highlights the bias present in the batch-on-batch single-sensor models. The models developed for the aerobic sensor modalities suggest a tendency of overestimation of bacterial counts, whereas the MSI and FTIR

models developed for the modified atmosphere packaging condition show poor generalisation to the test batch.

In a similar manner to the chicken spoilage scenario, repeated nested cross-validation was used to assess the generalised performance of the three tiers of fusion approach when compared to the corresponding single-sensor models, albeit with a lesser degree of confidence due to the wider standard deviation envelope. A general downward trend was visible in Figure 2-14 in mean and standard deviation of RMSE performance when moving from single sensors towards decision (late-) fusion methods. The performance metrics for the combined aerobic and MAP condition (Table 2-6) suggest the most convincing improvements to predictive performance by adopting a sensor fusion approach: MSI/FTIR/MSIF mid-/late fusion show an increase in mean accuracy of between 6 and 8% over FTIR alone.

Table 2-5: Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in beef mince samples using batch-on-batch validation (Aerobic n=46, Vacuum n=57, Combined n=103)

Model	CV Performance			Aerobic Performance			MAP Performance		
		RMSEcv (SD)	LVs (Meta LVs)	RMSEp	Acc %	R ²	RMSEp	Acc %	R ²
MSI	Aerobic	0.7170 (0.2614)	4	1.3048	45.6522	0.4326	1.8746	37.5000	-0.1164
	MAP	0.5578 (0.1258)	5	1.3074	56.5217	0.4303	1.4027	58.9286	0.3749
	A+M	0.7583 (0.1668)	5	1.4338	47.8261	0.3148	1.3526	55.3571	0.4188
MSIF	Aerobic	0.9630 (0.2673)	5	1.7839	30.4348	-0.0614	1.2002	51.7857	0.5422
	MAP	0.8023 (0.2010)	3	2.5896	19.5652	-1.2368	1.3251	42.8571	0.4419
	A+M	0.8441 (0.2095)	7	1.8263	30.4348	-0.1124	1.0808	62.5000	0.6288
FTIR	Aerobic	0.7764 (0.2573)	9	1.3851	52.1739	0.3606	2.4253	39.2857	-0.8688
	MAP	0.8309 (0.3005)	15	1.4396	47.8261	0.3093	1.6286	37.5000	0.1574
	A+M	0.7666 (0.1561)	16	1.5534	50.0000	0.1958	1.7130	44.6429	0.0678
<i>Late (Decision) Fusion</i>									
	Aerobic	0.5854 (0.2300)	4,9	1.1663	56.5217	0.5467	2.0272	44.6429	-0.3056
	MAP	0.5502 (0.1436)	5,15	1.2636	63.0435	0.4678	1.2846	62.5000	0.4757
	A+M	0.6220 (0.1271)	5,16	1.2273	56.5217	0.4980	1.4416	53.5714	0.3398
<i>Mid (Feature) Fusion</i>									
MSI +FTIR	Aerobic	0.5893 (0.2259)	4,9 (3)	1.1896	58.6957	0.5284	2.0653	41.0714	-0.3551
	MAP	0.5074 (0.1135)	5,15 (5)	1.2361	67.3913	0.4907	1.2968	66.0714	0.4657
	A+M	0.5145 (0.0929)	5,16 (10)	1.4185	54.3478	0.3290	1.6726	44.6420	0.1110
<i>Early Fusion</i>									
	Aerobic	0.6197 (0.2245)	5	1.8869	41.3043	-0.1866	1.7650	25.0000	0.0103
	MAP	0.5767 (0.1555)	7	2.4270	17.3913	-0.9631	1.9378	30.3571	-0.1930
	A+M	0.6814 (0.1456)	6	1.4526	45.6522	0.2968	1.3907	39.2857	0.3856
<i>Late (Decision) Fusion</i>									
	Aerobic	0.5907 (0.2308)	4,9,5	1.1186	58.6957	0.5827	1.8653	51.7857	-0.1058
	MAP	0.5535 (0.1447)	5,15,3	1.2515	56.5217	0.4776	1.2531	62.5000	0.5010
	A+M	0.6282 (0.1277)	5,16,7	1.1865	58.6957	0.5304	1.4244	53.5714	0.3551
<i>Mid (Feature) Fusion</i>									
MSI +FTIR +MSIF	Aerobic	0.5793 (0.1983)	4,9,5 (3)	1.1031	60.8696	0.5945	1.7455	51.7857	0.0321
	MAP	0.5457 (0.1266)	5,15,3 (5)	1.3075	60.8696	0.4302	1.3970	66.0714	0.3800
	A+M	0.6650 (0.1410)	5,16,7 (4)	1.2013	56.5217	0.5186	1.4952	55.3571	0.2894
<i>Early Fusion</i>									
	Aerobic	0.6411 (0.2394)	5	1.7345	43.4783	-0.0034	1.6092	26.7857	0.1770
	MAP	0.6035 (0.1595)	7	2.1515	21.7391	-0.5440	1.6677	35.7143	0.1161
	A+M	0.6963 (0.1450)	6	1.3967	52.1739	0.3494	1.3136	44.6429	0.4516

Table 2-6: Generalised Performance metrics for PLS-R models built for the purpose of estimation of bacterial Total Viable Count (log CFU/g TVC) in beef mince samples using repeated nested cross validation (50 folds).

Model	RMSE	RMSE	Acc %	Acc %	R2	R2
	Mean	SD	Mean	SD	Mean	SD
Aerobic Models						
PLS-Regression-MSI	1.0622	0.2255	69.4444	10.2883	0.5332	0.1809
PLS-Regression-FTIR	0.9530	0.1668	73.0526	10.3674	0.6258	0.1233
PLS-Regression-MSIF	0.9709	0.1719	70.2515	8.6537	0.6106	0.1326
PLS-MSI-FTIR-EarlyFusion	0.9149	0.1584	76.0117	9.7809	0.6539	0.1158
PLS-MSI-FTIR-MSIF-EarlyFusion	0.8739	0.1351	78.7368	6.6358	0.6859	0.1021
PLS-MSI-FTIR-MidFusion	0.9170	0.1752	75.6842	9.5596	0.6491	0.1290
PLS-MSI-FTIR-MSIF-MidFusion	0.9432	0.1610	74.3333	7.7958	0.6309	0.1333
PLS-MSI-FTIR-LateFusion	0.8444	0.1567	78.9181	9.7589	0.7057	0.1024
PLS-MSI-FTIR-MSIF-LateFusion	0.8284	0.1499	76.9649	9.4674	0.7181	0.0908
MAP Models						
PLS-Regression-MSI	0.9635	0.1831	74.7352	9.1533	0.6241	0.1577
PLS-Regression-FTIR	0.9419	0.1690	73.2846	9.3656	0.6419	0.1374
PLS-Regression-MSIF	0.9674	0.1402	70.5336	10.7800	0.6275	0.1141
PLS-MSI-FTIR-EarlyFusion	1.0374	0.2133	73.1383	9.5623	0.5593	0.1954
PLS-MSI-FTIR-MSIF-EarlyFusion	0.9287	0.1611	76.6680	9.0988	0.6476	0.1481
PLS-MSI-FTIR-MidFusion	0.8764	0.1532	76.9091	8.2672	0.6938	0.1011
PLS-MSI-FTIR-MSIF-MidFusion	0.7999	0.1389	79.9684	8.1167	0.7413	0.0943
PLS-MSI-FTIR-LateFusion	0.8151	0.1646	82.7747	7.7723	0.7315	0.1094
PLS-MSI-FTIR-MSIF-LateFusion	0.7924	0.1471	81.9130	8.9743	0.7491	0.0838
Aerobic+MAP Models						
PLS-Regression-MSI	0.9904	0.1162	73.0244	7.0759	0.6210	0.0883
PLS-Regression-FTIR	0.8992	0.1111	75.7561	5.9825	0.6898	0.0650
PLS-Regression-MSIF	1.0356	0.1283	69.8537	7.1320	0.5817	0.1172
PLS-MSI-FTIR-EarlyFusion	0.8972	0.1010	77.5122	5.6106	0.6906	0.0643
PLS-MSI-FTIR-MSIF-EarlyFusion	0.8635	0.0865	77.6098	6.0670	0.7125	0.0600
PLS-MSI-FTIR-MidFusion	0.8467	0.1067	79.2683	5.8384	0.7234	0.0677
PLS-MSI-FTIR-MSIF-MidFusion	0.8247	0.1257	81.4634	6.6528	0.7369	0.0726
PLS-MSI-FTIR-LateFusion	0.8138	0.1123	81.9024	5.5638	0.7457	0.0597
PLS-MSI-FTIR-MSIF-LateFusion	0.7867	0.1227	83.8049	6.2233	0.7599	0.0704

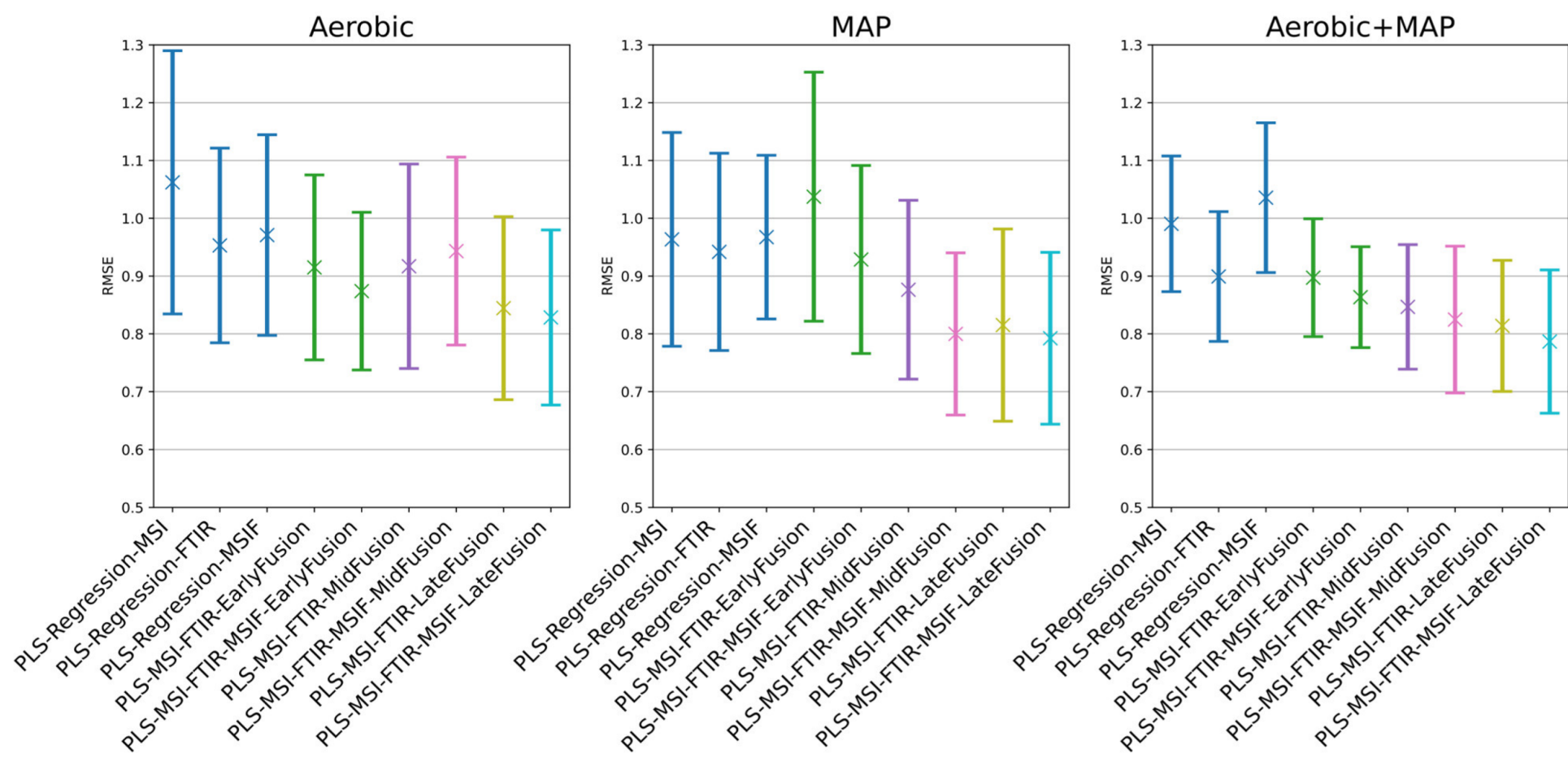


Figure 2-14: The root mean squared error of prediction of the repeated nested cross validation process for beef mince data.

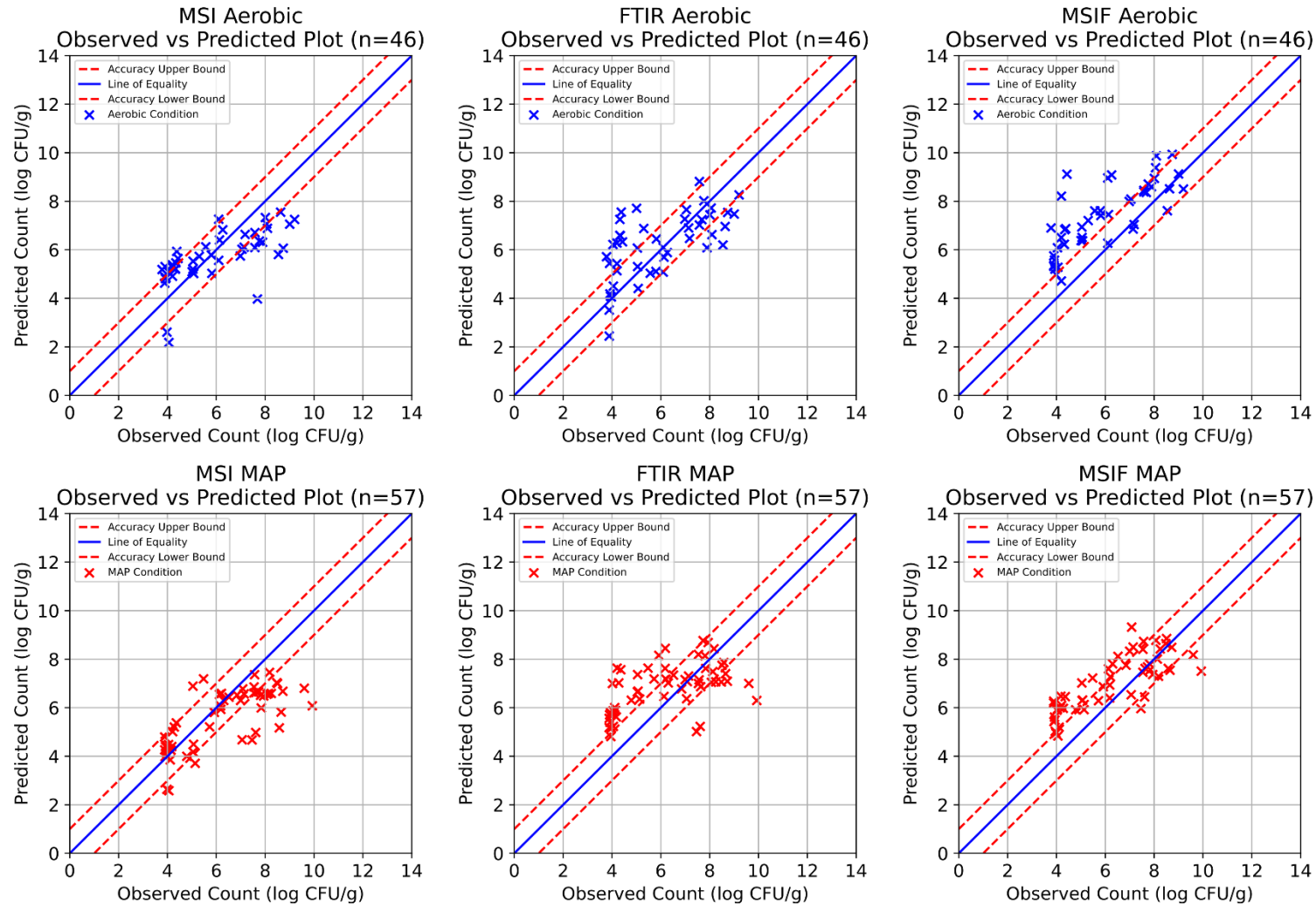


Figure 2-15: Performance plots of testing for single-sensor, single-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets.

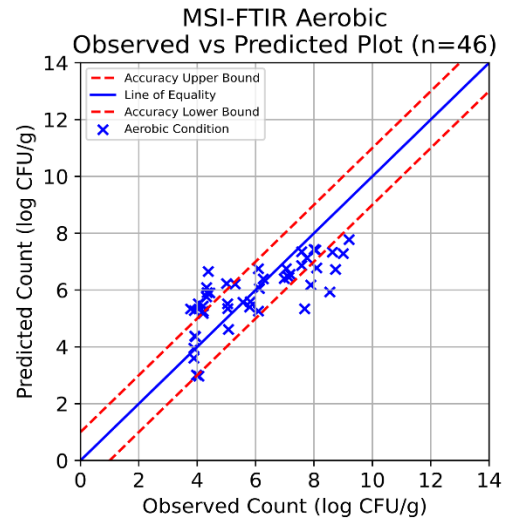
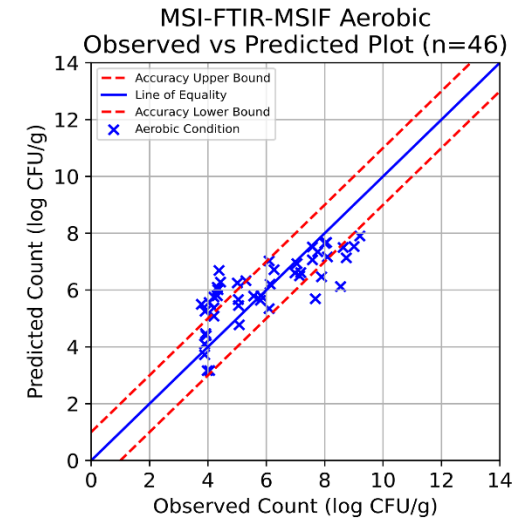
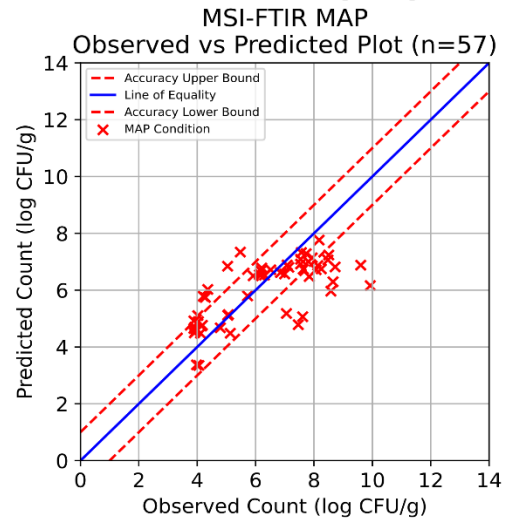
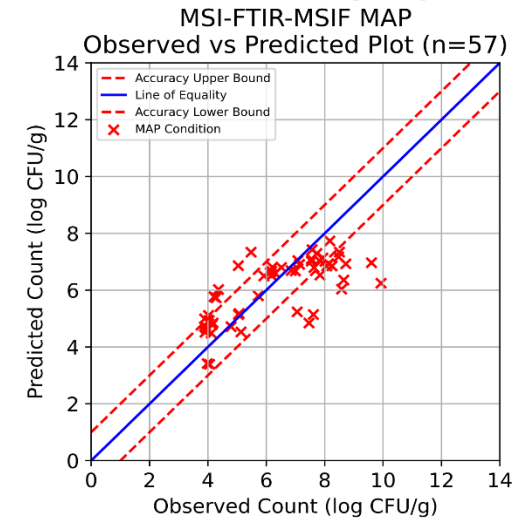


Figure 2-16: Performance plots of testing for two/three-sensor, single-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.

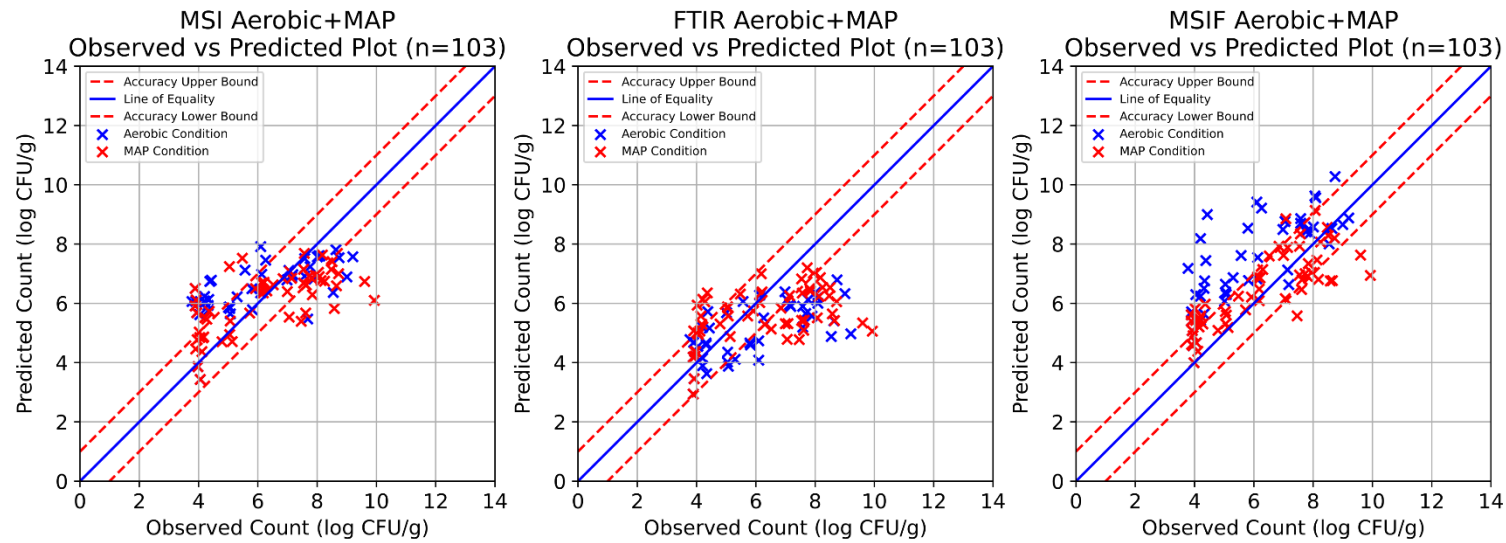


Figure 2-17: Performance plots of testing for single-sensor, two-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets.

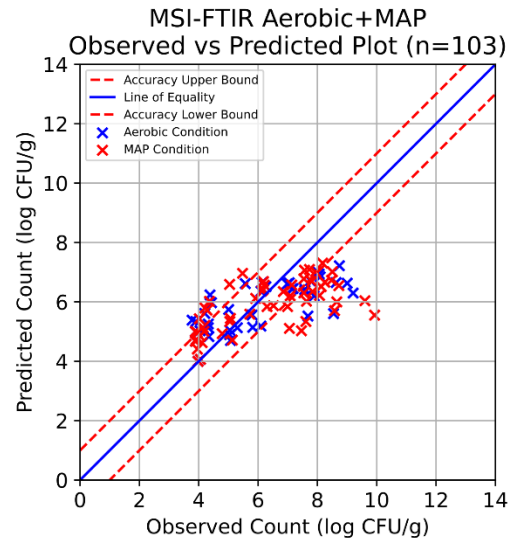
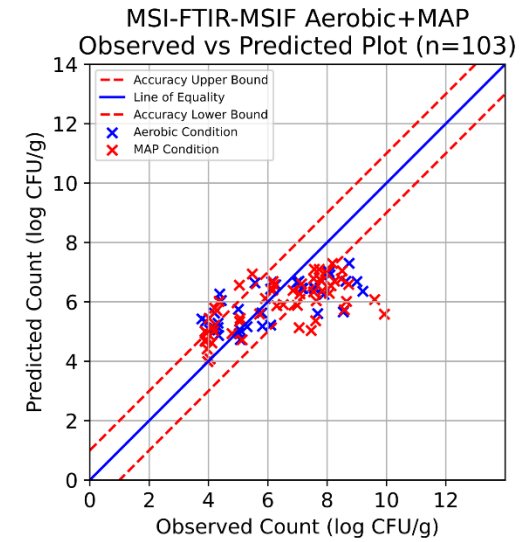


Figure 2-18: Performance plots of testing for two/three-sensor, two-packaging-condition beef mince spoilage prediction models built using MSI, FTIR and MSIF datasets, built using the late (decision) fusion approach.

2.4.6 Discussion and Conclusions

The results of this comparative performance assessment of different data fusion approaches in two meat spoilage scenarios suggest in a majority of cases enhanced predictive capability and generalisation to new data when compared to models built with individual constituent sensor device data. At the time of writing, fusion of multi-spectral imaging and Fourier-transform infrared in the context of meat spoilage assessment has not seen significant publication.

Results from the batch-validation feature fusion and decision fusion pipelines showed that the latter's results are in most cases on-par with feature fusion, and in the case of chicken spoilage, superior performance is observed. The performance of the fusion approaches was corroborated in the repeated nested cross validation results: a general trend is an improvement in mean RMSE performance when compared to the single sensor approaches, alongside a reduction in standard deviation, indicating more stable performance in the recorded fold results. For the individual sensor approaches explored for chicken spoilage analysis, MSI performance was in most cases superior to FTIR and MSI with fluorescence (MSIF), and in some cases performed similarly to the equivalent fusion approaches assessed. For the beef spoilage scenario, the individual sensor results were more unstable, with FTIR narrowly achieving better RMSE performance, with decision fusion methods indicated as offering the most significant performance improvement. It is anticipated that refinement of these feature and late-fusion methods, through more consideration of the fusion strategy, or in the case of the latter, additions to the first layer (expert) model library or reconfiguration of the meta-learning layer will result in improved performance, bringing the strategy more in line with stacked generalisation approaches. Modification of the hyperparameter tuning regime with alternative approaches such as Bayesian methods may accelerate convergence on optimal configurations.

Whilst widely explored in a variety of contexts outside of food analysis, sensor/data fusion has seen limited utilisation in the area of meat spoilage quantification. Previous attempts have been made in the analysis of minced meat

spoilage in Fengou *et al.* (Fengou et al., 2019b), which used naïve early fusion as previously described, the results of which suggested further research and optimisation was a potential avenue to explore for improvement of spoilage prediction. Additional data fusion efforts have been made in prediction of other meat quality metrics, such as moisture content in freeze-thawed pork (Cheng et al., 2022), in which adoption of decision fusion approaches similar to that implemented here achieved a notable enhancement of predictive utility over the models using single sensor modalities in isolation. The exploration of fusion methodologies in the context of food spoilage is timely, as increasing affordability of many high-throughput, minimally invasive sensor types is likely to prompt increased adoption throughout production and distribution pipelines. While unlikely to offer an accuracy matching established lab testing protocol in isolation, they offer clear benefits in terms of responsiveness, portability, and economic optimisation through minimisation of food wastage. Fusion of these sensor outputs is one possible avenue with which to improve upon predictive accuracy and reliability; interpretation of this growing wealth of multi-modal food monitoring data will likely lead to more informative, robust models for a variety of food control tasks.

The machine learning models generated in this work have corroborated existing research in the utilisation of sensor technologies for microbial count prediction, whilst also highlighting the potential performance improvements achievable by implementing a data fusion approach for combining their predictive power. The performance of these approaches serves as a suggestion of directions for optimisation of high-throughput sensor performance in meat analysis, as model training has been constrained with small sample sizes. Even so, the comparison of multiple architectures on an “even footing” suggest performance advantages obtainable by adopting fusion approaches in this domain. Regarding the efficacy of individual sensors applied in this research, it is likely that further work in the data preparatory stage may positively impact the predictive power of resultant models. Comparison with existing research shows similar data processing procedures were used for MSI and FTIR spectral preparation (Sahar and Dufour, 2014; Spyrelli et al., 2020, 2021), namely the application of SNV and filtering

techniques for preprocessing spectral datasets. However, some sources demonstrate optimisation and feature selection techniques for highly dimensional data, such as that from FTIR, that may improve results even for machine learning algorithms that are capable of handling some degree of collinearity, such as partial least squares techniques (PLS). Use of these methods must account for the data fusion context; some, such as recursive feature elimination and genetic algorithm (GA) derived techniques have shown promise in optimising the number of features available to subsequent algorithms (Leardi and Lupiáñez González, 1998), however, these methods result in removal of information that isn't informative in a solitary context, but may become informative with the addition of additional information provided by other sensor modalities (Smolinska et al., 2019).

In the interest of improving confidence in decisions increasingly informed by the use of high-throughput sensor technologies in the food analysis sector, more well-informed machine learning predictions may be made by incorporating the predictive power from multiple sensor modalities, reducing uncertainty by forming a more complete view of biological state through the union of each sensor's observational spaces. In practice, these sensor observations could be made throughout the food production and distribution pipeline, with a combination of devices installed in a permanent fashion with specialist oversight, and those deployed portably at the food business operator level (Nychas et al., 2021). The successful summation and fusion of this food state information in a real-world situation is dependent on the success of analysts in accounting for information collected across differences in sampling times, storage and distribution conditions. As such, appropriate decision support pipelines may be necessary to handle the influx of information and facilitate predictive model construction and inference in a semi-automated fashion.

Further work is required to refine fusion techniques suitable for the food analysis domain, through the exploration of different combinations of machine learning algorithms and methodologies, with larger quantities of experimental data with which to build more dependable models. In conclusion, the approaches

demonstrated here may serve as an indicator of potential utility of all three general types of sensor fusion in real world food analysis scenarios.

References

- Abdi, H. (2010) 'Partial least squares regression and projection on latent structure regression (PLS Regression)', *WIREs Computational Statistics*, 2(1), pp. 97–106.
- Alexandrakis, D., Downey, G. and Scannell, A.G.M. (2012) 'Rapid Non-destructive Detection of Spoilage of Intact Chicken Breast Muscle Using Near-infrared and Fourier Transform Mid-infrared Spectroscopy and Multivariate Statistics', *Food and Bioprocess Technology*, 5(1), pp. 338–347.
- Argyri, A.A., Jarvis, R.M., Wedge, D., Xu, Y., Panagou, E.Z., Goodacre, R. and Nychas, G.-J.E. (2013) 'A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage', *Food Control*, 29(2), pp. 461–470.
- Barnes, R.J., Dhanoa, M.S. and Lister, S.J. (1989) 'Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra', *Applied Spectroscopy*, 43(5), pp. 772–777.
- Bergstra, J. and Bengio, Y. (2012) 'Random Search for hyperparameter Optimization', *Journal of Machine Learning Research*, 13(10), pp. 281–305.
- Böcker, U., Ofstad, R., Wu, Z., Bertram, H.C., Sockalingum, G.D., Manfait, M., Egelanddal, B. and Kohler, A. (2007) 'Revealing covariance structures in fourier transform infrared and Raman microspectroscopy spectra: a study on pork muscle fiber tissue subjected to different processing parameters', *Applied Spectroscopy*, 61(10), pp. 1032–1039.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L. and Busto, O. (2015) 'Data fusion methodologies for food and beverage authentication and quality assessment – A review', *Analytica Chimica Acta*, 891, pp. 1–14.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., Calvo, A. and Busto, O. (2016) 'Prediction of olive oil sensory descriptors using instrumental data fusion and partial least squares (PLS) regression', *Talanta*, 155, pp. 116–123.
- Cawley, G.C. and Talbot, N. (2010) *On over-fitting in model selection and subsequent selection bias in performance evaluation - University of East Anglia*. Available at: <https://research-portal.uea.ac.uk/en/publications/on-over-fitting-in-model-selection-and-subsequent-selection-bias-> (Accessed: 24 January 2025).
- Cheng, J., Sun, J., Yao, K., Xu, M., Tian, Y. and Dai, C. (2022) 'A decision fusion method based on hyperspectral imaging and electronic nose techniques for moisture content prediction in frozen-thawed pork', *LWT*, 165, p. 113778.
- Clegg, I. (2020) 'Chapter 7 - Process analytical technology', in Riley, C. M., Rosanske, T. W. and Reid, G. (eds.) *Specification of Drug Substances and Products (Second Edition)*. Elsevier, pp. 149–173. Available at: 10.1016/B978-0-08-102824-7.00007-5 (Accessed: 3 September 2022).

