Machine Learning for Rapid Evaporative Ionization Mass Spectrometry for Marine Biomass Analysis

by

Jesse Wood

A thesis submitted to
Victoria University of Wellington
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Artificial Intelligence

2026

Abstract

This thesis advances seafood processing by applying deep learning to Rapid Evaporative Ionization Mass Spectrometry (REIMS) data, enabling automated and accurate marine biomass analysis.

This research addresses critical industry challenges in quality control, food safety, and fraud prevention. These include foundational tasks like species identification to combat mislabeling fraud and body part classification to optimize by-product utilization and prevent adulteration with offal. It also formalizes novel food safety problems, such as detecting oil contamination (a food safety hazard) and cross-species adulteration (a form of economic fraud). Finally, it tackles batch traceability, a task essential for rapid product recalls that currently relies on costly and impractical physical tagging methods.

To achieve this, the thesis first establishes a suite of advanced models for foundational tasks species and body part identification (Chapter 4), drawing on datasets described in Chapter 3. Specifically, this involves a binary classification dataset of 106 samples for Hoki and Mackerel identification and a multi-class dataset of 33 samples for classifying seven distinct body parts (fillets, heads, livers, skins, gonads, guts, and frames). The work then formalizes and solves novel food safety problems, including oil contamination (seven ordinal concentration levels from 50% to 0% in a 126-sample dataset) and cross-species adulteration (pure Hoki, pure Mackerel, and mixed Hoki-Mackerel samples from a 144-sample dataset) (Chapter 5), also based on Chapter 3 data. Finally, it introduces a self-supervised framework to address the challenge of batch traceability using a highly imbalanced, pairwise comparison dataset of 2,556 instances, derived from 72 fish samples across 24 distinct batches (Chapter 6). The high-dimensional nature of the REIMS data, with 2,080 features per spectrum for all tasks, necessitates sophisticated solutions.

Key contributions include the development of advanced Transformer architectures, which significantly outperform traditional methods by achieving up to 100% accuracy in species identification. The study also critically evaluates Mixture of Experts (MoE) Transformers to determine optimal configurations and investigates the asymmetric effects of transfer learning across different tasks, namely species identification, body part classification, oil contamination, and cross-species adulteration detection. Furthermore, a comparative analysis of ordinal classification techniques is included, demonstrating that models designed for ordered data significantly reduce error distance in graded contamination tasks. For batch traceability, a novel self-supervised contrastive learning method, SpectroSim, is introduced, which identifies sample origins with 70.8% accuracy without

using class labels. Explainable AI techniques are employed to ensure model predictions are interpretable, enhancing trust and providing chemically relevant insights.

This research is motivated by the need to overcome the limitations of current state-of-the-art analytical methods. The standard approach for REIMS analysis relies on traditional chemometric techniques like Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), which are often limited in their ability to model the complex, non-linear, and high-dimensional nature of mass spectra. For the foundational task of species and body part identification (Chapter 4), the challenge was to classify samples from complex and high-dimensional REIMS data. This was addressed by developing novel Transformer architectures that achieved near-perfect accuracy (up to 100%), establishing a new performance benchmark. For the novel tasks of oil contamination and cross-species adulteration (Chapter 5), the challenge was to detect subtle chemical signatures often masked by the sample's primary chemical profile. This was solved by adapting advanced models to these difficult classification problems and systematically applying transfer learning, which was found to consistently improve the detection of oil contamination. Finally, for batch traceability (Chapter 6), the challenge was to create a method that avoids costly physical tags and the need for impractical supervised labels. This was achieved by developing "SpectroSim", a self-supervised framework that accurately determines sample origin without labels, offering a practical and cost-effective solution for the industry.

Acknowledgments

To my beloved Mum, Linda Wood, who passed away while I was writing this thesis. Words cannot express my gratitude for the unwavering emotional support you offered me over the years. You were my biggest fan and my harshest critic, and you will be forever missed.

Thank you to my family: Dave, Jo, and Freya. I am grateful for your constant support and for patiently enduring my endless discussions about Artificial Intelligence (AI) at the dinner table. Your encouragement throughout my PhD journey has meant the world to me.

To my partner of six years, Millie, thank you for your love, invaluable advice, and enduring support, especially through the lean years of student life.

Beyond my immediate circle, I want to thank my friend Chris for being a constant sounding board for my rants about AI and for his encouragement that this work could have a positive impact. I also extend my gratitude to the wider group of friends and colleagues who provided camaraderie along the way: Andrew (x2), Anton, Eduardo, Mashfaq, Christian, Hayden, Jordan, Rimas, Carl, Kaan, Junhao, Peng, and Binh.

My sincere thanks go to Alastair Ross from AgResearch for generously providing the REIMS dataset that is the core subject of this thesis, as well as for his valuable feedback and essential insights into the practical applications of this work.

Finally, I wish to extend my deepest gratitude to my supervisors, without whose guidance this thesis would not have been possible: Dr. Bach Nguyen, Prof. Bing Xue, Prof. Mengjie Zhang, and Dr. Daniel Killeen. I am indebted to Bach for his meticulous attention to detail and deep expertise. I thank Bing for her consistent emotional support and inspiring dedication to her field. I am grateful to Meng for his impeccable grammar and breadth of knowledge, and to Daniel for his invaluable feedback and crucial insights into the chemical world.

This thesis was financially supported by the University Maori PhD Scholarship and the Cyber-Marine MBIE Endeavor Research Program under contract C11X2001 for both tuition fees and stipends.

List of Abbreviations

Each abbreviation is spelled out in full at its first occurrence in the main text.

Abbreviation	**Full Term**
AI	Artificial Intelligence
ANN	Artificial Neural Network
BERT	Bidirectional Encoder Representations from Transformers
CCE	Categorical Cross-Entropy
CLM	Cumulative Link Model
CNN	Convolutional Neural Network
CORAL	Consistent Rank Logits
DL	Deep Learning
DT	Decision Tree
ELISA	Enzyme-Linked Immunosorbent Assay
FFN	Feed-Forward Network
FTIR	Fourier Transform Infrared spectroscopy
GC-MS	Gas Chromatography–Mass Spectrometry
GELU	Gaussian Error Linear Unit
Grad-CAM	Gradient-weighted Class Activation Mapping
HACCP	Hazard Analysis and Critical Control Points
KAN	Kolmogorov–Arnold Network
KNN	K-Nearest Neighbours
LA-REIMS	Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry
LC-MS	Liquid Chromatography–Mass Spectrometry
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model-Agnostic Explanations

Continued on next page

Continued from previous page

Abbreviation	**Full Term**
LR	Logistic Regression
LSTM	Long Short-Term Memory network
Mamba	Selective State Space sequence model (Gu & Dao, 2023)
MCIFC	Multiple Class-Independent Feature Construction
ML	Machine Learning
MoE	Mixture of Experts
MSM	Masked Spectra Modelling
m/z	mass-to-charge ratio
NB	Naïve Bayes
NMR	Nuclear Magnetic Resonance spectroscopy
NoPE	No Positional Embeddings
NT-Xent	Normalised Temperature-Scaled Cross-Entropy loss (Chen et al., 2020a)
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
PAH	Polycyclic Aromatic Hydrocarbon
PCA	Principal Component Analysis
PCA-LDA	Principal Component Analysis–Linear Discriminant Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
REIMS	Rapid Evaporative Ionization Mass Spectrometry
RF	Random Forest
SOTA	State of the Art
SVM	Support Vector Machine
TIC	Total Ion Count
VAE	Variational Autoencoder
VUW	Victoria University of Wellington
XAI	Explainable Artificial Intelligence

Contents

1

Introduction

Chapter 1 provides a comprehensive overview of the entire thesis. It introduces the global seafood industry and its challenges, e.g., species mislabeling, cross-species adulteration, oil contamination, and batch traceability — all of which undermine consumer trust in seafood products. The chapter highlights the limitations of traditional analytical methods and posits Rapid Evaporative Ionization Mass Spectrometry (REIMS) combined with Machine Learning (ML) as a promising solution. In addition, the chapter outlines the major contributions of this thesis. Then it outlines the primary objectives of this research, including species and body part identification, oil contamination and cross-species adulteration detection, and contrastive learning for batch detection.

1.1 Introduction

Fish are a majestic and bountiful natural resource, serving as both an integral part of our marine ecosystems and food on our tables (Food and Agriculture Organization of the United Nations, 2024). The terminology marine biomass, frequently used throughout this thesis, often refers to the collection of these aquatic creatures. Fishing, when humans harvest these natural resources (Jennings et al., 2001), strives for greater efficiency, sustainability, and accurate monitoring, due to the diversity of fish species (for Economic Co-operation & Development, 2021), which makes evaluating fish stocks and the quality of harvested fish not a straightforward process. Following the catch, fish processing encompasses critical post-catch procedures such as sorting, grading, and packaging (Fellows, 2017). This chain transforms the raw catch into consumable products and is an area where modern advancements can offer transformative improvements.

Quality Assurance (QA) is essential to ensure that seafood products meet food safety requirements and customer expectations (Montgomery, 2019). QA employs various tools

for quality assurance in a systematic manner, e.g., tools ranging from simple checklists and manual inspection aids to sophisticated analytical techniques. One such advanced tool for quality assurance is Rapid Evaporative Ionization Mass Spectrometry (REIMS) (Schäfer et al., 2009; Cafarella et al., 2024), which allows for rapid chemical fingerprints of samples, providing detailed information that can be crucial for assessing fish species (Black et al., 2017; Shen et al., 2020, 2022), body parts, contaminants such as oil or cross-species adulteration (Premanandh, 2013), or tracing food origins (Lu et al., 2024; Gao et al., 2025; Gkarane et al., 2025). This data-rich environment creates a clear need for Automated QA, where intelligent systems are required to interpret complex chemical fingerprints, execute tests, and flag anomalies, thereby increasing efficiency and reliability in fish processing industrial settings (Xing et al., 2016).

One such advanced tool, which integrates data from tools like REIMS, can be enhanced greatly by Machine Learning (ML), a field of artificial intelligence (AI) that enables computers to learn from data without being explicitly programmed (Russell & Norvig, 2016). ML algorithms identify patterns and make predictions.

1.2 Problem Statement

The global seafood industry, which plays an important part in the food supply chains and world economy, faces troublesome challenges from fraudulent practices. Examples such as the mislabeling of products, waste utilization of byproducts, batch traceability, adulteration with lower-value products, and other forms of contamination pose significant food safety hazards and undermine consumer trust in seafood products by deceiving customers. Recent meta-analyses show this problem is current and widespread, with mislabeling rates reported as high as 39.1% in the US (Ahles et al., 2025) and 24.8% in the Eastern South Pacific (Marín, 2025). The current state-of-the-art approach to seafood quality control involves traditional analytical methods that involve time-consuming laboratory procedures and considerable domain expertise - necessitating the demand for rapid, accurate, and automated solutions that can be deployed efficiently within a factory setting.

These automated solutions must perform several crucial quality control tasks to ensure the efficient, safe, and effective monitoring of fish processing factories, including fish species and body part identification, oil contamination and cross-species adulteration detection, and batch traceability. For example, species identification prevents species substitution fraud (Australia, 2016), body part classification facilitates efficient waste utilization by enabling the re-use of byproducts (Ghaly et al., 2013), cross-species adulteration and oil contamination detection directly impacts food safety standards (Premanandh, 2013), and batch traceability allows for quick recalls if contamination occurs (Mai et al., 2010).

1.3 Motivations

1.3.1 Marine Biomass

Marine biomass is the total mass of all living marine organisms within a given area or ecosystem. Marine biomass is subject to mislabeling, species substitution, cross-species adulteration, and oil contamination. Studies have revealed an alarming amount of mislabeling within the global seafood industry (Pardo et al., 2016), a trend confirmed by recent large-scale meta-analyses in 2025 (Ahles et al., 2025; Marín, 2025). This fraud can be complex, with recent studies uncovering hidden mammal and avian species (pork, chicken) in processed fish balls, posing serious ethical and religious consumer challenges (Zhang et al., 2025). Whether these species substitutions are intentional and fraudulent, or systematic and prevalent, accurate methods for detecting fish species and cross-species adulteration within seafood products are essential for guaranteeing food authenticity in quality control in fish processing, especially in high-consumption regions like Asia (Do & Wong, 2025). One example of fraudulent species substitution was a restaurant that was found serving Vietnamese Catfish under the pseudonym Australian Dory (Australia, 2016). Less than half of a fish can be processed into fillets, leaving the byproduct to be repurposed into products such as fertiliser, fish meal, or omega-3 concentrates. One such concentrate is omega-3 supplements (Simopoulos, 2011), which are an essential, but often lacking, ingredient in Western diets (FAO, 2020). Methods for differentiating between fish body parts in marine biomass byproducts are crucial for maximizing waste utilization in fish processing (Stevens, 2018).

Contamination — contamination is the presence of any harmful or undesirable foreign substance — such as chemicals, pollutants, or other species—that renders the biomass impure, unsafe for consumption, or otherwise unfit for its intended use — whether it be oil pollutants or cross-species adulteration, presents another challenge. Adulteration is where a high-value product is mixed with a cheaper species or body part (e.g., offcuts/offals), which can undermine the integrity of the product and pose health risks (Black et al., 2019). Again, this emphasises the importance of biomass species and body part identification systems. One such famous example would be the European Horse Meat Scandal of 2013 (Premanandh, 2013), where beef mince was fraudulently and intentionally mixed with horse meat and served to customers under the label prime beef. Oil contamination, another harmful contaminant, can be introduced from engine oil from the boats, or processing equipment in the factory (Moens, 2003). It is pivotal that effective systems for contamination detection — be it oil pollutants or cross-species adulteration — are developed for fish processing, to catch contaminants early, well before they make it to your plate.

Not only do we need methods for contamination detection, but also **batch detection**, which we define as the task of using REIMS to determine if two fish samples originate

from the same processing batch, allowing for quick recalls if contamination occurs (Mai et al., 2010). This capability is critical for combating Illegal, Unreported, and Unregulated (IUU) fishing (Helyar et al., 2014) and enabling modern, transparent supply chains (Turkson et al., 2025). The current state-of-the-art is rapidly moving toward digital traceability systems based on Blockchain (Dahariya et al., 2025), Digital Product Passports (Jiang et al., 2025), and mobile data-entry platforms (Untal et al., 2025; Gastaldi García, 2025). However, these systems are vulnerable to fraudulent data entry and face significant adoption barriers, such as cost and infrastructure (Untal et al., 2025). This creates a critical need for an alternative, intrinsic verification method that can validate a sample's identity analytically, complementing these digital systems rather than relying on vulnerable physical tags like RFID chips (Mai et al., 2010).

1.3.2 Rapid Evaporative Ionization Mass Spectrometry

REIMS (Schäfer et al., 2009; Cafarella et al., 2024) is a direct-to-analysis technique that allows for the near-instantaneous chemical analysis of a sample with minimal to no preparation (Henderson et al., 2025; Pruekprasert et al., 2025). REIMS has the potential to revolutionize the field of seafood quality control (Cafarella et al., 2024; Black et al., 2017) by providing detailed molecular profiles that can differentiate between fish species (Black et al., 2017; Shen et al., 2020, 2022), fish body parts (Black et al., 2019), or trace food origins (Gao et al., 2025; Gkarane et al., 2025; Lu et al., 2024). It can also be used to identify contaminants, detect adulterations (Premanandh, 2013), and, as this thesis proposes, perform batch traceability (Mai et al., 2010). The nature of REIMS data presents challenges; firstly, REIMS sample preparation requires significant domain expertise and resources, resulting in limited training data. Secondly, the limited training data has many features, e.g., 2,080 to be exact. These first two challenges, coupled together, lead to a high-dimensional dataset that suffers from the curse of dimensionality (Köppen, 2000).

A critical point is that a single measured mass-to-charge (m/z) feature reflects an ion (such as a deprotonated molecule or adduct), which may represent multiple isomers or compounds. Because REIMS typically does not use mass fragmentation, the identification of the underlying chemical compounds is tentative, relying only on high-resolution mass measurements to narrow down possibilities. This data is best described as a complex chemical fingerprint or molecular profile rather than a definitive list of identified compounds.

In addition, REIMS data, due to the nature of sample preparation, instrumentation, and environmental factors, has noise inherent in the data. Traditional supervised techniques for analysis, such as PCA (Schäfer et al., 2009), PCA-LDA (Balog et al., 2010, 2013, 2016), and OPLS-DA (Bylesjö et al., 2006; Boccard & Rutledge, 2013)—which remain the standard in many recent studies (Black et al., 2017, 2019; Verplanken et al., 2017; Gkarane et al., 2025; Shen et al., 2020, 2022)—struggle to discern the signal from

the noise and capture meaningful patterns in the data. Noise is such a prevalent issue that practitioners often discard principal components with relative standard deviations above a certain threshold outright, a practice seen in (Black et al., 2017, 2019).