Cozzolino, D. (2022) 'Sensors for the Food Industry: An Introduction' Available at: 10.1039/9781839167966-00001 (Accessed: 13 March 2024).

Draper, N.R. and Smith, H. (1998) *Applied Regression Analysis*. 3rd edn. Wiley and Sons. Available at: <https://www.wiley.com/en-gb/Applied+Regression+Analysis%2C+3rd+Edition-p-9780471170822> (Accessed: 25 March 2024).

Ellis, D.I., Broadhurst, D., Kell, D.B., Rowland, J.J. and Goodacre, R. (2002) 'Rapid and Quantitative Detection of the Microbial Spoilage of Meat by Fourier Transform Infrared Spectroscopy and Machine Learning', *Applied and Environmental Microbiology*, 68(6) American Society for Microbiology, pp. 2822–2828.

Fengou, L.-C., Lianou, A., Tsakanikas, P., Gkana, E.N., Panagou, E.Z. and Nychas, G.-J.E. (2019a) 'Evaluation of Fourier transform infrared spectroscopy and multispectral imaging as means of estimating the microbiological spoilage of farmed sea bream', *Food Microbiology*, 79, pp. 27–34.

Fengou, L.-C., Lytou, A.E., Tsekos, G., Tsakanikas, P. and Nychas, G.-J.E. (2024) 'Features in visible and Fourier transform infrared spectra confronting aspects of meat quality and fraud', *Food Chemistry*, 440, p. 138184.

Fengou, L.-C., Spyrelli, E., Lianou, A., Tsakanikas, P., Panagou, E.Z. and Nychas, G.-J.E. (2019b) 'Estimation of Minced Pork Microbiological Spoilage through Fourier Transform Infrared and Visible Spectroscopy and Multispectral Vision Technology', *Foods*, 8(7), p. 238.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021) 'Resampling Methods', in James, G., Witten, D., Hastie, T. and Tibshirani, R. (eds.) *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer US, pp. 197–223.

Kandeeppan, G. and Tahseen, A. (2022) 'Modified Atmosphere Packaging (MAP) of Meat and Meat Products: A Review', *Journal of Packaging Technology and Research*, 6(3), pp. 137–148.

Kuzelov, A., Vasilev, K., Naseva, D., Taskov, N. and Dusica, S. (2015) 'Impact of packaging in modified atmosphere of sustainability of minced meat products', *Bulgarian Journal of Agricultural Science*, 21(6), pp. 1172–1176.

Laan, M.J. van der, Polley, E.C. and Hubbard, A.E. (2007) 'Super Learner', *Statistical Applications in Genetics and Molecular Biology*, 6(1) De Gruyter Available at: 10.2202/1544-6115.1309 (Accessed: 7 September 2024).

Leardi, R. and Lupiáñez González, A. (1998) 'Genetic algorithms applied to feature selection in PLS regression: how and when to use them', *Chemometrics and Intelligent Laboratory Systems*, 41(2), pp. 195–207.

- Liu, Y., Lyon, B.G., Windham, W.R., Realini, C.E., Pringle, T.D.D. and Duckett, S. (2003) 'Prediction of color, texture, and sensory characteristics of beef steaks by visible and near infrared reflectance spectroscopy. A feasibility study', *Meat Science*, 65(3), pp. 1107–1115.
- Naimi, A.I. and Balzer, L.B. (2018) 'Stacked Generalization: An Introduction to Super Learning', *European journal of epidemiology*, 33(5), pp. 459–464.
- Nychas, G.-J., Sims, E., Tsakanikas, P. and Mohareb, F. (2021) 'Data Science in the Food Industry', *Annual Review of Biomedical Data Science*, 4(1), pp. 341–367.
- Nychas, G.-J.E., Panagou, E.Z. and Mohareb, F. (2016) 'Novel approaches for food safety management and communication', *Current Opinion in Food Science*, 12, pp. 13–20.
- Pampoukis, G., Lytou, A.E., Argyri, A.A., Panagou, E.Z. and Nychas, G.-J.E. (2022) 'Recent Advances and Applications of Rapid Microbial Assessment from a Food Safety Perspective', *Sensors*, 22(7), p. 2800.
- Panagou, E.Z., Papadopoulou, O., Carstensen, J.M. and Nychas, G.-J.E. (2014) 'Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage', *International Journal of Food Microbiology*, 174, pp. 1–11.
- Pu, H., Kamruzzaman, M. and Sun, D.-W. (2015) 'Selection of feature wavelengths for developing multispectral imaging systems for quality, safety and authenticity of muscle foods-a review', *Trends in Food Science & Technology*, 45(1), pp. 86–104.
- Rahman, U. ur, Sahar, A., Pasha, I., Rahman, S. ur and Ishaq, A. (2018) 'Assessing the capability of Fourier transform infrared spectroscopy in tandem with chemometric analysis for predicting poultry meat spoilage', *PeerJ*, 6, p. e5376.
- Rodrigues, S.S.Q., Dias, L.G. and Teixeira, A. (2024) 'Emerging Methods for the Evaluation of Sensory Quality of Food: Technology at Service', *Current Food Science and Technology Reports*, 2(1), pp. 77–90.
- Ropodi, A.I., Pavlidis, D.E., Mohareb, F., Panagou, E.Z. and Nychas, G.-J.E. (2015) 'Multispectral image analysis approach to detect adulteration of beef and pork in raw meats', *Food Research International*, 67, pp. 12–18.
- Rouger, A., Tresse, O. and Zagorec, M. (2017) 'Bacterial Contaminants of Poultry Meat: Sources, Species, and Dynamics', *Microorganisms*, 5(3) Multidisciplinary Digital Publishing Institute, p. 50.
- Sahar, A. and Dufour, É. (2014) 'Use of Fourier transform-infrared spectroscopy to predict spoilage bacteria on aerobically stored chicken breast fillets', *LWT - Food Science and Technology*, 56(2), pp. 315–320.

Smolinska, A., Engel, J., Szymanska, E., Buydens, L. and Blanchet, L. (2019) 'General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences', in *Data Handling in Science and Technology*. Elsevier, pp. 51–79. Available at: 10.1016/B978-0-444-63984-4.00003-X (Accessed: 26 March 2023).

Spyrelli, E.D., Doulgeraki, A.I., Argyri, A.A., Tassou, C.C., Panagou, E.Z. and Nychas, G.-J.E. (2020) 'Implementation of Multispectral Imaging (MSI) for Microbiological Quality Assessment of Poultry Products', *Microorganisms*, 8(4) Multidisciplinary Digital Publishing Institute, p. 552.

Spyrelli, E.D., Ozcan, O., Mohareb, F., Panagou, E.Z. and Nychas, G.-J.E. (2021) 'Spoilage assessment of chicken breast fillets by means of fourier transform infrared spectroscopy and multispectral image analysis', *Current Research in Food Science*, 4, pp. 121–131.

Stelmach, A. and Cwiek-Ludwicka, K. (2006) '[Organoleptic analysis of food packaging]', *Roczniki Panstwowego Zakladu Higieny*, 57(4), pp. 355–360.

United Nations (2023) *Food Loss and Waste Reduction.*, United Nations Available at: <https://www.un.org/en/observances/end-food-waste-day> (Accessed: 9 April 2024).

US Institute of Medicine (2002) *Escherichia coli O157:H7 in Ground Beef: Review of a Draft Risk Assessment*. Washington (DC): National Academies Press (US), Institute of Medicine (US) Committee on the Review of the USDA E. coli O157:H7 Farm-to-Table Process Risk Assessment. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK221098/> (Accessed: 8 May 2024).

USDA (2024) *What is the most consumed meat in the world? - US Department Of Agriculture*. Available at: <https://ask.usda.gov/s/article/What-is-the-most-consumed-meat-in-the-world> (Accessed: 11 March 2024).

Videometer, Denmark (2021) *VideometerLab - multispectral imaging technology*. Available at: <https://videometer.com/videometerlab/> (Accessed: 18 March 2024).

Wolpert, D.H. (1992) 'Stacked generalization', *Neural Networks*, 5(2), pp. 241–259.

Yang, B., Qi, L., Wang, M., Hussain, S., Wang, H., Wang, B. and Ning, J. (2020) 'Cross-Category Tea Polyphenols Evaluation Model Based on Feature Fusion of Electronic Nose and Hyperspectral Imagery', *Sensors*, 20(1) Multidisciplinary Digital Publishing Institute, p. 50.

Yousefi, H., Su, H.-M., Imani, S.M., Alkhaldi, K., M Filipe, C.D. and Didar, T.F. (2019) 'Intelligent Food Packaging: A Review of Smart Sensing Technologies for Monitoring Food Quality', *ACS sensors*, 4(4), pp. 808–821.

3 SorfML2: A New Web Platform for Analysis of High Throughput Sensor Data in Food Spoilage and Authentication Tasks

3.1 Introduction

Growing populations have made clear the stresses and limitations of existing agri-food infrastructure. By 2050, the world population is set to near 10 billion (UN DESA, 2017) and with that, governmental regulatory bodies are likely to face challenges in the scaling of existing food management practices: conventional food safety and authentication analysis typically involving the utilisation of temporally and monetarily expensive chemical or microbiological destructive batch testing of foodstuff samples.

The utility of high-throughput, minimally-invasive and/or non-destructive sensor technologies is increasingly being explored within the agri-food industry as a consequence. Whilst many of these technologies have shown candidate status as supplements for some of the more well-established laboratory-based or qualitative approaches employed in various food standards practices worldwide, their usage forms part of a bipartite pipeline, relying on a subsequent analytical stage driven by twenty-first century information-technological advancements.

Abreast of these developments within the agri-food sector, emphasis is being placed on the ability of producers and suppliers to maintain a trail of accountability for the safety, quality, and verifiability of produce and its constituent elements, at a biological and chemical level.

Additionally, a desire to improve the dissemination of food nutrition and provenance information to consumers is emerging as a primary issue in the agri-food network as increasingly well-informed and connected consumer bases pursue health and nutrition concerns more pro-actively, partly due to the wide availability of health and nutritional education to populations encouraging the adoption of healthier lifestyles, but also due to the erosion of public trust in the food industry, due to propagated knowledge of stakeholder transgression against food authenticity and quality standards. Corollary aspects to the points made

about food safety, waste and security are economic in nature i.e., the desire to reduce expenditure for both food producers and regulatory bodies through the optimisation of standards compliance (Nychas et al., 2021).

These issues in concert form the landscape of objectives in development of domain-specific analytical platforms and decision-support systems. Effective developments that fit this profile may boost the utility and support the deployment of high-throughput sensor devices in the food production and distribution network, allowing the user base to pool domain expertise, whilst providing the means to analyse data in a close-to-real-time fashion, supporting the ingestion of data emerging from parametrised monitoring and feeding it into predictive modelling/machine learning steps, as the second half of the aforementioned bipartite pipeline. The high dimensionality, collinearity and quantity of data typically expected from such sensors necessitates the usage of statistical and computational processing and modelling to extract any useful inferences for influencing safety and authentication decisions, with specialised models being necessary to account for the high variability in associated experimental metadata.

A challenge is posed by the possibility and specifics of implementation of a means to summate data collected from a wide range of associated disciplines within food hazard detection and safety, such as metagenomics, metabolomic and other microbiological assays pertaining to a particular foodstuff, process or production chain, in order to further validate any rapid predictive models. An accessory issue to these two challenges is the current limitations or lack of standardisation of both ontological frameworks within the food safety sector and standards and protocols that limit the interchangeability of data sources and output formats in field-related pipelines and workflows. Consensus between wet-lab research, data science and development operations may optimise data flow within a large-scale food safety and validation framework, and provide a common interface for collaboration with other bioinformatics communities in the health and lifestyle sectors, as their demarcation continues to blur and their interactivity is further explored.

3.1.1 Existing Platforms

Domain specific projects have been developed to undertake some of the tasks involved in processing the types of data emerging from high throughput sensor devices, however not all were developed for the specific tasks of facilitating foodstuff monitoring.

Combase is a well-established food research platform, intended for the collation, and presentation of microbial growth and survival information and support for subsequent modelling and prediction tasks. At the time of writing, over 60,000 records have been deposited into the repository, reflecting academic research into the effects of environment on the viability of food products. These include temperature, pH, moisture content and other introduced factors, such as preservative agents and deliberate modifications to storage conditions, such as vacuum- or modified-atmosphere packaging. The inbuilt modelling functionalities allow the user to forecast microbial growth under various conditions explored in the corpus of academic research available to the platform (Baranyi and Tamplin, 2004). Users are able to “donate” their data to the project through a data annotation process: experimental procedural information is entered through a series of web forms and is used to index the data and facilitate data recall in subsequent platform usage. As a whole, the Combase platform is targeted at mathematical modelling and prediction tasks in food microbiology, and does not offer facilities for processing real-time food monitoring data.

Other such projects that encompass some of the features expected from a sensor data-driven platform include BreedBase/SolGenomics (Fernandez-Pozo et al., 2015), providing analytical tools for the Solanaceae family and relatives; over time, it has come to support the analysis of data deriving from a growing number of family members including tomato, potato, and aubergine (eggplant). The platform hosts genomic and phenotypic data regarding these species, and provides a web portal for users to upload their own phenotyping and loci information. Over time (and of specific relevance to this project), the platform and its derivatives have been extended with support for data matching the visible and near-infrared spectroscopy (Vis-NIRS) profile, for the purposes of phenotyping of

plants in breeding programs, effected through the implementation of the *Waves* package (Hershberger et al., 2021). The configuration of this package allows the user of the platform to process each spectroscopic upload, using various filtering and smoothing techniques, and input into a variety of machine learning algorithms, to associate the data with reference measurements and assess the performance of the sensors in prediction/classification of phenotypic traits. The authors of the package acknowledge the complexity of existing packages within the R and Python ecosystem and thus the barrier to entry for employing them in the analysis of Vis-NIR spectral data. To further lower the programming background necessary for employment of *Waves* in phenotypic prediction, they integrated the software into the Breedbase component of the Solgenomic Platform. In addition, the authors acknowledged that these developments represented one, if not the only domain specific tool available at the time incorporating these features that was free and open source (FOSS). This determination indicates the universality of the software, including for those that commercial software may not be financially viable.

An interesting feature of the Breedbase data platform is that instances are linked closely to controlled ontologies, based on the Crop Ontology system (Shrestha et al., 2012). This sets the combined platform apart from other, more generic solutions to spectral data processing, as the adherence to a controlled ontology simultaneously increases the flexibility of the platform in analysis of data emerging from various sensor device types and configurations, through annotation of registered data with common terminologies and definitions, whilst also helping to guarantee the consistency of output of the platform, without having to “engineer in” specific pipelines designed for individual sensor device specifications.

The SorfML platform was developed as the constituent machine learning element of the Symbiosis Online Research Framework project (Estelles-Lopez et al., 2017). This platform emerged as a benchmarking tool for ranking machine learning models trained on data emerging from a variety of high-throughput

sensor technologies, originally for data emerging from the monitoring of food sample state.

This platform made use of the MeatReg R library (and later the MeatClass library) to preprocess, train and assess the performance of a variety of machine learning algorithms in regression tasks including microbial growth prediction, comparing performance against conventionally established laboratory metrics such as Total Viable Count (TVC), Lactic Acid Bacteria (LAB) and Plate Count Agar (PCA) methods, whilst also supporting classification tasks such as authenticity analysis and two-class objectives such as the determination of “spoiled” or “non-spoiled” food status.

The platform is built around the concept of a key-value store, implemented in the non-relational MongoDB system. Uploaded datasets are saved as “documents” within the database, the data being retrieved as such when machine learning methods are applied from within the web interface. The results from any applied methods are then formatted and provided to the user as a “report” detailing the performance of each method, in a comparative assessment, detailing the per-fold metrics during each algorithm’s cross-validation and/or grid-search hyperparameter tuning, and the metrics of prediction against the input data’s hold-out test set.

This database schema (or lack thereof) is both a strength and weakness of the platform’s design when compared to schema- or ontology-backed platforms such as the Combase or the Breedbase/Solgenomics platform; the key-value architecture offers the platform far greater flexibility in the ingestion of data emerging from a variety of different sensor modalities than platforms using a more rigid data schema backed by a conventional relational database. Stored documents in a key-value store are essentially “opaque” to the platform with regard to their contents and as such can support a variety of different field names, types and sizes, even within the same collection of documents. In the case of SorfML, this eases the accommodation of data from a variety of sensor types including spectroscopic devices, multispectral imaging (MSI) platforms and chemometric devices such as Electronic Nose/Electronic Tongue

(eNose/eTongue). The only requirements for data uploaded to the platform are that the tabular data associated with an upload have the correct headers, i.e. the “Sample ID” column. Subsequent input (independent) variables in the dataset can be named arbitrarily (as indicated in Table 3-1).

Table 3-1: This table represents an acceptable format for tabular data input into SorfML. The Sample ID column is correctly indicated as the first column. The input (independent) variable columns can be given any name, except those that indicate the Sample ID.

Sample ID	X1	X2	X3	Xn
Sample_1	...	...	...	...

While the platform maintains a semi-structured database schema, where users (authors) are associated to “experiments” and “datasets” in a one-to-many fashion, limited enforcement of content requirements or reinforcement of ontological concepts in these entities complicates any attempt to provide consistency in the output of any analyses run using the platform. There is no inbuilt way to see differences in configurations of sensor data uploaded to the platform, nor any facility to store these configurations on the platform. As a consequence, any analytical pipeline built with SorfML is extremely sensitive to the configuration of data expected as input. Any deviation from the variables seen at the time of model training will likely cause unforeseen errors during subsequent use of the pipeline. The “opaque” nature of the associated database entities also limits the extensibility of the platform beyond its originally intended use, as without a solid schema, any exchange of heterogeneous documents may propagate unforeseen issues caused by misconfigured data.

The lack of annotative functionalities further limits the transferability and interchangeability of any information managed by the platform, as without the ability to link experimental/procedural metadata to platform-generated entities (e.g. predictive models) using a common terminological framework, users cannot reliably store, recall and compare the results of analytical pipelines. While not necessarily an issue for users applying the software in an “in-house” fashion, where experimental procedure has been established and is known by all members of a research team, it may become such when experimental and model configuration procedure has not been established by consensus between all members of a larger research community.

As an example, if a predictive model built with spectroscopic data that has received a certain pretreatment or sensor calibration operation is used for inference with data that has not received the same operations, or indeed is not reflective of the same training data e.g., specific foodstuff, predictions are highly likely to be impacted in some way, either through reduced predictive accuracy, increased prediction errors, or diminished reliability of the outcomes due to the model being forced to extrapolate beyond its original scope of training. While this appears an intuitive conclusion in reasoned isolation and thus avoidable, the potential consequences for “silent” unreliable performance could be highly damaging in the context of a semi-automated, near real-time decision-support system if the inference request data matches the same structural “fingerprint”, (i.e. dimensionality) as the model is expecting, but not the same “meaning”.

At the other end of the range of platforms available for bioinformatic and chemometric analytical work are some that exist in a more generic space, not concerned with a specific decision support role, instead operating as platforms supporting reproducible workflow execution. These systems are modular in nature, allowing the user to configure an analytical pipeline from start to finish, with appropriate intermediate steps, such as data processing and machine learning steps, with some supporting further conditional logic to modulate the flow of a pipeline based on the encountered input data.

A well-known example that fits this profile is the Galaxy Platform (Afgan et al., 2018). Galaxy has emerged over many years as a collaborative effort to build a platform that supports accessible, reproducible and collaborative research, through a web application that supports a variety of tools and processes. Whilst originally developed as a genomics platform, its success has led users to extend the functionality of the project with plugins that support a variety of other domains, including cheminformatics, bioimaging informatics, and machine learning. The system is intended to allow users an accessible avenue into computational biological analyses, by providing wrappers around tools and functionalities typically accessed via command line interfaces (CLIs). Constructed workflows

can then be saved and exchanged with other users of the platform and executed elsewhere.

As a generic data and workflow management system, the platform is not inherently designed to handle specific configurations of data, nor does it provide any controls or checking of inputs outside of checks defined in pipelines. Furthermore, it does not possess any inbuilt support for annotation of semantic/ontological information. As such, the repurposing of the platform for use in food analysis tasks may represent needless over-engineering and while intended to simplify construction of data pipelines for the wider scientific community, it does not lend itself well to being consistently employed as a component in a wider domain-specific framework, and its multifaceted nature may actually add complexity when attempting to configure analytical pipelines with input checking, result storage and a user-friendly, streamlined/expedient interface as one might expect for decision support objectives.

However, an important strength of the Galaxy platform that those previously mentioned lack is its modularity and extensibility. By following instructional material for the project, users can modify example configuration files to build “wrappers” for new tools and functionalities to be used in the web application. These wrappers can be uploaded to the Galaxy Tool Shed (The Galaxy Community, 2024), an online repository of over 10000 different tool configurations that can be employed in instances of Galaxy. This represents a huge extension of functionality outside of reliance on the features built-in to the platform, or those added through distinct developments to the web-app itself i.e. updates to the main project.

Another noteworthy development in the bioinformatics space is the Arvados Platform (Amstutz et al., 2024). Similar to Galaxy in that it offers the user a graphical interface to work with a variety of complex bioinformatics data, it is designed with a strong data management foundation, incorporating a wide array of features such as version control for uploaded data, data provenance features, and a workflow execution system that can employ a variety of different execution backends. However, Arvados’s large emphasis on reproducibility may add a layer

of complexity to the system, and contribute towards a steeper learning curve, due to the large degree of configurability in the development of workflows that heavily integrate across cloud architectures.

The only platforms specifically designed for the handling of high throughput sensor data described here are Breedbase with the appropriate plugins, and SorfML; while Arvados and Galaxy can theoretically handle these sensor devices' output, there are no domain-specific features that make them attractive for work within a food monitoring setting, nor support for inbuilt inference endpoint design, a useful feature in a decision support context.

3.1.2 Determination of Desired Features for a Candidate Food Analysis Platform

A platform that fulfils the type of objectives described in the Introduction of this section needs to reproduce some of the strengths of existing software solutions described in Chapter 3.1.1: Existing Platforms, whilst also adding domain-specific features that none of them currently embody.

The description of a complete integrated Food Safety Management System (iFSMS) is extensive and encompasses many individual constituent elements that reflect the complexity of the interrelationship between food production, distribution and the various control procedures that occur throughout. Nychas *et al.* suggested that stakeholders and authorities in the agri-food industry should integrate their data into secure web/cloud enabled services with an open architecture for reasons of oversight, thus improving trust between producers, suppliers and consumers, whilst also tackling some of the issues of inefficiency and waste, by supporting the recording of sample state data throughout the production and distribution process with a high degree of resolution (Nychas *et al.*, 2021). This would in theory minimise reliance on destructive and slow batch-based testing of food, through the summation of high-throughput, non-destructive sensor device information, coupled with other information technologies such as geospatial information gathered from satellite imaging, media intelligence gathered through outbreak analysis and social media feeds, and Digital Twin (DT) modelling of virtual representations of food samples in various storage conditions. The ultimate goal of such a system would be to provide a surveillance and intelligence system for users of all types, to alert them to issues that directly affect them e.g., notifying purchasers of contaminated meat, or alerting farmers to animal feed batches contaminated with pesticides.

A determining factor in the success of a system as described would be in the simultaneous need for jurisdictional oversight of all food operator's operations in a given region, whilst also respecting their business interests and properties. Additionally, it would be required to address the privacy concerns of individual end-users of the platform i.e., the consumers, allowing them to make decisions

about their food purchasing and consumption, without fear of their personal information being disclosed to third parties, or otherwise being deanonymised.

3.1.2.1 Data Management and Reproducibility

A common stumbling block seen increasingly in modern data-driven endeavours is in the reproducibility of research, an important feature of high-quality science. Over 70 percent of biologists and chemists that responded to a 2016 Nature survey of over 1500 people reported that they had failed to reproduce published science (Baker, 2016). The feedback from the survey suggested that these failures can be attributable to multiple factors, including poor analysis/statistical power and the unavailability of methods, code and raw data.

Comparative meta-assessment of bioinformatics pipelines has shown in multiple use cases that variations in analytical tools used (including the specific release versioning of said tool and its configuration) and data-theoretical processing steps taken can significantly alter the outcomes of the study, due to differences in output from identical inputs influencing the critical reasoning of researchers. One example of this is a recent study of the application of various variant calling algorithms (VCAs) and short-read-alignment algorithm pipelines on human Whole Genome Sequencing (WGS) data (Hwang et al., 2019). The researchers demonstrated that differences in performance and concordance could be achieved through different combinations of tools, particularly achieving statistically discordant performance when calling rare/novel genomic variants. They also acknowledge the complexity of validating bioinformatics pipelines, and stress the importance of benchmarking them as a continuous necessity in the clinical context as genomic technologies evolve (Kong et al., 2018).

Whilst differences in performance are to be expected between pipelines due to differences in modes of operation, specialisation and configuration, the accommodation of packaged procedural/configuration information alongside raw experimental data in an effort to support the reproducibility of computationally-enabled research would be of benefit in a repository for food analytical data, but also in connected interests such as health sciences, as it would allow for a view of generalised performance across experimentation and in minimal-human-

oversight analysis expected from high-throughput decision support pipelines utilising minimally-invasive sensor data.

Any system developed for the purposes of analysis of food data should be built to store experimental data in line with the FAIR principles. FAIR an acronym for Findability, Accessibility, Interoperability and Reproducibility, a common set of data governance guidelines that describe various aspects to data handling at the point of ingestion into a platform and data at rest within a repository (Wilkinson et al., 2016).

These principles are intended to improve the machine-readability of data uploaded to a platform and to reduce the amount of human intervention required to find, access and reuse knowledge stored in such repositories. They were devised in apprehension of the increasing reliance on computational support for the processing of data in analytical fields due to the increase in volume, complexity and rate of generation. Findability emphasises that data and associated metadata must be readily discoverable. This requires assigning a globally unique and persistent identifier, ensuring that data can be distinctly located and referenced. Metadata must be explicitly linked to their corresponding data through this unique identifier, facilitating efficient retrieval. To fully actualise the findable repository, all data should be registered in searchable manner, enabling users to efficiently discover relevant datasets.

Accessibility does not necessarily imply “free access” - instead it concerns the effective retrieval of data once it has been located, using standardised protocols. These protocols must be open, universally implementable, and free to use. There must also be support for clear authentication and authorisation processes to balance data accessibility with appropriate security. Critically, metadata must remain accessible independently of the availability of the actual data, preserving context even when datasets themselves become unavailable.

Interoperability pertains to the ease with which data can be integrated with other datasets or systems. Achieving interoperability necessitates that data and metadata employ a formal, universally understood, and broadly applicable language or representation framework. This representation must rely on

vocabularies that themselves adhere to FAIR principles, ensuring integration across varied utilising systems. Effective interoperability demands that data and metadata fully qualify their relationships with other data or resources, enabling comprehensive contextualization and linkage across multiple resources for ease of navigation.

Reusability focuses on the idea that data should retain their value beyond their initial purpose. To support reusability, data and metadata must be extensively and accurately described using multiple relevant attributes. This detailed characterization includes clear, accessible licensing information governing data usage rights, alongside thorough provenance documentation detailing the data's origin and history. Moreover, datasets must align with established domain-specific community standards, ensuring that they remain useful and trustworthy for future researchers and applications.

Issues related to the poor resilience of platforms such as SordML when encountering new data as described in Chapter 3.1.1: Existing Platforms may be ascribable to the limited adherence to the FAIR principles, in particular the Interoperability and Reusability conditions. While the data is indexed uniquely, appropriate access protocols are used, and authentication and authorisation schemes are employed, the opaque nature of uploaded data functionally eliminates the possibility of any analytical processes being conducted in an automated fashion without the direct intervention of a human operator.

Minimum data standards have emerged as another measure in the biology and chemistry spheres as a result of consensus from interested parties on necessary inclusions in packaged data for unambiguous interpretation and reproduction of experimentation. Some domain-adjacent examples include MIABE (Minimum Information about a Bioactive Entity), a widely scoped standard focusing on bioactive molecular interactions e.g., in the non-context of drugs, herbicides and pesticides (Orchard et al., 2011), STRENDA (Standards for Reporting ENzymology DATA), a set of guidelines and database for upload of manuscripts and standard-compliant supporting data (Malzacher et al., 2024; Swainston et al., 2018).

Subsequent meta-assessment of these standards has resulted in the *fairsharing.org* portal (originally MIBBI: Minimum Information about Biological and Biomedical Investigations), a consortium-led project allowing for a curated reference for domain specific bioinformatics data repositories, and for recording standards updates/versioning, thereby allowing for a degree of visibility and unification of data standards across user and research communities (Emmerich and Harris, 2020). To the author's knowledge, no "standalone" food safety and quality domain-specific standards currently exist for the data expected from the summative output of high-throughput sensors coupled with conventional in-vitro methods; instead, researchers are required to pick and choose applicable standards situationally.

Ontologies, as mentioned, allow for the formalisation of the relationships between concepts found in data, by way of mapping them to concepts found in the ontology's controlled vocabulary. This has the effect of decomposing data into its constituent elements that can be modelled as individual instances of objects, classes or concepts that can be linked using axioms or properties that indicate some type of interrelationship.

OWL (Web Ontology Language) is one of the most prolific languages for ontology construction, and provides a high degree of Web interoperability, due to the World Wide Web Consortium (W3C) granting it "recommended" status for use as a Web standard (World Wide Web Consortium, 2004). There is no apparent overarching ontology that sufficiently covers the representation of sensor-driven food safety and quality assessment analyses, however others exist in more dedicated areas encompassed in this overarching context, including agricultural, food and information science. Examples include AGROVOC (Soergel et al., 2004), covering areas of interest of the Food and Agriculture Organization of the United Nations, FTTO (Food Track&Trace Ontology) (Pizzuti and Mirabelli, 2013), an ontology for modelling food supply scenarios across geographical and jurisdictional boundaries, and a variety of sensor-descriptive ontologies, each corresponding to a specific subset of sensors or use cases; a noteworthy example of an ontology that covers the application of sensors and their

observations is SSN (Semantic Sensor Network), having seen case study exploration in agricultural scenarios (Taylor et al., 2013).

3.1.2.2 Reproducibility of Scientific Pipelines

Whilst employment of the aforementioned data governance methodologies would go a long way toward ensuring adherence to the FAIR principles for data during the upload and static storage phases of data handling in a candidate sensor data analysis platform, further steps could be taken to maintain the integrity of any information output in the form of predictions or trained machine learning models.

As mentioned, an important consideration in the construction of any bioinformatics pipeline is its reproducibility, and by extension, its portability. In the wider IT community, significant progress has been made in solving similar issues in problem spaces outside biology. Many of these issues can be distilled to the scenario of “this application works on your computer, but not on mine” – i.e. problems arising from software failing to run in heterogeneous environments.

Whilst employment of virtual machine technologies (the virtualisation of guest environments functioning on a host machine, in order to run specific software) solves issues with functional reproducibility of scientific pipeline execution, there are associated drawbacks with regard to portability, as virtual machines used in scientific research (VMs) often have large computational and storage footprints due to the high-performance nature of many domain-specific tools and libraries.

The logical progression from the use of VMs in science is the use of containerisation technology, which allows operating system (OS)-level virtualisation, as opposed to typically hardware-level VM technologies. This means that individual containers can make use of many common OS-level essential libraries and functionalities, thus minimising the footprint of such packaged tools, and affording bioinformatics solutions a greater degree of portability. A prominent application of this technology within the Bioinformatics space can be found with the BioContainers Project (da Veiga Leprevost et al., 2017), wherein a community standard has emerged for creation of containerised bioinformatics tools for use within multiple “-omics” disciplines, to enable

deployment of said tools into pipelines of varying complexity, and across a variety of architectures for local, cloud, and high-performance computing (HPC) environments.