Finally, techniques like OPLS-DA are not capable of capturing the complex sequential patterns and feature interactions present in mass spectra. This has created an active debate in recent literature (Xue et al., 2025). Some studies find that traditional chemometrics remain more robust than conventional machine learning (ML), like Random Forest or SVM for large-scale industrial tasks (De Graeve et al., 2023). Other studies find conventional ML models like K-Nearest Neighbors (KNN) are superior for new tasks like geographical authentication (Lu et al., 2024). This indicates a clear need for more sophisticated machine learning techniques to be evaluated. This need is reinforced by the very latest studies, which show that deep learning methods like Artificial Neural Networks (ANNs) can outperform traditional OPLS-DA by better modeling the complex, non-linear relationships in REIMS data (Cardoso et al., 2025).

1.3.3 Deep Learning

Deep learning for REIMS data analysis often refers to deep neural network-based methods that learn hierarchical feature representations from raw data (Goodfellow et al., 2016). While traditional methods like Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) (Bylesjö et al., 2006) have been the standard for REIMS biomass analysis, their capabilities are limited when faced with the inherent challenges of the data. REIMS mass spectra are high-dimensional, inherently sequential, and often generated from a limited number of prepared samples, creating significant hurdles for traditional machine learning. To overcome these limitations, this thesis turns to deep learning, a collection of methods that learn hierarchical feature representations directly from raw data. The following sections outline the key challenges of REIMS data and motivate the specific deep learning techniques chosen to address them.

First, a primary challenge is capturing the complex, sequential patterns present in mass spectra at multiple scales, from broad, low-resolution patterns to fine-grained, high-resolution details. To address the sequential nature and long-range dependencies in the data, this thesis employs Transformers (Vaswani et al., 2017), as their self-attention mechanism is uniquely suited to weigh the importance of all features in a sequence. To further enhance this capability, a multi-scale Ensemble Transformer (Wolpert, 1992) was developed. This architecture combines shallow, medium, and deep Transformers, allowing it to simultaneously analyze broad, low-resolution patterns and fine-grained, high-resolution details within the spectra. This multi-scale approach creates a more robust and comprehensive model than a single architecture could achieve.

Second, the analysis is constrained by the scarcity of labeled data and the need to leverage knowledge across different but related tasks. The preparation of REIMS samples

is resource-intensive, resulting in limited training datasets. This is mitigated with Unsupervised Pre-training (Devlin et al., 2018), a technique where a model first learns the fundamental structure of mass spectra from large amounts of unlabeled data. This process provides a powerful foundation for fine-tuning on smaller, labeled datasets for specific downstream tasks. To further leverage existing knowledge, Transfer Learning (Bozinovski & Fulgosi, 1976; Tan et al., 2018) has also been explored, which allows insights gained from a source task to be applied to a related target task, enabling the model to build on previous learning.

Third, as analytical tasks become more complex, there is a need for architectures that can scale in capacity and solve novel problems. Mixture of Experts (MoE) Transformers (Jacobs et al., 1991; Kaiser et al., 2017) address the issue of scalability by allowing a model's parameters to increase efficiently; this is achieved by routing inputs to specialized expert sub-networks, which increases model capacity without a proportional rise in computational cost. For novel challenges like batch traceability, which requires a pairwise comparison, Contrastive Learning (Chen et al., 2020a) offers a more suitable framework than standard classification. The proposed "SpectroSim" method uses this approach to learn an embedding space where similar samples are grouped, enabling accurate pairwise comparison without relying on explicit class labels.

Finally, for these advanced models to be deployed in real-world settings, a crucial challenge is ensuring their interpretability and trustworthiness. A persistent limitation of deep learning models is their black-box nature, which can hinder adoption by domain experts who need to verify the reasoning behind a decision. This challenge has been explicitly identified in the latest food science literature applying ANNs to REIMS data (Cardoso et al., 2025). To ensure the models developed in this thesis are verifiable, post-hoc explainability methods like LIME (Ribeiro et al., 2016) and Grad-CAM (Selvaraju et al., 2017) are employed. These techniques provide crucial insights into which m/z features are driving classification decisions, effectively opening the black box and bridging the gap between complex models and practical, real-world applications by providing chemically relevant hypotheses for domain experts to verify.

1.4 Tasks

This thesis addresses three tasks in REIMS-based marine biomass analysis, with a specific focus on species critical to the New Zealand seafood industry. The selected species, Hoki and Mackerel, represent two major commercial fisheries. Hoki, in particular, is New Zealand's largest and most valuable fishery (Ministry for Primary Industries, 2024), making its accurate identification and quality control a top national priority. These tasks are correlated through their shared objective, data source, and analytical methods used to solve them, as outlined in this thesis. All three tasks directly address major challenges in the seafood industry, such as fraud, waste utilization, safety, and traceability. The three

tasks all rely on data generated by the same core technology, REIMS. This thesis systematically explores a set of advanced machine learning models to solve these problems. In particular, those three tasks are:

1. **Fish Species and Body Part Identification:** Firstly, a binary classification task between two species of fish: Hoki and Mackerel. Secondly, a multi-class classification task between seven body parts of fish: head, fillet, frame, skin, liver, guts, and gonads. The first two tasks are related to species substitution (fish species identification) and waste utilization of byproducts (fish body parts identification).

2. **Oil Contamination and Cross-Species Adulteration Detection:** These are two ordinal multi-class classification tasks. Firstly, one where there are 7 different concentrations of oil-contaminated fish ranging from 50% to 0%. Secondly, there are three classes of cross-species adulterated fish, including pure Hoki, pure Mackerel, and 50-50 mixed Hoki-Mackerel. Both of these tasks are related to contamination detection to (obviously) prevent oil contamination and cross-species adulteration in fish supply chains.

3. **Batch Detection:** This is a pair-wise comparison task, where given two fish, we wish to detect if they originate from the same (or different) batches. A batch is a group in which fish were processed during the REIMS-based sample analysis. Batch detection is a new method of batch traceability, which allows for quick and accurate recalls of new contaminants in products, should they occur.

Furthermore, these tasks are interconnected problems in real-world settings. In a fish processing plant, these are not isolated issues. An adulterated product (Task 2) is also a mislabeled one (Task 1); for example, a product sold as "Pure Hoki" that is found to be adulterated with Mackerel (Task 2) is by definition also mislabeled (Task 1). If contamination (Task 2) is found, batch detection (Task 3) is required to execute an effective recall. Distinguishing body parts (Task 1) is essential for detecting certain kinds of adulteration, such as mixing high-value fillets with cheaper offal (Task 2). Together, these tasks form an automated, holistic approach to solving quality control challenges in marine biomass analysis.

1.5 Research Goals

The overall goal of this thesis is to develop and validate a suite of machine learning methods to enhance the analytical capabilities of REIMS-based marine biomass analysis, targeting rapid, automated, and in situ application in fish processing plants. In achieving this, the thesis is structured around the following three objectives:

1. **Develop an Approach for Fish Species and Body Part Identification:** This objective tackles the critical need for accurate product labeling to prevent consumer fraud and optimize the use of byproducts. Existing analytical methods are often too slow for real-time factory use, and methods like the state-of-the-art OPLS-DA are limited in their ability to model complex, non-linear spectral data (De Graeve et al., 2023; Cardoso et al., 2025). The primary challenge lies in accurately classifying species and tissues from complex, high-dimensional REIMS spectra where biochemical differences can be subtle. To address this, the goal is to develop a potentially interpretable machine learning pipeline, exploring advanced deep learning architectures. Specifically, Transformer variants are chosen for their self-attention mechanism, which is hypothesized to be uniquely suited to capturing the complex, long-range dependencies between m/z features across the entire mass spectrum. A central aim is to ensure these models are verifiable by domain experts through post-hoc explainability methods (LIME and Grad-CAM) that identify the key biochemical features driving classification decisions. This work represents the first application of deep learning with Transformer models to REIMS biomass analysis.

2. **Formalize and Solve Novel Contamination and Adulteration Detection Tasks:** This objective addresses direct threats to food safety and economic integrity by formalizing two novel problems for REIMS analysis: oil contamination and cross-species adulteration. These tasks are crucial for consumer protection, but no rapid, in-situ methods currently exist to detect them. The state-of-the-art (SOTA) is reliant on slow, lab-based methods, and the core challenge for an automated approach is identifying the faint chemical signals of contaminants, which are often masked by the dominant biochemical profile of the biomass itself. The goal is to develop highly sensitive models. Transformer variations are leveraged again, as their learned representations are hypothesized to be sensitive enough to detect these subtle signals. Furthermore, transfer learning is explored as a method to overcome data scarcity—a key challenge in food safety—by transferring knowledge from a data-rich source task to these more difficult, data-scarce detection tasks. Unsupervised pretraining is applied to these tasks as an alternative approach to overcoming data scarcity by learning general-purpose representations that can be reused for downstream tasks. Crucially, this objective includes using LIME explanations to validate that the models learn chemically relevant features, ensuring their decisions are transparent and reliable. This research is the first to apply REIMS analysis to oil contamination detection in biomass and the first to use it for cross-species adulteration detection in marine biomass.

3. **Develop a Contrastive Learning Framework for Batch Detection:** This objective addresses the need for robust batch traceability, which is essential for enabling swift product recalls in the event of a safety issue. The SOTA for traceability

is no longer just costly physical tags (Mai et al., 2010), but an emerging field of digital systems like Blockchain (Dahariya et al., 2025) and Digital Product Passports (Jiang et al., 2025). However, these systems are vulnerable to fraudulent data entry (Turkson et al., 2025) and face high adoption barriers (Untal et al., 2025). The main challenge is to create an intrinsic, analytical verification method that can validate these digital records without relying on explicit labels. To overcome this, the goal is to develop "SpectroSim," a novel contrastive learning method. This framework is chosen specifically because it is designed to learn a similarity-based embedding space in a self-supervised manner. This directly addresses the core challenges by (a) performing a pairwise comparison ("are these two samples similar?") rather than standard classification, and (b) removing the need for impractical, manually-created labels for every new batch. The aim is for this model to outperform standard classification approaches, establishing the first machine learning-based solution for analytical batch detection in this field.

As a whole, this thesis aims to demonstrate the transformative power of ML models, in particular, those of deep learning (e.g., Transformers and MoE), with ML methodologies for enhanced representation learning (e.g., unsupervised pre-training, transfer learning), in conjunction with REIMS technology, can be utilized for fast, accurate and potentially explainable quality control in fish processing. By systematically addressing challenges, from simpler tasks such as species identification to more complex tasks such as oil contamination detection or batch detection, this work contributes to the development of rapid, accurate, and automated systems for ensuring food safety and quality by preventing fish fraud or contamination through mislabeling, adulteration, or other means. This helps consumer trust through the measurable integrity of seafood products in fish processing. The subsequent chapters will elaborate on the methodologies, experimental results, and implications of each of these contributions in detail.

1.6 Major Contributions

The major contributions of this thesis are the development and application of novel deep learning methods for REIMS-based marine biomass analysis, for quality control in the seafood industry, particularly in the context of New Zealand. This work is the first to systematically apply advanced deep learning architectures—including Transformers, Mixture of Experts (MoE), and self-supervised contrastive learning—to the specific tasks and datasets explored in this thesis, consistently outperforming traditional analytical methods on this data. The key contributions are organized by the specific challenges addressed:

1. **Novel Deep Learning Architectures for Foundational Biomass Classification Tasks:** This thesis contributes two new deep learning models — a Mixture of

Experts (MoE) Transformer and a multi-scale Ensemble Transformer—that significantly advance the state-of-the-art for identifying fish species and body parts from the REIMS data. The new knowledge generated shows that specialized Transformer-based architectures can achieve unprecedented accuracy on these foundational quality control tasks. The contents of this chapter are based on the work presented in (Wood et al., 2025).

- **Technical Contribution:** A new Mixture of Experts (MoE) Transformer, named "Gone Phishing," was proposed to enhance model capacity and specialization for this domain. In Addition, a Multi-scale Ensemble Transformer called "Autobots" is proposed for REIMS-based marine biomass analysis.
- **Results and Analysis:** The introduction of these models led to significant performance gains on the datasets tested. The results demonstrate a clear task-dependent trade-off between the two architectures. The MoE Transformer ("Gone Phishing"), which uses specialized sub-networks, proved superior for the binary fish species classification task, achieving a perfect 100% accuracy. Conversely, the Multi-scale Ensemble Transformer ("Autobots"), which combines models of different depths, was more robust for the more complex multi-class fish body part classification task, achieving the top accuracy of 74.13% (a significant improvement over the OPLS-DA benchmark of 51.17%).

2. **The First Framework for Detecting Contamination and Adulteration with REIMS:** This research contributes the first-ever application of REIMS analysis to solve two critical food safety issues using the provided datasets: oil contamination and cross-species adulteration. The contribution lies in formalizing these previously unaddressed issues as machine learning problems and demonstrating that deep learning can effectively solve them. Oil contamination was formulated as an ordinal classification task to identify its severity, while cross-species adulteration was framed as a multi-class problem to detect the type and presence of foreign species.

   - **Technical Contribution:** A novel unsupervised pre-training technique, Masked Spectra Modeling (MSM), was developed by adapting BERT's masked language modeling for sequential REIMS data. Transfer Learning approaches were explored for the MoE Transformer for REIMS-based data. The successful adaptation of the deep learning methods (i.e., MoE and Ensemble Transformers) from the previous chapter's classification tasks to these more complex problems is a key contribution, demonstrating the robustness and versatility of the proposed architectures for these data. An exploration of ordinal classification approaches to the oil contamination task is proposed.
   - **Results and Analysis:** The models demonstrated strong performance on the test data, showing that pre-training improved accuracy. A key finding was

that the difficult task of oil contamination detection consistently benefited from transfer learning, regardless of the source task, improving accuracy on this task by up to 3.59%. This highlights that transfer learning is most effective for more challenging tasks with subtle chemical signals within this dataset. The exploration of ordinal classification approaches for oil contamination concluded that for distance-aware metrics, ordinal classification approaches improved accuracy.

3. **A Novel Self-Supervised Framework for Chemical Batch Traceability:** This work contributes a new method for batch traceability, addressing this novel challenge with a self-supervised learning approach that does not require labeled data. The novelty lies in creating a system that can learn to identify the unique chemical fingerprint of a production batch from the raw spectra alone.

    - **Technical Contribution:** A novel contrastive learning framework named "SpectroSim" was developed. This method adapts the SimCLR framework by incorporating a Transformer-based encoder and a custom projection head, specifically designed for the pairwise comparison of mass spectra.
    - **Results and Analysis:** SpectroSim achieved 70.8% accuracy in the task of batch detection. Crucially, it accomplished this without using any class labels during training, demonstrating the power of self-supervised learning to identify subtle, batch-specific chemical signatures from an imbalanced dataset, far surpassing traditional binary classification approaches, which struggled to achieve over 60% accuracy on this dataset.

1.7 Thesis Overview

In this thesis, we explore the design and development of new state-of-the-art deep learning methods for REIMS-based marine biomass analysis.

- In **Chapter 2**, we discuss the foundational materials and background necessary for the topic of REIMS marine biomass analysis, where we focus on the application of ML techniques for analyzing REIMS data. Therefore, we discuss: (1) The Challenges of Seafood Integrity, (2) Analytical Techniques for Seafood Authentication, and (3) Machine Learning for Enhanced REIMS Data Analysis. This background material is sufficient to understand the key contributions that comprise this thesis.

- In **Chapter 3**, the datasets provided by AgResearch, New Zealand, are introduced. This includes four classification tasks: fish species, fish body part, oil contamination, cross-species adulteration, and batch detection - a pairwise comparison task. Consequently, this chapter covers the following aspects of the dataset: (1) REIMS Data Acquisition and Characteristics, (2) Specific Datasets for Classification Tasks, and

(3) Dataset Preprocessing: Normalization. The contents of this chapter encapsulate the ML tasks that are addressed throughout the remainder of this work.

- In **Chapter 4**, we address the fundamental tasks of *Fish Species and Fish Body Part Identification*. This chapter details the initial application of machine learning, establishing their effectiveness over traditional OPLS-DA benchmarks for these REIMS marine biomass analysis tasks. Model interpretability using LIME and Grad-CAM is also explored. Building on this, the chapter introduces how more advanced methodologies developed in this thesis enhance these identifications: the Mixture of Experts (MoE) Transformer and Ensemble Transformer architectures are applied, and their impact on these specific tasks is evaluated.

- In **Chapter 5**, we tackle the more challenging *Oil Contamination and Cross-species Adulteration Detection* tasks. The chapter outlines how methods like unsupervised pre-training and transfer learning with Transformers prove effective for these complex REIMS data scenarios, with LIME again used for model interpretability. This chapter further demonstrates the capabilities of the advanced techniques central to this thesis: the Mixture of Experts (MoE) Transformer and Ensemble architectures are applied to these difficult classifications, showcasing their performance benefits. Moreover, systematic investigations of transfer learning are presented, highlighting significant outcomes such as consistent performance improvements for oil contamination detection and confirming the task-dependent nature of transfer success. Finally, ordinal classification approaches are explored and evaluated with distance-aware metrics for comparison.