Containerisation of workflows in bioinformatics applications and a wider data science context enables increased complexity of construction, due to compartmentalisation allowing modularity - the ability to “swap out” components to suit specific tasks, whilst not affecting preceding or subsequent operations. However, this increased complexity can complicate the running and monitoring of individual operations running in this manner, and when considered alongside the high degree of variability in the running times of bioinformatics workflows, automation of these processes may be desirable to allow them to run in an unattended fashion, potentially incorporating elements of conditional programming allowing for on-the-fly modification of workflows based on intermediate outputs, in a similar manner to the pipelining functionalities of the Galaxy platform described earlier. So-called “off-the-shelf” solutions in the data science/management industries exist for generic scenarios, including open-source, self-hosted solutions like Apache Airflow, whilst other, more bioinformatics-oriented solutions have emerged, such as Toil, a pipeline/workflow engine written in Python (Vivian et al., 2017), and Nextflow (Di Tommaso et al., 2017).

Additional efforts have been made in a parallel fashion to standardise the “workflow languages” employed – domain-specific languages that are parsed within each project’s implementation to provide their respective orchestration and conditional functionalities. Two examples of these efforts are Common Workflow Language (CWL) (Crusoe et al., 2022) and Workflow Description Language (WDL) (Voss, Auwera and Gentry, 2017). However, the scope of employment of these solutions is difficult to gauge within the Bioinformatics/Agri-food problem space, and ensuring compliance with these standards may represent additional complexity facing bioinformatics practitioners when developing new pipelines, versus adherence to a single enabling technology. Existing software, such as the previously noted Nextflow, make use of their own scripting languages, which suit

their own operating characteristics. Nextflow's idiosyncratic language has emerged as a result of heavy extension to the Groovy programming language, which is itself based on the Java programming language and runs on its platform, the Java Virtual Machine (JVM). This affords it a great deal of familiarity within the bioinformatics and data science space, as many data science practitioners have had exposure to Java programming, representing less of a learning curve than a completely new pipelining paradigm.

3.1.3 Platform Architecture Design Considerations

3.1.3.1 Support for FAIR data management

The most commonly considered and oftentimes the simplest architectural choice for many web applications is that of the “monolith”: in this design, all elements that serve the various functionalities of the platform are contained in one programming “block”. This includes the end-user interfaces and the parts of the application that underlie and support this functionality, such as the unique “business logic”, (the parts providing the platform’s utility) and the non-unique necessities, such as maintenance of database connectivity.

This leads to a tight integration of all functionalities in a single cohesive unit. This design choice has distinct positives and negatives in the ease of web service and future development directions of the projects. Some of the advantages of pursuing a monolithic architecture include the comparative simplicity in maintenance of a single codebase, a unified development and deployment process in a single programming language, accelerating development. The performance of such a system is likely less impacted by communication between individual aspects of the platform, as this can be achieved in a manner overseen by the web framework or runtime environment, thus not relying on inter-service communication through a messaging service. Overall, the monolithic archetype is likely to be easier to manage in terms of testing, debugging, and version control, due to the bundling of all functionality into the same package.

A commonly stated disadvantage of a monolithic platform is related to scalability; as the system grows in scope or in utilisation, the architecture can become harder to scale, as adding new features may necessitate scaling the entire application, even if only one part of it needs more resources.

This disadvantage in the context of a domain-specific web application such as a food decision support and analytical platform may appear less problematic, simply because issues related to scalability are unlikely to be a noticeable issue when measured purely by the size of the user base.

This may not be the case when the intention to facilitate data processing and machine learning objectives within the platform is factored in, however. This is due to the potentially high computational power required on-demand by such a system to process large quantities of data in such tasks as ML/Deep Learning model training.

The aforementioned tight coupling of components within the application is likely another issue, as isolation and updating of a particular functionality is not necessarily possible without introducing breaking changes to the remaining components of the platform. Such a design also requires careful decisions on the web framework and languages used throughout the project, as some may lend themselves better towards certain tasks, such as high-performance web service, but may fall short when considering other requirements such as processing of tabular data or initialisation of machine learning tasks.

This was a noteworthy issue of the SorfML platform, as it was designed with NodeJS server architecture, which had limited inbuilt support for execution of the types of long-running tasks expected of a data analysis platform.

An alternative architectural design finding increasing utilisation in the high availability, high throughput web service domain is the concept of “Microservices”. In high level terms, this design decision breaks down the overall system into smaller, independent services that communicate with each other over a network, either locally or mediated via a virtual private network (VPN). Each service is responsible for a specific functionality within the overall functioning of the system, meaning that they can be developed, deployed and scaled independently.

This in theory gives a representative system a greater degree of scalability and flexibility, as each “microservice” can be independently scaled based on demand for the functionality. For example, if more availability is required in the data extract, transform and load (ETL) parts of a project, they can be replicated, either on the same machine or on separate machines in an orchestrated architecture, in a concept known as “horizontal scaling”.

A further benefit to such a system is the additional isolation and resilience brought about by splitting the functionalities into distinct components. Even if the development process results in breaking changes to a specific functionality, or an unforeseen data processing bug causes errors in a service, it wouldn't necessarily affect the functionality of the overall system. This could be considered a major benefit for a critical platform like a food safety decision support system.

A pertinent benefit of such an architecture is in the comparative ease of implementation of new features in the project, as whereas in a monolithic design, features are heavily limited by the initial design decisions of the project, a (micro)service-oriented architecture is less limited, on the condition that new services are designed to communicate correctly with the other components. This means that new frameworks or tools can be hosted in environments that natively support them.

The main disadvantages to such a design choice relate to the additional complexity of development, coordination, and operation. A challenging aspect is the management of inter-service communication, ensuring resiliency in the eventuality that issues occur with a specific component, and data consistency in the event of a system failure.

In a domain-specific application of the service-based architecture, the Berlin Institute of Health (BIH) developed a framework for the creation and deployment of cloud-based platforms for support of biomedical research projects (Jentsch et al., 2024). The developers state that the development of such a framework represents a shift in thinking from Software-as-a-Service (SaaS) offerings such as Galaxy or Anvil, which provision the entire application stack, towards Platform-as-a-Service (PaaS) or Infrastructure-as-a-Service (IaaS). This shift allows for the tailoring of a platform solution for a specific research objective using a selection of pre-engineered service “building blocks” implementing different aspects of data storage or platform logic, minimising development time and optimising deployment size and complexity. The developers also suggest that such a platform design may be in keeping with the FAIR philosophy, as through

commonality of data storage components and API design, interoperability of similarly-built platforms may be more easily achievable.

3.1.3.2 API Design Strategies

Failure to achieve data interoperability, in spite of claims of support to the contrary, has led to claims of user “lock-in” in the Bioinformatics cloud platform space. This phenomenon has been attributed to the desire for platform providers to prioritise feature development, with limited focus on support for established data management community standards. The “lock-in” effect potentially results in challenges for data exchange between researchers, or migration to new platform designs. However, the move towards service-oriented architectures has been suggested to naturally lead to greater interoperability through the necessity to clearly define the interface with which the individual services are accessible (Sheffield et al., 2022).

A well-designed application programming interface (API) is a crucial part of any software development project: without such, communication between components of the platform or the end user becomes challenging to coordinate. An important aspect of this design process is clear forethought into the responsibilities of the represented platform and how they are represented logically in the interface.

A common methodology to describe web service APIs is through the use of OpenAPI Descriptions (OADs), declarative descriptions of methods, parameters, and data models that can be used at various stages in design and development. These can enable “contract first” development, where the functionality of the API is built to an initial specification, rather than relying on documentation of the behaviour of an API in retrospect. These specifications can be shared ahead of time in a collaborative project, granting every member a common blueprint with which to coordinate efforts.

The specifics of a web API are context dependent, however an oft-desired implementation standard is that of REpresentational State Transfer (REST). REST endpoints are stateless by design, instead relying on “Hypertext as the

engine of application state” or HATEOAS, as described by Roy Fielding in his originating work (Fielding, 2000). This stateless design allows the user to navigate the API with no foreknowledge of its structure, instead relying on context-dependent hypermedia information relayed in the responses from the API itself to guide subsequent requests.

3.1.3.3 Data Processing

Beyond the comparatively simple management of data via an API, data analysis platforms are often equipped with the ability to perform Extract-Transform-Load (ETL) and “FAIRification” tasks automatically, in a similar fashion to the SolGenomics/Breedbase platform, wherein data outputs are annotated with terms from the associated ontologies.

In doing so, annotated data is granted a greater degree of machine readability, and allows the user to more effectively query the associated repository for data matching certain criteria, even across previous limitations such as differing sample naming schemes, or incomparable experimental protocol descriptions. Such annotated data may be stored for use by the end user or directly by the platform, through semantic data stores (Knowledge Graphs). These typically store data in the form of semantic triples, consisting of a subject, a predicate, and an object, an example of such being (Sensor_A isDeployedAt Institute_B). This method of representation facilitates the structuring of data relationships as nodes and edges in the graph.

Knowledge graphs are typically queryable through an endpoint supporting the SPARQL language. A recursive acronym for “SPARQL Protocol and RDF Query Language”, it allows for complex queries to be constructed of concepts that are congruous with the semantic triple nature of the graph.

With the combination of a knowledge graph and an appropriate query language such as SPARQL, inferential reasoning becomes a possibility, and queries that would otherwise be complex to construct with conventional databases become comparatively trivial and can be conducted *ad hoc*: a domain-specific example “Find all machine learning models trained with data collected using Sensor_A at temperatures between 0°C and 25°C” might take the equivalent SPARQL form shown in Figure 3-1.

```
SELECT DISTINCT ?mlModel WHERE {  
    ?mlModel :trainedWith ?dataset .  
    ?dataset :collectedBy :Sensor_A .  
    ?dataset :hasTemperature ?temp .  
    FILTER(?temp >= 0 && ?temp <= 25) .  
}
```

Figure 3-1: An example of a SPARQL query representing the request: “Find all machine learning models trained with data collected using Sensor_A at temperatures between 0°C and 25°C”

3.2 Specific Aims and Objectives of this Development Project

The specific objectives for this development work, influenced by the original PhD objectives and informed by the previous developments highlighted in Section 3.1.1: Existing Platforms were to:

- Develop a robust repository database schema, capable of handling the data types handled in the DiTECT project pilot studies, including Fourier-Transform Infrared (FTIR) spectroscopy, Multispectral Imaging, FreshDetect (Dietrich, 2018) and Videometer (Videometer, Denmark, 2021) spectra, in addition to electronic-nose output
- Implement a “FAIRification” system to automatically process data uploaded to the platform, annotate it correctly with ontological concepts, and store it in the aforementioned repository, in keeping with the FAIR data principles
- Develop a server architecture capable of handling the type of basic CRUD (Create, Read, Update and Destroy) requests expected of a data-handling network application, in addition to advanced functionality related to the particulars of high-throughput sensor data storage and analysis
- Devise a platform to handle the construction, execution and storage of data-processing and machine learning workflows/pipelines

Successful implementation of the standalone platform was to be demonstrated by its capability to reproduce the type of data-ingestion, preprocessing, data fusion and machine learning pipelines implemented as an outcome of the machine learning objectives in an interactive, high-level fashion supported by the graphical user interface, without requiring access to low level programming environments.

3.3 Technical Specifications

The necessity for the developed platform to be modular, as stated in the initial aims of the PhD project, influenced the final architectural design decision to pursue a limited microservice pattern. As stated earlier, issues regarding web traffic were determined to be unlikely to affect developments, but the flexibility to scale the more computationally intensive parts of the platform, such as those serving data processing pipelines, even on systems outside of the original machine serving the application, was attractive for development objectives.

Whilst a monolithic design would have likely represented a reduced workload as a single-developer project, it was unclear if such a design would support future extension with anticipated iFSMS platform features, such as media intelligence, food consignment tracking, blockchain functionalities etc.

Components determined to be necessary at the outset of the project included those related to the data processing API and the user experience (UX) as indicated in the initial objectives, the FAIRification functionalities of the project, and the data processing handler services of the platform, the latter two running asynchronously from the former, allowing for long-running tasks to be executed without impacting the overall functioning of the user facing services, something not possible in SorfML, where data-intensive processes left running on the backend impacted the responsiveness of the web application.

The decision to follow this architectural design influenced the choice of programming language and web server framework to employ in the development tasks. Java was chosen for its strengths in the field, namely its long-standing popularity in web development tasks, various large and well-managed repositories of libraries and frameworks, and the consequent perceived stability of the Java platform.

The Java Spring Framework (spring.io) was chosen for its eminence in the professional web development domain, in addition to its perceived stability.

MongoDB was chosen as the persistence solution for the project. Its non-relational design was determined to be simpler to work with when designing

database schemas to serve the project, rather than the tabular structure of SQL-type database designs.

MongoDB stores content in collections of “documents” in the BSON format, a type of “binarised JSON” object. Such documents can represent complex structures within, in a variety of different data types, with no enforcement of types even between documents in the same collection. As previously stated, this was both a positive and a negative point with regard to ensuring data consistency and reliable behaviour in a networked application. Special care was to be paid in enforcing data types at the application level, as Java had strong typing support, and a well-developed interface with MongoDB, deriving functionality from the Spring Data Commons project to enable complex querying of stored objects.

3.3.1.1 Semantic Functionalities

Semantic information was to be applied to uploaded data from ontologies developed as part of the DiTECT project, consisting of mappings of existing sensor and scientific ontologies, contextualised in the agri-food domain with the addition of previously unrepresented concepts (Yilmaz, 2023).

This feature was determined to require additional configuration from the user upon initial upload of data matching a specific profile; subsequent uploads of similar data would be handled automatically in a specific upload pipeline. This functionality was to be implemented in a standalone extract, transform and load (“ETL”) service, built to run asynchronously, receiving instructions on tasks to perform via a message queue structure. This frees up the main data API application to continue servicing other requests, without waiting for the resolution of FAIRification operations.

3.3.1.2 Machine Learning Functionalities

Machine learning support, alongside the “FAIRification” functionality built into the platform was considered to be the most novel aspect of the development. Having the ability to train, validate and ultimately employ machine learning models from within the web environment was determined to be a highly pertinent feature to include.

This was to work in synergy with the semantic functionalities built into the platform. This was to be achieved through development of a means to assign ontological concepts to the features within the datasets uploaded to the platform by associating them with “Sensor” entities that represented the specific configurations of sensor devices found within an experiment. During the model training process, this information would be persisted in the database, for recall during a time at which the model was to be used for inference purposes. The intention behind this design decision was to allow users of the analytical platform to rapidly identify stored models that fit their querying criteria, for example, when making inference requests with data collected using similar configurations of sensors as in the training process, thus ensuring that any predictions made would be statistically valid.

The implementation of such a system was anticipated to require some sort of orchestration, due to the potential complexity of conducting a machine learning model training procedure. A variety of options for such a system were explored, but the previously described “Nextflow” workflow execution system was decided upon, due to its large degree of popularity in the bioinformatics sphere, and its large body of documentation allowing for easier understanding of the underlying nature of implementation, in addition to its open-source licence permitting its use in other open-source projects.

In order to provide an environment for the Nextflow element to run, a second microservice structure was to be devised, in a similar nature to the Semantic ETL service, to allow for asynchronous processing of long-running requests, and to allow scalability of this element if more “workers” were to be required.

3.3.2 System Requirements

A pertinent question asked throughout the development process was related to the expected system requirements of the platform when running complex tasks, such as the amount of working memory expected to be necessary when serving a fully populated knowledge graph (fulfilling the semantic data store requirement of the project).

Over time, it became clear that such concerns were beyond the scope of the project being developed, as a consequence, development continued using a local hosting infrastructure.

A benefit of using the Nextflow system for the execution of data processing was the flexibility of environment upon which the execution of workflows could take place. Nextflow has built-in support for execution locally, on on-demand system-as-a-service architectures, such as AWS's EC2 and Elastic, in addition to high performance computational architectures via grid scheduling; thus, the decision to continue development locally did not affect the overall possibility of extending the platform to more capable hardware, as the support was available if necessary.

The individual components of the platform architecture were intended to be served within the Docker ecosystem, in containers. This containerisation would theoretically allow for flexibility in the choice of the hosting system, future integration with the components of the externally developed DiTECT platform, in addition to potential migration to more mature web service orchestration platforms, such as Kubernetes. Table 3-2 provides a general summary of the major frameworks and libraries used in the project.

Table 3-2: Tabular summary of the major technological frameworks and libraries that enable the functionalities of the SorfML2 platform.

Technology	Category	Role	Description
Spring (Java)	Backend	Application Framework	Provides File API logic, (handling REST APIs, security, and database interactions).
Angular (Typescript)	Frontend	Web Framework	Delivers the user interface/single-page application (SPA) by managing client-side components and routing
MongoDB	Database	Document Database	Non-relational database; uses JSON-like documents for storage of related data together in a denormalised fashion.
Virtuoso	Database/ Semantic Store	RDF Triplestore	Stores, manages, and queries semantic web data (RDF/semantic triples), supporting linked data operations and querying with SPARQL.
MinIO	Storage	Object Storage	Handles S3 storage for large (raw) data for the platform. Flexible, due to S3 compliance, so hosted/cloud S3 compliant alternatives could be employed in the future instead.
RabbitMQ	Messaging	Message Broker	Enables asynchronous communication and message delivery between components through queues.
ETL/Workflow Services (Go)	Backend/Data Processing	ETL Services	Performs Extract, Transform, Load (ETL) tasks and “FAIRification” of data before ingestion into Semantic Store, in addition to executing and handling workflows in conjunction with Nextflow.
Nextflow	Workflow Management	Pipeline Execution	Orchestrates reproducible, scalable scientific workflows and computational pipelines, managing task dependencies and execution.
FastAPI (Python)	Backend	Inference Server	Context-dependent server that conducts inference tasks with serialised Python machine learning models.

3.3.3 Data Handling and Storage

The platform was to make use of two main data repositories, the first provided by MongoDB (primary storage of dataset, sensor and experiment entities), and the second provided by the Knowledge Graph/Triple Store. Although multiple well-known triple store services exist, such as Ontotext's GraphDB, few exist in the free and open-source software (FOSS) community. Consequently, the open-source version of Openlink Virtuoso was selected as the knowledge graph server to use for the project.

In order to facilitate asynchronous communication across the various components of the project, RabbitMQ was chosen as a message queue. This would allow the platform's main data API, data processing workers, and the ETL service to communicate in a non-blocking fashion, as services would not be obliged to wait for responses to requests sent as described earlier.

ETL requests would be sent by the Data API to the ETL service, to initiate FAIRification of the concepts found within uploaded data files, allowing the API to continue normal operation. Similarly, requests sent by the user to initiate data processing or machine learning tasks would be relayed by the API to the appropriate worker via the message queue.

RabbitMQ possesses a certain degree of resiliency during periods of service disruption, and logs can be persisted for later retrieval if the platform is shut down, allowing services to continue processing data at a later time. A generalised view of this messaging arrangement is shown in Figure 3-2:

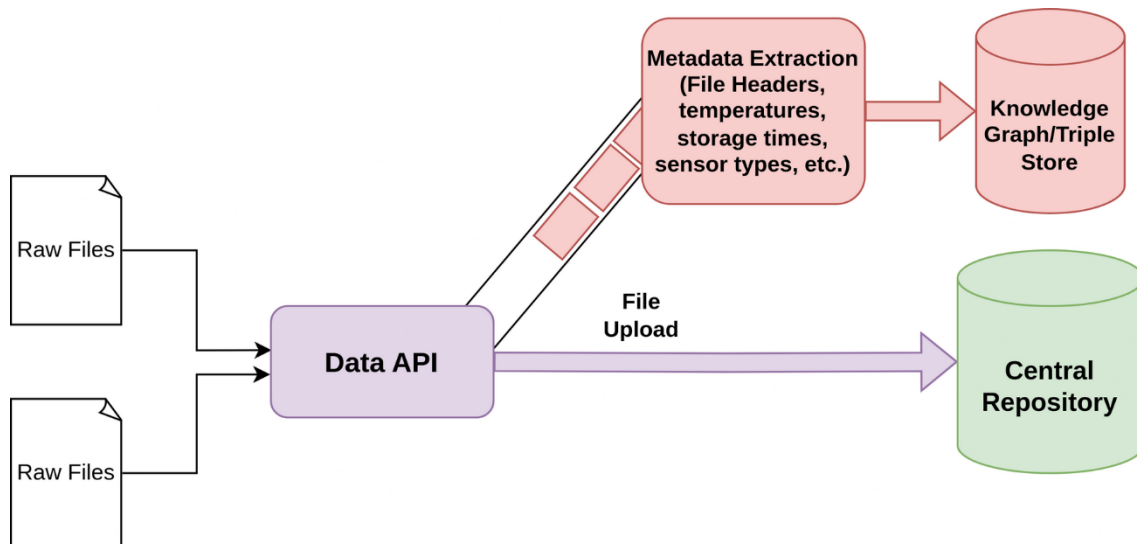


Figure 3-2: A generalised view of sensor data inputs into the new platform. The primary upload process happens in a similar fashion to SorfML, however a secondary FAIRification step is initiated by the API service asynchronously.

Figure 3-3 describes the abstracted topology of the SorfML successor platform, highlighting the relationship between each of the major functionalities of the platform. The ETL services were to receive information about the various entities deposited on the primary MongoDB server, including sensor, experiment and dataset information.

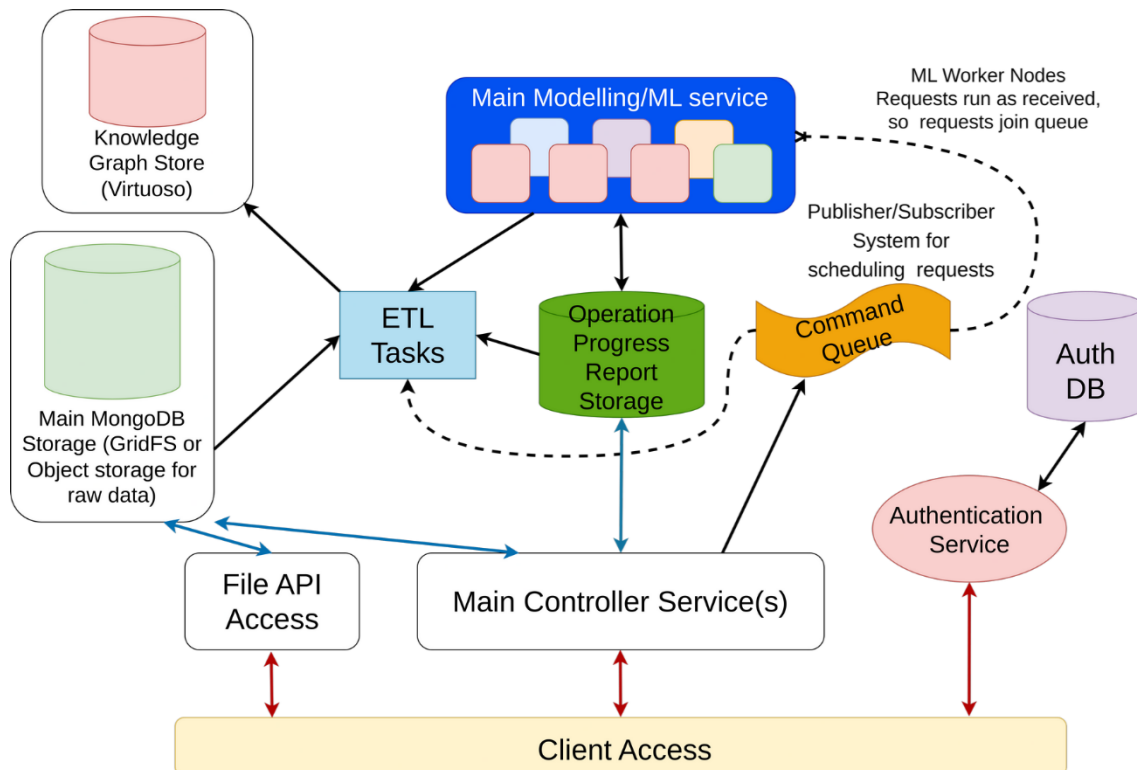


Figure 3-3: The initial intended topology of the “SorfML2 Platform” application. This highlights the interrelationship of the services expected to be necessary for the complete functionality of the platform. The dashed lines output from the “Command Queue” service are to indicate the asynchronous nature of their communication.

3.3.4 User Interfaces

Another important feature of an analytical platform/decision support system is an effective, informative user interface (UI) experience. SorfML made use of Angular, a single-page-application (SPA) framework developed by Google. This gave the platform an overall consistent and intuitive design. The decision was made to also develop the replacement platform’s user interface in Angular, due to the perceived maturity of the Angular project.

A key requirement of the new system was to be the ability to configure data processing and machine learning workflows using this interface in a graphical, intuitive manner, whilst maintaining an appropriate degree of reconfigurability of the represented Nextflow-backed processes that formed the overall workflow.

This was to be achieved with a system to render configurable processes on the platform dynamically. In other words, processes, either representing machine learning algorithms, data processing, or other steps would be defined using configuration files stored on the platform's database. These would be used to render appropriate entities on the web UI. Further process types could then be added to the platform at a later date, without having to update the platform itself, granting a great degree of extensibility not seen in the SorfML platform, instead seen, albeit in a complex form, on the Galaxy Platform with the “tool shed” concept (The Galaxy Community, 2024).

3.4 Implementation Details

This section of the manuscript will detail the characterisation and development of the components of the platform beyond the initial planning stages from the preceding sections, and provide a detailed description of the major modules and components within the system, their functions and how they contribute to the overall functionality of the platform.

3.4.1 Platform-wide Authentication and Authorisation

The new platform makes use of a JWT (JSON web token) – backed approach to authentication and authorisation. In an abstract sense, authentication refers to the condition that the user is who they claim to be, and authorisation refers to the condition that the user has the correct credentials to access the requested resource.

Having such a scheme in place is one of the stipulations of the FAIR Accessibility data principles, i.e. ensuring that a data repository has a clearly defined procedure to access data.

The functionality is provided by the Spring Security Module of the Spring Platform, offering an extremely mature and well documented manner with which to enforce such access control.

JWT are a compact means of representing claims to be transferred between the user and the platform. The token is issued after a successful login and includes encoded information (the user's ID and the email address and role information). The token is then sent by the user with each subsequent request, allowing for stateless authentication.

Authorisation is implemented in both a role-based manner, using the verified role accompanying a user's request to make available certain administration functionalities, and in an attribute-based manner, using matched fields stored in database artefacts to allow access, such as "authorId", "userPermission" (arrays of user IDs) and the "isPublic" flag.

3.4.2 Modules and Components

3.4.2.1 Main Data API

Design of the API progressed as expected with the use of Spring to construct the application, with the SorfML API used as a comparison tool, with the intention of replicating and then superseding the functionality offered by the previous platform. Much of this task rested on the update of the schema supporting the data concepts to be handled by the project. The SorfML platform made use of concepts, representing "Experiments", "Models", "Datasets" and "Users", modelled as documents in their associated MongoDB collections.

In this old schema, "Experiments" represent a discrete body of research, i.e. a collection of uploaded "Datasets" and associated findings. "Models" were used to represent a discretely configured machine learning model object, saved on the platform, with the intention of being available for inference tasks at a later time. The old entity relationships can be viewed in Figure 3-4.

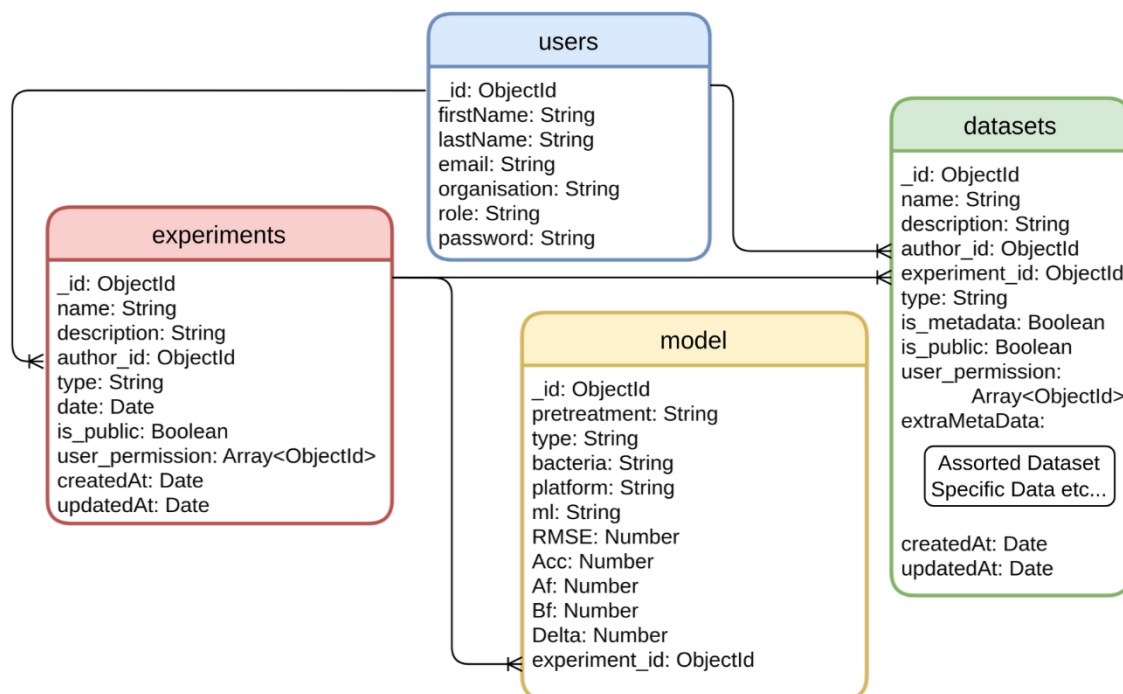


Figure 3-4: Entity Relationships in the SorfML platform. Simplistic data concepts limited the extensibility of the platform beyond the representation of singular models.

The new system utilises the same basic concept of “Experiments” and “Datasets”, but the advanced functionalities provided by the new platform needed several additional data concepts to function. The entity relationships of the new platform can be seen in Figure 3-5.

Firstly, instances of sensor devices were represented as “Sensor” documents in the primary data repository. This was to allow users to store specific instances of sensor devices of a known configuration, and metadata about that configuration (to be used in the FAIRification process), at a known point in time.

The web application includes a GUI to allow users to manually annotate “header concepts” within a dataset that is representative output of that sensor configuration, thereby concreting knowledge about the types of data expected in specific columns (or ranges of columns) in the data matrix for subsequent dataset uploads of that sensor type, such as Sample IDs, features (x) and response (y) variables. This dramatically simplifies referrals to these types of concepts within any pipelines built within the platform, and greatly reduces the chance of edge

case situations in dynamically generated Nextflow scripts where specific variable names differ.

By doing so, users are then able to upload datasets as child instances of that sensor configuration, adding a level of safeguarding against data erroneously being available in the wrong context, e.g. used as inputs in incompatible data processing pipelines stored on the system. Also, if the functionality is opted into upon upload, datasets can be automatically annotated with the sensor inheritance patterns in the knowledge graph system. This allows for subsequent querying of the knowledge graph with known parameters of interest across multiple datasets, for example, “request all observations made at a certain sample temperature, using data from a specific packaging condition, from a specific region”. The universally unique identifier (UUID) given to new “individuals”, representing sensors and datasets in the knowledge graph, are stored as keys in the MongoDB Dataset and Sensor representations’ “resourceId” field, to act as a “pointer” to the information in the knowledge graph, and also to act as an anonymisation feature, adding a degree of separation between the metadata modelled in the knowledge graph and the experimental data, referenced in MongoDB.

Dataset metadata are represented within the knowledge graph down to the sample (row) level of tabular data uploaded to the platform, however the corresponding measurement is not stored.