- In **Chapter 6**, we develop *Contrastive Learning for Batch Detection*, a contrastive learning framework with a custom encoder and project head for batch detection for REIMS marine biomass analysis. This contribution compares binary classification to contrastive learning for pair-wise comparison batch detection of marine biomass. The proposed method, *SpectroSim*, a SimCLR network with a transformer encoder and custom projection head, performs with the highest accuracy (70.8%) at the batch detection task, significantly outperforming binary classification methods.

- In **Chapter 7**, the thesis reaches its conclusion, with a review of the key contributions and findings, and remarks on the overall significance of the work to the field of REIMS marine biomass analysis. The limitations of the research, followed by future work directions, are addressed, and finally, some concluding remarks are given.

2

Literature Survey

Following on from the introduction, Chapter 2 comprehensively reviews the existing knowledge pertinent to this research by building a "Problem-Gap-Solution" narrative. First, we introduce the **Problem**, covering the challenges of the global seafood industry such as mislabeling and adulteration. The chapter then discusses the technological **Landscape**, reviewing existing analytical techniques for seafood authentication with a focus on Rapid Evaporative Ionization Mass Spectrometry (REIMS). Next, we establish the **Gap** by presenting a focused literature review on the state of REIMS data analysis, tracing its evolution from foundational PCA/LDA, through chemometric OPLS-DA, to the most recent applications of foundational machine learning and ANNs. This review identifies the limitations of current methods and establishes the need for more advanced models. Subsequently, the chapter details the **Solution**, introducing the suite of advanced deep learning architectures that form the core of this thesis, including Transformers, Mixture of Experts, and strategies like Transfer Learning and Contrastive Learning. Finally, we address the critical component of **Validation** by discussing Explainable AI as a tool for building trust in these complex models.

2.1 The Challenges of Seafood Integrity: Fish Processing, Fraud, and Contamination

The global fish processing industry is a critical part of our food supply chain; it transforms marine biomass (i.e., a collective term for every single living organism in the ocean) into a wide variety of consumer products. Fish processing is a complex process that involves sorting, cleaning, filleting, and packaging with rigorous quality control throughout — that is fraught with challenges that impact food safety, authenticity, and economic value. Pertinent to these challenges are seafood fraud (e.g., species mislabeling, substitution, and

adulteration) and various forms of contamination.

2.1.1 Mislabeling

Mislabeling is when a different species (usually less expensive) is sold as another (usually more expensive), which is a pervasive issue. A meta-analysis of 51 peer-reviewed papers (n = 4,500) found an average mislabeling rate of 30% (Pardo et al., 2016), an alarming rate that underscores the severity of this deception. This finding is reinforced by more recent large-scale analyses; a 2025 meta-analysis of U.S. studies found an overall mislabeling rate of 39.1%, with species substitution specifically accounting for 26.2% of samples (Ahles et al., 2025). A similar 2025 meta-review focusing on the Eastern South Pacific reported a regional mislabeling rate of 24.8% (Marín, 2025). The issue is a global one, with research proving mislabeling has occurred across several diverse markets. For instance, studies in the U.S. have utilized DNA barcoding to unmask seafood mislabeling, emphasizing the technology's role in food authentication and quality control (Khaksar et al., 2015). Similarly, in Canada, extensive market substitution has been revealed (Hanner et al., 2011). Mislabeling is prevalent in Asia (Chin et al., 2016; Kannuchamy et al., 2016; Changizi et al., 2013), a critical finding underscored by a 2025 review, which notes that Asia accounts for approximately two-thirds of global seafood consumption (Do & Wong, 2025). In Europe, this problem is rife, with two case studies in Western Europe (Miller et al., 2012), and studies detailing substitution rates in Italy (Filonzi et al., 2010) and France (Bénard-Capelle et al., 2015). Closer to home in Australia, specifically Tasmania, studies have employed DNA barcoding to assess labeling accuracy (Lamendin et al., 2015), aiding wider efforts to barcode Australia's fish species (Ward et al., 2005). Furthermore, in Egypt, studies have found significant levels of mislabeling (Galal-Khallaf et al., 2014), and in South Africa, this problem also presents itself (Cawthorn et al., 2015). Here we highlight the 2016 incident where a Melbourne restaurant was accused of serving a cheaper Vietnamese catfish and premium Australian dory (Australia, 2016) as both an example of mislabeling and species substitution. This fraud can go beyond simple fish-for-fish substitution; a 2025 study on Chinese fish balls sold online found a 65.9% mislabeling rate, uncovering hidden undeclared mammal and avian species, including pork, chicken, and beef, highlighting significant ethical and religious consumer challenges (Zhang et al., 2025). **Species substitution** is the fraudulent practice where a different species, usually a less expensive one, is sold as another, more expensive species to deceive customers. These fraudulent practices deceive consumers but also undermine fair market competition, creating economic repercussions for legitimate producers.

Mislabeling has implications that are far-reaching; these can be linked to failings in traceability within fish processing, creating opportunities for Illegal, Unreported and Unregulated (IUU) fishing (Helyar et al., 2014). This link is well-documented, with studies showing that mislabeling is used to sell endangered or threatened species. For exam-

ple, a 2025 meta-review of the Eastern South Pacific found a "substantial proportion" of mislabeled products were highly threatened shark species (Marín, 2025). A separate 2025 study in New England similarly used DNA barcoding to find that samples sold as 'swordfish' were, in fact, endangered mako and thresher sharks (Eppley & Coote, 2025). Gene-associated markers are a tool being developed to combat illegal fishing and the false eco-certification that is possible in such opaque fish processing supply chains (Nielsen et al., 2012). Take, for example, the mislabeling of Sablefish with Patagonian and Antarctic Toothfish in China's online market has opened the doors to IUU fishing (Xiong et al., 2016). Governmental regulatory forensic programs — e.g., those in South Brazil that use DNA barcoding — are pivotal for identifying commercialized seafood and combating these illicit activities (Carvalho et al., 2015). DNA authentication has shown that mislabeling is often associated with various stages of seafood processing (Muñoz-Colmenero et al., 2016).

2.1.2 Adulteration

Adulteration is where a high-value product is mixed with a lower-value product fraudulently for criminal gains. This can include cross-species adulteration where different species of different values are mixed, or low (e.g., offcuts/offal) and high-value (e.g., prime beef) biomass from the same species. One such example of cross-species adulteration in the seafood industry was demonstrated by the molecular identification of fish species in surimi-based products labeled as Alaskan pollock (Keskin et al., 2012). The problem, however, extends far beyond simple fish-for-fish substitution. A 2025 study on Chinese fish balls, for instance, found a 65.9% mislabeling rate, uncovering undeclared pork, chicken, and beef, which poses serious ethical, religious, and food safety concerns (Zhang et al., 2025). The 2013 Horse Meat Scandal (Premanandh, 2013) amplified the concern over food integrity. In Europe, widespread cross-species adulteration was found (i.e., beef mixed in with horse meat), which eroded consumer trust in food supply chains, and highlighted some gaps, or lack of, quality assurance.

This event was a major wake-up call for food supply chains, underscoring the urgent need for more robust quality assurance for food authentication and traceability. This has spurred research into a variety of rapid detection methods. Some approaches focus on chemical adulterants used as preservatives, with novel methods including high-sensitivity photonic crystal fiber sensors to detect formalin in real-time (Veluchamy et al., 2025) and "smart films" that change color to visually indicate the presence of formalin or ammonia (Moeinpour et al., 2025). Other state-of-the-art methods focus on species or ingredient adulteration by combining spectroscopy with traditional chemometrics. For example, Fourier Transform Infrared (FTIR) spectroscopy and Gas Chromatography-Mass Spectrometry (GC-MS) have been combined with chemometrics like Linear Discriminant Analysis (LDA) and Partial Least Squares (PLS) to detect pork oil in fish oil for halal authentication, achieving 100% classification accuracy (Lestari et al., 2025). Similarly, vibra-

tional spectroscopy (IR and Raman) coupled with Principal Component Analysis (PCA) and other chemometric algorithms has been effective in differentiating natural caviar from imitation samples (Amelin et al., 2025). These methods often seek to overcome the limitations of DNA-based techniques, which can be unreliable for highly processed or heated foods; one 2025 study proposed using stable N-glycan markers identified via LC-IM-MS to authenticate processed fish species (Walsh et al., 2025).

A common thread in these spectroscopic approaches is their reliance on traditional machine learning and chemometric models like PCA, PLS, and SVM. However, recent evidence suggests these methods are being surpassed by deep learning. A 2025 study on Antarctic Krill Oil adulteration directly compared these approaches, finding that a deep learning ResNet (CNN) model achieved superior classification accuracy and significantly outperformed traditional PCA-PLS methods and SVM for this exact type of spectral analysis (Ke et al., 2025). This finding strongly motivates the core premise of this thesis: that applying advanced deep learning architectures to spectral data is a necessary and superior approach for tackling the complex challenges of seafood adulteration. While the prevalence of mislabeling and adulteration is a critical issue, some research shows that there has been some improvement, with one study from Europe suggesting low mislabeling rates in their seafood market operations, perhaps to be attributed to the efficacy of increased regulatory scrutiny and improved practices in fish processing (Mariani et al., 2015).

2.1.3 Contamination

Contamination is where undesirable substances are introduced to fish products, making them unsafe for human consumption. This can occur at any stage of fish processing, including harvesting and handling, to processing, packaging, and storage. Contamination can be categorized into three main types: (1) biological - the presence of harmful microorganisms, (2) chemical - the presence of harmful chemical substances, and (3) physical - the presence of foreign objects. Effective control of contamination in fish processing relies on food safety management systems, such as Hazard Analysis and Critical Control Points (HACCP) (Bremner, 2002). This thesis focuses on chemical and biological contamination, which pose severe, large-scale risks to both human health and ecosystem stability.

A clear example of chemical contamination is from oil. This can be from processing equipment or boat engines (Moens, 2003), but also from large-scale environmental disasters. For instance, a 2019 oil spill in Brazil led to the bioaccumulation of carcinogenic Polycyclic Aromatic Hydrocarbons (PAHs) in local seafood, which in turn caused DNA damage and lipoperoxidation in fish and posed a significant health risk to traditional human communities (Mello et al., 2025).

Heavy metal contamination represents another pervasive chemical threat. Recent global studies have found alarming concentrations of arsenic (As) and mercury (Hg) in fish, with many samples exceeding WHO permissible limits (Cobbinah et al., 2025). This

is a worldwide issue, with studies identifying arsenic as a primary contaminant in dried fish from India (Sonone et al., 2025), cadmium (Cd) bioaccumulating in fish from mineral-rich ecosystems (Kumari et al., 2025), and industrial pollution in Bangladesh leading to high levels of As, Cr, Pb, and Zn in local fish (Choudhury et al., 2025). These metals pose significant carcinogenic and non-carcinogenic risks, particularly to children (Cobbinah et al., 2025; Sonone et al., 2025; Onoja et al., 2025).

Furthermore, emerging physical and chemical contaminants like microplastics (MPs) present a complex and growing challenge. MPs are now ubiquitous, with one 2025 review finding them in over 450 wild freshwater fish species, including 35 on the IUCN Red List of threatened species (de Araújo et al., 2025). The risk to humans is direct, as MPs (like PE, PP, and PET) are found not only in the gills and gastrointestinal tracts but also in the edible **muscle tissue** (Pingki et al., 2025; Jamal et al., 2025; Shakik et al., 2025). One study estimated a potential human intake of over 1.2 million MPs per person per year from seafood consumption (Jamal et al., 2025). Critically, a 2025 meta-analysis highlights that MPs also act as vectors, absorbing other emerging contaminants, and this "combined pollution" has a more severe adverse impact on fish reproduction, development, and neurotoxicity than single contaminants alone (Wu et al., 2025).

Finally, biological contamination in the form of cross-species adulteration (Premanandh, 2013) is also prevalent in fish processing. As these examples of severe, multifaceted contamination demonstrate, there is a clear and urgent need for rapid and automated tools. Effective systems are required to detect not only chemical pollutants like oil and heavy metals, but also biological and physical contaminants like microplastics and fraudulent cross-species adulteration.

2.1.4 Batch Detection

Batch detection is the process of identifying the specific production batch from which a sample originates. This capability enables batch traceability, a system for ensuring food safety, preventing fraud, and managing the supply chain (Lu et al., 2024; Xue et al., 2025). Traceability is pivotal for minimizing risks, as it allows for the quick and targeted recall of products should contamination occur (Mai et al., 2010). In order to meet safety standards and improve fish processing efficiency, regulations will often mandate batch-level traceability. This is because traceability is a key tool to combat fraud, such as the mislabeling and adulteration that plague complex global seafood supply chains (Do & Wong, 2025; Walsh et al., 2025). Such fraudulent practices, which can be systemic, not only deceive customers but also enable illicit activities; failures in traceability, for instance, are directly linked to laundering fish from Illegal, Unreported, and Unregulated (IUU) fishing into legitimate markets (Helyar et al., 2014), a problem that modern digital traceability systems are now being designed to solve (Turkson et al., 2025). The scale of this problem is significant, with global meta-analyses showing mislabeling rates as high as

30% (Pardo et al., 2016), and regional studies in the US and the Eastern South Pacific confirming this trend (Ahles et al., 2025; Marín, 2025). High-profile incidents like the 2013 horse meat scandal have demonstrated how quickly such failures erode consumer trust (Premanandh, 2013).

Therefore, verifiable batch traceability is essential for building and maintaining that trust. However, the implementation of current state-of-the-art traceability systems faces significant limitations. The field is currently focused on developing digital and logistical frameworks (Gastaldi García, 2025), such as blockchain-based systems (Dahariya et al., 2025), 'Digital Food Product Passports' (Jiang et al., 2025), and mobile data-entry platforms that connect fishers directly to the supply chain (Turkson et al., 2025; Untal et al., 2025). While powerful, these systems can be complex and prohibitively costly, especially for smaller fish processing firms (Thompson et al., 2005). Furthermore, their adoption faces significant on-the-ground challenges, including fishers' technical literacy, access to smartphones, and reliable internet infrastructure (Untal et al., 2025). Lastly, and most critically, these logistical systems are fundamentally vulnerable to deliberate fraud at the point of data entry; a system is only as reliable as the label it carries, and an incorrectly or fraudulently labeled batch renders the entire tracking process ineffective (Helyar et al., 2014).

This motivates the need for an intrinsic, chemical-based verification method to complement these digital systems. While REIMS has been successfully applied to traceability in other food sectors like pork and lamb (Gkarane et al., 2025; Gao et al., 2025), its application to seafood batch detection remains a critical gap. This approach aligns with other emerging research that uses analytical chemistry and machine learning (such as for mercury determination) to enhance and validate traceability monitoring systems (Piroutková et al., 2025). This thesis, therefore, explores an intrinsic method that does not rely on external labels, but on the chemical fingerprint of the sample itself.

2.2 Analytical Techniques for Seafood Authentication

To ensure seafood authenticity and combat fraud, the industry requires analytical techniques that are faster and more accurate than traditional, time-consuming laboratory methods. This section reviews the technologies used for this purpose, focusing on Rapid Evaporative Ionization Mass Spectrometry (REIMS), a transformative method that allows for the near-instantaneous chemical analysis of a sample with minimal preparation. We will discuss how REIMS generates a detailed chemical fingerprint that reflects a sample's unique molecular composition.

2.2.1 Traditional Methods

Traditional analytical methods for marine biomass analysis are often labor-intensive, time-consuming, and necessitate extensive sample preparation, creating a bottleneck for the rapid, automated quality assurance required in modern fish processing. These historical approaches, such as traditional chemical assays, contrast starkly with new state-of-the-art analytical techniques and machine learning paradigms now being explored to enhance efficiency and tackle complex fraud and contamination challenges. The table below summarizes these different analytical options, ranging from the well-established methods often used for complex biomass analysis to the emerging, high-speed techniques being investigated as alternatives. This list has been adapated from the book by Vaz Jr (2016), and also includes the methods discussed in this literature review.

- **Capillary Electrophoresis** The principle of this technique is based on the migration of ions or charged particles through a capillary when an electric field is applied. Its primary advantage lies in its high separation efficiency, particularly for polar compounds resulting from biomass degradation. However, its use is limited when analyzing nonpolar compounds.

- **Differential Scanning Calorimetry (DSC)** DSC measures enthalpy (heat flow) changes associated with transitions and reactions in materials. It is advantageous because it requires only a small sample quantity and offers high sensitivity, enabling the determination of physicochemical changes in materials.

- **Mass Spectrometry (MS)** This fundamental technique works by measuring the ratio of mass-to-charge (m/z) of ionized fragments after a sample undergoes molecular fragmentation. It is widely used for the structural identification and quantification of organic compounds based on their m/z ratio. A limitation is that it often requires preliminary separation techniques (like chromatography) for optimal resolution when used alone.

- **X-ray Fluorescence (XRF) Spectroscopy** The principle of XRF is based on the emission of characteristic X-rays when a sample is irradiated. This allows for multielemental quantification in both solid and liquid samples derived from biomass residues. However, the technique can be susceptible to interference from the physical and chemical composition of the sample.