The raw data files are stored upon upload and inserted into the storage platform of choice when running the platform. Administrators can configure either local storage of dataset files, where data is stored in a directory structure on the host platform, or in a configured object storage (S3) compatible store. For the purposes of development, object storage was found to be the most suitable, due to the simplicity of managing file upload references: files are stored in a flat structure with no need to manage complex file hierarchies. MinioS3 was chosen as the storage platform, due to its permissive open-source licence and its API compatibility with other S3 providers, such as Amazon Web Services, the originators of the S3 protocol. References to uploaded files are stored in the

“rowFileLink” field of the dataset entity. These uploads are dynamically requested by the various service components of the platform when required to complete a processing task.

The “headersHash” field of the Sensor entity allows for a rapid sanity check when uploading datasets via the graphical UI. The header of a representative dataset file is hashed with SHA256 when configuring a sensor entity. When subsequent data is uploaded, the headers from the dataset are similarly hashed, and the platform is able to automatically query the system database to find exactly matching sensor configurations. The header string is normalised by removing parts that do not affect its logical meaning, such as quotation marks, capital letters and whitespace, in addition to the various invisible line terminator characters used by different operating systems (combinations of carriage return and line feed). This makes the comparison process less stringent, avoiding multiple avenues for comparison errors to occur.

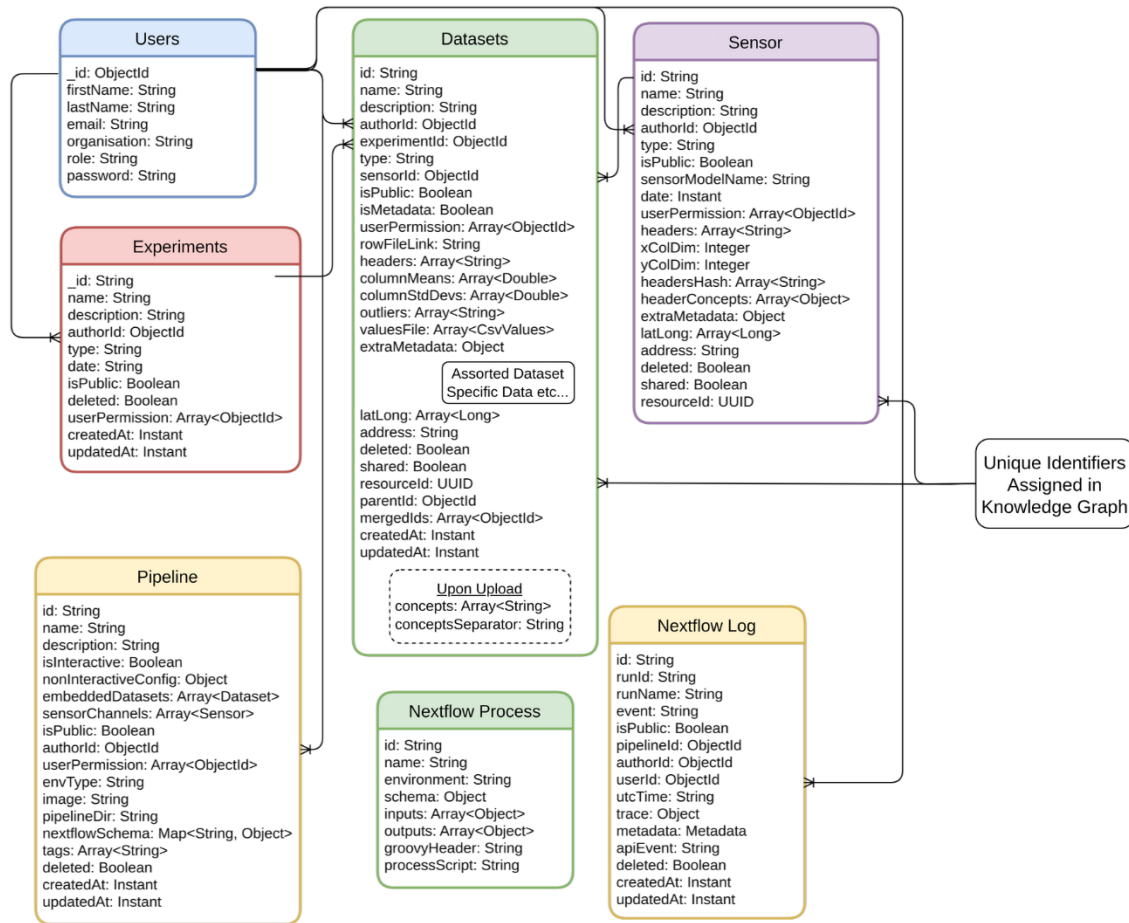


Figure 3-5: Entity relationship diagram of the new platform. This design makes use of new concepts, particularly “Sensor”, “Pipeline”, “Nextflow Process” and “Nextflow Log” to model many more data processing scenarios than were available in SorfML.

In addition to the modelling of sensor devices on the platform, the database concepts of “Pipeline”, “Nextflow Process” and “Nextflow Log” are used to enable the far greater degree of configuration available on the new platform. Rather than relying on hardcoded data processing pipelines to output “Model” entities in the database as was the case in the SorfML system, the Sensor, Dataset and Nextflow Process entities are used in concert by the new platform to dynamically generate Pipelines, which are then saved, to be run by the Nextflow worker services.

The platform’s API is fully capable of parsing these entities, and generating a Nextflow pipeline package that is fully executable outside of the web platform, i.e.

in a standalone fashion, on architectures that support Nextflow. This adds a great deal of convenience to the platform, as users have the flexibility to configure an analysis pipeline using the graphical interface, and then either run it in-place utilising the hardware serving the analysis platform instance, or download it to run elsewhere, e.g. non-interactively, using High Performance Computing (HPC) schedulers.

The specifics of this functionality, including graphical representations of these features will be discussed in greater detail in the following sections.

3.4.2.2 FAIRification/ETL Service

The FAIRification/Extract-Transform-Load service was developed in the Go programming language, due to its lightweight nature and high performance lending itself to microservice-oriented design.

The service itself was designed to listen to a RabbitMQ channel, and then upon receiving a request for annotation of an upload of tabular data, cache the file locally, and then proceed to iterate through the dataset row-by-row, “destructuring” the naming scheme found in the data, using the “concepts” and “conceptsSeparator” properties specified during the upload process. This functionality was enabled by the *rdf2go* library (github.com/deiu/rdf2go), and allowed for the construction of the additions to the knowledge graph hosted in the Virtuoso server.

The web UI can intelligently determine upon upload when a naming scheme appears in tabular data, and request for the user to intervene to assign concepts. One can view various datatype and object properties from the ontologies available to the system, and assign them to concepts found in the naming scheme of the data being uploaded, i.e. the coding given to observations in the typically named “Sample ID” column of a dataset. An example of a configured concepts string is shown in Table 3-3. These “concepts” strings can then be used to de-construct the sample IDs as stated above.

Table 3-3: Example of a coded Sample ID, and the concepts string describing the ontological concepts represented within it.

Sample ID	0C_120h_map_b_b3
Concepts	hasTemperature_time_hasPackageType_hasAdditionalID_hasBatchNumber

3.4.2.3 Pipeline Execution Environment

The decision to use Nextflow as the pipeline orchestrator for the project necessitated the development of a new service structure to support its execution. Using a similar design philosophy as the FAIRification platform, Go was selected as the programming language to build the Nextflow worker service. The service was designed in such a way to allow execution of pipelines in a variety of environments, either natively in the worker environment, or in container (Docker) environments. This facility was useful, as it allowed workflows/pipelines to be constructed in any appropriate language including Python and R, and to use specific versions of libraries and frameworks in the construction of models. As described in Chapter 3.1.2.2: Reproducibility of Scientific Pipelines, the ability to containerise scientific workflows greatly reinforces their portability and scientific utility, when implementations can be disseminated amongst peers and used to reproduce results (Hwang et al., 2019; Kong et al., 2018).

Pipelines scheduled to run in specific containers will as a consequence use only the library versions available in the specific image used to construct the pipeline, thereby aiding with the consistent processing of outputs, provided that the inputs remain the same. This resolves a sizeable issue found with SorfML, which had no support for containerisation – when configuring a machine learning benchmarking run, any change to the server environment could drastically affect the outcome, due to unintended modifications of libraries and configurations.

Data processing pipelines are executed with Nextflow, and the outputs are stored in a directory hierarchy determined by the type of analysis, detailed in Figure 3-6; pipeline runs that are related to specific experiment instances are separated from runs that refer to specific instances of datasets, including those for exploratory

data analysis tasks. The Data API is able to navigate this directory structure and present most types of common text and tabular files to the user (.CSV, JSON, .TXT, .MD), by parsing files and sending them via a query/transfer system of the author's design to the web browser. This allows for interactive viewing of pipeline outputs in real time, without having to necessarily resort to download of output files.

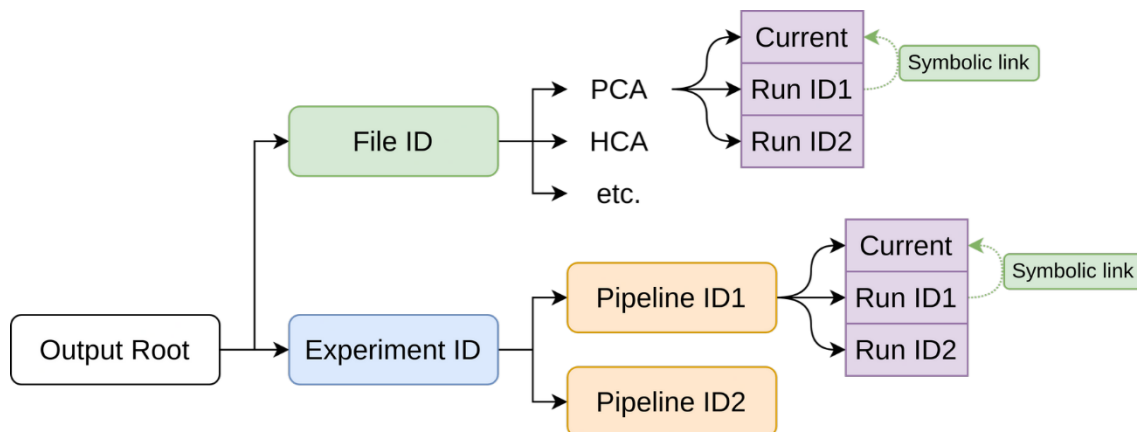


Figure 3-6: The output hierarchy, determined by the type of analysis run. This allows the platform to display information interactively in the browser about the pipeline run outputs, using a Data API-mediated approach that doesn't expose the complete directory structure of the host.

3.4.3 Final User Interface Design

The final design of the user interface of the new platform was informed by the key entities modelled by the server-sided application. As such, key concepts such as experiments, sensors, datasets, and pipelines are all configurable in the browser, through the use of interactive forms.

MSI

Device Installation Date

02/05/2018

Visibility

Public

File

Choose files

representative_sensor_output_file.csv

1. Selection of representative sensor data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Sampl	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	Comp	TVC

Name?

component

Row Type - TBD

Mean

Start Cell

B

End Cell

R

+

2. Header Concepts are annotated and recorded as metadata using these menus

Property Name	Type	Start Column	End Column	
<N/A>	SampleName	0	0	×
component	Mean	1	17	×
bacterial_count	TVC	18	18	×

3. ...and appear here

Figure 3-7: The Sensor Configuration Page in the new platform. Users select a representative Dataset tabular file (1), and use it as a reference to annotate the header concepts found within it. Any subsequent uploads of datasets matching this Sensor configuration will be automatically processed with the ontological concepts annotated in this menu (2 and 3).

Dataset Name: MSI Chicken Aerobic_and_Vacuum Training Data

Description: MSI Training Batch for Platform Evaluation.

Type: MSI

Extra Metadata: {}

Sensor ID:
Dataset and Sensor Hashes Equal Videometer Mean/Std.Dev.
ID: 67b40293cd2f132f5cfc3d52
Local Header Hash: ae788c93a37152a862e22dfa36310634e60810dcd37de50362671dc2fcfb0249
Sensor Header Hash: ae788c93a37152a862e22dfa36310634e60810dcd37de50362671dc2fcfb0249

Search for a Sensor Specification

☒ AND ☐ OR

Name	type
Videometer Mean/Std.D...	MSI

Figure 3-8: The Dataset upload Page in the new platform. The user selects a dataset to upload (1), and matches it to a stored Sensor Configuration (2). The Dataset input is sanity checked with the stored Sensor Configuration, by matching the stored Sensor’s header hash with the hash of the Dataset.

For the interactive design of data processing and machine learning pipelines, the design of the UI elements that would enable interactive configuration were informed by the nature of the machine learning concepts explored in both the introduction of this thesis, and the design of the machine learning pipelines implemented manually in the previous chapters.

Particular emphasis was paid to the concept of data- and sensor- fusion and to how this could be represented in the browser, and more importantly, how this could be used for the construction of reusable pipelines and models that make use of fusion concepts, and as a result expect multiple sensor inputs to function.

The final design of this was influenced by the concept of “channels”. During configuration of a machine learning pipeline (the model training phase), these channels are used to load sensor configurations (see Figure 3-9). Representative datasets can be assigned to these sensor channels, using advanced querying based on the ontological concepts assigned to the data upon upload. Data can

then be preprocessed using multiple filtering and preprocessing algorithms, such as Standard Normal Variate (SNV) and Savinsky-Golay. Additionally, exploratory data analysis (EDA) tasks can be performed on assigned data, to be used for outlier annotation and user-directed feature selection (see Figure 3-10 and Figure 3-11).

After these configurations are set, the user can select any model type or analysis pipeline currently available on the platform (stored in the MongoDB database as “Nextflow Processes”), and similarly configure these. The forms for setting up the analysis pipelines are rendered dynamically from configuration files stored in the appropriate Nextflow Process document, in a similar manner to how Galaxy (Afgan et al., 2018) is able to dynamically render forms for its tool shed plugins.

At this stage, the user is able to click the “Export” button on the GUI menu, giving access to a JSON configuration file that upon “Import” into a new instance will repopulate the GUI with the configuration options selected at export time, accelerating productivity when instantiating repetitive tasks.

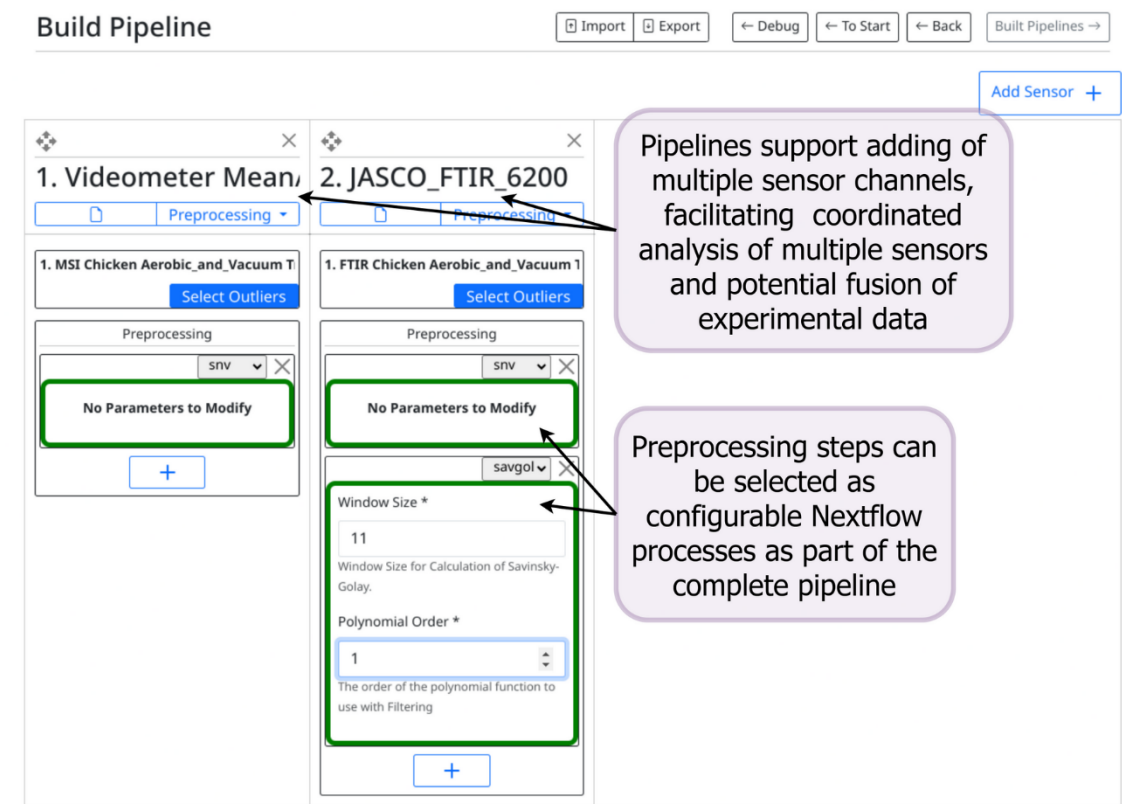


Figure 3-9: The pipeline configuration page within the web application. The user configures “channels” of sensors and associated datasets, and subsequent data analysis or machine learning pipelines. Valid configurations are also given visual feedback, in this case indicated by a *solid* green border around the process form – invalid configurations will result in a *dashed* red border.

Dataset Information:

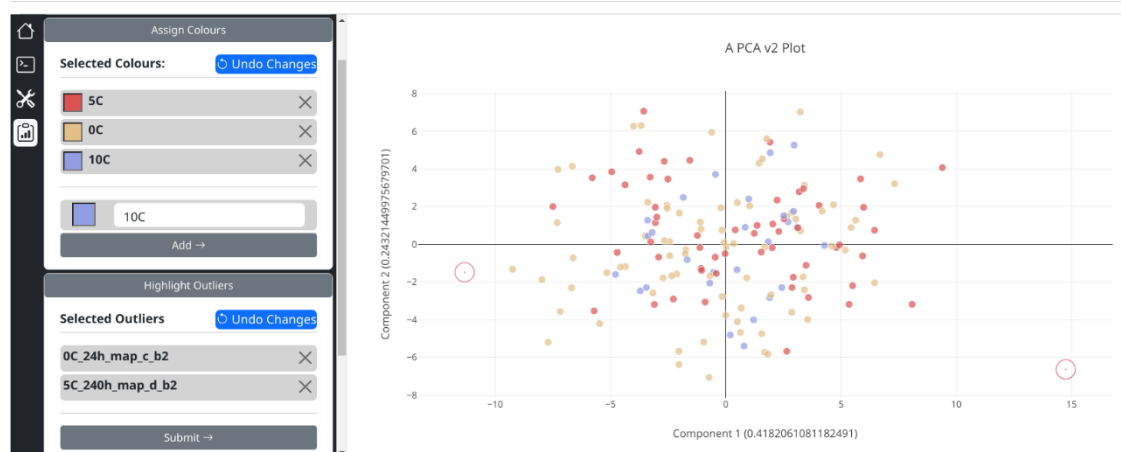


Figure 3-10: Once datasets are assigned to a sensor channel, the user can conduct a variety of exploratory data analysis steps, to view the structure of the data and annotate outliers for exclusion from subsequent data processing.

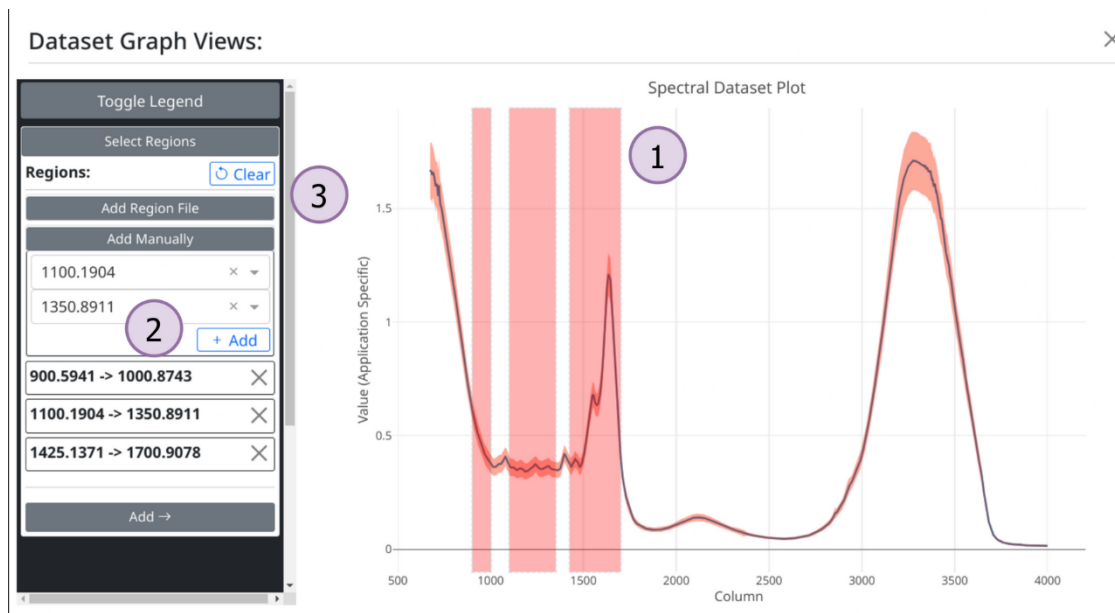


Figure 3-11: For complex spectral data, the user is able to manually direct the process of feature selection, through selection of included regions on a representative data trace (1) - informed by prior research or publications, the feature ranges (2) will be used exclusively in any following pipeline steps or inferencing tasks. Alternatively, comma separated value (CSV) files of spectral ranges from prior work can be supplied to automate the process of range selection (3).

Machine learning algorithms can be added to processing pipelines using the graphical interface similar to the preprocessing steps (Figure 3-12). These are modular components that can be added to the platform and reused, with a standard method of defining inputs/parameters and dynamically rendering appropriate forms in the web application.

Fourteen different regression model algorithms are currently available from within the pipeline builder GUI: Linear, Ridge, Lasso, ElasticNet, Support Vector Machine, Principal Component, Partial Least Square, Bayesian Ridge, Decision Tree, Random Forest, Gradient Boosting, Extra Trees and Adaptive Boosting (AdaBoost). A further two model evaluation/hyperparameter processes are also implemented: the first implements nested-cross validation for all of the mentioned regression algorithms for a relatively unbiased assessment of model suitability for a specific machine learning task, and the second performs hyperparameter

tuning operations in using random grid search with repeated k-fold cross validation for optimisation of model performance and comparison between tuned models. The platform makes the support of further algorithms possible by defining an appropriate Nextflow process snippet and associated configuration form that defines the GUI form fields in the model builder. Processes can also be configured to run in isolation e.g., in Docker containers or Conda environments, to ensure that the employed algorithm has access to the appropriate libraries and frameworks to allow successful execution.

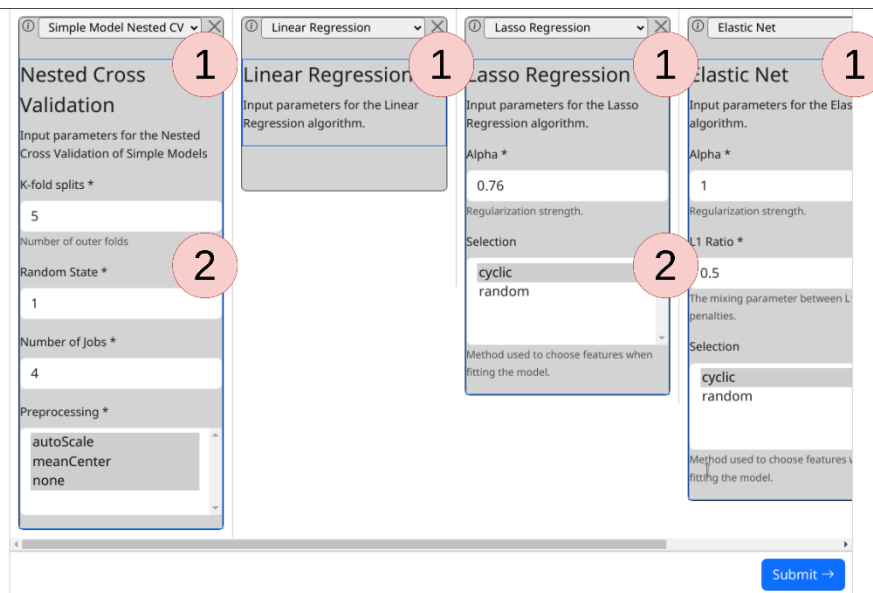


Figure 3-12: A demonstration of the configuration of an analysis pipeline. In this example, the user has selected four modular processes (1), and is contextually able to modify each of their parameters, if supported (2).

A pipeline configuration request is sent to the accompanying API component, and automatically a Nextflow pipeline directory is configured, by parsing the request from the web application. The directory is populated with the datasets that were requested in the user interface, then the entire configured pipeline is compressed into an archive format, and stored in the server file system or object store service. These configured pipeline archives can then be downloaded directly by the user and utilised in a standalone fashion, away from the platform's architecture.

Once a pipeline is configured, it can be run as a complete pipeline, and the user is kept informed of its current status via the pipeline run status page. If a pipeline fails, the user is able to view the error log emitted from the Nextflow execution,

thus facilitating debugging of unexpected errors. This separates the effects of an error from the overall functioning of the platform: other users or other pipeline executions are unaffected by these types of occurrences.

Upon successful completion of a pipeline run, the user can view the outputs in their native output directory structure interactively, using the built-in viewer, capable of representing most plain-text types, as stated in Chapter 3.4.2.3: Pipeline Execution Environment. The user also has the option to compress the entire output directory into an archive for download.

The system will detect when selected output file types are representative of supported serialised model formats – “pickle” or “cloudpickle” is supported. These models must be accompanied by a configuration file with the naming format “{name}.{serialisation_extension}config”; for example, “plsmodel.pklconfig”. These files are autogenerated in the Nextflow Process code body.

This configuration file contains all of the relevant information used to generate the model, and includes the references to the sensor “channels” used in its construction, alongside any information about its particular configuration, i.e., the parameters used in model training. Currently, Scikit-Learn (Pedregosa et al., 2011) models and pipelines (nb. the pipeline concepts from the Scikit-Learn library, consisting of chained transformers and predictor models), are well-supported with this functionality, due to the relative ease of serialisation of custom model and pipeline designs, especially in conjunction with Cloudpickle.

A request can be sent to the Data API for the model to be saved on the platform as a model “endpoint”, along with access specification (private or public). The serialised models themselves are saved either on the local filesystem, or stored in a configured object store. The database entity model used can be found in Figure 3-13.

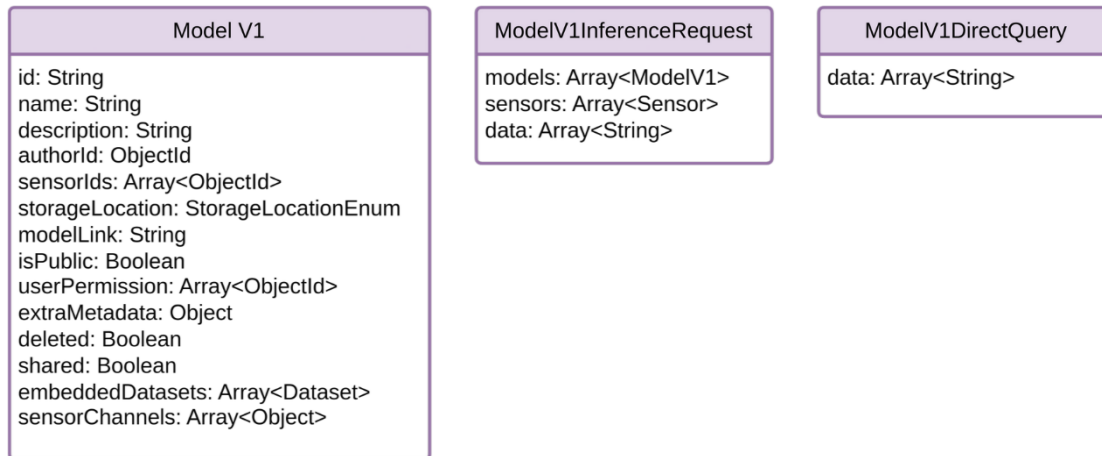


Figure 3-13: Entity diagram for a Model Endpoint. This supplies pertinent information to the platform about the specification of models saved by a user. **ModelV1InferenceRequest** is used in the GUI to request inference from multiple model configurations as specified in the “models” field, and data is supplied as an array of strings, with each entry in the array representing a single sensor’s input (if the model expects more than one channel), in the order specified in the “sensors” field. **ModelV1DirectQuery** is used to make requests directly via the REST API. Only singular models can be specified in this approach, and data order must exactly match that expected by the requested model.

The web application includes an inference API, that allows the user in a prediction context to cue predictor variables representing the sensor channels used in the model, and query the model endpoint. Users can select multiple models to be queried in parallel, and receive results in real (or close to real) time. Users can also query the API directly, either using publicly available models, or privately available ones, provided that the user is authorised to use them. This is represented graphically in Figure 3-14.

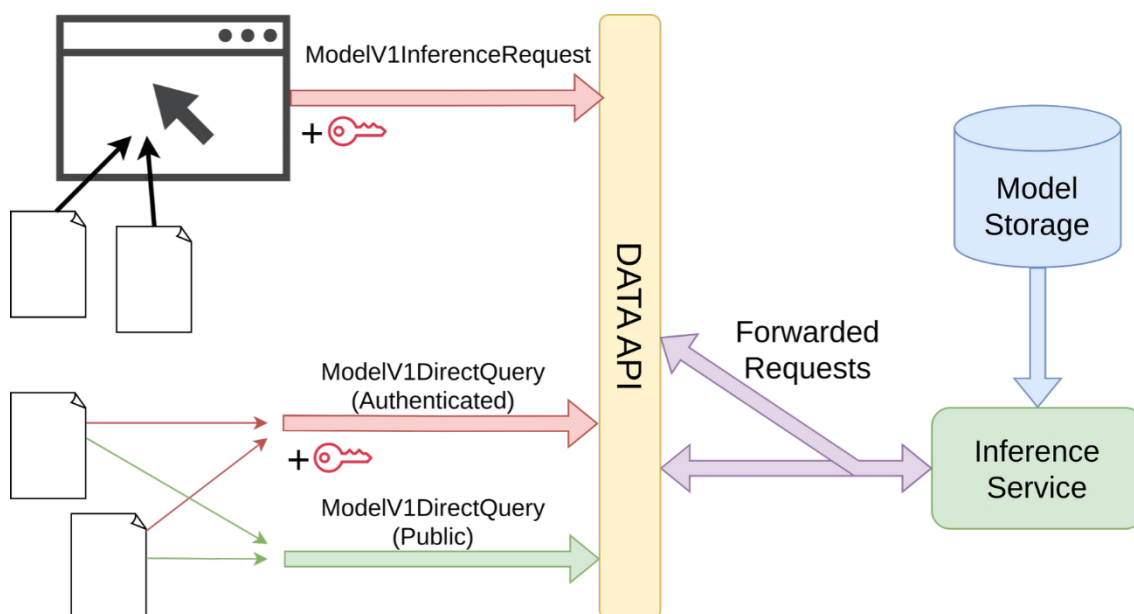


Figure 3-14: The ways in which the new platform's model inference functionality can be actuated. All requests made through the GUI must be authenticated regardless of the requested model's public visibility, whereas requests made via the API can either be public or authenticated, depending on the requested model.

An archive containing the built pipeline, processes, and input files can be downloaded separately, as stated above, to allow the user to reproduce the workflow on any platform with the correct software, allowing for greater reproducibility and adaptability to new data.

3.4.3.1 Model Inference Service

Once the model has been saved and an authorised inference request has been sent to the main data API, it is forwarded to the appropriate model server via HTTP (synchronous) requests – the server then retrieves the model from storage using the resource identifier, and pushes them into an LRU (Least Recently Used) cache, which temporarily stores them for the eventuality of repeated inference requests with the same models.

The model is then used for the prediction task, and the prediction array is sent back to the API. If multiple model predictions have been requested, the API then aggregates all predictions made and sends them back to the user, in a JSON object with the Model V1 IDs as keys.

3.4.4 Semantic Data Querying

The incorporation of FAIRification functionalities into the project address approximately one half of the requirements of a FAIR data repository, in general terms, the assignation of unique identifiers to instances of (meta)data, clear definition of authentication and authorisation of the implemented access protocol(s) and annotation of data with ontological concepts to enhance its interoperability and reusability.