- **Infrared (IR) and Near-Infrared (NIR) Spectroscopy** These methods measure energy absorption resulting from molecular vibrational changes. NIR is considered an easy and efficient technique for the structural identification of lignocellulosic components. A key disadvantage is that it may suffer from low resolution for compounds that share similar functional groups, though this can often be addressed using chemometrics.

- **X-ray Diffractometry (XRD)** XRD works by measuring the intensity of diffracted X-rays. It is a critical tool for determining the crystallinity and chemical composition of cellulose, providing essential structural information for materials science. The main limitation is the long acquisition time required (often hours or days), making it impractical for real-time process control.

- **Scanning Electron Microscopy (SEM)** The principle involves generating an image and analysis by scanning the sample surface with a primary electron beam. SEM provides highly valuable physical information for analyzing the surface and internal structure of materials like natural fibers and polymers. Similar to XRD, it is limited by a long acquisition time.

- **Nuclear Magnetic Resonance (NMR) Spectroscopy** NMR is based on detecting energy changes following the transition of nuclear spin states inside atomic nuclei when exposed to a magnetic field. It offers superior resolution for complex molecular structures and is vital for structural identification in biomass. Its widespread application is sometimes constrained by the long acquisition time required for analysis.

- **Voltammetry** This electrochemical technique measures changes in electric current as a function of an applied potential. Its advantage is the rapid response it offers for chemical speciation and quantification. However, determining the optimal supporting electrolyte or voltammetric technique to use can be a time-consuming step.

- **Light-based** One approach is to use analytical techniques based on light, e.g., UV or fluorescence spectrophotometry, or vibrational spectroscopy (infrared, near-infrared, or Raman spectroscopies). These techniques have been applied in combination with Genetic Programming (GP) for nutrient assessment in horticultural products.

- **DNA Sequencing** This is limited due to extremely low sample size, and very high-dimensional data, e.g., the average human genome contains 3 billion base pairs and 30,000 genes. The dimensionality, and consequently the computation required to process it, rules out genomics data for real-time fish contamination detection. DNA identification methods were examined in a meta-analysis, which revealed an average mislabeling rate of 30% in seafood processing. DNA methods are limited, as they only differentiate between species, and are not useful for determining different body parts from the same species, or non-organic matter (e.g., engine oil).

- **Gas Chromatography-Mass Spectrometry** Previous work demonstrated that GC-MS can identify fish species with high accuracy. However, GC-MS techniques require significant time and domain expertise to prepare and analyze samples. This does not apply to real-time fish contamination detection.

2.2.2 Rapid Evaporative Ionization Mass Spectrometry (REIMS)

Rapid Evaporative Ionization Mass Spectrometry (REIMS) (Schäfer et al., 2009), a key technique in the rapidly expanding field of Ambient Ionisation Mass Spectrometry (AIMS) (Henderson et al., 2025), is a direct-to-analysis technique that allows for the near-instantaneous chemical analysis of a sample with minimal to no preparation (Schäfer et al., 2009; Cafarella et al., 2024; Henderson et al., 2025; Eckel et al., 2025). REIMS has emerged as a transformative technology in response to the need for faster and more accurate analytical methods for biomass analysis (Schäfer et al., 2009; Cafarella et al., 2024). A notable demonstration of this technique was its capability in interoperative tissue identification (Balog et al., 2013), a capability that continues to be validated with high accuracy in complex surgical settings as recently as 2025 (Hendriks et al., 2025; Pruekprasert et al., 2025). REIMS typically involves an electrosurgical knife, sometimes referred to as an iKnife (Balog et al., 2013; Tzafetas et al., 2020; Hendriks et al., 2025), although other sampling devices are also used, such as CO2 lasers (Pruekprasert et al., 2025), laser-assisted REIMS (LA-REIMS) (Cardoso et al., 2025), and novel plasma-based cutters like the "arc iKnife" for analyzing plant tissues (Cai et al., 2025). These devices rapidly heat a sample to generate an aerosol of ionized molecules (Schäfer et al., 2009). That aerosol is then fed into a mass spectrometer, which measures the mass-to-charge (m/z) of the ions, and their relative abundance, illustrated in Figure 3.1. This results in a mass spectrum, which provides a detailed chemical fingerprint of the sample that reflects its unique molecular composition (Golf et al., 2015; Cafarella et al., 2024), often referred to as a metabolic fingerprint (Gkarane et al., 2025). A single m/z value is explicitly referred to as a 'feature' because it represents an ion (such as a deprotonated molecule or adduct) and may not definitively identify a single compound without further fragmentation data, such as tandem mass spectrometry.

Compared to these traditional laboratory methods, REIMS offers significant advantages, especially with its ability to produce near-instantaneous results, making it ideal for in-situ (i.e., in the original place or on-site) applications (Hendriks et al., 2025; Cai et al., 2025; Henderson et al., 2025). As a 2024 review by Cafarella et al. (2024) highlights, its two primary application areas are clinical diagnostics and the agri-food sector. In clinical settings, it is used for real-time tissue classification in neurosurgery (Hendriks et al., 2025), differentiating degrees of fatty liver disease (Eckel et al., 2025), and evaluating venous leg ulcers in outpatient settings (Pruekprasert et al., 2025). Its use in the agri-food sector is a major, rapidly expanding field aimed at safeguarding food quality and security (Cafarella et al., 2024; Henderson et al., 2025). Recent food science applications include fish processing (Black et al., 2017), real-time pork breed and boar taint classification (Gkarane et al., 2025; Verplanken et al., 2017), and accurate lamb origin traceability (Gao et al., 2025). The versatility of REIMS is further shown by its use in entomology, where it can reliably determine the age and species of malaria-vector mosquitoes (Wagner et al., 2025).

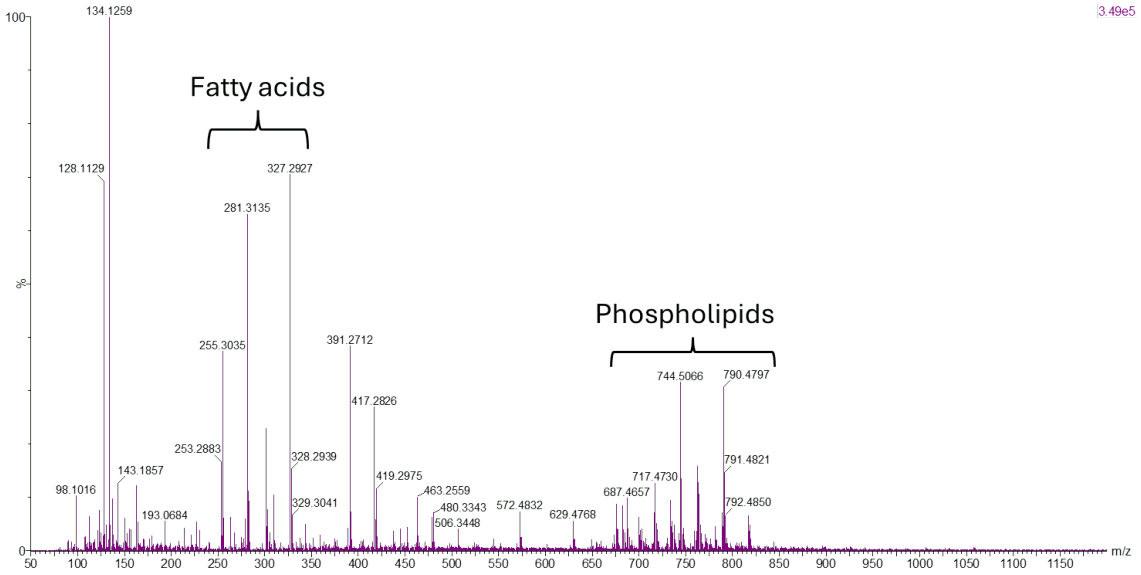

Figure 2.1: A representative REIMS mass spectrum acquired from a marine fish tissue sample using an electrosurgical probe, as displayed in the Waters MassLynx instrument software. The x-axis represents the mass-to-charge ratio (m/z), spanning the analytical range of m/z 77–999, and the y-axis represents the raw ion intensity, with a maximum of 3.49×10^5 counts for this acquisition. Each peak corresponds to the detection of a discrete ionic species produced during electrosurgical ablation of the tissue. The dense cluster of high-intensity peaks in the low m/z region (approximately m/z 77–450) is characteristic of lipid-derived ions, particularly fatty acid and glycerophospholipid fragments, which dominate the REIMS spectral profile of fish tissue. The sparser peaks in the higher m/z region (approximately m/z 700–999) correspond to intact phospholipid species.

REIMS has consistently shown high accuracy in these various applications, with recent studies reporting accuracies of 80-94% for medical diagnostics (Pruekprasert et al., 2025; Wagner et al., 2025), over 89-90% for pork breed identification (Gkarane et al., 2025), 99.14% for lamb origin traceability (Gao et al., 2025), and 100% for in-vivo tumor identification (Hendriks et al., 2025). Other established applications include species of origin identification for meat products (Balog et al., 2016), fish species identification, and catch type identification (Black et al., 2017). Furthermore, it has detected adulteration with offal contamination in minced beef samples at concentrations as low as 1% (Black et al., 2019).

A critical component of these applications, as highlighted in the 2024 review by Cafarella et al. (2024), is the combination of REIMS data with chemometrics and machine learning to build classification models. The vast majority of current studies, in both clinical and food science, rely on traditional multivariate statistical methods. For example, high-accuracy classifications in glioblastoma, leg ulcer, and fatty liver analyses were achieved using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) (Hendriks et al., 2025; Pruekprasert et al., 2025; Eckel et al., 2025). Similarly, recent high-performance models for pork and lamb traceability were built using Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), Support Vector Machines (SVM), and Random Forest (Gkarane et al., 2025; Gao et al., 2025). However, these traditional models are not always sufficient for modeling the highly complex, non-linear relationships within a full mass spectrum. This limitation is a key motivator for the work in this thesis. Recent 2025 research in food science has begun to demonstrate the superiority of deep learning; a study on coffee trait prediction found that Artificial Neural Networks (ANNs), a form of deep learning, significantly outperformed traditional PLS-DA by better capturing the complex relationships between the REIMS profiles and the modelled properties (Cardoso et al., 2025).

2.3 Literature Review: The State of REIMS Data Analysis

The technique of REIMS analysis is relatively new, with existing literature on its data analysis methods falling into two main categories. The first group comprises foundational studies that introduced the REIMS technology and demonstrated its potential in medical applications, primarily for real-time tissue analysis using basic dimensionality reduction and linear classification. The second group builds on this foundation, applying REIMS to the field of food science, which led to the adoption of more advanced chemometrics and, very recently, foundational machine learning models. This section reviews these key works, analyzing their strengths and limitations to contextualize the contributions of this thesis.

2.3.1 Foundational Studies: Establishing REIMS for In-Situ Tissue Analysis

The initial development of REIMS focused on its application as a real-time analytical tool for biological tissue, particularly in medical settings. The pioneering work by Schäfer et al. (2009) introduced REIMS as a novel technique for in vivo, in situ biomass analysis, demonstrating its ability to distinguish between malignant tumor cells and healthy tissue using Principal Component Analysis (PCA). This was advanced by Balog et al. (2010), who presented the first experimental setup for in vivo analysis in rats, achieving over 97% accuracy in tissue identification using a PCA-Linear Discriminant Analysis (PCA-LDA) algorithm. The most high-profile application from this period was the development of the "iKnife" by Balog et al. (2013), which used REIMS to analyze surgical smoke in real-time, matching histological diagnoses with 100% accuracy and bringing significant publicity to the technique.

These foundational studies successfully established REIMS as a revolutionary technology capable of near-instantaneous analysis with no sample preparation. Their primary achievement was demonstrating extremely high accuracy and utility in a demanding real-world application like intraoperative diagnostics.

The analytical methods used in these early stages were limited to PCA and PCA-LDA. These techniques are primarily for dimensionality reduction and linear classification, and they are not well-suited for capturing the complex, non-linear patterns present in high-dimensional mass spectra. This reliance on simpler models represented a key area for future improvement.

2.3.2 Advancements in Food Science: From Chemometrics to Machine Learning

Building on the success in medical diagnostics, researchers began applying REIMS to address critical challenges in the food industry. Balog et al. (2016) Demonstrated REIMS as a fast and effective method for meat product identification, achieving 100% species identification accuracy and detecting adulteration at concentrations as low as 5%. This application to food safety was further explored by Verplanken et al. (2017), who used an Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) model to detect boar taint with over 99% accuracy. The work of Black et al. were particularly influential in this area. Their 2017 study (Black et al., 2017) showed REIMS with OPLS-DA was a rapid, preparation-free alternative to genomic profiling for fish species identification, achieving 98.99% accuracy. In 2019 (Black et al., 2019), they further demonstrated the technique's sensitivity by using it to detect offal adulteration in minced beef at concentrations as low as 1-5%. This application of OPLS-DA for binary classification became a standard, with further studies successfully using it and similar multivariate techniques for identifying fish species like cod vs. oilfish (Shen et al., 2020) and basa vs. sole (Shen et al., 2022).

This body of work successfully transitioned REIMS from a medical tool to a powerful instrument for food fraud and safety applications. The adoption of OPLS-DA represented a methodological step forward from PCA-LDA, providing a more sophisticated supervised learning approach for chemometric analysis.

However, while effective for specific tasks, the reliance on OPLS-DA and similar linear models was a limitation. Recognizing this, more recent research, as chronicled in reviews like Xue et al. (2025), began to explore more advanced machine learning (ML) models to better capture non-linear patterns. This led to direct comparisons between traditional ML (e.g., Random Forest, SVM) and established chemometrics (e.g., PCA-LDA). A key study by De Graeve et al. (2023) on large-scale fish speciation found that while ML models were promising, traditional PCA-LDA could remain more robust and 'industry-proof' on new test sets, highlighting an active debate in the field. Other work pairing ML with REIMS focused on new tasks and instrumentation. For example, Lu et al. (2024) successfully applied a K-Nearest Neighbors (KNN) model for the geographical authentication of fish, demonstrating a new application.

Most recently, this trend has progressed to deep learning, specifically Artificial Neural Networks (ANNs). In a study highly relevant to this thesis, Cardoso et al. (2025) demonstrated that ANNs outperformed traditional PLS-DA models for predicting coffee traits from LA-REIMS data. This work not only showed the superior predictive power of ANNs but also introduced novel methods for interpreting their model weights, beginning to address the 'black-box' problem.

To the best of the author's knowledge, the work of Cardoso et al. (2025) and the present thesis represent the only two studies to have applied deep learning methods directly to REIMS data, making a detailed comparison necessary rather than optional. The two studies are complementary, differing along several key dimensions. First, with respect to instrumentation: Cardoso et al. employ laser-assisted REIMS (LA-REIMS), whereas this thesis uses conventional electrosurgical REIMS. Second, with respect to substrate: Cardoso et al. analyse coffee beans, while this thesis analyses marine biomass (Hoki and Mackerel). Third, with respect to analytical task: Cardoso et al. address a regression problem — predicting continuous sensory trait scores from mass spectra — and demonstrate that foundational Artificial Neural Networks (ANNs) outperform PLS-DA for this purpose. This thesis addresses a suite of classification problems using substantially more advanced architectures: encoder-only Transformer variants, Mixture of Experts (MoE) Transformers, and a self-supervised contrastive learning framework (SpectroSim). The MoE Transformer used here is conceptually related to the sparse MoE paradigm explored in large language models Kaiser et al. (2017); Guo et al. (2025); Jaech et al. (2024), but differs fundamentally: it is an encoder-only model for one-dimensional mass spectrometry sequences, not an autoregressive text-generation architecture, and its experts are multi-scale Transformers of varying depth rather than identically structured feed-forward networks. Fourth, with respect to interpretability: both studies identify the black-box

nature of neural networks as a challenge, but address it differently — Cardoso et al. analyse ANN weight matrices, while this thesis employs LIME and Grad-CAM, which are applicable to the more complex Transformer architectures used here. Taken together, the two works reinforce one another: Cardoso et al. establish that neural methods can outperform chemometrics for REIMS-based regression in food science, while this thesis extends that finding to a broader set of classification tasks in marine biomass analysis, using more advanced architectures and a richer suite of interpretability tools. The claim to novelty is therefore robust against comparison with Cardoso et al.: the substrates, instruments, tasks, architectures, and analytical objectives are distinct.

2.3.3 Connecting to the Contributions of This Thesis

The limitations and recent advancements of the existing literature directly motivate the core contributions of this thesis. While prior work established REIMS as a powerful data acquisition tool and recently began applying foundational ML and ANNs, it highlighted a clear need for state-of-the-art analytical methods to unlock its full potential. This thesis addresses these gaps in three key ways:

1. Addressing the Analytical Gap: Where previous studies relied on traditional linear models (PCA-LDA, OPLS-DA) and, more recently, conventional ML (KNN, SVM) and foundational ANNs (Cardoso et al., 2025), this thesis introduces a suite of state-of-the-art deep learning architectures (e.g., Transformers, Mixture of Experts) capable of learning the complex, sequential, and non-linear patterns inherent in REIMS data.