Beyond this, the platform required a mechanism to actually make use of this enhanced data repository, by supporting advanced querying mechanics, and linking back to the experimental/sensor data stored within the primary (MongoDB) repository.

A methodology was devised to implement this functionality beyond what was offered by the SPARQL endpoint inherently available from the knowledge graph service itself (Virtuoso). It was felt that the use of this service was beyond the technical knowledge of the typical food scientist/operative, and required an extra level of abstraction to streamline the process of data access, through a semi-graphical addition to the web application that would allow users to construct SPARQL SELECT queries through interaction with context-driven query “building blocks”, thus recursively constructing the complete query. This was achieved through the addition of new Data API messaging concepts: at every addition of a context-dependent block representing a property about the preceding one (or the top level resource identifier), auto-complete options (applicable object and datatype properties) were queried from the SPARQL endpoint via the Data API using prepared statements, and presented to the user for subsequent selection. Once a statement is fully formed by the user, the nested JSON output is sent to the Data API, where it is checked for forbidden operations and parsed into a SPARQL query, and sent to the hosted endpoint. The Data API interacts with the endpoint’s result, and sends it back to the web application for display to the user. The graphical interface is shown in Figure 3-15, and the corresponding request object sent to the Data API, and the resultant generated SPARQL query are shown in SPARQL Functionalities.

Select Property: + ✖

Select Property: Enter Value: Regex? ☒ ✖

Select Property: Enter Value: Regex? ☒ ✖

Select Property: Enter Value: ✖

Add New Condition

Add New Condition

Figure 3-15: The process for making a graphical SPARQL query using the Web Application. In this example the user is requesting all Datasets, matching the temperature regular expression “10” and the packageType regular expression “map”

3.4.5 Final Service Architecture

Figure 3-16 represents the completed topology of the new platform. The graphic highlights the separation of functionalities into three categories of operation: *primary data handling*, responsible for serving uploading, downloading and other interactions with analytical sensor data, the *semantic functionalities*, responsible for transformation of the ontological concepts within the sensor data into formats compatible with knowledge graph storage, and *task handling*, responsible for execution of machine learning and data transformational tasks.

Each subdivision is equipped with its own MongoDB database collection for persistence and as previously described, the system as a whole makes use of asynchronous message queues to decouple the performance of the user-facing API service from the back-end functionalities of the platform. This allows for future scalability of these components, as the queue allows for tasks to be distributed amongst replicated stateless services if the platform is saturated with tasks, such

as those for Nextflow task handling, and reduces load on the API service which is no longer responsible for long-running job execution, as was an issue for the SorfML platform. Dataset entity uploads are pushed to either the hosting file system or an external S3-compatible object storage service, to avoid exceeding the ~20MB size limitations imposed on individual MongoDB documents. MongoDB's inbuilt GridFS solution proved unsuitable for data storage due to its relatively undocumented support between programming languages, complicating data transfer between components.

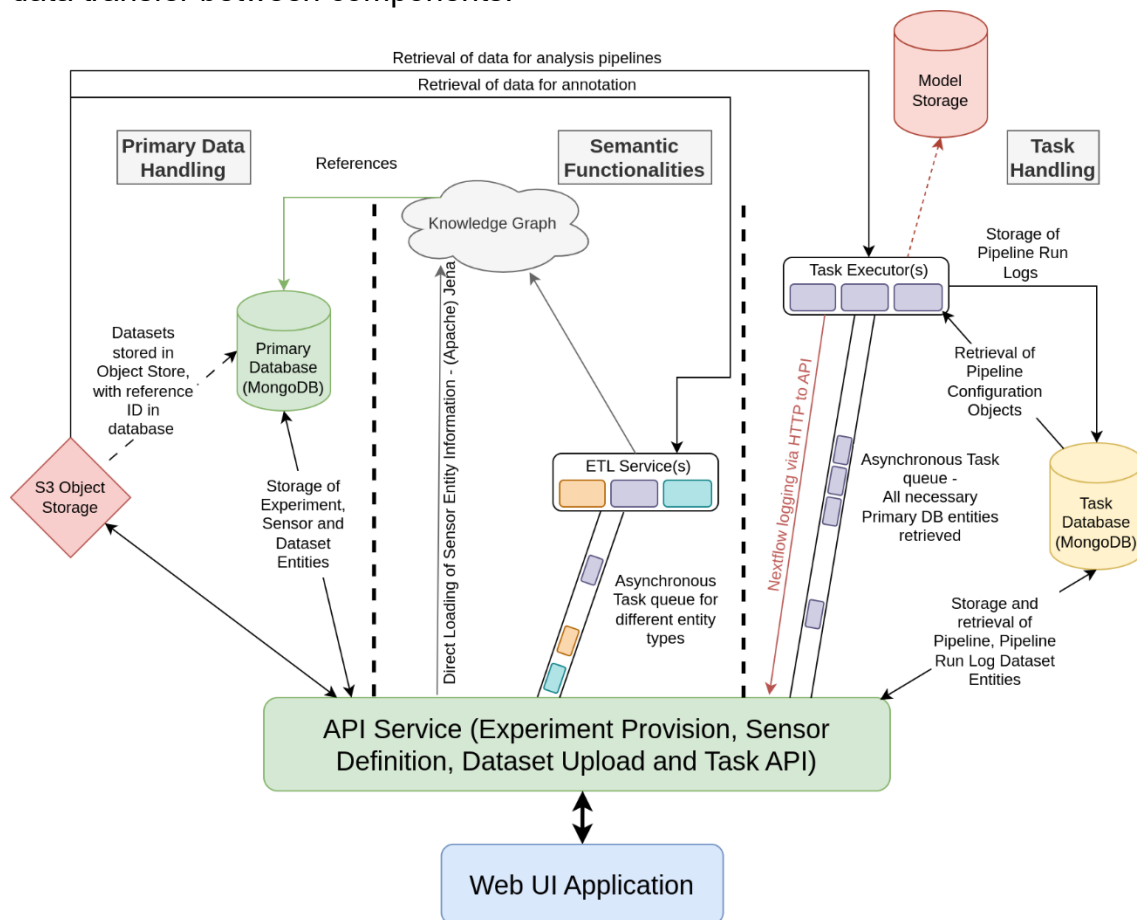


Figure 3-16: The final architecture of the platform, divided with dashed lines into services for primary data handling of uploaded and processed data, semantic functionalities related to annotation of data (FAIRification) and task handling, which manages the Nextflow execution environment.

3.5 Platform Evaluation

The first aim of this platform evaluation was to demonstrate the utility of the new platform in iterating on the original SorfML/MeatReg platform's machine learning model benchmarking capabilities, through the assessment of multiple configurations of model algorithm and data pretreatment. The second, equally important aim was to use the platform to explore alternative algorithm choices for the prediction of bacterial load in chicken samples,

To demonstrate the new platform's ability to perform machine learning model benchmarking and ranking functionalities, chicken spoilage regression MSI (VideometerLab) and FTIR datasets were used in the first case study to conduct a nested cross validation assessment of fourteen different regression model algorithms: Linear, Ridge, Lasso, ElasticNet, Support Vector Machine, Principal Component, Partial Least Square, Bayesian Ridge, Decision Tree, Random Forest, Gradient Boosting, Extra Trees and Adaptive Boosting (AdaBoost).

Upon completion of this task, a conventional hyperparameter grid search and cross validation procedure was conducted to further fine-tune the best performing models selected from the initial nested cross validation.

The second case study demonstrated the other aspect of the new platform's intended purpose: complete models were built and served using the web interface and queried for inferencing tasks using test data from separate experimental batches. Following this, test batch performance was evaluated using the model auditing and comparison facilities built into the web application.

3.5.1 Materials and Methods

3.5.1.1 Meat Sample Preparation

Fresh chicken thigh samples were sourced from Greek meat distributors rapidly after preparation. Chicken thigh fillets were provided by the food company KOTINO (Athens, Greece). Sample storage and preparation is fully described elsewhere (Fengou et al., 2024). Chicken samples were prepared and stored in either vacuum-packaged or an aerobic state.

The samples were stored at one of three isothermic temperatures for both packaging conditions, 0°C, 5°C (safe) or 10°C (marginally abusive), and monitored at 24-48h intervals, until visual deterioration of the samples became apparent. Data for each storage condition was recorded in two batches, corresponding to meat sourcing on separate days, with a variable number of replicates (2-11) for each sampling condition in each batch. Maximum storage time varied depending on the storage temperature and packaging condition: for the vacuum-packaged samples, those stored at 0°C and 5°C were monitored with all sensor configurations for a maximum of 240h after collection, while those stored at 10°C were monitored for a maximum of 120h. For the aerobically packaged samples, those stored at 0°C were monitored for a maximum of 216 h, those stored at 5°C were monitored for a maximum of 192h, and those at 10°C for a maximum of 96h.

3.5.1.2 Spectral Acquisition

MSI spectra were captured using the VideometerLab platform (Videometer, Denmark, 2021), which captures information in a spread of 18 monochromatic wavelengths ranging from 405nm to 970nm, at intervals between 15 and 150 nm. The imaging data produced from the sensor is representative of two spatial dimensions and one wavelength dimension in a layered *width×height×18* data cube, with each layer representing the reflected response from illumination with each wavelength.

To convert the data into a tabular form that can facilitate analysis with conventional data processing methods, a segmentation and averaging process is undertaken, wherein Canonical Discriminant Analysis (CDA) is performed, using the accompanying VideoMeter software version 2.12.39. In an averaging process described in previous work, the mean and standard deviation intensity of reflectance across the target area was recorded (Fengou et al., 2019).

FTIR spectra were recorded between 4000 and 400cm⁻¹ using a zinc selenide (ZnSe) HATR (Horizontal Attenuated Total Reflectance) crystal and an FTIR-6200 spectrometer (JASCO, Tokyo, Japan) equipped with a standard sample chamber, a triglycine sulphate (TGS) detector and a Ge/KBr beamsplitter. The

HATR crystal had a refractive index of 2.4 and a penetration depth of 2 μm at 1000cm^{-1} .

Using the Spectra Manager™ Code of Federal Regulations (CFR) software version 2, spectra were collected by accumulating 100 scans with a resolution of 4 cm^{-1} and a total integration time of 2 min.

3.5.1.3 Data Partitioning

As the experimental data consisted two sets of two aerobic and two vacuum stored batches, corresponding to observations made with each sensor, the MSI and FTIR data was collated into three separated train-test scenarios: aerobic, vacuum storage and combined aerobic and vacuum storage conditions. In the first two scenarios, one batch formed the train partition and the other formed the test partition. In the combined scenario, the aerobic and vacuum train batches were combined, the test batches combined in a similar fashion separately. The sizes of the training and testing partitions for each scenario are detailed in Table 3-4.

Table 3-4: Table of the training and testing partition sizes for the three scenarios explored in this study.

Partition	Aerobic condition	Vacuum condition	Combined conditions
Train	70	65	135
Test	66	72	138

3.5.1.4 Platform Data Upload and Processing

The Videometer (Multi-spectral Imaging - MSI) and Fourier-transform infrared (FTIR) sensors were configured on the platform, using the training data to map ontological concepts to the headers found in the training batch data. Once the sensor entities were configured, the training datasets were uploaded with a reference to the appropriate sensor. This process required annotation of the observation names found within the dataset. This process was intended to facilitate downstream analysis on the platform by granting the associated ETL (“FAIRification”) service additional information to link to concepts found within the observation naming scheme used by the researcher, and make the information available for ingestion into the knowledge graph. The Sensor/Dataset upload workflow is shown in Figure 3-17 and Figure 3-18 respectively, using the MSI aerobic training data as the example, however the process is no different for the other training data inputs including that from FTIR. Figure 3-19 shows the optional annotation of the MSI entity with geographical information detailing the location of origin of the dataset. This functionality was intended to record the location of the experimental/monitoring facility, and not for recording of the distribution chain or provenance of the original foodstuff.

Add Sensor Specification
×

Sensor/Device Instance Name

Videometer Mean/Std.Dev.

Description

Videometer Chicken Thigh Data subject to Canonical Discriminant Analysis in VideoMeter software version 2.12.39.

Sensor Model Name

VideometerLab

Sensor Type

MSI

Device Installation Date

01/01/2024

Visibility

Public

File

Choose files msi_air_b2.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		405	435	450	470	505	525	570	590	630	645	660	700	850	870	890	910	940	970	StdDe	StdDe	StdD

Name?

Row Type - TBD

Start Cell

End Cell

+

✕

Name

RESPONSE

AL

AL

Property Name	Type	Start Column	End Column	
Sample ID	SAMPLE_NAME	0	0	✕
<N/A>	MEAN	1	18	✕
<N/A>	STANDARD_DEVIATION	19	36	✕
<N/A>	RESPONSE	37	37	✕

Figure 3-17: Configuration of the MSI Sensor Entity used for the MSI data in this study. The representative file is supplied only as a reference for the interactive header annotation wizard, and is not stored. Header ranges are selected either graphically via the tabular view or via the form underneath. Previously selected ranges are highlighted in the tabular view, and are listed and removable in the bottom section of the wizard. The “Type” field is populated with one of a variety of concepts pre-loaded onto the platform, in this case the significance of the numerical values in columns 1-18 (mean values) and 19-36 (standard deviations from the segmentation process used in conversion of Videometer data from a graphical to tabular data representation).

msi_air_train.csv

Dataset Name
MSI Chicken Aerobic Training Data
Type
MSI

Description
MSI Chicken Aerobic Training Data.
Extra Metadata
{}

Sensor ID

Dataset and Sensor Hashes Equal
Videometer Mean/Std.Dev.
ID: 67b40293cd2f132f5fc3d52
Local Header Hash: ae788c93a37152a862e22dfa36310634e60810dcd37de50362671dc2fcfb0249
Sensor Header Hash: ae788c93a37152a862e22dfa36310634e60810dcd37de50362671dc2fcfb0249

Search for a Sensor Specification

AND
OR

name
like
Videometer Mean/Std.

+ Rule
X

Name
type
Videometer Mean/Std.D...
MSI

Submit

Sample Naming Scheme

Define Sample Naming Scheme *Naming Scheme Correct*

+
Selected Separator:

Block 1:
hasTemperature
Remove
+
Block 2:
time
Remove
+
Block 3:
hasPackageType
Remove
+
Block 4:
hasAdditionalId
Remove
+
Block 5:
hasBatchNumber
Remove

Generate String

Generated Naming Concepts:
hasTemperature_time_hasPackageType_hasAdditionalId_hasBatchNumber

Current Regex Scheme:
^[^_]+_[^_]+_[^_]+_[^_]+_[^_]+\$

Name
OC_168h_air_a_b1
OC_168h_air_b_b1
OC_168h_air_c_b1
OC_168h_air_d_b1
OC_168h_air_e_b1
OC_168h_air_f_b1
OC_192h_air_a_b1
OC_192h_air_b_b1
OC_192h_air_c_b1
OC_192h_air_d_b1

Figure 3-18: Upload procedure for Videometer (MSI) aerobic training data. Individual tabular files are selected and annotated with basic metadata. Following this, the appropriate sensor configuration is selected, and sample/observation labelling takes place, using ontological concepts from the associated ontology. Following upload to the file API, the ETL service will split and store the observation identifier concepts using this schema.

Select Data Origin

Agricultural University of Athens, Ιερά Οδός, Βοτα

37.982451600000005

23.70444440333074

Agricultural University of Athens, Ιερά Ο

Clear

Visibility

Public

Is it Metadata?

No

Submit

Figure 3-19: As an optional step, users can further annotate the uploaded dataset with geographical information detailing the location of origin of the dataset. Users can select this location directly from the map view, or via the location search or latitude/longitude inputs. In this instance and as an example, the MSI data was annotated with the details of the Agricultural University of Athens, the originating institution of the experimental data used in this work.

3.5.2 Case Study 1 – Chicken Spoilage Model Benchmarking and Ranking

3.5.2.1 Case Study Introduction

To provide a generalised assessment of algorithm performance across the training dataset(s), nested cross-validation was conducted. Nested cross validation is an adaptation of conventional cross validation and is typically used in a different fashion. Non-nested cross-validation is commonly used in the tuning phase of machine learning model construction, often alongside an exploration of the hyperparameter space of the model, in a *grid search*. The specific implementation of the grid search process can take many forms, such as exhaustive searching, random searching or more elaborate

probabilistic/Bayesian methods, however the cross-validation process in general involves an assessment of generalised performance of the model using some type of repeated splitting of the training set. K-fold is commonly employed, in which the training set is subdivided into k folds and candidate models are repeatedly trained on different combinations of $k-1$ folds and tested with the held out fold. The per-fold performance is averaged and is used as a metric for worthiness of hyperparameter selection. This process has utility when the training set is small, as it maximises the degree of exploration of unseen data used for train/validation splits.

The general non-nested strategy is sometimes also used in a benchmarking capacity for the direct comparison of generalised model performance, however this practice is considered by some to be unsound: Cawley and Talbot describe how this practice is susceptible to optimistic bias through overfitting of models to training data when hyperparameter tuning is conducted in the same step as performance estimation, with inadequate separation between training and validation data. They argued for the necessity of nested cross-validation (NCV) in obtaining an *unbiased estimate* of model performance, wherein the inner loop is used for hyperparameter tuning and the outer loop is used to obtain validation results from the inner loop's best-fitted models (Cawley and Talbot, 2010)

3.5.2.2 Methods

The following work utilised a two-step workflow: first, a nested cross validation procedure was conducted using the platform, to assess the general suitability of a wide range of regression algorithms (Linear, Ridge, Lasso, ElasticNet, Support Vector Machine, Principal Component (PCR), Partial Least Square (PLS), Bayesian Ridge, Decision Tree, Automatic Relevance Determination (ARD), Random Forest, Gradient Boosting, Extra Trees and Adaptive Boosting (AdaBoost)). This process used a five-fold outer loop, with an inner five-fold hyperparameter optimisation procedure making use of the outer four training folds of each outer combination. Within the inner loop, the size of the default hyperparameter search grid, coupled with the computational complexity of exhaustive or Bayesian grid search methods prompted the use of random search,

which was used to balance likelihood of selection of well-performing parameters with performance, an important consideration in the context of a web application potentially serving multiple users simultaneously.

The inner grid search process utilised one hundred random iterations, which covered at least 20% of the total search space for the tree-based models and at most more than 100% for others i.e. equivalence to exhaustive search.

This process was repeated for each preprocessing/pretreatment strategy selected, in this case, standard scaling, mean centring, and raw data/ no scaling. The hyperparameter space is detailed in Table 3-5.

Table 3-5: Table of hyperparameter spaces explored in the nested cross-validation process for all MSI and FTIR training data. This process was conducted for the three preprocessing methods selected (standard scaling (Z-score normalisation), mean centring and raw/no preprocessing).

Model	hyperparameter Values
PCRRegression	n_components: [1-10]
PLSRegression	n_components: [1-10]
Ridge	alpha: [0.1, 1-9], solver: [auto, svd, cholesky, lsqr, sparse_cg, saga]
Lasso	alpha: [0.1, 1-9], selection: [cyclic, random]
Lars	n_nonzero_coefs: [5, 10, 15], eps: [1e-6, 1e-7, 1e-8]
Elastic Net	alpha: [0.1, 1-5], l1_ratio: [0.0-0.9 (step 0.1)]
SVR	C: [0.1, 1, 5, 10], gamma: [scale, auto], kernel: [linear, poly, rbf, sigmoid]
BayesianRidge	alpha_1: [1e-6, 1e-7, 1e-8], alpha_2: [1e-6, 1e-7, 1e-8], lambda_1: [1e-6, 1e-7, 1e-8], lambda_2: [1e-6, 1e-7, 1e-8]
ARDRegression	alpha_1: [1e-6, 1e-7, 1e-8], alpha_2: [1e-6, 1e-7, 1e-8], lambda_1: [1e-6, 1e-7, 1e-8], lambda_2: [1e-6, 1e-7, 1e-8]
DecisionTreeRegressor	max_depth: [1-10], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4]
RandomForestRegressor	n_estimators: [10, 50, 100], max_depth: [None, 10, 20], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4]
GradientBoostingRegressor	n_estimators: [100, 200, 300], learning_rate: [0.01, 0.05, 0.1], subsample: [0.5, 0.7, 1.0], max_depth: [3, 4, 5]
AdaBoostRegressor	n_estimators: [50, 100, 150], learning_rate: [0.01, 0.1, 0.5, 1.0]
ExtraTreesRegressor	n_estimators: [100, 200, 300], max_depth: [None, 10, 20, 30], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4]

Following the nested cross-validation process, the resultant “leaderboard” was used to inform the selection of algorithm strategies for a cross-validation/hyperparameter tuning operation. This process used repeated k-fold cross validation as a more robust and less computationally intensive alternative to leave-one-out (LOOCV) cross validation, (a special case of k-fold where k is equal to sample size), which has been indicated as commonly resulting in a high degree of variance in validation performance and for larger training set sizes, the effect of holding one validation observation out at each iteration is minimal on the model, leading to correlated results. K-fold derived methods may result in higher bias than LOOCV, however the impact of this trade-off is negligible for the purposes of hyperparameter tuning, as the bias will likely affect all candidate configurations in a similar fashion (Wainer and Cawley, 2017). For this process, three repetitions of five-fold nested cross validation were used. In addition, limiting model evaluation to a subset of all algorithms allowed for a greater default exploration of the hyperparameter space for a roughly equivalent computational budget on the platform, as such one thousand iterations of random search were configured for every cross validation run by default.

3.5.2.3 Platform Model Configuration

3.5.2.3.1 Nested Cross Validation

Figure 3-20 shows the configuration of a nested cross validation process in the graphical interface. The user first selects the sensor channel, in this case, the VideometerLab platform corresponding to the sensor used in the collection of the experimental data. The MSI aerobic training partition, shown in the upload process in the preceding sections and the MSI vacuum training partition (not shown), were selected as the data to use for the run to represent combined aerobic/vacuum conditions. Then, the “SKLEARN_Regression_Nested_CV” process in the method panel underneath was selected. Standard Normal Variate was then configured in the “Preprocessing” section of the Sensor GUI component. This process was replicated for the FTIR combined condition scenario, with additional Savinsky-Golay filtering, using a first-degree polynomial and an eleven point window.

The `SKLEARN_Regression_Nested_CV` process conducts nested cross validation of the model types listed above, tuning the model hyperparameters in the inner k-fold loop and assessing the best performing configuration against the outer held-out set. The process outputs an interactive report detailing the per-outer-fold RMSE and R^2 performance of each algorithm, alongside the overall mean and standard deviation of these metrics in tabular and graph form.

Additionally, the process produces a JSON file that can be parsed by the web application and viewed in an interactive graph format, which in addition to the former information includes a record of the per-fold best performing hyperparameter combinations, which may help inform hyperparameter ranges in follow-up tuning operations.

1. Videometer Mea

Preprocessing ▾

1. MSI Chicken Aerobic Training Da
Select Outliers

2. MSI Chicken Vacuum Training Da
Select Outliers

Preprocessing
+

Nested Cross Validation

Nested Cross Validation of Regression Models in Scikit-Learn.

Inner K-fold splits *

5

Number of folds in Randomised Search CV

K-fold splits *

5

Number of outer folds

Randomised Search Iterations *

100

Number of picks of hyperparameters in Randomised Search CV

Number of Processing Jobs *

4

Preprocessing *

autoScale
meanCenter
none

Random State *

1

Figure 3-20: Configuration of a Nested Cross Validation Benchmarking Run. The user first selects the Sensor Channels, assigns data, and then selects the **SKLEARN_Regression_Model_CV** process, configuring the number of K-Fold splits and the preprocessing steps to compare.

Upon configuration and submission of the pipeline, the user is transferred to the Pipeline Run Log overview page (Figure 3-21), this allows users to download the completely configured pipeline (top-left icon in left pane) and cue new runs of the pipeline (“Start New Run” button on top row of second pane). Completed runs can be seen in the Logs section of the second pane and downloaded with the log download button. By clicking the inspection button (magnifying glass icon on log entity menu), users are transferred to the run log output viewer as shown in Figure 3-22. From this view, users can inspect the output files from a pipeline. By selecting the output files, users can present the data in context-dependent modes, such as rendered markdown, HTML reports or parsed JSON data in the form of interactive plots (Figure 3-23).

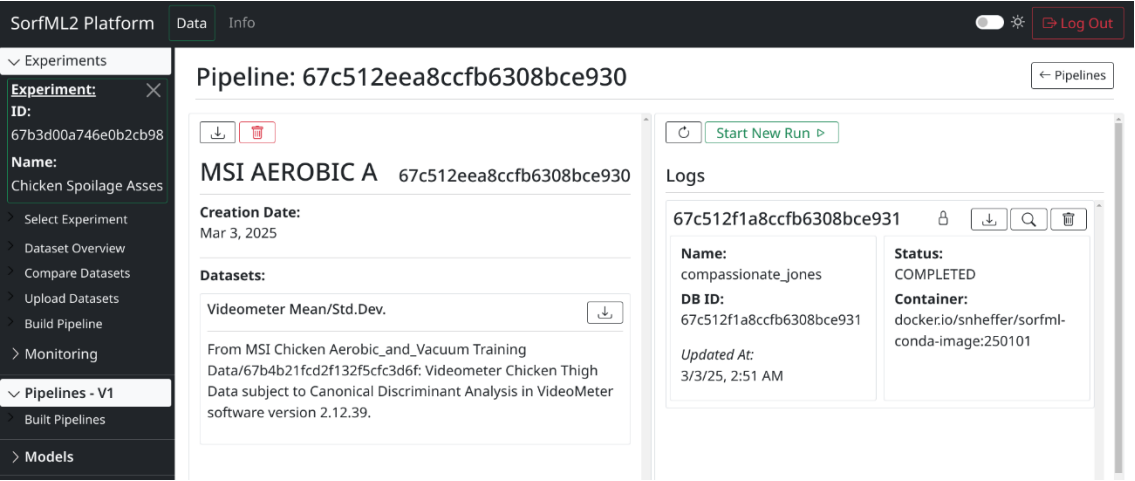


Figure 3-21: Pipeline Run Log page showing the completed run log, with the auto-generated name "compassionate_jones", which contains the benchmarking performance reports for the selected data.

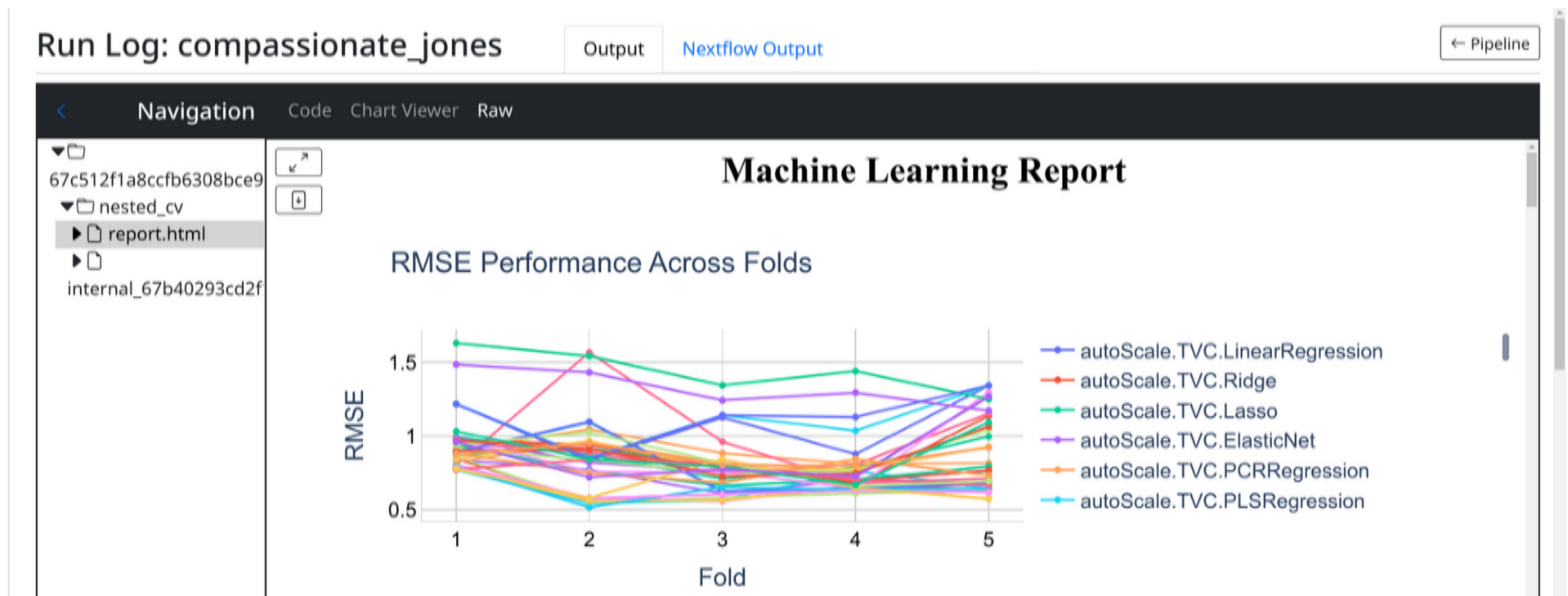


Figure 3-22: Interacting with the report generated by the analysis pipeline: in the first panel, the work flow output directory is explorable from the browser, the view panel showing the interactive HTML report from the SKLEARN_Regression_Nested_CV process. Other representative plots are included in the report, including an RMSE performance heat map, which is shown below.

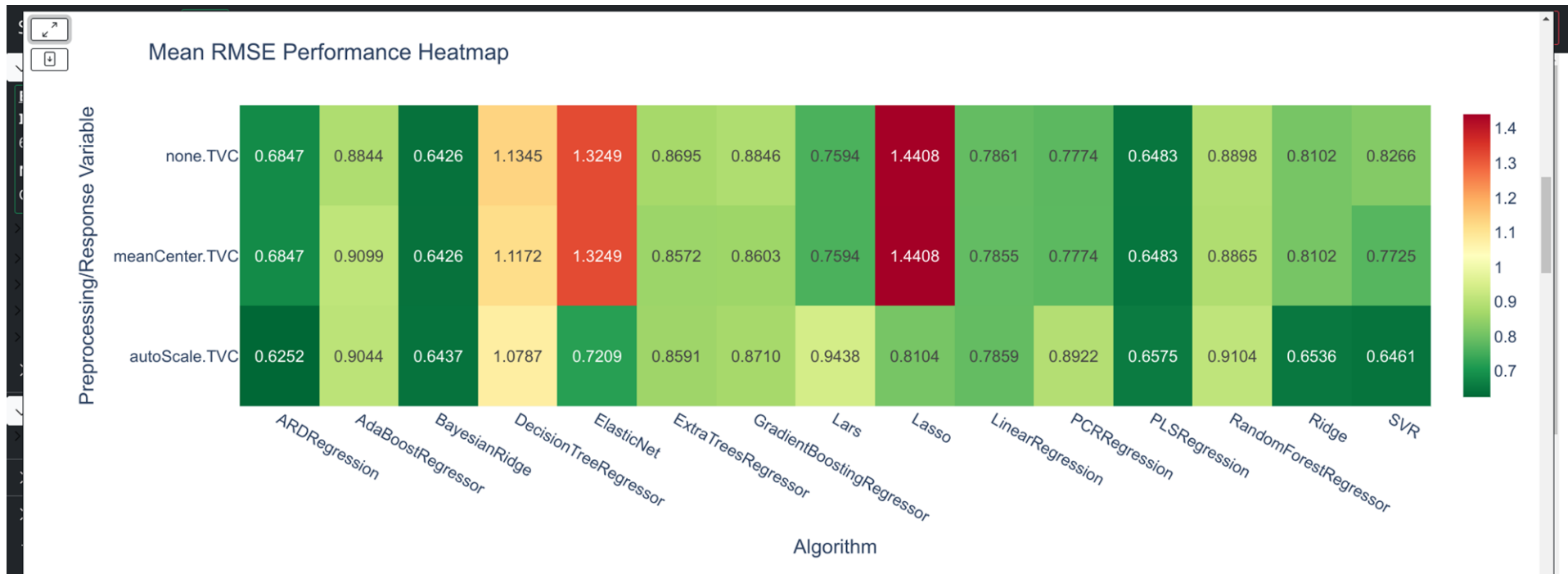


Figure 3-23: Cross Validation RMSE performance Heat Map for the combined MSI Aerobic and Vacuum condition data. In this example, the best mean performing model trends across preprocessing strategies are clearly visible as ARD Regression, Bayesian Ridge and PLS Regression.