2. Addressing the Task Gap: This thesis moves beyond foundational classification (species identification, geographical origin) by formalizing and solving novel and more challenging problems, such as detecting graded levels of oil contamination and developing a completely new, self-supervised method for batch traceability ("SpectroSim").

3. Enhancing Trust and Interpretability: While recent work has begun to develop methods for interpreting ANNs (Cardoso et al., 2025), this remains a significant challenge for more complex architectures. To overcome the black-box nature of these models, this thesis integrates advanced Explainable AI (XAI) techniques (LIME and Grad-CAM), ensuring that model decisions are transparent and verifiable by domain experts, a crucial step for real-world adoption.

2.4 Advanced Deep Learning Architectures for REIMS Analysis

The limitations of traditional statistical and foundational ML methods, as outlined in Section 2.3, have driven the exploration of sophisticated deep learning techniques to unlock the full potential of REIMS data. This section introduces the advanced architectures and strategies that form the core of this thesis's solution.

2.4.1 Deep Learning

Deep learning is a specialized subset of machine learning often based on artificial neural networks with multiple layers (i.e., deep architectures) to automatically learn hierarchical feature representations from raw data. The term deep refers to the depth of these layered networks, which can contain dozens or even hundreds of interconnected layers. Deep learning has been incredibly successful in handling complex, high-dimensional data across various domains (Goodfellow et al., 2016). The aforementioned capability to learn hierarchical feature representations from raw data makes them ideal for capturing the complex relationships found in REIMS mass spectra.

In order to appreciate it fully, here is a brief bibliographical history of deep learning, in chronological order. This is not an exhaustive list, but a representative sampling of some key papers that inform deep learning methodology and inspired the methods used in this thesis. In particular, those papers are:

- **Linnainmaa (1970)** Seppo Linnainmaa's Master's thesis was not initially about neural networks. However, it laid the mathematical foundation for the backpropagation algorithms. He proposed the reverse mode of automatic differentiation, a technique for efficiently calculating the gradient of a composite function by accumulating derivatives backwards from the output. This method proved most useful when you needed to compute the gradient of a function with a large number of input variables and a small number of output variables. Take, for example, a neural network, where the goal is to minimize a scalar loss function concerning millions of network weights. It would take over a decade for the application of neural networks to be found and popularized, Linnainmaa's work provided the foundational algorithm behind backpropagation for training deep, complex neural architectures.

- **Rumelhart et al. (1985)** This paper helped popularize the backpropagation algorithm for training Multi-layer Perceptrons (MLPs), or deep feedforward neural networks. While the core idea of backpropagation has already been discovered, this paper articulates the method and demonstrates its potential for learning complex internal representations in hidden layers. Rumelhart, Hinton, and Williams showed that error could be efficiently propagated backward through the network to adjust

the weights. This paper is widely credited with starting the connectionist revival of the 1980s, and it lays the foundation for subsequent advances in deep learning.

- **LeCun et al. (1989b)** Yann LeCun and his colleagues at Bell Labs provided one of the earliest and most convincing applications of backpropagation, particularly when combined with their Convolutional Neural Networks (CNNs). Their application of backpropagation to CNNs was for recognizing handwritten zip codes for the U.S. Postal Service. It demonstrates the ability of these networks to learn hierarchical features directly from raw pixel data, performing automatic feature extraction, it achieving high accuracy and robustness. They create a system that could outperform traditional pattern recognition techniques by integrating concepts like local receptive fields, shared weights, and backpropagation for training. The success of both validated backpropagation for real-world tasks, and also demonstrated the efficacy of the CNN architecture.

- **Bromley et al. (1993)** Jane Bromley, Yann LeCun, and his colleagues developed Siamese networks for pair-wise comparison tasks. Siamese networks were first developed for signature verification, a task that involves comparing a known authentic signature with a query signature to determine if they were written by the same person. The goal was to predict whether a signature was genuine or forged. Siamese networks achieve this by using two identical neural networks that share the same weights and architecture. One network processes a reference input (e.g., an authentic signature or a known sample), while the other processes the query input (e.g., the signature being tested or an unknown sample). The outputs from both networks are then combined using a distance metric to produce a similarity score. For instance, the original paper used Euclidean distance; a smaller distance indicated higher similarity, suggesting the query was genuine or from the same source, while a larger distance suggested it was forged or from a different source.

- **Hochreiter & Schmidhuber (1997)** Sepp Hochreiter and Jürgen Schmidhuber introduced the Long Short-Term Memory (LSTM) architecture. The LSTM addresses the vanishing gradient problem that plagued simple Recurrent Neural Networks (RNNs) and produced unstable training dynamics for tasks involving long-term dependencies. LSTMs integrate specialized memory cells and gating mechanisms (i.e., input, output, and forget gates) that allow the network to selectively remember or forget information over time. This design enables LSTMs to have long-term memory, allowing them to capture and exploit context from distant past inputs. This made them exceptional for a wide variety of sequence modelling tasks, e.g., speech recognition, machine translation, and time series analysis, where they were a dominant approach for many years.

- **LeCun et al. (1998)** Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick

Haffner designed the LeNet-5 architecture, a Convolutional Neural Network (CNN) that was highly optimized for document recognition tasks, i.e., handwritten digit and character recognition. LeNet-5 integrated convolutions with learnable filters, subsampling (i.e., pooling) layers for spatial invariance, and fully connected layers, all trained with end-to-end gradient-based learning. The paper highlighted the importance of automatic feature extraction. LeNet-5 is a foundational model that influenced later architectures and breakthroughs in computer vision.

- **Krizhevsky et al. (2012)** Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton developed "AlexNet," a deep convolutional neural network that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, beating the other traditional computer vision models of the time. AlexNet's success was attributed to its deep architecture, the use of Rectified Linear Units (ReLUs) activation function, dropout for regularization, and efficient GPU training. This watershed moment convinced the computer vision community of the power of deep learning. This paper, the advent of big data, and GPU training, together unfroze the AI winter and ignited the modern deep learning revolution.

- **Kingma & Welling (2013)** Diederik Kingma and Max Welling introduced Variational Autoencoders (VAEs), providing an influential framework for deep generative modelling by combining concepts from variational Bayesian methods and neural networks. VAEs learn a probability distribution of the input data, allowing them to generate synthetic data samples that were not present in the training set. VAES learn a probabilistic encoder that maps input data to a distribution in the latent space (e.g., typically Gaussian), and a probabilistic decoder that maps points from the latent space back to the data space, i.e., the decoder reconstructs the data from the latent representation.

- **Goodfellow et al. (2014)** Ian Goodfellow and his collaborators introduced Generative Adversarial Networks (GANs) as a new framework for generative AI. A GAN has two neural networks, a generator and a discriminator, trained simultaneously in a min-max game to reach a Nash equilibrium: the generator learns to produce data that matches the training set, and the discriminator learns to distinguish the training set from the synthetic data. This adversarial process leads to the generator producing more realistic samples over time. GANs quickly became one of the most popular ideas in deep learning. It leads to rapid progress in generating high-fidelity images, videos, and more. GANs inspired a vast amount of research in generative AI and unsupervised learning.

- **Kingma & Ba (2014)** Diederik P. Kingma and Jimmy Ba introduced Adaptive Moment Estimation (Adam), a highly popular and effective optimization algorithm used in deep learning. Adam computes adaptive learning rates for each parameter.

By integrating estimates of both the first moment (i.e., mean) and the second moment (i.e., uncentered variance) of the gradients, Adam combines the key ideas of Momentum and RMSprop, two other popular optimization algorithms. It is computationally efficient, requires little memory, and is well-suited for large-scale machine-learning problems with many parameters. Due to strong empirical performance and ease of use, Adam quickly became the most popular and often default optimization algorithm for training deep learning models.

- **Srivastava et al. (2014)** Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov introduced Dropout, a regularization technique to combat overfitting in large neural networks. Enabled only during training, Dropout randomly turns off a subnetwork of neurons for each training batch, preventing neurons from co-adapting too much and encouraging them to learn more robust features. Conceptually, it efficiently approximates a bagged ensemble of sub-neural networks. Dropout significantly improved the generalization performance in most cases and has now become an indispensable tool for deep learning practitioners.

- **He et al. (2016)** Kaiming He and his colleagues introduced Deep Residual Networks (ResNets) to address the problem of degradation in very deep networks (where deeper models performed worse). ResNet, an extension of Convolutional Neural Networks (CNNs), employs skip connections or shortcuts between layers. This creates gradient superhighways that allow gradients to propagate more easily through extremely deep architectures. This made it feasible to train neural networks with hundreds of layers. On computer vision benchmarks like ImageNet, ResNet achieved record-breaking results, demonstrating that, when managed correctly, network depth is crucial for improving model performance. This has influenced future network designs profoundly.

- **Vaswani et al. (2017)** Vaswani and his colleagues introduced the Transformer architecture, which revolutionized sequence modeling. Departing from traditional deep learning architectures such as convolutional and recurrent layers for sequence-to-sequence tasks, the Transformer uses attention mechanisms, specifically self-attention, to draw global dependencies between input and output. Its ability for parallelism allows for training on much larger datasets when compared to RNNs. Some more key innovations for the transformer were multi-head attention and positional encodings. The Transformer's effectiveness, efficiency, and scalability quickly made it the default architecture for NLP tasks, and this architecture has also been successfully applied to many other domains.

- **Devlin et al. (2018)** Jacob Devlin and his colleague introduced Bidirectional Encoder Representations from Transformers (BERT), building upon Transformers and pre-training. BERT was designed to pretrain deep bidirectional representations by

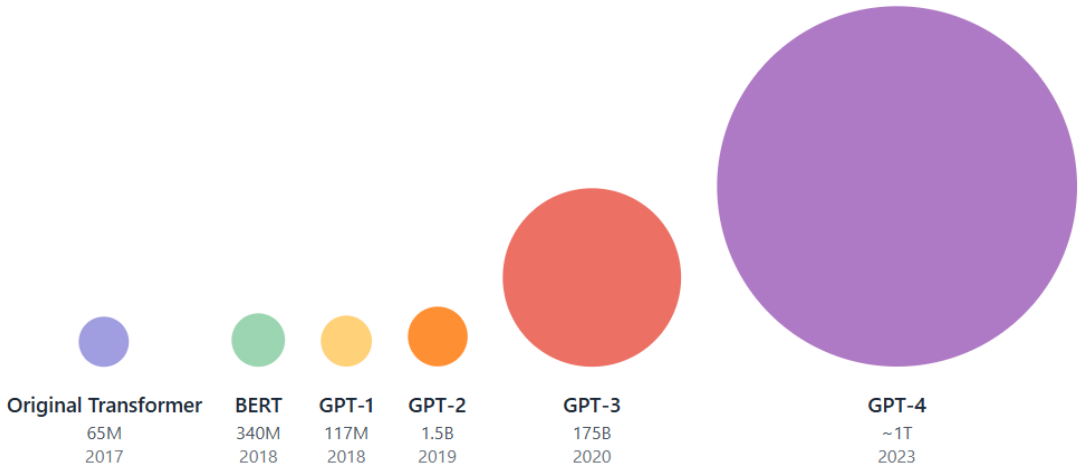

Figure 2.2: Illustrated here is a comparison of Transformer models, with each circle's size representing its approximate number of parameters. This visualization highlights the prominent scaling laws observed in large language models, particularly the exponential increase in parameters across successive iterations of the GPT series (GPT-1, GPT-2, GPT-3, GPT-4). For context, the original Transformer and BERT models are also included.

joint conditioning on both left and right context in all layers, using novel pre-training tasks, such as Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). Once pretrained on a large text corpus, BERT could be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of NMLP tasks without substantial task-specific architecture changes. This ushered in the era of large-scale pretrained language models.

- **Brown et al. (2020)** Brown and his colleagues unveiled Generative Pre-Trained Transformer 3 (GPT-3). A large language model with 175 billion parameters, proving the extreme scaling capabilities of the Transformer architecture. The authors demonstrated that a dramatic increase in model size, dataset, and computation could achieve remarkable performance on many NLP tasks without any fine-tuning, with GPT-3 matching or exceeding state-of-the-art fine-tuned models. The so-called emergent capabilities of large language models (LLMs), such as few-shot or in-context learning (i.e., where models could perform a new task without retraining by being provided a natural language description and a few examples as an input prompt), have inspired a new wave of deep learning research into the scaling of LLMs. The scaling laws of LLMs are demonstrated in Figure 2.2.

- **Ho et al. (2020)** Jonathan Ho and his collaborators introduced Denoising Diffusion

Probabilistic Models (DDPMs), a new class of generative AI models based on diffusion. These diffusion models achieved state-of-the-art results in image synthesis, dethroning GANs from their previous perch. Diffusion models work by systematically destroying an image through a forward diffusion process (i.e., gradually adding noise) and then learning a reverse diffusion process that restores the image. This new approach, with a specific training objective and sampling procedure, demonstrated high efficacy for generating high-fidelity images. DDPM's success led to a large interest in diffusion models within the deep learning community. Now, many state-of-the-art image, audio, and video generation systems consist of diffusion models.

- **Gu & Dao (2023)** Albert Gu and Tri Dao introduced Mamba, a type of State Space Model (SSM) that incorporates a selection mechanism inspired by gating in RNNs and the efficiency of convolutional approaches. It serves as an efficient alternative to Transformer-based architectures, especially for very long sequences. Its key contribution is the capability to model long-range dependencies with linear time complexity with regard to sequence length, a large improvement over the quadratic complexity of standard Transformers. This is achieved through a hardware-aware design that leverages parallel scans for efficient computation during training and inference.

- **Liu et al. (2024)** Inspired by the Kolmogorov-Arnold representation theorem, Ziming Liu, Max Tegmark, and colleagues recently introduced Kolmogorov-Arnold Networks (KANs). This novel neural network architecture distinguishes itself by placing learnable activation functions (modeled as splines) on its edges, while nodes perform simple summations. A key advantage of KANs is their enhanced interpretability, allowing a direct mathematical expression of the network to be derived, thereby mitigating the black-box nature of traditional neural networks. Their design also grants them high efficiency in approximating complex functions.

2.4.2 Stacking Ensembles

Ensemble theory, through the lens of Wolpert's Stacked Generalization (Wolpert, 1992), is a framework for improving the predictive accuracy of machine learning by combining multiple models. The central goal is to create a more powerful and robust classifier than any single model could achieve on its own by leveraging the wisdom of the crowds. This approach is particularly effective when the individual models are independent, complementary, and diverse.

The architecture proposed by Wolpert consists of two levels. The first level contains multiple independent base models, referred to as level-0 generalizers. The predictions from these base models are then used as input features for a single, higher-level level-1 general-

izer, or meta-model. In a stacked voting ensemble, this meta-model is often implemented as a learnable weighted combination of the predictions from the level-0 models.

The objective of this framework is for the meta-model to learn how to optimally combine the outputs from the base models. By training the weights of the voting classifier, the ensemble learns to correct for the individual errors and biases of each base model. This intelligent combination of diverse models, each with its own strengths and weaknesses, ultimately reduces the overall generalization error and leads to a more robust final prediction. This principle is most effective when the errors of the base models are not perfectly correlated, as their individual idiosyncrasies can then cancel each other out during the aggregation process.

2.4.3 Transformers

Transformers (Vaswani et al., 2017) have gained significant prominence among deep learning architectures. Initially, they were only applied to the field of Natural Language Processing (NLP) (Ansar et al., 2024; Fields et al., 2024), but now they have been applied to a variety of domains (Lai-Dang, 2024; Li et al., 2021; Lee et al., 2024). Some key contributions of the Transformer are the self-attention mechanism and multi-head attention (note: positional embeddings are excluded as we opt for No Positional Embeddings (NoPE) (Wang et al., 2024) in the transformer models presented in this thesis).

Self-Attention

An **attention** function transforms a query vector and a collection of key-value vector pairs into an output vector. This output is calculated as a weighted sum of the value vectors, with the weight for each value derived from a compatibility measure between the query and its associated key.

For each input element, three learned linear transformations are applied to generate:

- **Query** (`Q`): What an element is looking for.
- **Key** (`K`): What an element offers.
- **Value** (`V`): The actual information to be aggregated.