3.5.2.3.2 Cross Validation/hyperparameter Tuning

Once the nested cross-validation process was completed for the MSI and FTIR combined aerobic and vacuum scenarios, the best performing model/preprocessing combinations were considered for hyperparameter tuning in anticipation of model construction in the web interface.

Given their strong performance with MSI data and comparable performance with FTIR data (shown in Table 3-7 and Table 3-8 respectively, below), Automatic Relevance Determination, Bayesian Ridge and Support Vector Machine were considered for further evaluation, alongside PLS Regression, which had been used by the author in previous related research (Section 2). Two preprocessing methods were carried forward for this stage: Standard Scaling (Z-score normalisation) was selected as a common preprocessing strategy for the models in the hyperparameter tuning stage, alongside the default baseline performance of internal mean centring in Scikit-Learn's Bayesian Ridge and ARD implementations, which followed the initial row-wise transformations applied in the data preparatory stage (Standard Normal Variate (SNV) and Savinsky-Golay filtering for FTIR, with a first-degree polynomial and a window size of eleven).

The focused nature of the cross-validation procedure allowed for more in-depth exploration of the hyperparameter space, as noted previously. The adapted search grid is given in Table 3-6. As described in the Methods section, the hyperparameter search process consisted of one thousand iterations of random search, with three repeats of five-fold cross validation per hyperparameter selection. Figure 3-24 highlights the configuration of the SKLEARN_Regression_Random_Grid_Search_CV process.

As Bayesian Ridge and Automatic Relevance Determination are probabilistic with similarity to L2-regularised linear regression, i.e. Ridge regression, the effect of optimising the hyper-priors Alpha 1, Alpha 2, Lambda 1 and Lambda 2 was expected to have little impact on the final convergence of the models, because of the large number of training iterations (one thousand) supplied as default for both in the process configuration, coupled with the relatively small amount of training data. Furthermore, the number of random search iterations was able to

adequately cover the SVM search space, effectively resulting in exhaustive grid search.

Table 3-6: Table of hyperparameter spaces explored in the hyperparameter grid search operation for all MSI and FTIR training data. This process was conducted for the three preprocessing methods selected (standard scaling (Z-score normalisation), mean centring and raw/no preprocessing).

Model	Hyperparameter Values
SVR	C: [0.1, 1, 5, 10, 50, 100], gamma: [scale, auto, 0.01, 0.1, 1], kernel: [linear, poly, rbf, sigmoid]
BayesianRidge	alpha_1: [1e-6, 1e-7, 1e-8, 1e-9], alpha_2: [1e-6, 1e-7, 1e-8, 1e-9], lambda_1: [1e-6, 1e-7, 1e-8, 1e-9], lambda_2: [1e-6, 1e- 7, 1e-8, 1e-9]
ARDRegression	alpha_1: [1e-6, 1e-7, 1e-8, 1e-9], alpha_2: [1e-6, 1e-7, 1e-8, 1e-9], lambda_1: [1e-6, 1e-7, 1e-8, 1e-9], lambda_2: [1e-6, 1e- 7, 1e-8, 1e-9]

Cross Validation Configuration

Configuration parameters for cross-validation and hyperparameter tuning of regression models.

☒ Has Index
Whether the dataset has an index column.

Cross Validation Repeats *

3

Number of times to repeat k-fold cross-validation.

K-Fold Splits *

5

Number of folds for cross-validation.

Regression Methods *

ElasticNet

BayesianRidge

ARDRegression

SVR

PCRRRegression

List of regression methods to evaluate.

Number of Iterations *

1000

Number of parameter settings sampled in hyperparameter search.

Number of Jobs *

4

Number of parallel jobs to run during model training.

Preprocessing Strategies *

standardScaler

meanCenter

none

List of preprocessing strategies to apply to the data.

Random State *

1

Random state seed for reproducibility.

Figure 3-24: Configuration of a hyperparameter search in the GUI. The user first selects the Sensor Channels, assigns data similarly to the nested cross-validation procedure, and selects the SKLEARN_Regression_Random_Grid_Search_CV process, configuring the number of k-fold splits, repetitions and the preprocessing steps to compare.

3.5.2.4 Results and Discussion

The best-performing results from the nested cross validation for the VideometerLab (MSI) and FTIR combined aerobic and vacuum conditions are shown in Table 3-7 and Table 3-8 respectively. The ranking of the models shows that Automatic Relevance Determination (ARD), Bayesian Ridge, Partial Least Squares and Support Vector Machine regression demonstrated consistent high performance when applied to MSI data.

For the MSI results, ARD, Bayesian Ridge, SVR and PLS algorithms occupy the first four positions in the leaderboard. However, variation between top-performing configurations is not stark: the difference in mean cross-validation RMSE performance between the highest performing configurations of each algorithm is roughly 0.02 log CFU/g TVC. With regard to model stability across folds, the top performing ARD, Bayesian Ridge and PLS methods produced standard deviation of cross validation RMSE performance below 0.10 log CFU/g TVC.

For the FTIR data, the results show poorer cross-validation performance for all algorithms, with the best-performing model, Bayesian Ridge, producing a mean cross validation RMSE value of 0.793 log CFU/g TVC with a standard deviation of 0.165 log CFU/g, a larger degree than that seen with the equivalent MSI data. The difference in mean cross validation RMSE between the best performing MSI model types when applied to the FTIR data was even less, with at most 0.012 log CFU/g separation.

Results for the hyperparameter tuning stage are shown in Table 3-9 and Table 3-10 for the MSI and FTIR training partitions, and are largely in agreement with those seen in the nested cross validation for the combined aerobic and vacuum conditions. As suggested earlier, the Bayesian Ridge and ARD methods were able to converge regardless of the shape and weight parameters for α and λ provided in the search grid, as the top candidate models for each scenario showed the same RMSE metric results. Interestingly, the default configuration of SVM performed consistently well with standard scaling (linear kernel and $C=1$ or $C=0.1$) and was able to produce similar results to the other methods.

Table 3-7: The twenty best performing algorithms from Nested Cross Validation using the VideometerLab (MSI) combined aerobic and vacuum training partition, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 5 outer and 5 inner folds. Results were generated with the *SKLEARN_Regression_Nested_CV* process. Note that the inclusion of a mean centring step in the Scikit-Learn implementations of Bayesian Ridge regression and PLS regression leads to consistent results regardless of user-specified transformation.

Model	RMSE Mean	RMSE Std Dev	R2 Mean	R2 Std Dev
standardScaler.TVC.ARDRRegression	0.6252	0.0921	0.8018	0.0324
meanCenter.TVC.BayesianRidge	0.6426	0.0817	0.7901	0.0324
none.TVC.BayesianRidge	0.6426	0.0817	0.7901	0.0324
standardScaler.TVC.BayesianRidge	0.6437	0.0959	0.7894	0.0393
standardScaler.TVC.SVR	0.6461	0.1227	0.7872	0.0541
meanCenter.TVC.PLSRegression	0.6483	0.0983	0.7837	0.0526
none.TVC.PLSRegression	0.6483	0.0983	0.7837	0.0526
standardScaler.TVC.Ridge	0.6536	0.1158	0.7821	0.0526
standardScaler.TVC.PLSRegression	0.6575	0.1070	0.7767	0.0645
meanCenter.TVC.ARDRRegression	0.6847	0.1212	0.7562	0.0887
none.TVC.ARDRRegression	0.6847	0.1212	0.7562	0.0887
standardScaler.TVC.ElasticNet	0.7209	0.0869	0.7363	0.0352
none.TVC.Lars	0.7594	0.0731	0.7039	0.0587
meanCenter.TVC.Lars	0.7594	0.0731	0.7039	0.0587
meanCenter.TVC.SVR	0.7725	0.1053	0.6951	0.0653
meanCenter.TVC.PCRRegression	0.7774	0.0825	0.6921	0.0475
none.TVC.PCRRegression	0.7774	0.0825	0.6921	0.0475
meanCenter.TVC.LinearRegression	0.7855	0.2014	0.6787	0.1494
standardScaler.TVC.LinearRegression	0.7859	0.2033	0.6781	0.1515
none.TVC.LinearRegression	0.7861	0.2031	0.6779	0.1512

Table 3-8: The twenty best performing algorithms from the Nested Cross Validation using the FTIR combined aerobic and vacuum training partition, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 5 outer and 5 inner folds. Results were generated with the SKLEARN_Regression_Nested_CV process. Note that the inclusion of a mean centring step in the Scikit-Learn implementations of Bayesian Ridge regression and PLS regression leads to consistent results regardless of user-specified transformation.

Model	RMSE Mean	RMSE Std Dev	R2 Mean	R2 Std Dev
standardScaler.TVC.BayesianRidge	0.7933	0.1653	0.6681	0.1446
standardScaler.TVC.Ridge	0.7948	0.1931	0.6614	0.1752
none.TVC.BayesianRidge	0.8035	0.1712	0.6594	0.1519
meanCenter.TVC.BayesianRidge	0.8035	0.1712	0.6594	0.1519
standardScaler.TVC.ElasticNet	0.8038	0.1671	0.6579	0.1542
standardScaler.TVC.ARDRRegression	0.8053	0.1763	0.6554	0.1574
standardScaler.TVC.SVR	0.8075	0.2095	0.6444	0.2029
meanCenter.TVC.PLSRegression	0.8136	0.1680	0.6472	0.1640
none.TVC.PLSRegression	0.8136	0.1680	0.6472	0.1640
none.TVC.ARDRRegression	0.8154	0.1839	0.6461	0.1662
meanCenter.TVC.ARDRRegression	0.8154	0.1839	0.6461	0.1662
standardScaler.TVC.PLSRegression	0.8240	0.1426	0.6394	0.1501
none.TVC.Ridge	0.8729	0.1279	0.6100	0.0979
meanCenter.TVC.Ridge	0.8729	0.1279	0.6100	0.0979
meanCenter.TVC.SVR	0.8760	0.1350	0.6078	0.0991
none.TVC.SVR	0.8761	0.1348	0.6076	0.0990
standardScaler.TVC.PCRRegression	0.9651	0.1524	0.5245	0.1206
standardScaler.TVC.LinearRegression	0.9931	0.2025	0.4840	0.1994
meanCenter.TVC.PCRRegression	1.0123	0.1315	0.4779	0.1044
none.TVC.PCRRegression	1.0123	0.1315	0.4779	0.1044

Table 3-9: Best performing configurations of ARD, Bayesian Ridge and SVM Regression following hyperparameter tuning using the VideometerLab (MSI) aerobic, vacuum and combined condition training partitions, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 3 repeats of 5 fold cross-validation.

Model	Mean RMSE	Std.Dev. RMSE	Selected Parameters
Combined			
(StandardScaler) SVR	0.6316	0.1053	kernel: linear, C: 1
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.6322	0.1067	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.6456	0.1172	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	0.6316	0.1053	kernel: rbf, gamma: 1, C: 50
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.6384	0.1080	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.6755	0.1079	alpha_2: 1e-06, alpha_1: 1e-06
Aerobic			
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.7604	0.1319	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler) SVR	0.7667	0.1175	kernel: linear, C: 1
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.7702	0.1174	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.7176	0.1206	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	0.7238	0.1130	kernel: rbf, gamma: 1, C: 50
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.7899	0.1113	alpha_2: 1e-06, alpha_1: 1e-06
Vacuum			
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.5288	0.1092	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler) SVR	0.5525	0.1084	kernel: linear, C: 1
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.5740	0.1300	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.5244	0.0990	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	0.5284	0.0776	kernel: rbf, gamma: 1, C: 50
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.5799	0.1200	alpha_2: 1e-06, alpha_1: 1e-06

Table 3-10: Best performing configurations of ARD, Bayesian Ridge and SVM Regression following hyperparameter tuning using the FTIR aerobic, vacuum and combined condition training partitions, when sorted by RMSE performance in Total Viable Count prediction (log CFU/g), with 3 repeats of 5 fold cross-validation.

Model	Mean RMSE	Std.Dev. RMSE	Selected Parameters
Combined			
(StandardScaler) SVR	0.7759	0.1274	kernel: linear, C: 1
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.8010	0.1063	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.8040	0.1310	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.8055	0.1094	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.8193	0.1228	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	0.9409	0.1036	kernel: linear, C: 5
Aerobic			
(StandardScaler) SVR	0.8910	0.1522	kernel: linear, C: 1
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.8924	0.1523	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.9837	0.1558	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.8851	0.1450	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.9483	0.1731	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	1.1103	0.1573	kernel: linear, C: 5
Vacuum			
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.6259	0.0947	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.6310	0.0800	alpha_2: 1e-06, alpha_1: 1e-06
(StandardScaler) SVR	0.6740	0.1533	kernel: linear, C: 0.1
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
ARDRegression	0.6206	0.0919	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring)			max_iter: 1000, lambda_2: 1e-06, lambda_1: 1e-06,
BayesianRidge	0.6616	0.0736	alpha_2: 1e-06, alpha_1: 1e-06
(MeanCentring) SVR	0.7819	0.0827	kernel: linear, C: 5

3.5.3 Case Study 2 – Chicken Spoilage Model Development and Prediction

In order to demonstrate the new machine learning model construction functionalities implemented in the platform, it was used to replicate a Partial Least Square (PLS) machine learning pipeline implemented in prior research, with the data from the preceding case study. The aerobic and vacuum packaged meat training data from the batch-validation strategy was used to train the PLS decision-fusion model shown here. This model type was shown to demonstrate the intuitive nature of configuration of pipelines supporting sensor fusion.

Figure 3-25 shows the pipeline configuration page, after addition of two sensor entities to the platform, specifying the configurations of the VideometerLab and JASCO FTIR-6200 sensor devices used in the preceding DiTECT pilot study research. The peak selection functionality has been used to select the 2000-900cm⁻¹ range from the FTIR data, as before. The user has already configured each sensor with the RESPONSE variable set to correspond to the “TVC (Total Viable Count)” column in the tabular data. The “Chicken Aerobic_and_Vacuum Training Data” datasets have been confirmed by the platform as matching the VideometerLab and FTIR_6200 sensor entities’ specifications respectively.

[Add Sensor +](#)

1. Videometer Mean/...

Preprocessing

1. MSI Chicken Aerobic_and_Vacuum Trz

Select Outliers

Preprocessing

snv

No Parameters to Modify

+

2. JASCO_FTIR_6200

Preprocessing

1. FTIR Chicken Aerobic_and_Vacuum Trz

Select Outliers

Preprocessing

snv

No Parameters to Modify

savgol

Window Size *

11

Window Size for Calculation of Savinsky-Golay.

Polynomial Order *

The order of the polynomial function to use with Filtering

+

Figure 3-25: The pipeline configuration page, showing the MSI and FTIR sensor configurations. Visual feedback notifies the user to misconfigured aspects of the pipeline, in this case the lack of specified polynomial order in the Savinsky-Golay Preprocessing step.

Upon configuration and submission of the pipeline, the user is transferred to the Pipeline Run Log overview page (the same page shown in Figure 3-21), this allows users to download the completely configured pipeline (top-left icon in left pane) and cue new runs of the pipeline (“Start New Run” button on top row of second pane). A completed run can be seen in the Logs section of the second pane. By clicking the inspection button (magnifying class icon on log entity menu), users are transferred to the run log output viewer as shown in Figure 3-26. From this view, users can inspect the output files from a pipeline. By selecting the serialised model produced as output from the machine learning model process, the user can configure a machine learning endpoint and “host” this model for predictions.

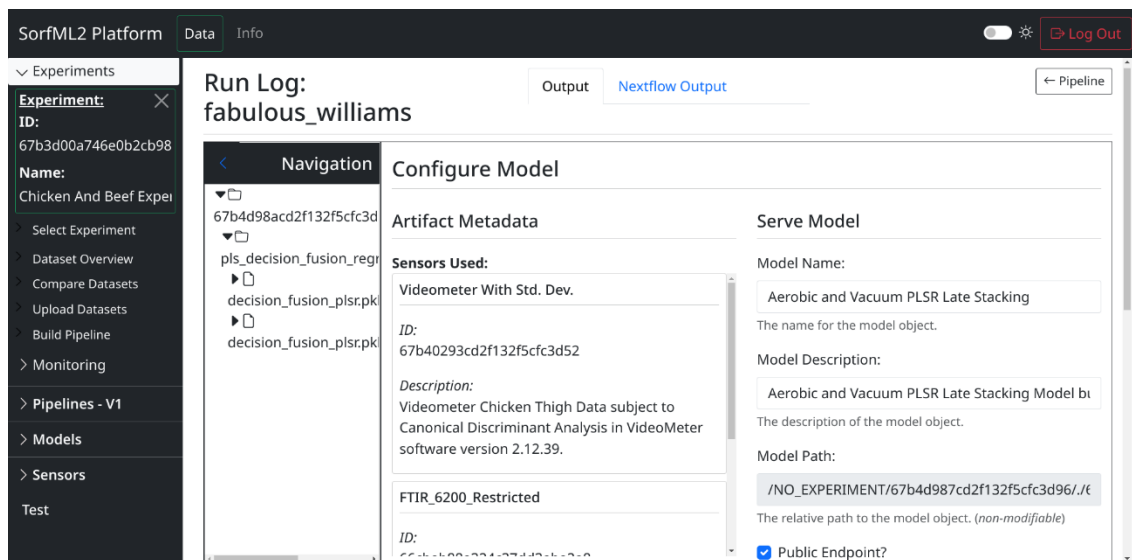


Figure 3-26: Selection of a serialised model object in the File Explorer UI. This allows for easy configuration of a machine learning endpoint on the platform.

Figure 3-27 shows the inference GUI. From this menu, the user can select data with which to request predictions, by dragging it into the left pane, or by using the file browser button. As the user “matches” the inference data to the intended sensor concepts, the middle pane will update with models registered on the platform that fit the current sensor criteria. The final pane shows information about the currently selected model in the middle pane. Clicking the “Predict” button will open the results page for all currently selected models and allow users to save specific predictions using the button in the top-right corner of every prediction summary.

Once a prediction is requested by the user, the platform’s API forwards the request to the appropriate inference service, receives a response, and returns it to the user, who is presented with the inference results for every compatible model selected: in this example two models are selected – the combined aerobic and vacuum condition late fusion model, and equivalent MSI model, however only one prediction result is visible in the captured results view (Figure 3-28).

Observed values can be supplied in the interface, to allow metrics regarding individual model performance to be displayed interactively in the browser to allow expedient comparison of test performance. These metrics are eligible for saving

on the platform, for the purposes of tracking of model performance across a deployment.

Figure 3-29 shows the model “auditing” functionalities built into the web application – regression-applicable metrics are calculated from the reported results and the supplied observed data displayed to the user. By selecting the “plot metrics” tab in Figure 3-30, users can then interactively compare this model performance in a graphical fashion.

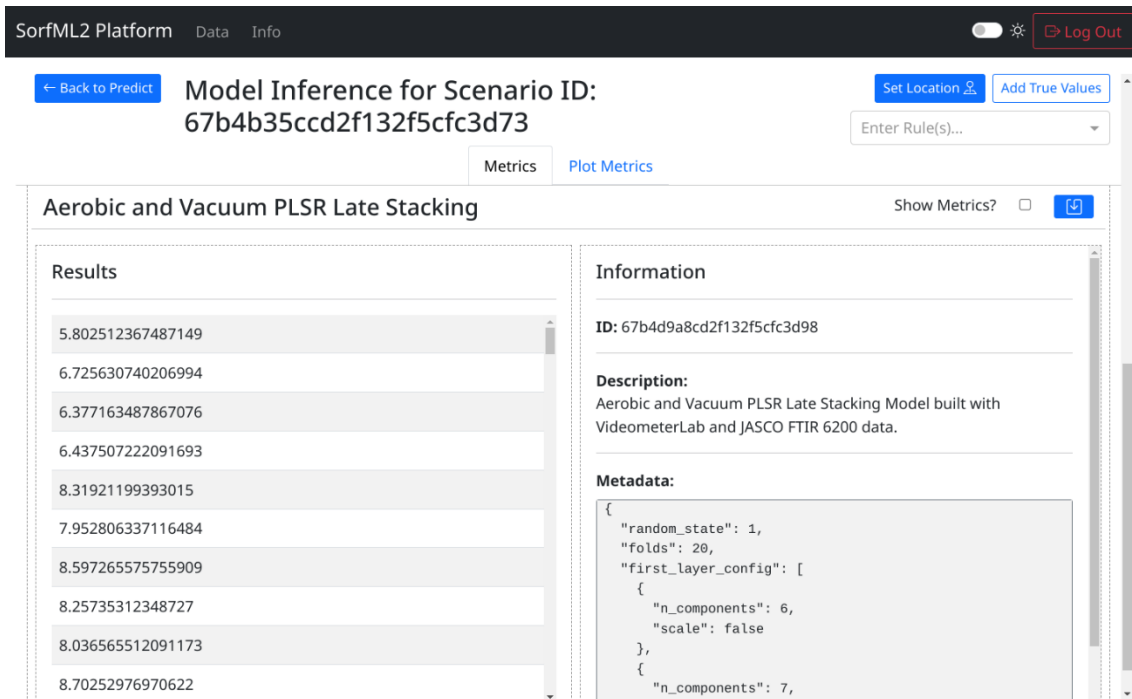


Figure 3-28: The results page, after clicking the “Predict” button in the inference GUI, containing a per-model result for every row supplied in the inference data.

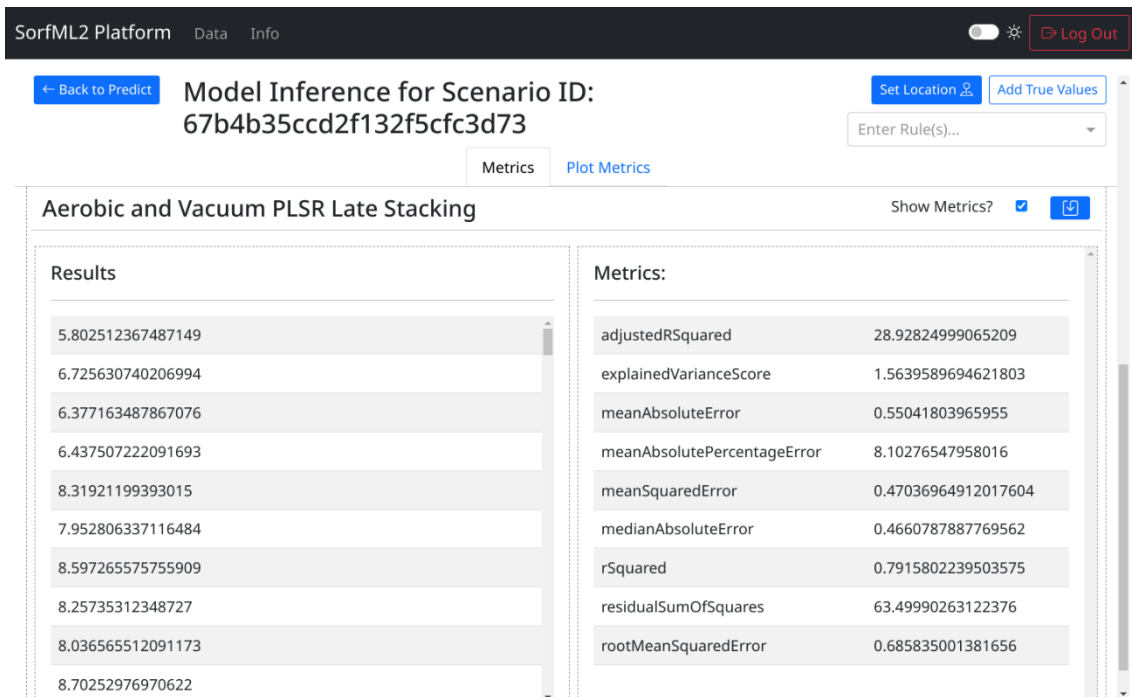


Figure 3-29: With observed values supplied in the interface (top right) the platform acts in a model “auditing” capacity, allowing users to compare the performance of many models simultaneously.

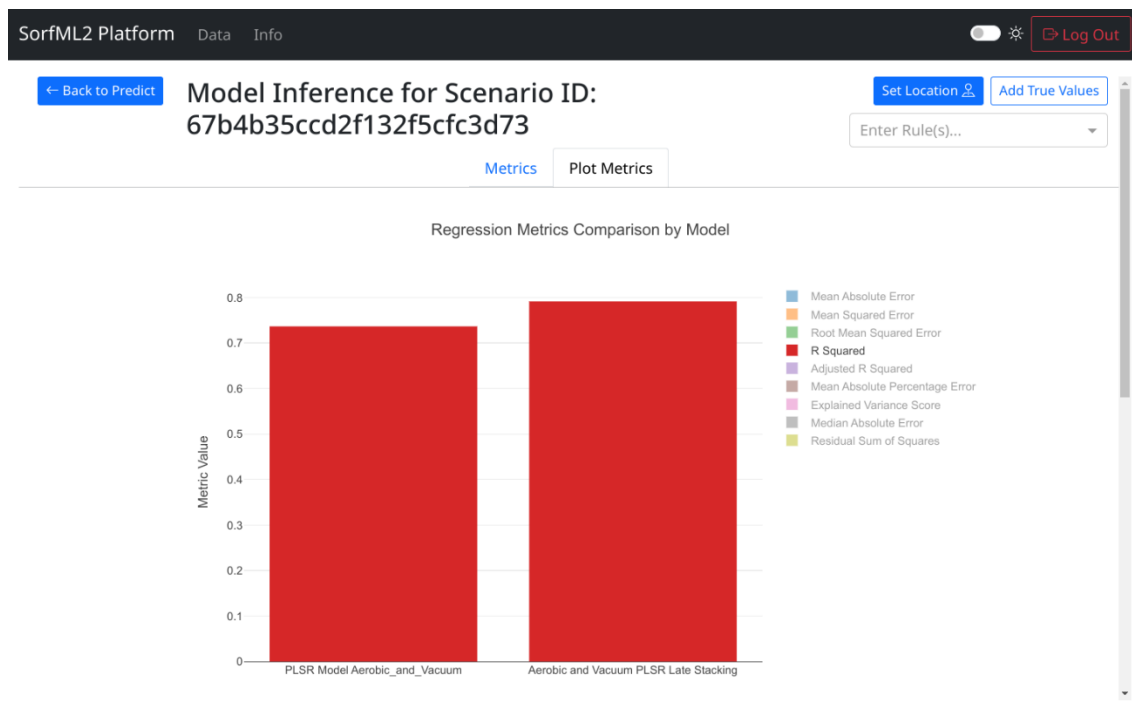


Figure 3-30: With observed values supplied in the interface (top right), selecting the “plot metrics” tab allows for interactive graphical comparison of model performance, including bar plots and observed/predicted plots

The models were tested with their corresponding test batch in an identical manner to that described in the original research using the inference interface in Figure 3-27. As expected, the test results from models configured using the new platform were identical to those configured manually. These results can be found in Table 3-11 and Table 3-12.

Table 3-11: Results from the employment of four different PLS regression derived machine learning approaches in the new platform, for prediction of Total Viable Count values using aerobic or vacuum packaged chicken spoilage data from either MSI (VideoMeter) and/or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72)

Sensor	RMSEp	R2
MSI Aerobic	0.6147	0.8604
FTIR Aerobic	1.6038	0.0498
MSI-FTIR Aerobic Late Fusion	0.5465	0.8897
MSI Vacuum	0.7933	0.6312
FTIR Vacuum	1.2924	0.0212
MSI-FTIR Vacuum Late Fusion	0.6973	0.7151

Table 3-12: Results from the employment of four different PLS regression derived machine learning approaches in the new platform, for prediction of Total Viable Count values using the combined aerobic and vacuum packaged chicken spoilage training data from either MSI (VideoMeter) and/or FTIR sensors. (Combined aerobic and vacuum test batch size: 138)

Sensor	RMSE	R2
MSI Aerobic and Vacuum	0.7727	0.7355
FTIR Aerobic and Vacuum	1.0796	0.4837
MSI-FTIR Aerobic and Vacuum Late Fusion	0.6873	0.7907

3.5.4 Case Study 3 – Evaluation of Bayesian Framework and Support Vector Machine Models for Chicken Spoilage Prediction

The PLS-derived chicken spoilage prediction models were reproduced using the machine learning provisioning and inference functionalities displayed in Section 3.5.3: Case Study 2 – Chicken Spoilage Model Development and Prediction, trained on the appropriate batch of aerobically- or vacuum-packaged chicken thigh samples, or a combination of both. These were either single-sensor PLS regression models or stacking late-fusion models, trained on data from both sensor types. The data received the same preparatory treatment as detailed in Section 3.5.1: Materials and Methods.

The nested cross-validation results from Section 3.5.2: Case Study 1 – Chicken Spoilage Model Benchmarking and Ranking showed that Bayesian Ridge, Automatic Relevance Determination (ARD) and Support Vector Machine (SVM) models produced marginally superior or comparable performance to PLS regression when applied to the combined aerobic and vacuum scenario. Follow-up hyperparameter tuning with the platform was conducted in order to confirm applicability to all investigated scenarios. With the tuned configurations determined in this process, out-of-batch testing was considered the final step in their evaluation, due to the strong degree of statistical validity offered through testing with a separate experimental batch.

To achieve this objective, models were constructed using the model builder interface shown previously; Figure 3-31 shows the model builder user interface when selecting all three algorithm types used in this study.

The test results for the Bayesian Ridge, ARD and SVM regression single- and dual-packaging condition models with standard scaling and mean centring are given in Table 3-13/Table 3-15 and Table 3-14/Table 3-16 respectively.

3.5.4.1 Results

The Videometer-derived Bayesian Ridge models achieved results similar in performance to the equivalent optimised PLS regression models described in Table 3-11 and Table 3-12, however this performance did not carry over to the FTIR derived models, with the mean-centred Bayesian Ridge models marginally missing the performance set by the associated PLS models.

Although both ARD and SVM performed well in the initial nested cross validation and hyperparameter optimisation procedure, the test batch results were not satisfactory: although the ARD models with standard scaling proved roughly on-par with the PLS and Bayesian Ridge equivalents for the MSI scenarios, the FTIR performance was inferior to PLS and more unstable than Bayesian Ridge when standard scaling was omitted, resulting in extreme results in the case of the vacuum condition, where an RMSEp of 2.82 log CFU/g was observed, more than twice that of the mean-centred vacuum-condition Bayesian Ridge model.

This unstable behaviour was also exhibited by the FTIR vacuum-condition SVM model, with a RMSEp of 2.32 log CFU/g. Performance across the other FTIR and the MSI test batches was also inferior to PLS to a lesser extent, with the exception of the MSI vacuum-condition model, which was able to marginally improve upon the PLS results, with an RMSEp of 0.75 log CFU/g.

Table 3-13: Results from the employment of Bayesian Ridge, ARD and SVM Regression with Z-score normalisation, for prediction of Total Viable Count (log CFU/g) values, using aerobic or vacuum packaged chicken thigh datasets from MSI (VideoMeter) or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72)

Model Algorithm	Sensor	RMSE	R2
(StandardScaler) Bayesian Ridge	MSI Aerobic	0.5851	0.8735
	FTIR Aerobic	1.6802	-0.0429
	MSI Vacuum	0.7585	0.6628
	FTIR Vacuum	1.7443	-0.7829
(StandardScaler) Automatic Relevance Determination (ARD)	MSI Aerobic	0.6226	0.8568
	FTIR Aerobic	1.7861	-0.1786
	MSI Vacuum	0.7892	0.6350
	FTIR Vacuum	1.5239	-0.3608
(StandardScaler) Support Vector Machine Regression (SVR)	MSI Aerobic	0.7129	0.8123
	FTIR Aerobic	1.8119	-0.2129
	MSI Vacuum	0.7608	0.6609
	FTIR Vacuum	2.3156	-2.1422

Table 3-14: Results from the employment of Bayesian Ridge, ARD and SVM Regression with mean centring, for prediction of Total Viable Count (log CFU/g) values, using aerobic or vacuum packaged chicken thigh datasets from MSI (VideoMeter) or FTIR sensors. (Aerobic Test Batch Size: 66, Vacuum test batch size: 72)

Model Algorithm	Sensor	RMSE	R2
(Mean Centring) Bayesian Ridge	MSI Aerobic	0.5669	0.8813
	FTIR Aerobic	1.6460	-0.0001
	MSI Vacuum	0.7568	0.6644
	FTIR Vacuum	1.3255	-0.0295
(Mean Centring) Automatic Relevance Determination (ARD)	MSI Aerobic	0.7790	0.7758
	FTIR Aerobic	1.6587	-0.0165
	MSI Vacuum	0.8759	0.5505
	FTIR Vacuum	2.8158	-3.6461
(Mean Centring) Support Vector Machine Regression (SVR)	MSI Aerobic	0.5756	0.8776
	FTIR Aerobic	1.3910	0.2852
	MSI Vacuum	0.8017	0.6233
	FTIR Vacuum	1.2597	0.0701

Table 3-15: Results from the employment of Bayesian Ridge, ARD and SVM Regression with Z-score normalisation, for prediction of Total Viable Count (log CFU/g) values, using combined aerobic and vacuum packaged chicken thigh datasets from either MSI (VideoMeter) or FTIR sensors. (Combined aerobic and vacuum test batch size: 138)

Model Algorithm	Sensor	RMSE	R2
(StandardScaler) Bayesian Ridge	MSI Aerobic and Vacuum	0.7827	0.7286
	FTIR Aerobic and Vacuum	1.2109	0.3504
(StandardScaler) Automatic Relevance Determination (ARD)	MSI Aerobic and Vacuum	0.7537	0.7483
	FTIR Aerobic and Vacuum	1.3043	0.2464
(StandardScaler) Support Vector Machine Regression (SVR)	MSI Aerobic and Vacuum	0.8169	0.7044
	FTIR Aerobic and Vacuum	1.1436	0.4207

Table 3-16: Results from the employment of Bayesian Ridge, ARD and SVM Regression with mean centring, for prediction of Total Viable Count (log CFU/g) values, using combined aerobic and vacuum packaged chicken thigh datasets from either MSI (VideoMeter) or FTIR sensors. (Combined aerobic and vacuum test batch size: 138)

Model Algorithm	Sensor	RMSE	R2
(Mean Centring) Bayesian Ridge	MSI Aerobic and Vacuum	0.7589	0.7449
	FTIR Aerobic and Vacuum	1.2091	0.3524
(Mean Centring) Automatic Relevance Determination (ARD)	MSI Aerobic and Vacuum	0.7541	0.7481
	FTIR Aerobic and Vacuum	1.3581	0.1829
(Mean Centring) Support Vector Machine Regression (SVR)	MSI Aerobic and Vacuum	0.7470	0.7528
	FTIR Aerobic and Vacuum	1.0568	0.5053

3.5.4.2 Discussing Model Selection and Performance

The consistent performance of Bayesian Ridge in the cross-validation assessment conducted on the new platform in Section 3.5.2 suggested that its extended examination may have offered an alternative algorithm to partial least squares regression for the analysis of chicken spoilage sensor data, at least for the purposes of MSI (Videometer) data analysis. PLS methods are widely utilised in the cheminformatics field due to their aptitude for handling highly collinear data, as one might expect to emerge from spectroscopic and imaging analyses.

Bayesian Ridge is an extension of conventional L2-regularised linear (Ridge) regression, which uses a probabilistic framework for the estimation of uncertainty in the model's parameters. It uses a shrinkage term to perform regularisation of the model, to reduce the likelihood of the model overfitting and to reduce the impact of highly collinear features in the data. The hyperparameters in the model are treated as random variables, with the model performing internal optimisation of these in an iterative process when passed initialising values. These values are used to inform the model of prior belief about the shape and scale (rate) of the prior distribution of precision of the coefficients (α - the regularisation term used in the model) and the noise (λ - the uncertainty in the data assumed by the model). Adequate fitting of both is required to produce a model that does not exhibit traits of over- or under-fitting the training data.

The result is of interest, as Ridge and Bayesian Ridge regression models have not seen heavy utilisation for the analysis of spectral and sensor data in the cheminformatics/food analysis space; other methods are regularly shown to offer sufficient performance when dealing with highly collinear or dimensional data. Michel *et al.* claim that through the identical regularisation of all features of the dataset, Bayesian Ridge regression isn't suitable for analyses where only a few features are relevant (as might be seen with highly collinear data), and highlights the strengths of other methods, such as Automatic Relevance Determination (ARD) Regression, a similarly derived method that instead assumes priors for α and λ from an elliptical Gaussian rather than a spherical Gaussian distribution,

which can impose sparsity in the coefficients (weights) that are calculated (Michel et al., 2011).

However, the evaluation of ARD performance using the SorfML2 platform in this study has shown it to produce mixed results across the combinations of sensors and conditions/scenarios explored here. In the case of performance comparison between ARD and Bayesian Ridge, this is likely due to the probabilistic assumptions made: Bayesian Ridge assumes a global prior over all regression coefficients, leading to an equal regularisation weighting, whereas ARD applies independent priors to each of the regression coefficients. This drives the feature selection effect, allowing some feature contributions to be driven towards zero, effectively removing them from the final model. However, it has been shown that ARD can be prone to under-fitting when the model contains too many features (Michel et al., 2011; Qi et al., 2004). In the case of the FTIR scenarios explored in this work, the high dimensionality, coupled with the relatively low number of observations may be responsible for ARD's poor generalisation performance with the test data.

Furthermore, the choice of data preprocessing has been shown to have considerable impact on the final predictive accuracy of the resultant models. The more complex regularisation strategy employed in ARD when applied to unstandardised data likely resulted in more aggressive pruning of feature contributions and instability in the final model when compared with the global strategy in Bayesian Ridge, which is more likely to distribute the impact of the regularisation process across all features. Application of standardisation (Z-score normalisation) did appear to remove the occurrence of unstable performance in ARD, but in the case of Bayesian Ridge did appear to marginally increase RMSEp values. This may arise from the nature of the spectral datasets being explored: the necessity of column-wise standardisation procedures is debatable for such data, as the spectral values are not independent, and are considered to be in the same scale. As such row wise normalisation is typically performed (as is in this work with the Standard Normal Variate procedure). By performing a further column-wise standardisation, low-contribution variables are given greater

importance, and the high-contribution (e.g. peak/trough) variable importance is reduced accordingly. This may have led to increased noise in the resultant data, and a more complex optimisation task for the Bayesian methods.

The best results out of the three model architectures were mostly observed with Support Vector Machines: mean centring appeared sufficient to allow SVM regression to perform well in the scenarios explored, its results in some cases were superior to both the equivalent Bayesian Ridge and single-sensor PLS models, although missing the performance of the PLS decision fusion approaches. Similar to Bayesian Ridge, the absence of column-wise standardisation appeared to affect the kernel choice and regularisation process during grid search. Further optimisation of the performance of both Bayesian Ridge and SVM may be obtainable by including a preliminary feature transformation step in the feature reduction pipeline; previous research has shown that incorporating dimensionality reduction methodologies such as Principal Component Analysis (PCA) or PLS is effective when paired with SVM for the estimation of microbial quality of pork (Fengou et al., 2020).

Other algorithms, such as LASSO (Least Absolute Shrinkage and Selection Operator) may display satisfactory performance for this type of data, due to their ability to shrink some features' contributions to zero, by employing strong L1 regularisation. This acts as an inbuilt feature selection strategy, and can simplify resultant models. This is in contrast to other regularised methods such as Ridge, that uniformly modify the coefficients (weights) for the features, thereby not performing a feature elimination step. Higashi *et al.* highlighted potential applications for the algorithm as a feature selection step in a pipeline when compared with other regression methods; for their study, the efficacy of LASSO (group LASSO and sparse group LASSO) derived methods were evaluated on fish and pork datasets and compared with PLS and Ridge methods, with the outcome showing improvements in both sparsity and in prediction accuracy for the fish dataset, and marginally inferior predictive accuracy for the pork dataset, but simplification of the overall model for negligible performance cost, decreasing prediction times (Higashi, ElMasry and Nakauchi, 2016). Alshejari *et al.* also

incorporated LASSO into their ensemble-based feature selection and neuro-fuzzy regression framework, alongside six other feature selection algorithms, for the prediction of TVC in beef fillet samples, using a similar procedure with multispectral imaging data (using the same type of VideometerLab sensor as in this research). In their testing, their new framework outperformed established models such as PLS regression, support vector machine (SVM) and multi-layer perceptron (MLP) methodologies (Alshejari, Kodogiannis and Leonidis, 2023). With that said, the testing of their models could be argued to have relatively weak biological and statistical validity, due to the small testing partitions employed, and the apparent lack of cross-batch testing.

Another algorithm that attempts to bridge the gap between the L1-regularised LASSO and the L2-regularised Ridge regression is Elastic Net. This utilises both types of regularisation, modifying its behaviour when selecting features. An issue that can arise with LASSO derives from the fact that it can shrink features' contributions to zero in isolation, without having impact on surrounding variables. The incorporation of both types of regularisation in Elastic Net can lead to shrinkage of contributions from groups of correlated variables (Zou and Hastie, 2005), which is an obvious advantage for analysis of low rank data, and where the number of predictors greatly exceeds the number of observations, otherwise known as a small n , large p problem, such as that arising from FTIR analysis, as seen in this study. In their methodology to detect adulteration of wheat flour with talcum powder using near-infrared (NIR) spectroscopy, Du *et al.* employed a method of feature selection in which either Elastic Net, or a combination of Elastic Net and Genetic Algorithm (GA) were employed prior to passing the selected features into a Gradient Boosting Decision Tree (GBDT). For the Elastic Net feature selection, the results showed a negligible effect on the prediction accuracy (R^2 of 0.9770 versus 0.9724 and 0.9778 respectively), however the feature selection simplified the resultant models, lowering the number of predictors used from 1050 without selection, to 155 with EN selection, or 55 with EN and GA selection (Du *et al.*, 2022).

The performance of SVM regression and the apparent deviation of Bayesian Ridge regression performance from published observations on its use in cheminformatic analysis is evidenced by the extensive cross validation assessments conducted in Section 3.5.2. The hyperparameter grid used for the nested cross-validation can be found in Table 3-5 . The cross-batch test results further corroborate these findings in a biologically independent test set, adding further statistical validity.

3.5.4.3 Case Study Conclusion

In the single-storage condition scenario for MSI and FTIR, the Bayesian Ridge and Support Vector Machine regression algorithms were able to produce predictions on par with the optimised PLS regression models built in Section 3.5.3. The comparative simplicity of the Bayesian Ridge model, versus the comparative complexity of SVM and the combined dimensionality reduction and predictive steps taken in PLS regression may offset the marginal performance difference seen in the single-storage-condition models and the dual-storage-condition MSI model. The extensive cross-validation and full model testing facilitated with the new platform offered insight into the performance of the algorithm for multiple storage conditions and sensor technologies, against other algorithms that are more well characterised in the cheminformatics and food spoilage prediction domains, such as LASSO and Elastic Net.

3.6 Discussion

3.6.1 Comparison With Existing Platforms

The platform developed as part of this project is considered a domain-specific machine learning tool for rapid inspection of high-throughput sensor data, construction of data processing and machine learning pipelines, and service of machine learning models for inference queries, either through the associated graphical user interface, or through the API.

Few other platforms fit this domain specific space: SorfML, also developed at Cranfield University, is an entry into the sensor analysis space, allowing for rapid prototyping of machine learning methods in order to rank the most successfully performing examples. However, the SorfML platform is superseded with the additional developments made in this project, particularly with regard to the key aims set out as part of the originating DiTECT project.

The first of these is in the design of the new platform's data repository, designed to address the shortcomings of the original SorfML system, particularly in the data upload process. The old system's data repository consisted of a simple key-value store, implemented in the MongoDB system, which stored data objects, either representing experimental or sensor data, as functionally opaque entities in their respective collections. As a consequence of this, very little machine interoperability was exercised with the uploaded data, as little information was stored about data structure alongside primary data. This design impacted the extensibility of the original SorfML platform, thus making it, and any derivatives, unsuitable candidates for inclusion in the envisioned DiTECT integrated food safety management system (iFSMS) architecture.

The new platform enforces the integrity of any data uploaded to the platform on multiple levels. The first of these is through the implementation of data concepts not represented in the SorfML system. Via the concept of inheritance, sensor data uploaded to the new platform is required to be a "descendant" of a sensor device. These sensor entities are configured on the platform to represent *specific* instances of sensor devices utilised in the collection of the data. This information

is crucial to allow the platform to deduce which sensor datasets are directly comparable, and which are representative of completely different concepts, even if the dimensional properties of the data matrices appear similar. This “type checking” during upload of data and inference tasks solves an issue found in SorfML when employing it in the inference mode after model construction, where minute changes in the structure of data, or meanings behind certain concepts, could impart unforeseen consequences on any predictions made, and drastically decrease their validity. Further to this, experiment entities are now used in a more concrete fashion, to represent distinct, self-contained bodies of research. These experiment entities can be used to store further information about the specific experimental conditions, or production and distribution conditions encountered in a real-world scenario, and allow users to partition datasets based on these metadata. Knowledge of experimental conditions and their effects on collected sensor data in the model training phase is key to assessment of the re-applicability/transferability of machine learning models in new experimental situations.

The second key development in the new data repository structure is the addition of a secondary “FAIRification” facility to the platform. FAIR data concepts were not enforced in the SorfML platform, instead, the onus was placed on the user to exercise good data handling procedure – any deviation from the intended mode of operation often introduced unforeseen issues that had to be resolved with administrator access to the platform. The issue with extensibility of SorfML caused by the functionally opaque database entity representation structure was compounded by the lack of ontological reinforcement of the concepts found within the database. This lack of commonality between data and metadata uploaded by different users of the platform severely limited the prospect of interoperability and data-sharing between users. Whilst not necessarily an issue when considering the original purpose of the SorfML platform – a small-scale web application to allow for on-demand sensor data analysis, the overall state of the agri-food data sharing ecosystem may suggest that further developments in the domain must provide provisioning for FAIR data, or its FAIRification. Kumar *et. al.* conducted an extensive review of the FAIRness state of data in projects funded by

Horizon2020, and found that at the time of review (2024), less than 10% of multi-actor-projects were employing FAIR practices in the conduct of their research (Kumar et al., 2024). Of particular interest is the state of data generation, software, and tool development; between each of the explored “findable”, “accessible”, “interoperable” and “reusable” categories, each of the three showed differing combinations of exemplary and average implementation statuses.

The new platform is equipped with a data FAIRification facility, through the use of services that examine uploaded sensor specifications and datasets and annotate them with concepts found within the FSO ontology developed as part of the DiTECT project (Yilmaz, 2023). This extension to the interoperability and re-usability of the data hosted on the platform allows for additional functionality to be added that exploits the data in a more domain-agnostic fashion and lays the groundwork for future data integration, unification, analytics and sharing facilities to be supported by the platform. Boyles *et. al.* give insight into the utility of ontologies in research (in the author’s case, in the field of toxicology where data is often siloed and requires significant curation), particularly in the integration of knowledge across research groups and the wider scientific community. By providing a structured and shared framework for knowledge exchange, data previously considered heterogeneous by consideration of vocabularies and other metadata are more likely integrable across domains (Boyles et al., 2019). The paper’s authors also argue that wider adoption of ontologies within the scientific community will make the rapidly growing body of data-driven scientific research more navigable and analysable to individuals and research groups with differing research objectives and expertise, thus driving novel insights, envisioning a future holistic knowledge graph that will encompass concepts from across the entire biomedical domain.

The new analysis platform enables SPARQL (SPARQL Protocol and RDF Query Language) querying functionalities that can be used to conduct advanced querying upon the data within the knowledge graph, and map it to outside knowledge in other ontologies and knowledge graphs. Galgonek and Vondrášek demonstrate the utility of ontological querying within the context of a

bioinformatics/cheminformatics project (Galgonek and Vondrášek, 2021). They were able to demonstrate within their research that by implementing a SPARQL endpoint, they could aggregate queries across PubChem, ChEMBL and ChEBI small-molecule datasets into the so-called Integrated Database of Small Molecules (IDSM) that could be used for querying compound substructure and conducting similarity searches in concert with the other components of their project; they note that the functionality, while publicly available via an endpoint, is intended for programmatic access by other services, mentioning federated querying, wherein clients making use of a single SPARQL endpoint can query other SPARQL/RDF endpoints.

Interactive querying features developed as part of this project further to the described SPARQL functionalities are currently at a nascent stage. These features further streamline the process of making ontological queries by greatly simplifying the process when compared to SPARQL, by offering autocompletion of datatype and nesting object properties for a top level resource identifier type, and allow for direct integration within the web application, without having to expose endpoints for general public use. This solves one of the limitations of semantic, context-dependent querying: the user being required to have at least a working understanding of the querying language in use i.e., SPARQL in this example. In this new approach, the user can employ contextual querying to request information about datasets, sensors and other entities on the platform, in an intuitive way that requires no foreknowledge of how to construct SPARQL queries.

The third major feature of the new platform is the manner in which machine learning and data processing objectives are met. In contrast to SorfML, the new platform has incorporated a number of features that facilitate reproducible, reliable and portable pipelining of data science. A large part of the data handling orchestration is passed over to handler services running in the platform network, that provide a working environment for Nextflow, a Domain Specific Language (DSL) that enables construction and execution of workflows. Nextflow's inherent flexibility allows the user the choice of execution environment for their respective

data science tasks, by breaking workflows down into modular “processes”, containing the actual data processing tool or element code snippet, which are assignable to differing execution environments, including Docker containers, Python/Conda environments, or on-demand infrastructure. The majority of the processes and pipelines constructed in the course of developing this new platform were set to run in Docker containers, configured with all of the relevant tools and programming languages for their relevant tasks; as an example, execution of Scikit-Learn backed pipelines on the platform were conducted with a Docker image configured with Python 3.11.7 and Scikit-Learn 1.3.0, (available at <https://docker.io/snheffer/sorfml-conda-image:250101> for the case studies shown). Nextflow was chosen because of its good degree of community support, in addition to its large body of documentation. Recent literature suggests that Nextflow is one of the more popular DSLs in the bioinformatics sphere, alongside Snakemake, Common Workflow Language and Workflow Description Language (Wratten, Wilm and Göke, 2021).

In the support of modular and reproducible workflows, the new platform has more in common with domain-agnostic tools available to the bioinformatician and data scientist, such as the aforementioned Galaxy and Arvados platforms. However, the new system bridges the gap between these more generic pipeline management systems, and domain specific data handling functionalities offered by SorfML. The modularity afforded by the new platform allows for the construction of pipelines that implement SorfML’s hard-coded model benchmarking and ranking functionalities, and supersedes it with its greater support for deployment of models for inference tasks.

3.6.2 Limitations of Current Implementation

In the design of the new platform, it was determined that a balance had to be struck between platform complexity (and thus configurability), and usability, in the same way that SorfML abstracts many of the particulars of the machine learning design process away from the user to facilitate an intuitive experience. As a consequence, the modular pipeline design process currently only supports the service of Python-derived model architectures in the inference platform. This is

mainly due to time constraints limiting implementation of a server system capable of serving alternative model configurations in other languages, such as R – it is not a limitation imposed by design decisions.

Additionally, the platform's file handling could face hardware and web browser limitations when dealing with raw data files exceeding sizes supported by the main API (in the multi-gigabyte range), upon upload. While not an issue for the encountered tabular data, this could become a problem when dealing with the types of files emerging from such technologies as Hyper-Spectral Imaging (HSI), where raw captured data files can exceed several gigabytes in size.

There is currently no support for proprietary sensor formats upon upload of data, and requires intermediary conversion of these types into generic or well-supported tabular formats for input into the platform, such as comma-separated values (CSV) or Microsoft XLS(X) spreadsheet files. This is often achievable through the use of software that accompanies an installation of such a sensor device.

3.6.3 Future Work

There are several avenues that may be explored to further improve the platform's utility, beyond what has been described and demonstrated in this document.

These would first be to address the shortcomings described in the previous section. By supporting more programming languages such as R in the model hosting and inference system, a greater pool of algorithms and methodologies could theoretically be accessed by the user of the platform. This would require the programming of a new inference service to handle these types of serialised models, or extension of the existing one to function in a more “universal” fashion. This may be achieved by devising a new containerisation strategy for deployed models – for example, by allowing the user to generate model inference Docker images, with an inbuilt inference API, they could be downloaded and employed in locations outside of the platform's environment, greatly increasing the portability and utility of predictive models in production settings.

Improving the platforms handling of large files, and including support for proprietary formats would reduce the amount of “context switching” required from the operator of the platform, by removing the need to convert data into more open formats. This may not be supportable in the browser, due to the limitations placed on conventional HTTP operations and may require the development of alternative channels to facilitate data upload. This may be achieved through the development of a command-line interface (CLI) application to facilitate uploads through technologies such as FTP, in a similar fashion to that supported by Galaxy and Arvados.

3.7 Conclusion

The new platform in its current state supports multiple use-cases that meet the original objectives of the PhD project, in summary, the provisioning of a modular, resilient data repository for the storage of high-throughput sensor data and an accompanying well-defined API supporting it, in addition to the machine learning and data processing functionalities offered by the new web application, the functionality of which is demonstrated in this document through the case studies, showing both machine learning model construction/inference and machine learning model ranking/benchmarking in a similar fashion to SorfML. The functionality of SorfML has been superseded in most, if not all regards by the new platform, through the development of modular pipelines that can replicate the hard-coded functionalities of the prior system. The modular design of the platform itself offers a solid foundation for extension with new services and integration with extended or collaborative developments with other partners towards a new, more well integrated data landscape in the agri-food community.

The new platform offers novelty in the form of support for complex machine learning workflows, including concepts such as sensor/data fusion, interactive inspection of pipeline outputs, and the provisioning of machine learning models on the platform for later deployment, bridging the gap between domain-specific decision support tools and more complex, but generic bioinformatics workflow management systems.

References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A. and Blankenberg, D. (2018) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update', *Nucleic Acids Research*, 46(Web Server issue), pp. W537–W544.
- Alshejari, A., Kodogiannis, V.S. and Leonidis, S. (2023) 'Combining Feature Selection Techniques and Neurofuzzy Systems for the Prediction of Total Viable Counts in Beef Fillets Using Multispectral Imaging', *Sensors*, 23(23) Multidisciplinary Digital Publishing Institute, p. 9451.
- Amstutz, P., Bértoli, J., Cesar, N., Clegg, T., Di Pentima, L., Kutyla, D., Li, J., Smith, S., Vandewege, W., Wait Zaranek, A., Wait Zaranek, S. and Knox, L. (2024) *Arvados*. Available at: 10.5281/zenodo.12694166 (Accessed: 23 August 2024).
- Baker, M. (2016) '1,500 scientists lift the lid on reproducibility', *Nature*, 533(7604) Nature Publishing Group, pp. 452–454.
- Baranyi, J. and Tamplin, M.L. (2004) 'ComBase: A Common Database on Microbial Responses to Food Environments', *Journal of Food Protection*, 67(9), pp. 1967–1971.
- Boyles, R.R., Thessen, A.E., Waldrop, A. and Haendel, M.A. (2019) 'Ontology-based data integration for advancing toxicological knowledge', *Current Opinion in Toxicology*, 16, pp. 67–74.
- Cawley, G.C. and Talbot, N. (2010) *On over-fitting in model selection and subsequent selection bias in performance evaluation - University of East Anglia*. Available at: <https://research-portal.uea.ac.uk/en/publications/on-over-fitting-in-model-selection-and-subsequent-selection-bias-> (Accessed: 24 January 2025).
- Crusoe, M.R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., Ménager, H., Soiland-Reyes, S., Gavrilović, B., Goble, C. and Community, T.C. (2022) 'Methods included: standardizing computational reuse and portability with the Common Workflow Language', *Communications of the ACM*, 65(6), pp. 54–63.
- Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. and Notredame, C. (2017) 'Nextflow enables reproducible computational workflows', *Nature Biotechnology*, 35(4) Nature Publishing Group, pp. 316–319.
- Dietrich, O. (2018) *FRESHDETECT – food safety – fast and reliable | FRESHDETECT Project | Fact Sheet | H2020., CORDIS | European Commission* Available at: <https://cordis.europa.eu/project/id/822732> (Accessed: 11 December 2023).

Du, C., Sun, L., Bai, H., Zhao, Z., Li, X. and Gai, Z. (2022) 'Quantitative detection of talcum powder in wheat flour based on near-infrared spectroscopy and hybrid feature selection', *Infrared Physics & Technology*, 123, p. 104185.

Emmerich, C.H. and Harris, C.M. (2020) 'Minimum Information and Quality Standards for Conducting, Reporting, and Organizing In Vitro Research', in Bespalov, A., Michel, M. C. and Steckler, T. (eds.) *Good Research Practice in Non-Clinical Pharmacology and Biomedicine*. Handbook of Experimental Pharmacology. Cham: Springer International Publishing, pp. 177–196. Available at: 10.1007/164_2019_284 (Accessed: 10 October 2022).

Estelles-Lopez, L., Ropodi, A., Pavlidis, D., Fotopoulou, J., Gkousari, C., Peyrodie, A., Panagou, E., Nychas, G.-J. and Mohareb, F. (2017) 'An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling', *Food Research International*, 99, pp. 206–215.

Fengou, L.-C., Lytou, A.E., Tsekos, G., Tsakanikas, P. and Nychas, G.-J.E. (2024) 'Features in visible and Fourier transform infrared spectra confronting aspects of meat quality and fraud', *Food Chemistry*, 440, p. 138184.

Fengou, L.-C., Mporas, I., Spyrelli, E., Lianou, A. and Nychas, G.-J. (2020) 'Estimation of the Microbiological Quality of Meat Using Rapid and Non-Invasive Spectroscopic Sensors', *IEEE Access*, 8, pp. 106614–106628.

Fengou, L.-C., Spyrelli, E., Lianou, A., Tsakanikas, P., Panagou, E.Z. and Nychas, G.-J.E. (2019) 'Estimation of Minced Pork Microbiological Spoilage through Fourier Transform Infrared and Visible Spectroscopy and Multispectral Vision Technology', *Foods*, 8(7), p. 238.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A. and Mueller, L.A. (2015) 'The Sol Genomics Network (SGN)—from genotype to phenotype to breeding', *Nucleic Acids Research*, 43(D1), pp. D1036–D1041.

Fielding, R.T. (2000) *Architectural styles and the design of network-based software architectures*. phd. University of California, Irvine.

galaxyproject.org (n.d.) *Galaxy ToolShed*. Available at: <https://galaxyproject.org/toolshed/#philosophy> (Accessed: 24 August 2024).

Galgonek, J. and Vondrášek, J. (2021) 'IDSM ChemWebRDF: SPARQLing small-molecule datasets', *Journal of Cheminformatics*, 13(1), p. 38.

Hershberger, J., Morales, N., Simoes, C.C., Ellerbrock, B., Bauchet, G., Mueller, L.A. and Gore, M.A. (2021) 'Making waves in Breedbase: An integrated spectral data storage and analysis pipeline for plant breeding programs', *The Plant Phenome Journal*, 4(1), p. e20012.

Higashi, H., ElMasry, G.M. and Nakauchi, S. (2016) 'Sparse regression for selecting fluorescence wavelengths for accurate prediction of food properties', *Chemometrics and Intelligent Laboratory Systems*, 154, pp. 29–37.

Hwang, K.-B., Lee, I.-H., Li, H., Won, D.-G., Hernandez-Ferrer, C., Negron, J.A. and Kong, S.W. (2019) 'Comparative analysis of whole-genome sequencing pipelines to minimize false negative findings', *Scientific Reports*, 9(1) Nature Publishing Group, p. 3219.

Jentsch, M., Schneider-Lunitz, V., Taron, U., Braun, M., Ishaque, N., Wagener, H., Conrad, C. and Twardziok, S. (2024) 'Creating cloud platforms for supporting FAIR data management in biomedical research projects.', *F1000Research*, 13, p. 8.

Kong, S.W., Lee, I.-H., Liu, X., Hirschhorn, J.N. and Mandl, K.D. (2018) 'Measuring coverage and accuracy of whole-exome sequencing in clinical context', *Genetics in Medicine*, 20(12) Elsevier, pp. 1617–1626.

Kumar, P., Hendriks, T., Panoutsopoulos, H. and Brewster, C. (2024) 'Investigating FAIR data principles compliance in horizon 2020 funded Agri-food and rural development multi-actor projects', *Agricultural Systems*, 214, p. 103822.

Malzacher, S., Meißner, D., Range, J., Findrik Blažević, Z., Rosenthal, K., Woodley, J.M., Wohlgemuth, R., Wied, P., Nidetzky, B., Giessmann, R.T., Prakinec, K., Chaiyen, P., Bommarius, A.S., Rohwer, J.M., de Souza, R.O.M.A., Halling, P.J., Pleiss, J., Kettner, C. and Rother, D. (2024) 'The STRENDa Biocatalysis Guidelines for cataloguing metadata', *Nature Catalysis*, 7(12) Nature Publishing Group, pp. 1245–1249.

Michel, V., Eger, E., Keribin, C. and Thirion, B. (2011) 'Multiclass Sparse Bayesian Regression for fMRI-Based Prediction', *International Journal of Biomedical Imaging*, 2011, p. 350838.

Nychas, G.-J., Sims, E., Tsakanikas, P. and Mohareb, F. (2021) 'Data Science in the Food Industry', *Annual Review of Biomedical Data Science*, 4(1), pp. 341–367.

Orchard, S., Al-Lazikani, B., Bryant, S., Clark, D., Calder, E., Dix, I., Engkvist, O., Forster, M., Gaulton, A., Gilson, M., Glen, R., Grigorov, M., Hammond-Kosack, K., Harland, L., Hopkins, A., Larminie, C., Lynch, N., Mann, R.K., Murray-Rust, P., Lo Piparo, E., Southan, C., Steinbeck, C., Wishart, D., Hermjakob, H., Overington, J. and Thornton, J. (2011) 'Minimum information about a bioactive entity (MIABE)', *Nature Reviews Drug Discovery*, 10(9) Nature Publishing Group, pp. 661–669.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011) 'Scikit-learn:

Machine Learning in Python', *Journal of Machine Learning Research*, 12(85), pp. 2825–2830.

Pizzuti, T. and Mirabelli, G. (2013) 'FTTO: An example of Food Ontology for traceability purpose', *2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*., Vol.01, pp. 281–286.

Qi, Y. (Alan), Minka, T.P., Picard, R.W. and Ghahramani, Z. (2004) 'Predictive automatic relevance determination by expectation propagation', *Twenty-first international conference on Machine learning - ICML '04*. Banff, Alberta, Canada: ACM Press, p. 85. Available at: 10.1145/1015330.1015418 (Accessed: 6 March 2025).

Sheffield, N.C., Bonazzi, V.R., Bourne, P.E., Burdett, T., Clark, T., Grossman, R.L., Spjuth, O. and Yates, A.D. (2022) 'From biomedical cloud platforms to microservices: next steps in FAIR data and analysis', *Scientific Data*, 9, p. 553.

Shrestha, R., Matteis, L., Skofic, M., Portugal, A., McLaren, G., Hyman, G. and Arnaud, E. (2012) 'Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice', *Frontiers in Physiology*, 3 Frontiers Available at: 10.3389/fphys.2012.00326 (Accessed: 18 August 2024).

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. and Katz, S. (2004) 'Reengineering thesauri for new applications: The AGROVOC example', *Journal of Digital Information*, 4

Swainston, N., Baici, A., Bakker, B.M., Cornish-Bowden, A., Fitzpatrick, P.F., Halling, P., Leyh, T.S., O'Donovan, C., Raushel, F.M., Reschel, U., Rohwer, J.M., Schnell, S., Schomburg, D., Tipton, K.F., Tsai, M.-D., Westerhoff, H.V., Wittig, U., Wohlgemuth, R. and Kettner, C. (2018) 'STRENDAB: enabling the validation and sharing of enzyme kinetics data.', *The FEBS journal*, 285(12) England, pp. 2193–2204.