Given a query, key, and value, we compute the Scaled Dot Product Attention as follows:

$$\texttt{Attention}(Q, K, V) = \texttt{softmax}(\frac{QK^\top}{\sqrt{d_k}})V$$

Self-attention operates within a single sequence. It allows each element of the sequence to learn its relationship and importance to all the other elements within the same sequence. In self-attention, the query, key, and value vectors are all derived from the same

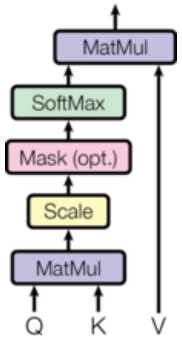

Figure 2.3: This figure illustrates the Scaled Dot-Product Attention mechanism (Rush, 2018), a core component of self-attention. It takes queries (Q) and keys (K) of dimension d_k, and values (V) of dimension d_v. Attention weights are computed by taking the dot product of Q with all **K**s, scaling by $\sqrt{d_k}$, and applying a softmax function. These weights are then applied to the **V**s to produce the output.

sequence. Unlike Recurrent Neural Networks (RNNs) (Rumelhart et al., 1985) that process sequentially, self-attention can be done in parallel. Therefore, it can efficiently model long-range dependencies and contextual relationships regardless of their distance.

Multi-head Attention

Multi-Head Attention is an extension of the self-attention mechanism, designed to enhance the model's ability to focus on different parts of the input sequence from various perspectives or representation subspaces.

While single self-attention may restrict a model to learning only one type of relationship, Multi-Head Attention addresses this by performing the attention mechanism multiple times in parallel, each with its own independent set of learned linear projections. This allows the model to jointly attend to information from diverse representation subspaces at different positions, thereby learning richer and more varied relational patterns.

Specifically, for H heads, the input Query (Q), Key (K), and Value (V) matrices are first linearly projected into distinct lower-dimensional subspaces for each head. The standard scaled dot-product attention is then applied independently to these projected Q_h,K_h, and V_h sets. Finally, the outputs from all H heads are concatenated and linearly transformed back to the original model dimension, allowing the model to collectively leverage diverse contextual information. This process is illustrated in Figure 2.4.

2.4.4 Mixture of Experts (MoE)

The **Mixture of Experts (MoE)** architecture is a neural network design that combines multiple specialized subnetworks, known as experts, coordinated by a trainable gating network (Jacobs et al., 1991). The gating network dynamically routes input data (or

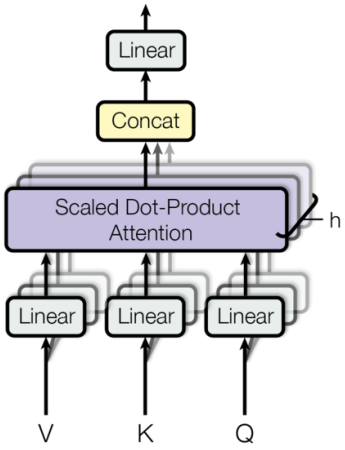

Figure 2.4: Multi-head attention (Rush, 2018) enhances a model's ability to capture complex relationships by allowing it to jointly focus on distinct representation subspaces at different positions. A single attention head's averaging effect would otherwise obscure such nuanced information.

parts of it) to the expert(s) deemed most suitable for processing that specific information. This selective activation allows MoE models to scale to a very large number of parameters without a proportional increase in computational cost per inference, as only relevant experts are activated for any given input. When integrated into Transformer architectures (MoE Transformers) (Kaiser et al., 2017; Guo et al., 2025; Jaech et al., 2024), typically by replacing standard feed-forward network layers with MoE layers, this approach can enhance model capacity and adaptability, allowing the model to develop specialized representations for different aspects of the data or tasks (Kaiser et al., 2017). This makes MoE Transformers particularly promising for handling diverse and complex datasets like those generated by REIMS across various fraud detection tasks.

2.4.5 Pretrained Transformers

Unsupervised pretraining is a powerful technique where a Transformer is pretrained to learn general-purpose embeddings from data that doesn't explicitly require labels (i.e., unsupervised) before being fine-tuned on a specific downstream task. This technique allows the Transformer to capture the underlying structure of the data, which is helpful when training data is scarce. Two prominent approaches to pre-training are demonstrated by GPT (Radford et al., 2018) and BERT (Devlin et al., 2018). The Generative Pretrained Transformer (GPT) ((Radford et al., 2018)) was trained to predict the next word in sequence, learning a left-to-right unidirectional representation. Alternatively, BERT (Devlin et al., 2018) introduced Masked Language Modelling (MLM) as a pretraining task.

In MLM, some tokens in the input sequence are randomly masked, and the model is trained to predict the masked tokens using the context. This forces the model to learn deep bidirectional relationships within the data, creating robust embeddings to transfer to downstream tasks.

2.4.6 Transfer Learning

Transfer learning (Bozinovski & Fulgosi, 1976; Tan et al., 2018) is a machine learning paradigm where knowledge gained from solving one problem (the source task) is applied to a different but related problem (the target task). In practice, a model is often pre-trained on a large dataset for a general task and then its learned parameters and feature representations are transferred to initialize a new model for a target task, which is then fine-tuned on a smaller, task-specific dataset. This technique can significantly reduce the amount of labeled data needed for the target task, shorten training times, and often improve performance, especially when the source and target domains share underlying patterns. However, the effectiveness of transfer learning can be asymmetric and task-dependent. While it can yield significant improvements in some cases, it may offer little benefit or even degrade performance in others. Understanding these dynamics is crucial for their effective application.

2.4.7 Contrastive Learning

Contrastive learning is a machine learning technique that learns effective representations by contrasting positive and negative pairs of data instances—distinct from class labels—mapping similar instances closer in the embedding space and dissimilar ones further apart. It excels in supervised and self-supervised settings, notably in computer vision (Bromley et al., 1993; Jing et al., 2022) and natural language processing (Zhu et al., 2022). In supervised contrastive learning, labeled data are used to train models to distinguish similar from dissimilar instances, while in self-supervised learning, models are trained using unlabeled data, forming pairs via augmentation to capture specific features like edges or semantic similarities, enhancing performance over traditional methods in tasks like classification.

SimCLR (Chen et al., 2020a), exemplifies self-supervised contrastive learning, using a backbone like ResNet (He et al., 2016) to process augmented views of the same input. It applies NT-Xent loss to align representations of these views (positive pairs) while separating different inputs (negative pairs), identifying positive pairs as augmented versions of the same data point without needing labels. This enables pre-training on large unlabeled datasets for downstream tasks like image classification and object detection, leveraging augmentation for effective representation learning.

However, SimCLR faces several limitations. It demands high computational resources, necessitating large batch sizes and extensive training for optimal performance (Chen et al.,

2020b). The method's effectiveness is heavily dependent on data augmentations, as poor choices—like random cropping—can significantly impair the quality of learned representations (Chen et al., 2020a). Furthermore, SimCLR struggles with small datasets, often requiring specific tweaks to be effective in limited data regimes (Kinakh et al., 2021). Finally, without robust augmentation, SimCLR can lose its self-supervised nature, inadvertently relying on labeled data for effective representation learning (Chen et al., 2020a).

2.5 Explainable AI for Model Trust and Interpretability

A significant challenge in deploying advanced deep learning models, as noted in recent REIMS literature (Cardoso et al., 2025), is their "black-box" nature. To address this and build trust for real-world adoption, this thesis integrates Explainable AI (XAI) techniques.

Ribeiro et al. (2016) introduced **Local Interpretable Model-agnostic Explanations (LIME)**, a post-hoc explainable AI technique that can explain the predictions of any black-box model, particularly applicable to those from deep learning. LIME approximates a complex model with a simpler interpretable one (for example, a linear regression model) in the local area of a specific instance. LIME creates many perturbed versions of an input instance to see how the predictions change. Through this process, it identifies the important features that influence the model's decisions for that instance. This thesis uses LIME explanations to explore the black-box Transformer models and their variants proposed.

Selvaraju et al. (2017) proposed **Gradient-weighted Class Activation Mapping (Grad-CAM)**, another technique for producing visual explanations of the decision process behind deep learning models. Grad-CAM works by using the gradients of the target class that flow through the final feature extraction layer of the model during backpropagation. This process generates an "average correct Grad-CAM," which highlights important features that consistently contribute to accurate classification. This thesis implements a 1-dimensional Grad-CAM for Transformers that utilizes the gradients of the final attention layer to identify the important features behind correct classifications.

2.6 Summary

The challenge of food safety in fish processing requires a move beyond traditional analytical methods. The literature shows a clear progression: from foundational PCA/PCA-LDA and OPLS-DA to more recent explorations of conventional ML and foundational ANNs. While REIMS provides a powerful tool for rapid data acquisition, its true potential can only be realized by coupling it with advanced machine learning techniques that can model the complex, high-dimensional, and sequential nature of its data. This thesis directly addresses this gap. It proposes state-of-the-art deep learning architectures, particularly

Transformers and their MoE variants, to solve novel, complex tasks like graded contamination and self-supervised batch traceability. Furthermore, by integrating explainable AI (XAI), it addresses the critical need for model interpretability, paving the way for robust, automated, and trustworthy systems for quality assurance in fish processing.

3

Datasets and Processing

With the foundational knowledge established, Chapter 3 details the empirical basis of this research. This chapter explains the data pipeline in three stages. First, Section 3.1 describes the **Data Source and Acquisition**, detailing the origin of the REIMS data from AgResearch, New Zealand, and the specifics of its laser-assisted generation. Second, Section 3.2 explains the **Data Curation and Preprocessing**, showing how the raw spectral signal was converted into a high-dimensional data matrix and then normalized for consistency. Finally, Section 3.3 provides a detailed breakdown of the **Task-Specific Datasets**, explaining how the final data matrix was curated and split to create the five distinct classification tasks that form the basis of this thesis's experiments.

3.1 Data Source and Acquisition

The research described in this thesis utilizes a series of datasets derived from Rapid Evaporative Ionization Mass Spectrometry (REIMS) analysis of seafood samples. This data was provided by AgResearch, which aims to develop quality assurance systems for marine biomass processing (AgResearch, 2025). An example of a REIMS mass spectrograph is given in Figure 3.1. The utility of REIMS extends to a wide range of food analyses, enabling researchers to make complex and rapid measurements of meat composition to assess quality and detect fraud (Ross et al., 2021), discriminate between milk samples fermented with different starter cultures (Murphy et al., 2021), and even differentiate lamb meat based on various aging methods and dehydration levels (Zhang et al., 2023).

The REIMS data were collected at AgResearch (now Bioeconomy Science Institute). All samples were collected following standard AgResearch animal ethics and welfare protocols. The analysis utilized a Laser-Assisted REIMS setup, coupling a CO_2 laser interface to a Xevo G2 XS quadrupole time-of-flight mass spectrometer (Waters Ltd, Wilmslow, UK)

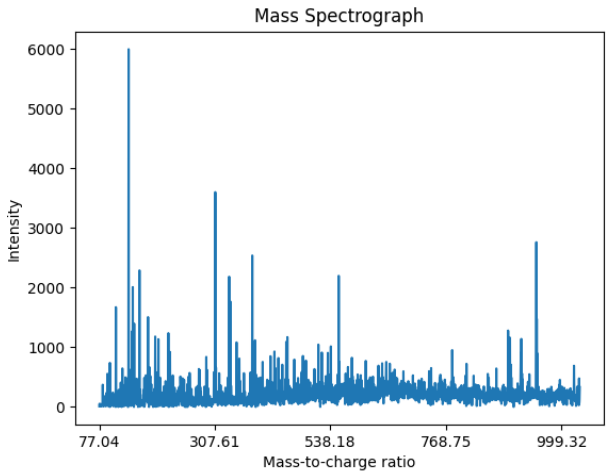

Figure 3.1: This is the artifact generated by Rapid Evaporative Ionization Mass Spectrometry (REIMS). The x-axis depicts the mass-to-charge ratio (m/z), and the y-axis depicts the relative abundance or intensity. The peaks on the chart represent the detection of ions. A taller peak indicates a greater abundance of ions with that particular mass-to-charge ratio detected by the instrument. This example is drawn from the AgResearch New Zealand dataset used in this thesis, and is representative of the spectral profiles from which all experimental results are derived.

for ionization and analysis. Samples were received frozen, and all analyses were performed within 10 minutes of removal from the freezer to prevent lipid oxidation.

During acquisition, the MS was run in negative ionization mode (MS^1 only, with no collision energy applied). This mode is commonly used in REIMS analysis as it preferentially ionizes and detects deprotonated species of lipids and fatty acids, which are key biomarkers in biological tissues. The MS scanned between m/z 50–1200 at a scan rate of 2 Hz. Isopropanol was infused at a rate of 200 µL/min to enhance ionization.

3.2 Data Curation and Preprocessing

Data processing began with ProGenesis Bridge (Waters Ltd), which performed baseline removal and lock mass correction (using oleic acid, C18 : 1) to maintain mass accuracy. Further processing with ProGenesis QI generated the final spectra matrix by picking individual mass features across the range. Each spectrum comprises $2,080$ distinct m/z features, generally spanning a range from approximately m/z 77.04 to m/z 999.32. Triplicate measurements per biological replicate were collected, averaged, and merged with the metadata to create the final data matrix. The high dimensionality (2,080 features), coupled with often limited sample sizes for specific tasks, presents a “curse of dimensionality” challenge

for many traditional machine learning algorithms (Köppen, 2000).

3.2.1 Normalization

A standard preprocessing step applied across all datasets is normalization. Specifically, the raw REIMS spectral data underwent a two-stage normalization process. First, the data were normalized based on total intensity Total Ion Count (TIC) across the mass spectral range (van den Berg et al., 2006). This crucial step accounts for sample-to-sample variations in ionization efficiency due to subtle differences in tissue characteristics or instrument performance.

Following this, the m/z intensity values were further scaled to fall within the range of $x \in [0, 1]$ using min-max normalization. The formula for this normalization is:

$$X' = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{3.1}$$

Where X_i is the intensity value after TIC normalization, $X_{\min}$ is the minimum value of that feature in the training set, and $X_{\max}$ is the maximum value of that feature in the training set. This final scaling helps to prevent features with larger absolute values from dominating distance calculations or gradient updates in certain machine learning algorithms.

3.3 Task-Specific Datasets

The final curated and normalized data matrix was used to create five distinct datasets, each tailored to a specific classification task. The experimental methodology differed between tasks based on dataset size and structure.

For the first four tasks (Species, Body Part, Oil Contamination, and Adulteration), we employed a stratified k-fold cross-validation strategy. This approach is ideal for these datasets as it ensures that each fold (i.e., each train/test split) contains a representative distribution of all classes, which is especially important for the imbalanced "Body Part" dataset. For most tasks, a 5-fold split was used (80% train, 20% test); for the small "Body Part" dataset, a 3-fold split was used (∼67% train, ∼33% test).

For the fifth task, "Batch Detection," a fixed stratified split was used due to its relatively much larger number of samples. The dataset was permanently divided into 60% training, 20% validation, and 20% testing sets, with the extreme class imbalance (2.82% positive class) being preserved across all three sets.

3.3.1 Species Identification

The task is to distinguish between two key New Zealand fish species: Hoki (New Zealand's largest fishery) and Mackerel, illustrated in Figure 3.2, based on 2080 features derived

(a) Mackerel (b) Hoki

Figure 3.2: Mackerel (left), Hoki (right) fish species.

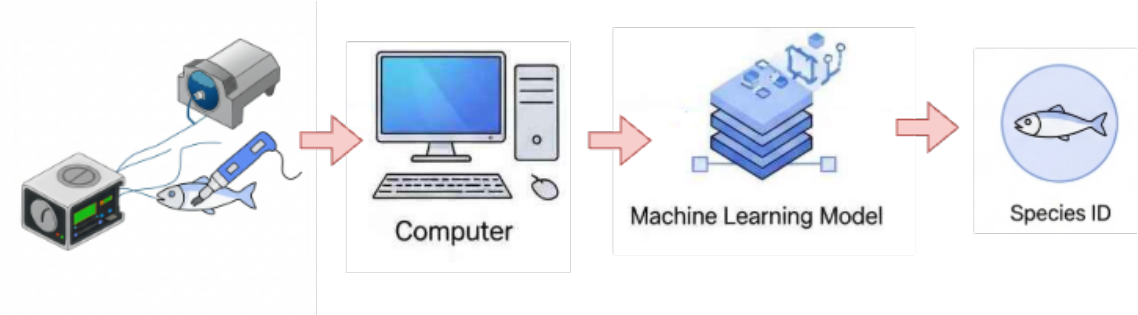

Figure 3.3: This diagram illustrates a machine learning model running on a computer to identify fish species. First, the fish is sampled with an electrosurgical knife, turned into an aerosol, and fed into a mass spectrometer. That information is converted into an Excel spreadsheet on a computer. That Excel spreadsheet is preprocessed and normalized, then fed to a machine learning model. That machine learning model is a classifier that performs species identification between hoki and mackerel.

from REIMS analysis. Figure 3.3 illustrates the fish species classification workflow. This classification is crucial for food authentication and quality control in the seafood industry, helping prevent species substitution fraud and ensure accurate product labelling. We focus on pure (non-contaminated/non-mixed) samples to establish a reliable baseline for species identification. The dataset contains 106 samples, with a relatively balanced class distribution of 44.44% Hoki and 55.56% Mackerel. These proportions reflect the natural availability of samples while maintaining sufficient representation for both species to train a robust classifier for species verification.