Taylor, K., Griffith, C., Lefort, L., Gaire, R., Compton, M., Wark, T., Lamb, D., Falzon, G. and Trotter, M. (2013) 'Farming the Web of Things', *IEEE Intelligent Systems*, 28(6), pp. 12–19.

The Galaxy Community (2024) 'The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update', *Nucleic Acids Research*, 52(W1), pp. W83–W94.

UN DESA (2017) *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100* | UN DESA | United Nations Department of Economic and Social Affairs. Available at: <https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html> (Accessed: 15 August 2022).

da Veiga Leprevost, F., Grüning, B.A., Alves Aflitos, S., Röst, H.L., Uszkoreit, J., Barsnes, H., Vaudel, M., Moreno, P., Gatto, L., Weber, J., Bai, M., Jimenez, R.C., Sachsenberg, T., Pfeuffer, J., Vera Alvarez, R., Griss, J., Nesvizhskii, A.I. and Perez-Riverol, Y. (2017) 'BioContainers: an open-source and community-driven framework for software standardization', *Bioinformatics*, 33(16), pp. 2580–2582.

Videometer, Denmark (2021) *VideometerLab - multispectral imaging technology*. Available at: <https://videometer.com/videometerlab/> (Accessed: 18 March 2024).

Vivian, J., Rao, A.A., Nothhaft, F.A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A.D., Musselman-Brown, A., Schmidt, H., Amstutz, P., Craft, B., Goldman, M., Rosenbloom, K., Cline, M., O'Connor, B., Hanna, M., Birger, C., Kent, W.J., Patterson, D.A., Joseph, A.D., Zhu, J., Zaranek, S., Getz, G., Haussler, D. and Paten, B. (2017) 'Toil enables reproducible, open source, big biomedical data analyses', *Nature biotechnology*, 35(4), pp. 314–316.

Voss, K., Auwera, G.V. der and Gentry, J. (2017) 'Full-stack genomics pipelining with GATK4 + WDL + Cromwell', *F1000Research*, 6 Available at: 10.7490/f1000research.1114634.1 (Accessed: 24 August 2024).

Wainer, J. and Cawley, G. (2017) 'Empirical Evaluation of Resampling Procedures for Optimising SVM Hyperparameters', *Journal of Machine Learning Research*, 18(15), pp. 1–35.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3(1) Nature Publishing Group, p. 160018.

World Wide Web Consortium (2004) *OWL Web Ontology Language Reference*. Available at: <https://www.w3.org/TR/owl-ref/> (Accessed: 12 October 2022).

Wratten, L., Wilm, A. and Göke, J. (2021) 'Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers', *Nature Methods*, 18(10) Nature Publishing Group, pp. 1161–1168.

Yilmaz, A. (2023) 'FSO: Food Safety Monitoring Ontology', *14th International Conference On Semantic Web Applications And Tools For Health Care And Life Sciences (Swat4Hcls 2023)*, Basel, Switzerland, Vol. 3415, pp. 42–51.

Zou, H. and Hastie, T. (2005) 'Regularization and Variable Selection Via the Elastic Net', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), pp. 301–320.

4 Discussion and Conclusions

The overall motivation for this work was the desire to facilitate transparent, reliable monitoring practices in the food supply chain, through the implementation and utilisation of state-of-the-art sensor technologies, coupled with an equivalent degree of informatics support through the development of a platform capable of ingesting and processing raw data and producing output suitable for supporting decisions regarding the authenticity, quality and safety of food products moving throughout the production and distribution pipeline. The preceding chapters covered the two major themes covered in this project, the work in both largely progressing simultaneously owing to their synergistic nature.

The first component of this work was the development of predictive mathematical models to detect and quantify food characteristics. High throughput and minimally invasive sensor devices were utilised to produce input data for Machine Learning pipelines for the prediction of microbial load, a common indicator of food spoilage. Sensor monitoring of food samples was conducted with two main optical methodologies: Fourier-Transform Infra-red Spectroscopy (FTIR) and Multispectral Imaging (MSI). The pipelines were validated using two separate raw food types, representing two commonly consumed meat products in respective modes of preparation: raw chicken thigh and raw beef mince. The results from these Machine Learning pipelines were compared to results gathered in a conventional, lab-based manner, i.e. bacterial colony counting. These pipelines simultaneously evaluated the performance of each sensor modality in an individual context, while also exploring the potential of data/sensor fusion to enhance the predictive power of predictive models built with access to their combined observational space. Increased accuracy, sensor redundancy and reliability of predictions are often considered as potential benefits of utilising a fusion workflow (Hall and Llinas, 1997). Therefore, investigation of these practices in food monitoring was warranted given the significant potential consequences that inaccurate food metric predictions could have for consumers, food business operators, and food and agricultural authorities.

The second major component of this work was formed by the development of a bioinformatics/cheminformatics platform suitable for ingestion, manipulation and presentation of data gathered from sensor devices representative of a modernised food monitoring framework, including those devices encountered in the preceding component (FTIR and MSI). The platform was designed to enable development of analytical workflows that support the FAIR principles in science (Findability, Accessibility, Interoperability, and Reproducibility (Goble et al., 2020)) through the enforcement of data consistency and typing with ontological backing from previous semantic research in the agrifood domain (Yilmaz, 2023), and support for modular, extensible tools with containerisation allowing for easily-disseminated, reproducible workflows that can be executed on a variety of supported hardware.

Validation of the platform was considered to be achievable through evaluation of its functionalities that enabled sensor data analysis. Multiple case studies were used to demonstrate the various aspects of a typical food analysis workflow, from upload and annotation of concepts within sensor data files, to exploratory data analysis, model benchmarking and evaluation, and construction of complete machine learning models for inference tasks. The manually-conducted food spoilage research described previously was successfully reproduced in the platform, including the development and deployment of fusion models representative of the published results.

4.1 Meat Spoilage Detection with Minimally Invasive Sensors

4.1.1 Interpretation of Results

Section 2 of this thesis explored the application of high throughput sensor technologies to the task of meat spoilage quantification: Fourier-transform Infrared (FTIR) spectroscopy and Multispectral Imaging were employed for the quantification of the surface deterioration of chicken thigh and beef mince samples, which has been shown to be indicative of the action of endogenous enzymes and spoilage-causing and potentially pathogenic microbes (Alexandrakis, Downey and Scannell, 2012; Liu et al., 2003; Panagou et al., 2014).

The sensor fusion methodologies applied in this work were shown in most cases to improve predictive performance when compared to equivalent single-sensor approaches, however the improvement was particularly apparent when fusion techniques were applied to the chicken spoilage scenario. The strong performance of single-sensor MSI models, and consequently of feature-level and decision-level fusion methods, may partly be due to the nature of the meat samples. While the beef mince was produced in a macerating fashion that disrupted the previously homogeneous morphology of the muscle and fatty tissue, the chicken thigh samples were formed of contiguous tissue, and may have given the MSI and FTIR sensors an even optical basis for their respective modes of operation.

The improvement in performance seen in the chicken spoilage scenario through fusion indicates that MSI and FTIR were performing in adequately complementary observational spaces, where the two sensors were capturing a greater variety of chemical/physical spoilage indicators, and thus generating a richer training dataset. The addition of the fluorescence monitoring data (MSIF) did not appear to offer a significant benefit to the performance of the fusion pipelines - the common VideometerLab sensor device (Videometer, Denmark, 2021) operating in different modes (MSI and MSIF) may not have offered a significant degree of complementarity, potentially due to their close spectral ranges (270-405nm for MSIF, 405-970 for MSI) and potential collinearity of observations.

Diversity is a concept often discussed in the interrelated context of machine learning fusion approaches and stacked generalisation ensemble methods, which employ some method of combining and ideally surpassing the predictive utility of multiple individual predictive “expert models”. Diversity is crucial because combining predictions or features from expert models that consistently agree is unlikely to substantially improve overall performance. The strength of such a pipeline typically arises from the ability for the “meta-layer” of the model to gauge the level of agreement or disagreement between experts; in the case of a stacked generalisation ensemble, this agreement/disagreement bias information is the

relevant information passed through the model to make final predictions (Wolpert, 1992). In stacked generalisation, the diversity is typically supplied by employing a heterogeneous library of expert model algorithms – even poorly performing algorithms are useful in this process, and are potentially ideal when expert predictions are independent or negatively correlated with each other (Zhang and Ma, 2012).

It follows that sensor devices demonstrated to perform poorly in isolation may perform better in concert with other sensor devices, due to the complementarity of their respective observational spaces, errors and resultant predictive models. This may have contributed to the superior performance of the feature and decision fusion pipelines explored in the work of Chapter 2, where individual sensor performance was imbalanced, as seen with the chicken scenario and the consistent superior performance of MSI. This is especially relevant to the decision fusion approach employed, as it may be considered a special case of stacked generalisation, where the homogeneity of the individual sensor model's algorithm choice of Partial Least Squares was offset with the heterogeneity of the input features. This concept of generation of diversity through training expert models on individual subsets of training data has been discussed in the context of ensemble models in (Zhou, 2012).

A consistent trend in the work of Chapter 2 was the consistent under-performance of the early fusion methodologies. This is likely due in part to the nature of the data combined in this type of pipeline: the large disparity in dimensionality between MSI and FTIR (36 variables vs ~1200 variables) has been considered likely to impact the overall efficacy of the resultant machine learning model, as FTIR's larger data matrix effectively "swamps" the contributions made by MSI. This phenomenon has been discussed in a generalised sense by Smolinska *et al.* in their review of the application of the three main archetypes of sensor/data fusion (early-, mid- and late fusion) to tasks within the life sciences (Smolinska *et al.*, 2019). Fengou *et al.* employed a type of early fusion which was replicated as the early fusion approach in the work of Chapter 2, wherein the FTIR and MSI variables were z-score normalised before being concatenated into a single

dataset (Fengou et al., 2019). This fusion method generally produced minor improvements in performance in the nested cross-validation approach for the chicken data, and generally on-par performance to the single-sensor approaches in beef, whereas no benefit was found in either meat type's batch validation approach. The normalisation method was employed at the variable level, rather than the MSI/FTIR data block level, which likely provided no counter to the dominating effect of FTIR, as the meaningful ratio of the variance between any two variables inside each MSI or FTIR block was not preserved (Smolinska et al., 2019).

While the strong variation in the beef mince batches negatively impacted the respective batch validation results, the results gained from the alternative repeated cross-validation sampling approach did indicate that employment of the feature- and decision-level fusion of MSI, FTIR and MSIF data did offer an improvement over the single sensor approaches.

In the chicken scenario's repeated nested cross-validation testing, the benefits of employing a fusion methodology are most apparent in the merged aerobic and vacuum-storage condition, where all three fusion types were able to produce improved generalised (mean) performance over the single-sensor equivalent models. This was combined with a visibly smaller envelope of standard deviation in RMSE_{cv} performance from the repeated nested cross validation procedure relative to the single-sensor models. This smaller envelope, particularly visible with the mid-level fusion approaches, was another indicator of improved generalisation relative to the single sensor models – while MSI had shown good performance in the aerobic condition, when applied to a scenario containing both aerobic *and* vacuum data, the resultant PLS model appeared less able to adequately characterise the compounded optical properties emerging from different storage conditions in addition to the effects of meat spoilage.

The validation of late- and mid-level approaches for meat spoilage regression tasks via microbial count quantification is significant, as little published research has explored fusion methodologies in this context; Cheng *et al.* explored the equivalent of early, mid and late fusion approaches for the purposes of quality

estimation via the determination of pork moisture content, determining that a Partial Least Squares Regression (PLSR) - based decision fusion approach offered improved accuracy over individual sensor models built with hyperspectral imaging (HSI) and electronic nose (E-nose) data (Cheng et al., 2022). Robert *et al.* presented a similar three-tier comparison of fusion approaches for Raman and FTIR data in the estimation of beef, lamb and venison quality using pH and percentage of intra-muscular fat (%IMF) as metrics, determining that despite previous under-exploration of late fusion in food analysis, it provided superior results to single-sensor, early- and mid-level fusion approaches for pH estimation (Robert et al., 2021). Interestingly, their early fusion approach delivered the best performance for estimation of %IMF, despite commonalities to the early fusion approaches employed in this work. This was likely due to the similarity in dimensionality between the Raman and FTIR matrices used (314 and 399 variables respectively).

4.1.2 Practical Considerations for Sensor Fusion Deployment

4.1.2.1 Real-time computation

While the sensor fusion methodologies applied in the course of Chapter 2 were shown to offer improved predictive power over equivalent single-sensor alternatives in a majority of cases, a clear enabling factor in the utilisation of such techniques in real-world, “production” scenarios is the ability to adequately host and deploy fusion pipelines for on-demand inference tasks. Such pipelines must incorporate appropriate data pre-processing, matrix concatenation and feature extraction steps prior to feeding the inference data into the respective fusion models. Modern data analysis and machine learning programming libraries have moved towards supporting this type of behaviour, such as with Python’s Sci-kit Learn, which supports construction of machine learning model pipelines incorporating transformers, elements that transform data inputs, feeding into classifiers and regressors which are used to make predictions. Similarly for neural network development, libraries such as PyTorch, Tensorflow and the interface library Keras support development and serialisation of model types for transferability and storage, leading to deployment in the case of Tensorflow’s

serving module. However, these solutions are not immediately useful for the fusion architectures explored in this work or the typical food scientist, as they typically require bespoke arrangements combining multiple trained sensor-specific models or separate feature extraction steps prior to their integration into the final fusion layer. The complexity of mid- and late fusion methodologies may require proven deployment pipelines implementing established methodologies, intermediate data aggregation, and logic to handle outputs from calibrated sensors. This necessitates domain-specific software solutions or frameworks, perhaps employing containerisation technologies such as Docker to ensure consistency, reproducibility, and modular deployment of fusion pipeline configurations across varied computational environments.

High computational overhead may be expected from some sensor technologies that may complicate real-time inference tasks - in this case FTIR presented the most complex data to process due to its dimensionality, however some imaging sensors employed in food research can produce data outputs exceeding Gigabytes (GB) in size, for example Hyperspectral Imaging, which records spectral and spatial data over a wide band of wavelengths. Practical deployment of such sensor fusion systems may need optimisation strategies including dimensionality reduction and efficient data handling, such as by offloading tasks to on-demand computing with high-performance computers (HPCs), or edge-computing solutions capable of localised data processing. Localised approaches could also include federated or “distributed” learning, which allows for machine learning model training across decentralised resources. Each sensor-equipped location can locally train models using their data, and then collaboratively aggregate model parameters without transmitting raw data. This approach has been explored in the health sector in concepts such as the European “Personal Health Train”, for promotion of decentralised analysis of health data (Bonino da Silva Santos et al., 2023). In this case, the transmission of patient diagnostic information, such as medical imaging data, to a central location for processing would have severe privacy and security implications, aside from the network overhead imposed by offloading all data processing to a centralised location. If the same practices were applied to the processing of food monitoring data, many

of the same benefits would apply: much of the data processing overhead would be removed from a centralised inference endpoint, while enhancing privacy and security across international food chains, as raw food monitoring data would not necessarily have to leave the jurisdiction of food business operators (FBOs) and national food authorities, while the benefits of well-generalised predictive models would be shared amongst all partners in the federated project (Durrant et al., 2022).

4.1.2.2 Sensor Calibration and Standardisation

As described in the preceding research chapters, the efficacy of fusion pipelines is heavily dependent on the quantity, quality and processing of training data: while some issues in this regard have been shown to affect single-sensor pipelines, such as models not generalising well across differences in comparatively small batches of biological data (Spyrelli et al., 2021), issues facing long-term deployment of machine learning models are likely to be compounded in the context of sensor fusion, due to the additional complexity of supporting multiple sensor modalities. If not correctly accounted for, variations in sensor calibration (i.e. sensor drift) may propagate through predictive models, potentially causing deterioration in fusion pipeline performance. Maintaining sensor-specific calibration procedures is essential for preserving the utility of fusion models. This is likely dependent on the types of sensors used and the context-dependent specific recalibration routines that must be employed.

Further to this, implementation of machine learning pipelines built with experimental data will invariably encode information about the scientific techniques employed and the operational environments in which the data was collected, and this will impact the specific behaviour of the resultant model. The additional metadata must also be accounted for and stored alongside the model, and may include information on sample preparation procedure, sensor configuration information, or environmental factors such as temperature and humidity. Standardised metadata capture regarding environmental and operational conditions, coupled with appropriate quality control and model auditing functionality may improve model robustness under conditions of sensor

drift or environmental variability, and simplify knowledge exchange in the scientific community.

Additionally, providing an interface to allow end-users of a supported decision support framework to query stored models using standardised search parameters will enhance the applicability of data fusion in food control. Users could select models specifically calibrated to their operational contexts or experimental conditions, such as matching sample types and sensor calibration settings, enhancing predictive reliability.

4.2 Development of the SorfML2 Platform

In Chapter 3, an online data-analysis platform (SorfML2) was developed, specifically designed to address practical challenges identified in the sensor fusion research outlined in Chapter 2. This platform enables streamlined, reproducible analysis and real-time inference with high-throughput sensor data, with particular focus on support for complex analytical pipelines. By adopting containerisation (Docker), workflow management tools (Nextflow), and applying FAIR (Findable, Accessible, Interoperable, Reusable) data principles, the platform ensures robust and standardised analytical processes, addressing previously identified concerns in Section 4.1.2: Practical Considerations for Sensor Fusion Deployment around data consistency, dimensionality disparity, and operational complexity.

This was a multi-stage process that was initiated with the evaluation of existing bioinformatics platforms, and the identification of a technological gap in which subsequent development efforts would be aimed to fill: in the context of sensor-enabled food monitoring and data management, it was determined that no sufficient representation of a free software agri-food-specific platform existed that supported typical informatics-enabled workflows, from data ingestion and exploratory data analysis to machine learning model evaluation and model deployment/inferencing tasks. Previous developments, such as Cranfield University's Symbiosis Online Research Framework (SorfML) platform had been shown to be of great utility for rapid evaluation of machine learning approaches for working with high-throughput, minimally invasive sensor data in the context of

food research (Estelles-Lopez et al., 2017; Spyrelli et al., 2021), however its underlying data handling approach and architecture did not lend itself to extension with additional functionalities, such as support for exploration of the aforementioned sensor fusion methodologies.

The developed web platform provides users the capability to dynamically construct and execute custom analysis pipelines, thereby supporting practical deployment of complex fusion methodologies. It facilitates sensor-specific preprocessing, model training, and inference pipelines, acting to address many of the sensor calibration variability and standardisation needs raised earlier. Moreover, the adoption of proven software architectures, such as the Spring framework and support for documented API-based inference querying enhances reliability and accessibility, and may in future support the automated operation of sensors deployed in-, at- or on-line in a Process Analytical Technology (PAT) enabled food monitoring strategy.

Practical validation through benchmarking case studies demonstrates the platform's effectiveness in accurately handling diverse datasets and sensor modalities, confirming its potential as a decision-support tool within food safety monitoring contexts. While limitations were acknowledged in Chapter 3, most were related to limitations in development time: future work could further improve usability, scalability, and integration with domain specific libraries and frameworks for supporting increasingly complex analysis tasks, the modular nature of the platform's "business logic" facilitating this process.

Ultimately, the development and implementation of SorfML2 significantly advances the practical applicability of sensor-driven approaches in food analysis, supporting platform users between initial research and real-world deployment of predictive models. Compared to existing analytical platforms within food analysis and sensor fusion contexts, SorfML2 combines dynamic pipeline construction, real-time inference capabilities, and reproducibility through containerisation and workflow management, significantly streamlining the process of deploying complex predictive models.

4.3 Recommendations for Future Work

As the work in the preceding chapters and recently published research has shown, the application of high throughput, non- or minimally-destructive sensor devices to food monitoring tasks is a viable addition to the arsenal of testing methodologies available to agri-food operatives for the purposes of quality, authenticity and safety analyses. Such techniques offer a rapid means to quickly assess the state of food samples at an individual level in production and distribution, rather than simply relying on batch-to-batch testing. The work in this thesis has simultaneously contributed to the exploration of methodologies to enhance the predictive power of such sensors, through the comparative use of sensor fusion techniques, and also to facilitate the use of such techniques in real-world scenarios, through the development of a web platform that supports high-throughput sensor data ingestion, transformation, and utilisation in predictive modelling applications. To build upon the findings of this work, areas identified for further study or validation are discussed below.

The first and most intuitive suggestion, as with any data-driven analytical project, is further validation of the methodologies explored in this work with additional foodstuffs and preparations. Such data may be representative of food moving through different stages of a production line. To use the example of meat products, this could include data from different stages of processing raw cuts of meat into mince or other preparations, before being packaged, or indeed further processed into (cooked) ready-to-eat products. A pertinent suggestion in this context is the exploration of additional sensor device types for possible integration into decision-support / sensor fusion pipelines. This would have the potential benefit of enhancing the observational space recorded about a food sample. This could include moving towards the addition of non-optical sensors such as electronic noses (E-nose), which are relatively low-cost devices that record the volatile aromas from samples in a manner emulating human sensory experience. For meat-based analyses, their predictive accuracy has been recorded variably from mixed to well-performing when compared to other sensor types in an isolated sense, (Estelles-Lopez et al., 2017; Han et al., 2020; Roy and Yadav, 2022), however in a suitably configured fusion pipeline, such sensors may serve

to increase the recorded coverage of sample state, adding redundancy and thus reliability to predictions.

Another avenue to improve the combined predictive power of non-invasive food sensors with a greater quantity of training data may be to explore and implement different algorithms. The previously discussed similarity of the employed late fusion method to stacked generalisation may suggest that addition of heterogeneity to the base-learner “expert” library may improve the predictive power of the approach. Such methods may include different combinations of dimensionality reduction and regression algorithms, such as Principal Component Analysis (PCA), Support Vector Machine (SVM) or Random Forest (RF). Further to this, greater focus on performing feature reduction or engineering with high-dimensional datasets may help to improve predictive utility in an independent and sensor-fusion context. Non-parametric deep learning methods like Convolutional Neural Networks (CNNs), may be able to capture non-linear interactions between modalities and perform a degree of automatic selection of relevant data features. Recurrent Neural Networks (RNNs) have also shown utility for forecasting purposes, when they have access to temporal patterns related to sample monitoring progression, and in compensating for sensor drift, which can affect sensors like the aforementioned E-nose (Zou and Lv, 2020).

Greater numbers of sensors included in multi-modal analyses will likely increase operational complexity, both in pipeline construction and utilisation. In the context of SorfML2, more robust methods for model deployment and maintenance may be required to enhance the long-term utility of inference endpoints. This may include greater support for model versioning, where older models are brought off-line to be replaced by newer iterations with updated parameters. An alternative strategy for model updates is through the concept of “online learning”, where models learn incrementally from new data, rather than in the conventional batch-based manner. This type of training method may be more applicable where model complexity makes repetitive batch learning unfeasible, but may have implications for model reliability and explainability if performance is not regularly audited (Riksrevisjonen, 2023). The platform may benefit from a way to better support

such efforts, and detect deviations in performance over time by keeping a record of benchmarks, aiding in detection of machine learning phenomena such as drift, where changes in statistical distributions in the population are not correctly reflected in the sample upon which models are trained (Sahiner et al., 2023).

5 Conclusion

Collectively, the meat spoilage research presented in Chapter 2, combined with the development of a web-based platform described in Chapter 3, simultaneously provides validation for employment of high-throughput monitoring technologies in food analysis, in both singular and multi-sensor approaches, and provides a framework for advancing sensor-driven analysis of food through the implementation of pipelines for data preprocessing, exploratory data analysis, and integration of inputs into machine learning pipelines, before deployment for inferencing tasks. The techniques shown in this work suggest promising results when applied to real-world food monitoring scenarios, however further work is required to validate the results of the techniques described here with data reflective of the real-world journey of food products through the supply chain. Additionally, access to new sensor modalities in integrative predictive pipelines will likely enhance the accuracy and reliability of predictions.

References

- Alexandrakis, D., Downey, G. and Scannell, A.G.M. (2012) 'Rapid Non-destructive Detection of Spoilage of Intact Chicken Breast Muscle Using Near-infrared and Fourier Transform Mid-infrared Spectroscopy and Multivariate Statistics', *Food and Bioprocess Technology*, 5(1), pp. 338–347.
- Alshejari, A., Kodogiannis, V.S. and Leonidis, S. (2023) 'Combining Feature Selection Techniques and Neurofuzzy Systems for the Prediction of Total Viable Counts in Beef Fillets Using Multispectral Imaging', *Sensors*, 23(23) Multidisciplinary Digital Publishing Institute, p. 9451.
- Bonino da Silva Santos, L.O., Ferreira Pires, L., Graciano Martinez, V., Rebelo Moreira, J.L. and Silva Souza Guizzardi, R. (2023) 'Personal Health Train Architecture with Dynamic Cloud Staging', *SN computer science*, 4(1), p. 14.
- Cheng, J., Sun, J., Yao, K., Xu, M., Tian, Y. and Dai, C. (2022) 'A decision fusion method based on hyperspectral imaging and electronic nose techniques for moisture content prediction in frozen-thawed pork', *LWT*, 165, p. 113778.
- Durrant, A., Markovic, M., Matthews, D., May, D., Enright, J. and Leontidis, G. (2022) 'The role of cross-silo federated learning in facilitating data sharing in the agri-food sector', *Computers and Electronics in Agriculture*, 193, p. 106648.
- Estelles-Lopez, L., Ropodi, A., Pavlidis, D., Fotopoulou, J., Gkousari, C., Peyrodie, A., Panagou, E., Nychas, G.-J. and Mohareb, F. (2017) 'An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling', *Food Research International*, 99, pp. 206–215.
- Fengou, L.-C., Spyrelli, E., Lianou, A., Tsakanikas, P., Panagou, E.Z. and Nychas, G.-J.E. (2019) 'Estimation of Minced Pork Microbiological Spoilage through Fourier Transform Infrared and Visible Spectroscopy and Multispectral Vision Technology', *Foods*, 8(7), p. 238.
- Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M.R., Peters, K. and Schober, D. (2020) 'FAIR Computational Workflows', *Data Intelligence*, 2(1–2), pp. 108–121.
- Hall, D.L. and Llinas, J. (1997) 'An introduction to multisensor data fusion', *Proceedings of the IEEE*, 85(1), pp. 6–23.
- Han, F., Huang, X., H. Aheto, J., Zhang, D. and Feng, F. (2020) 'Detection of Beef Adulterated with Pork Using a Low-Cost Electronic Nose Based on Colorimetric Sensors', *Foods*, 9(2) Multidisciplinary Digital Publishing Institute, p. 193.
- Liu, Y., Lyon, B.G., Windham, W.R., Realini, C.E., Pringle, T.D.D. and Duckett, S. (2003) 'Prediction of color, texture, and sensory characteristics of beef steaks

by visible and near infrared reflectance spectroscopy. A feasibility study', *Meat Science*, 65(3), pp. 1107–1115.

Panagou, E.Z., Papadopoulou, O., Carstensen, J.M. and Nychas, G.-J.E. (2014) 'Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage', *International Journal of Food Microbiology*, 174, pp. 1–11.

Riksrevisjonen, N. (2023) *Auditing machine learning algorithms - The Supreme Audit Institutions of Finland, Germany, the Netherlands, Norway and the UK*. Available at: <https://www.auditingalgorithms.net/> (Accessed: 25 March 2025).

Robert, C., Jessep, W., Sutton, J.J., Hicks, T.M., Loeffen, M., Farouk, M., Ward, J.F., Bain, W.E., Craigie, C.R., Fraser-Miller, S.J. and Gordon, K.C. (2021) 'Evaluating low- mid- and high-level fusion strategies for combining Raman and infrared spectroscopy for quality assessment of red meat', *Food Chemistry*, 361, p. 130154.

Roy, M. and Yadav, B.K. (2022) 'Electronic nose for detection of food adulteration: a review', *Journal of Food Science and Technology*, 59(3), pp. 846–858.

Sahiner, B., Chen, W., Samala, R.K. and Petrick, N. (2023) 'Data drift in medical machine learning: implications and potential remedies', *British Journal of Radiology*, 96(1150), p. 20220878.

Smolinska, A., Engel, J., Szymanska, E., Buydens, L. and Blanchet, L. (2019) 'General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences', in *Data Handling in Science and Technology*. Elsevier, pp. 51–79. Available at: 10.1016/B978-0-444-63984-4.00003-X (Accessed: 26 March 2023).

Spyrelli, E.D., Ozcan, O., Mohareb, F., Panagou, E.Z. and Nychas, G.-J.E. (2021) 'Spoilage assessment of chicken breast fillets by means of fourier transform infrared spectroscopy and multispectral image analysis', *Current Research in Food Science*, 4, pp. 121–131.

Videometer, Denmark (2021) *VideometerLab - multispectral imaging technology*. Available at: <https://videometer.com/videometerlab/> (Accessed: 18 March 2024).

Wolpert, D.H. (1992) 'Stacked generalization', *Neural Networks*, 5(2), pp. 241–259.

Yilmaz, A. (2023) 'FSO: Food Safety Monitoring Ontology', *14th International Conference On Semantic Web Applications And Tools For Health Care And Life Sciences (Swat4Hcls 2023)*, Basel, Switzerland, Vol. 3415, pp. 42–51.

Zhang, C. and Ma, Y. (2012) *Ensemble Machine Learning: Methods and Applications*. Springer New York.

Zhou, Z.H. (2012) *Ensemble Methods: Foundations and Algorithms*. CRC Press, Chapman & Hall/CRC Machine Learning & Pattern Recognition.

Zou, Y. and Lv, J. (2020) 'Using Recurrent Neural Network to Optimize Electronic Nose System with Dimensionality Reduction', *Electronics*, 9(12) Multidisciplinary Digital Publishing Institute, p. 2205.

Appendices

Ethical approval letter



26 March 2025

Dear Mr Heffer ,

Reference: CURES/24674/2025

Project ID: 15917

Title: Sensor Approaches for Non-Destructive Safety and Authentication Analysis of Food Products

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your project title change request, reference CURES/24674/2025 has been approved .

If you have any queries, please contact CURES Support.

Regards,

CURES Team

Appendix A:

SPARQL Functionalities

```
[
  {
    "value": [
      ""
    ],
    "_meta": {
      "valueRange": false,
      "valueRegex": false
    },
    "_nested": [
      {
        "value": [
          "10"
        ],
        "_meta": {
          "valueRange": false,
          "valueRegex": true
        },
        "_nested": null,
        "property": {
          "property": "https://purl.archive.org/purl/fsmon/Ontology#hasTemperature",
          "_type": "datatypeProperty",
          "range": "http://www.w3.org/2000/01/rdf-schema#Literal",
          "label": "hasTemperature"
        }
      }
    ],
    {
      "value": [
        "map"
      ],
      "_meta": {
        "valueRange": false,
        "valueRegex": true
      },
      "_nested": null,
      "property": {
        "property": "https://purl.archive.org/purl/fsmon/Ontology#hasPackageType",
        "_type": "datatypeProperty",
        "range": "http://www.w3.org/2000/01/rdf-schema#Literal",
        "label": "hasPackageType"
      }
    }
  ],
  {
    "value": [
      ""
    ],
    "_meta": {
      "valueRange": false,
      "valueRegex": false
    },
    "_nested": null,
    "property": {
      "property": "",
      "_type": "datatypeProperty",
      "range": "",
      "label": ""
    }
  }
],
  {
    "property": {
      "property": "http://www.w3.org/ns/sosa/hasSample",
      "_type": "objectProperty",
      "range": "http://www.w3.org/ns/sosa/Sample",
      "label": "hasSample"
    }
  }
]
```

A query generated with the Graphical Query Tool in the developed Web Application. In this example, the user has requested all datasets that match a temperature value of "10" and a packageType of "map" (Modified Atmosphere Packaging).