3.3.2 Body Part Identification

This multi-class classification task aims to identify seven distinct fish parts (fillets, heads, livers, skins, gonads, guts, and frames) using REIMS data. Figure 3.4 illustrates the fish parts classification task workflow. The classification supports process automation by enabling automated sorting and processing in seafood production lines, helps maximize the value of each fish part (e.g., using fillets for premium products and frames for fish meal),

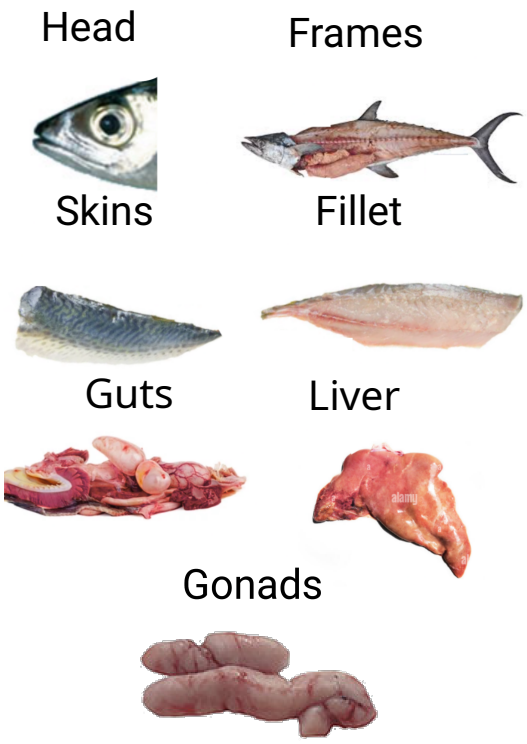

Figure 3.4: Fish body parts.

and ensures proper tracking and documentation of different fish components throughout the supply chain. The dataset consists of 33 samples, presenting a multi-class classification challenge with some class imbalance: fillets, heads, livers, skins, and guts each represent approximately 16.66%, while gonads and frames each represent 8.33%. The relatively small sample size per class is attributed to a limited number of annotated samples for each class of body part.

3.3.3 Oil Contamination Detection

Oil contamination detection is crucial in fish processing as contamination can occur at multiple stages: from boat engine oil during harvesting, processing equipment lubricants, or mechanical systems in the factory. Figure 3.6 gives an exaggerated example comparing healthy fish to those contaminated with oil. For assessing product safety, a dataset of 126 samples was used to detect oil contamination at seven distinct concentration levels: 50%, 25%, 10%, 5%, 1%, 0.1%, and 0% (control group). The seven concentration levels were chosen to cover the full range from heavy contamination (50%) to trace amounts (0.1%), allowing the model to detect contamination across all practically relevant scenarios. Each class was equally represented (14.28% of samples), allowing for the training of models to identify varying degrees of contamination with equal importance.

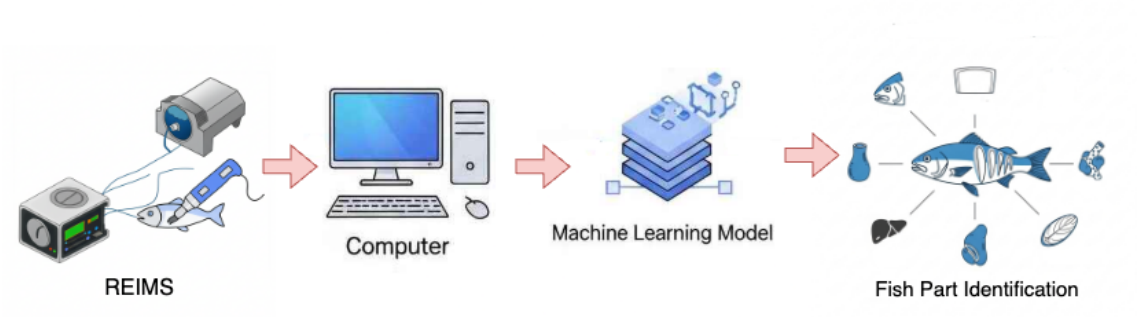

Figure 3.5: This diagram illustrates a machine learning model running on a computer to identify fish body parts. Similar to the previous diagram for fish species identification, we follow the same process. REIMS mass spectrometry data is captured from a fish sample. It is converted into an Excel spreadsheet. Then, it was preprocessed and fed into a machine learning model. That machine learning model is a classifier that performs fish part identification.

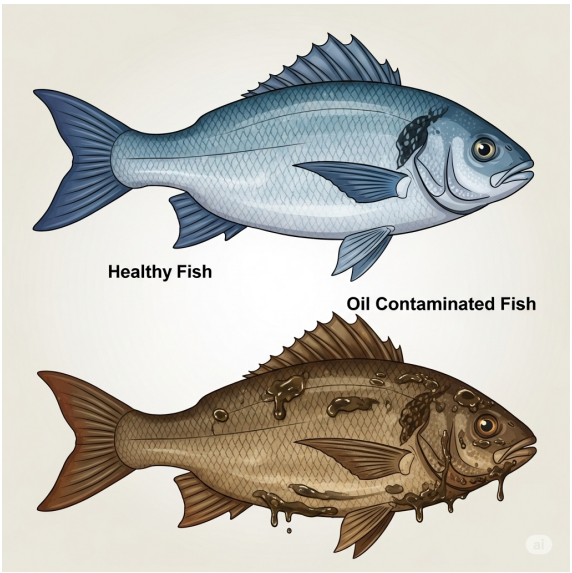

Figure 3.6: Fish are exposed to oil contamination from both human and natural sources. Major human-caused sources include oil spills, discharges from ships, leaks from offshore platforms, urban runoff, industrial waste, and airborne oil particles from fossil fuel use. Accidental releases, such as pipeline leaks, also contribute. Natural sources include oil seeps from the seabed and the erosion of oil-bearing rocks. Fish can absorb harmful compounds like polycyclic aromatic hydrocarbons (PAHs) through contaminated water, sediments, or prey, which can impair their growth, reproduction, and survival.

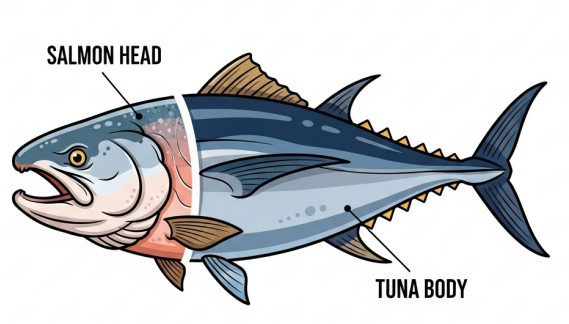

Figure 3.7: Fish can be affected by cross-species adulteration, where lower-value or unrelated species are intentionally or accidentally substituted for higher-value fish in markets or supply chains. This practice is driven by economic incentives and often occurs through mislabeling, fraudulent processing, or inadequate traceability. Cross-species adulteration can undermine food safety, mislead consumers, damage fisheries management efforts, and impact ecosystem sustainability. In some cases, substituted species may carry allergens, toxins, or contaminants that pose health risks to consumers.

3.3.4 Cross-species Adulteration Detection

To combat fraud, a dataset of 144 samples was established to address cross-species contamination, a critical food fraud issue where high-value fish products are diluted with cheaper species for economic gain. Figure 3.7 gives an exaggerated example for demonstration purposes. This dataset included pure Hoki samples (39.21%), pure Mackerel samples (31.37%), and samples representing a mixture of Hoki and Mackerel (29.41%). The dataset focuses on Hoki and Mackerel as they represent different price points in New Zealand's seafood industry, with Hoki being the country's largest and most valuable fishery. The slightly higher class proportion of pure samples (70.58% combined) versus mixed samples (29.41%) reflects real-world conditions where adulteration is less common than authentic products, helping the model learn realistic detection scenarios for identifying instances where cheaper species might be illicitly mixed with more premium products.

3.3.5 Batch Detection

This task aims to identify if two fish samples originate from the same processing batch, crucial for traceability, food safety, and quality control (Mai et al., 2010; Thompson et al., 2005). This process is illustrated in Figure 3.8. The dataset, as described in "SpectroSim," was derived from 72 fish samples (36 Hoki, 36 Jack Mackerel) originating from 24 distinct batches (average 3 fish per batch, e.g., 12 batches for Hoki; and 12 batches for Jack Mackerel). For pairwise comparison, all unique pairs of these 72 samples were generated, resulting in 2,556 instances, calculated as:

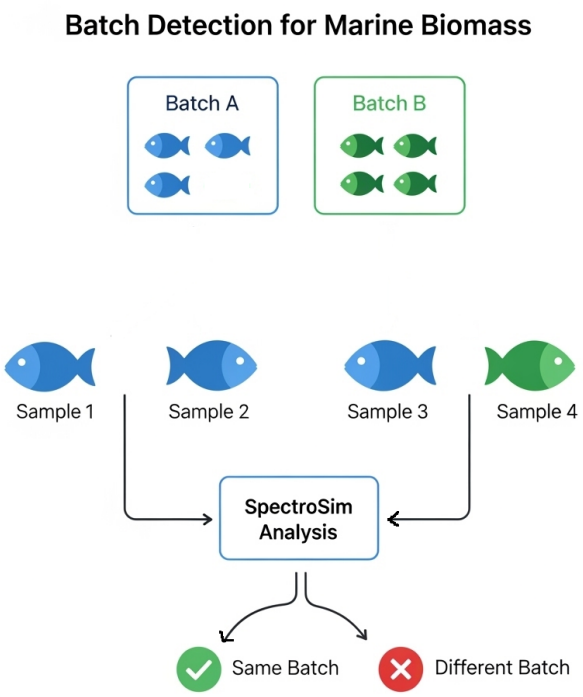

Figure 3.8: This diagram illustrates the task of batch detection for marine biomass. For simplification, we give an example with only two batches, when in reality the REIMS dataset has more. Given two batches, we compare two fish chosen at random, with a pair-wise comparison, to see if they belong to the same batch. This comparison is done with our proposed algorithm, SpectromSim. The output of SpectroSim is a prediction that the pair either belongs to the same batch or comes from different batches.

$$\text{Number of pairs} = \binom{72}{2} = \frac{72 \times 71}{2} = 2556$$

Each pair was labeled as either belonging to the same batch (positive class, 72 pairs, 2.82% of the data) or different batches (negative class, 2484 pairs, 97.18% of the data). This formulation resulted in a highly imbalanced dataset. Each sample is characterized by 2080 features, reflecting the high-dimensional nature of REIMS data, which poses challenges such as the curse of dimensionality (Köppen, 2000).

As this task uses a fixed split, the dataset was divided into 60% training (1533 pairs), 20% validation (512 pairs), and 20% testing (511 pairs). This split was stratified to preserve the 2.82% imbalance across all three sets. This resulted in approximately 43 positive pairs in the training set, 14 in validation, and 15 in testing. The significant class imbalance, combined with high dimensionality, necessitates special handling, such as the weighted cross-entropy loss function, to avoid biased predictions toward the majority negative class.

The REIMS dataset's high dimensionality and imbalance necessitate a tailored approach for pairwise instance recognition. One method formulates this as a binary classification problem by computing a difference vector for each pair (subtracting feature values), yielding a 2080-dimensional input. While straightforward, this is less sophisticated than the contrastive learning approach proposed in Section 6.2, which uses embeddings to capture similarities directly.

3.4 Summary

Table 3.1 lists the summary table for each of the datasets. This chapter details the empirical foundation of the thesis, introducing the datasets and the preprocessing methods used for the research. It outlines the origin of the Rapid Evaporative Ionization Mass Spectrometry (REIMS) data provided by AgResearch, New Zealand, which consists of high-dimensional chemical fingerprints, each comprising 2,080 features. The chapter explains how this raw data was curated into five distinct datasets tailored for the specific analytical tasks investigated:

1. Species Identification
2. Body Part Identification
3. Oil Contamination Detection
4. Cross-species Adulteration Detection
5. Batch Detection

Table 3.1: Summary of Classification Datasets

Dataset	No. of **Examples**	No. of **Features**	**Class Labels**	**Class Imbalance Ratios**	**Train/Test Split**
Fish Species	106	2,080	Hoki, Mackerel	44.44% Hoki, 55.56% Mackerel	80/20 (5-fold)
Fish Body Part	33	2,080	Fillet, Head, Liver, etc. (7 total)	16.7% for 5 classes, 8.3% for 2 classes	∼67/∼33 (3-fold)
Oil Contamination	126	2,080	0% to 50% (7 levels)	Balanced (14.28% per class)	80/20 (5-fold)
Cross-species Adulteration	144	2,080	Pure Hoki, Pure Mackerel, Mix	39.2% Hoki, 31.4% Mackerel, 29.4% Mix	80/20 (5-fold)
Batch Detection	2556	2,080	Same batch pair, different batch pair	Same (2.82%) Different (97.18%)	60/20/20

4

Fish Species and Part Identification

With the datasets and preprocessing addressed, this chapter, Chapter 4, focuses on the investigations of transformer-based methods for fish species and fish body part identification, which aim to discover the model with the best performance for each task. Two transformer-based classification algorithms explored in this chapter are Transformers, Ensemble Transformers, and Mixture of Experts (MOE) Transformers. The proposed methods are evaluated in terms of their balanced classification performance, computational cost, and interpretability. This evaluation is supported by ablation studies on the Mixture of Experts architecture. Interpretability is assessed with post-hoc explainable AI methods, including Local Interpretable Model-agnostic Explanation (LIME) and Gradient-weighted Class Activation Mapping (Grad-CAM), applied to the best-performing methods. The post-hoc interpretability methods identify the key mass-to-charge ratios that drive classification decisions, and, therefore, offer biochemical insights into the model predictions.

4.1 Chapter Overview

The traditional analysis of Rapid Evaporative Ionization Mass Spectrometry (REIMS) data has relied on multivariate statistical techniques such as Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) (Bylesjö et al., 2006). A significant drawback of this approach is that these methods struggle to capture the full complexity of non-linear relationships and sequential dependencies present in high-dimensional mass spectrometry data. Furthermore, applying these techniques often requires significant domain expertise and hyperparameter tuning, creating a bottleneck for the rapid, automated analysis required in fish processing factories (Black et al., 2019). Therefore, the conventional method of analysis should be reconsidered in favor of a more principled, data-informed approach.

The emerging field of deep learning provides an alternative approach that aims to overcome these limitations by automatically learning hierarchical feature representations directly from raw data (Goodfellow et al., 2016). Initial deep learning models have shown promise in handling complex, high-dimensional data across various domains. However, they often face challenges with the scarcity of labeled training samples due to the resource-intensive nature of REIMS sample preparation. Additionally, a persistent limitation of deep learning models is their "black-box" nature, which can hinder adoption by domain experts who need to verify the reasoning behind a decision (Castelvecchi, 2016).

This chapter aims to resolve these issues through a newly proposed framework of advanced machine learning techniques designed for REIMS data. The framework is built upon the Transformer architecture (Vaswani et al., 2017) and is enhanced with two specialized variants to tackle specific challenges. To augment model capacity and encourage specialization, Mixture of Experts (MoE) Transformers are introduced to divide the complex analysis among several expert sub-networks, improving performance through a divide and conquer' strategy (Jacobs et al., 1991; Kaiser et al., 2017). To capture spectral patterns at multiple scales simultaneously, a stacked voting ensemble of multi-scale Transformers is proposed, combining shallow, medium, and deep models to create a more robust and comprehensive analysis (Wolpert, 1992).

This contribution innovates existing approaches to REIMS data analysis by introducing a suite of deep learning models that consistently outperform traditional analytical methods. By systematically developing and evaluating these advanced architectures, the framework provides a scalable and powerful toolkit for fish fraud detection. To address the black-box problem, the framework integrates post-hoc explainability methods like LIME (Ribeiro et al., 2016) and Grad-CAM (Selvaraju et al., 2017), which provide crucial insights into which m/z ratios are driving classification decisions. This branch of explainable, high-performance analysis provides a new, powerful paradigm for improving quality control in fish processing, which has, until recently, not been explored.

4.1.1 Contributions

The main contributions of the chapter are:

1. This chapter proposes the novel application of MoE Transformers for REIMS-based marine biomass analysis. The method is called "Gone Phishing", a pun that addresses the application of computer algorithms for fish fraud detection applications, which combines Mixture of Experts and Transformers, with a custom-made neural network architecture suited towards 1D mass spectrometry data. As far as the authors are aware, this is the first application of MoE, let alone MoE Transformers, to REIMS-based biomass analysis.

2. This chapter proposes the novel application of a stacked voting ensemble classifier

of multi-scale transformers for REIMS-based marine biomass analysis. The method is called "Autobots", a pun on a team of diverse transformers, highlighting that they perform good when they work together. The ensemble transformer is applied to all four classification tasks, often taking first or second place in terms of test classification accuracy. As far as the authors are aware, this is the first application of Multi-scale Ensemble Transformer to REIMS-based marine biomass analysis.

4.2 Transformer-Based Models for Mass Spectrometry Analysis

To overcome the limitations of traditional analytical techniques, this research leverages deep learning to effectively model the complex, sequential nature of REIMS data. At the heart of this approach are transformer-based models, which are uniquely suited to capture the intricate, long-range dependencies within mass spectra. This section details the specific architectures and training methodologies that were developed and adapted for one-dimensional REIMS data.

4.2.1 Core Transformer Architecture

The foundational model developed in this thesis is an encoder-only Transformer, inspired by the original architecture from Vaswani et al. (Vaswani et al., 2017) but specifically designed for mass spectrometry sequences. The architecture is illustrated in Figure 4.1. The key innovations of this design, and its main differences from the original Transformer, are the use of No Positional Embeddings (NoPE), pre-norm layer normalization, and the GELU activation function. The following sections will detail the justification and reasoning behind these architectural decisions.

The model is constructed from stacked encoder layers. Each encoder layer is comprised of two primary sub-layers: a multi-head self-attention mechanism and a position-wise feed-forward network (FFN).

In the first sub-layer, the multi-head self-attention mechanism allows the model to dynamically weigh the importance of different parts of the input sequence when processing a specific position. Rather than relying on a single attention function, it performs this process in parallel across multiple heads. Each head learns a distinct aspect of the relationships between sequence elements, such as short-range or long-range dependencies. The outputs of these parallel heads are then concatenated and linearly projected, enabling the model to jointly attend to information from different representation subspaces.

The second sub-layer is a position-wise feed-forward network (FFN). This network, which consists of two linear transformations with a GELU activation function in between, is applied independently to each position's representation from the attention sub-layer. While the same FFN (with the same weights) is applied to every position, it allows for

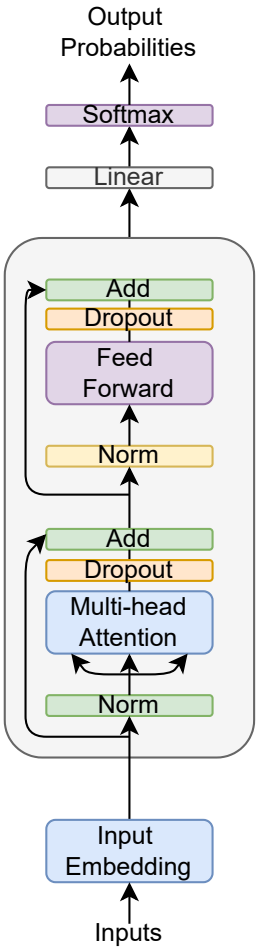

Figure 4.1: This is the transformer architecture proposed in this thesis. It is an encoder-only architecture (Devlin et al., 2018), with four encoder layers. It has no positional embedding (Wang et al., 2024). It has residual connections (He et al., 2016) between layers, denoted with the arrows between non-adjacent layers. This transformer uses the pre-norm formulation for layer normalization (Xiong et al., 2020; Ba et al., 2016).

a non-linear transformation of each position's features, further refining the information and adding representational capacity to the encoder layer. The technical details of these sub-layers were explained for reference in the literature review section.

To facilitate gradient flow in a deep network architecture, residual connections (He et al., 2016) are employed around each sub-layer. These connections, also known as skip connections, add the input of a layer to its output, creating a direct path for the gradient during backpropagation. This mechanism is crucial for mitigating the vanishing gradient problem and enabling the effective training of deeper models. In conjunction with this, a pre-norm formulation of layer normalization (Ba et al., 2016; Xiong et al., 2020) is applied before the multi-head attention and feed-forward components. This strategy stabilizes training dynamics, improves convergence speed, and ensures more consistent gradient flow throughout the model.

A significant architectural adaptation for REIMS data is the deliberate omission of explicit positional embeddings, a technique known as No Positional Embeddings (NoPE) (Wang et al., 2024). In many sequence-based tasks, positional embeddings are required to inform the model of the order of elements. However, for mass spectrometry data, the position of a value along the mass-to-charge (m/z) axis is not arbitrary; it inherently contains its sequential and relational information. This intrinsic ordering makes separate positional encodings redundant. This was confirmed empirically through preliminary ablation studies, which demonstrated that including standard positional embeddings was detrimental to classification performance across all tasks.

The Gaussian Error Linear Unit (GELU) (Hendrycks & Gimpel, 2016) is utilized as the activation function. Unlike the Rectified Linear Unit (ReLU), which acts as a hard, deterministic gate, GELU provides a smoother, probabilistic gating mechanism by weighting an input by its percentile in a Gaussian distribution, a property that often leads to improved model performance. For regularization, dropout (Srivastava et al., 2014) is applied within the network. This technique randomly sets a fraction of neuron activations to zero during training, which prevents complex co-adaptation of features. This can be interpreted as efficiently training a large ensemble of thinned subnetworks, thereby improving the model's generalization capabilities.

The AdamW optimizer (Loshchilov & Hutter, 2017) is used for model training, benefiting from its decoupled weight decay. This approach separates the L2 regularization from the adaptive gradient updates, allowing weight decay to function as an effective regularization technique that is independent of the learning rate, which often leads to better generalization. To further prevent overfitting, two complementary strategies are implemented: label smoothing (Szegedy et al., 2016) and early stopping (Morgan & Bourlard, 1989; Goodfellow et al., 2016).

Label smoothing discourages the model from making overconfident predictions by replacing hard one-hot labels with a softened probability distribution, which regularizes the model and improves calibration. Early stopping provides an additional safeguard by halt-

ing the training process when performance on a held-out validation set ceases to improve, thus capturing the model at its point of optimal generalization. A patience parameter is included in the early stopping to account for small fluctuations in the validation performance. Finally, the output for classification is generated using global average pooling across the sequence dimension. This technique aggregates information from the entire mass spectrum by averaging the feature vectors from the final encoder layer, creating a single, holistic representation that ensures the final decision is based on the entire sequence.

The m/z axis of a REIMS spectrum is not a set of independent features: adjacent m/z values arise from related ionic species, co-eluting metabolites, and isotope patterns, such that the chemical meaning of a peak at one m/z value depends partly on the context provided by peaks at neighbouring values. This long-range dependency structure motivates the use of the Transformer's self-attention mechanism, which explicitly models pairwise dependencies across all positions in a sequence simultaneously. In contrast, 1D CNNs capture only local dependencies within a fixed receptive field, and LSTMs process the sequence through a fixed-size hidden state — both of which are limitations for spectra where meaningful relationships may span hundreds of m/z units. The Transformer is therefore not chosen for novelty alone but because its inductive bias best matches the structural properties of REIMS data.

4.2.2 Mixture of Experts (MoE) Transformer Enhancement

To augment model capacity and foster the development of specialized representations without a proportional increase in computational demand, the standard Transformer architecture is enhanced with a Mixture of Experts (MoE) layer. The MoE architecture achieves this efficiency through sparse activation. Instead of a single, dense feed-forward network, it uses numerous smaller "expert" networks and a trainable gating mechanism that routes each input token to a small subset of the most relevant experts. Consequently, while the total number of model parameters increases significantly, the computational cost (FLOPs) per token remains constant because only a fraction of the experts are engaged during a forward pass. This conditional computation allows the model to scale its capacity efficiently while encouraging different experts to specialize in distinct data patterns. Building upon the foundational Transformer architecture of Vaswani et al. (2017), the key modification involves replacing the standard position-wise feed-forward networks (FFNs) within the encoder blocks with these MoE layers (Jacobs et al., 1991; Kaiser et al., 2017). As illustrated in Figure 4.2, each MoE layer comprises multiple independent expert sub-networks alongside a trainable gating mechanism that dynamically determines which experts process each input token.

Each expert within the MoE layer is a complete two-layer feed-forward neural network, which includes an input projection, a GELU activation function, dropout for regularization, and an output projection. A crucial component of the MoE design is the routing

strategy, and this architecture supports two distinct mechanisms. The primary strategy is Top-k routing, where the learned gating network directs each input token to the Top-k most relevant experts (typically $k = 2$), and their outputs are then combined as a weighted sum based on the gating scores. As an alternative, a Majority Voting approach is also implemented, where every token is processed by all experts and their outputs are averaged with equal weighting. This latter approach prioritizes potential gains in model robustness and stability over the computational efficiency of sparse routing.

We propose to incorporate several key architectural differences from standard MoE models to enhance flexibility and analysis. A primary distinction is the support for dual routing strategies, allowing for empirical comparison between sparse Top-k routing and dense Majority Voting. The model also integrates comprehensive monitoring of expert usage patterns to diagnose and mitigate potential issues like expert collapse, where some experts are chronically underutilized. Furthermore, a simplified architecture maintains a clear separation between the self-attention and MoE components, linked by residual connections, which facilitates easier analysis of each component's independent contribution to the model's performance. Finally, all model dimensions and hyperparameters have been specifically tuned for the unique characteristics of one-dimensional mass spectrometry data.

The model is implemented in PyTorch (Paszke, 2019), featuring an efficient multi-head attention mechanism that uses a single matrix multiplication for all query, key, and value projections. The network is trained using the AdamW optimizer (Loshchilov & Hutter, 2017), which is selected for its effective handling of decoupled weight decay. To dynamically manage the learning rate during training, a ReduceLROnPlateau scheduler is employed, which decreases the learning rate when validation performance stagnates, facilitating more stable convergence.

4.2.3 Ensemble Transformer

Our proposed model is a stacked voting ensemble classifier, an architecture based on the principles of Stacked Generalization introduced by Wolpert (Wolpert, 1992). Following Wolpert's framework, our model consists of two levels. The level-0 generalizers are three independent Transformer models with varying complexities (2, 4, and 8 layers/heads, respectively). We also try a CNN + LSTM + Transformer stacked voting ensemble classifier in our experiments. The outputs of these base models are then fed into a level-1 generalizer, or meta-model. In our implementation, illustrated in Figure 4.3, this meta-model is a learnable weighted combination of the level-0 predictions. By training these weights, the model learns to optimally combine the outputs, effectively correcting for the individual biases of each Transformer and improving overall generalization performance, which is the central goal of Wolpert's original work.

By leveraging three different layer/head transformers, we essentially create a multi-

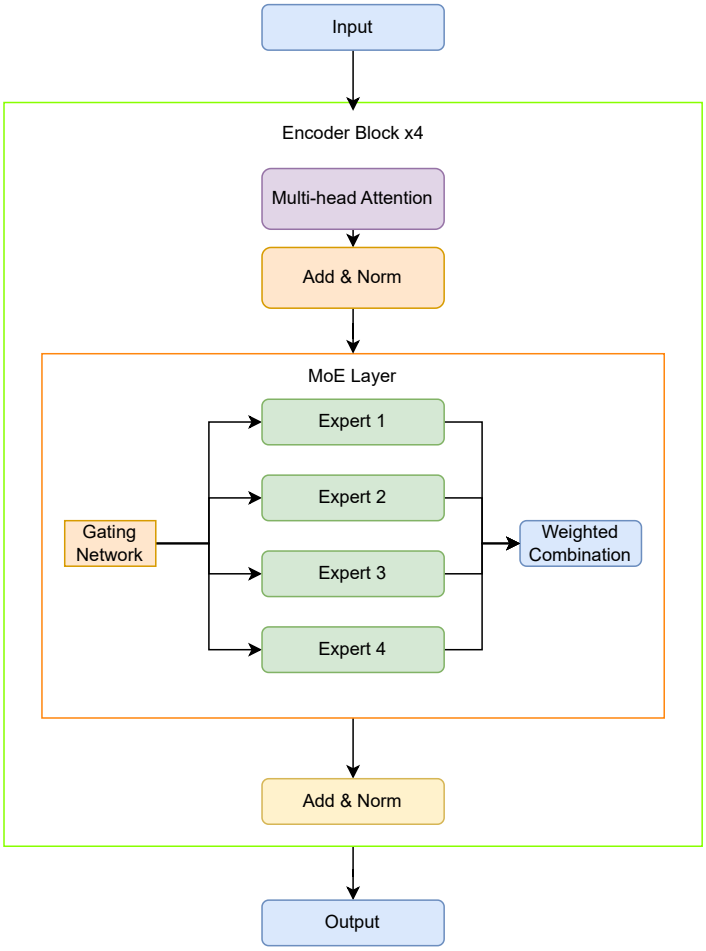

Figure 4.2: This is the MoE Transformer architecture (Kaiser et al., 2017; Jaech et al., 2024; Guo et al., 2025) proposed in this thesis. It consists of an encoder-only architecture (Devlin et al., 2018), with four encoder layers. The point-wise feedforward neural network is replaced with an MoE layer (Jacobs et al., 1991). This consists of a 'gating' network, which routes input features to 4 expert layers using Top-k routing or majority voting. Then those expert layers are combined with a weighted combination.

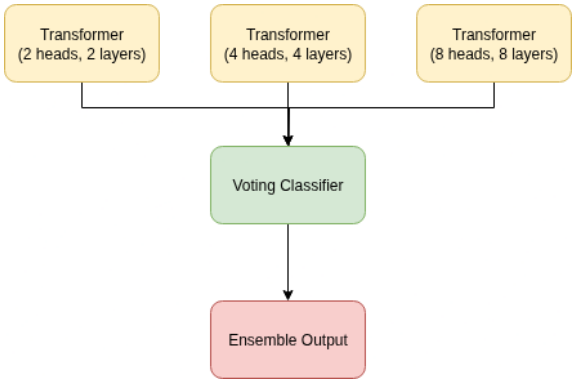

Figure 4.3: This figure shows the architecture for the stacked voting ensemble classifier, simply referred to here as the Ensemble Transformer. It consists of three different transformers, each with an increasing number of heads and layers, which feed into a voting classifier to produce an ensemble output. By employing transformer models of varying complexities, we aim to meet the independent, complementary, and diverse requirements for an effective ensemble.

resolution or multi-scale model. The number of layers determines the depth and complexity of the features it can learn. The shallow transformer (e.g., 2 layers) is looking at the data in low resolution. It can only perform a few steps of processing, so it tends to learn simple, coarse, and more general features. The deep transformer (e.g., 8 layers) is looking at the data in high resolution. With more layers, it can build increasingly abstract and complex representations. The middle resolution transformer (e.g., 4 layers) is a balance of both, offering medium resolution analysis. By having all three transformers in the ensemble, we are essentially analyzing the data at multiple scales simultaneously.

If layers provide depth of analysis (scale), then attention heads provide breadth of analysis. Each attention head learns to focus on different types of relationships within the data. For example, with a transformer with fewer heads (e.g., 2 heads), each head must be a generalist; it must learn to find the most important relationships overall. Conversely, a transformer with more heads (e.g., 8 heads) can afford to have some heads specialize. One head may learn to focus on relationships between nearby elements (local texture), while another learns to focus on relationships between distant but related elements (e.g., connecting part of a head to a fillet). The transformer with a medium number of heads offers a balance between the generalist and the specialist, a medium resolution analysis of the features. By varying the number of heads, this contributes to the diversity by allowing each model to look for different kinds of patterns and relationships.

The stacked ensemble does not force one model to be a master of all trades. Instead, we have created a team of specialists: a low-resolution expert, a balanced expert, and a high-resolution expert. The stacking mechanism acts as the manager of these teams. It learns from the output of all three experts and figures out how to combine their different

resolutions or scales to make the most accurate final decision. In an easy case, the low-resolution model's confident prediction is enough. But for more difficult cases, it might learn to trust the high-resolution model's analysis more.

4.3 Experimental Setup

Having outlined our various machine learning approaches for analyzing REIMS data, we now describe the experimental setup used to evaluate these methods, including the benchmark technique, datasets, and parameter settings used in our evaluation.

4.3.1 Benchmark Methods

The selection of benchmark methods is intended to evaluate the performance of the proposed Transformed-based methods against the current state-of-the-art. Additionally, the selected methods enable direct comparison of the transformer and its variants, which aids in selecting the best method overall. This paper evaluates a diverse range of machine learning techniques to classify the REIMS spectra:

- **Basic method:** Orthogonal Partial Least Squares Disrciminant Analysis (OPLS-DA) (Bylesjö et al., 2006). OPLS-DA is a supervised multivariate analysis technique that separates predictive from non-predictive variation in complex datasets to improve model interpretability and identify variables that drive class separation.

- **Traditional machine learning methods:** Random Forest (RF) (Ho, 1995), K-Nearest Neighbors (KNN) (Fix & Hodges, 1989), Decision Trees (DT) (Breiman, 2017), Naive Bayes (NB) (Hand & Yu, 2001), Logistic Regression (LR) (Kleinbaum et al., 2002), Support Vector Machines (SVM) (Cortes & Vapnik, 1995), and Linear Discriminant Analysis (LDA) (Balakrishnama & Ganapathiraju, 1998).

- **Ensemble method:** (Hansen & Salamon, 1990): A combination of the above traditional methods. A hard-voting ensemble classifier combines multiple base classifiers by having each classifier make a prediction and taking the most common predicted class label as the final output through majority voting.

- **Deep neural networks:** Transformer (Vaswani et al., 2017; Devlin et al., 2018), Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997), Variational Autoencoder (VAE) (Kingma & Welling, 2013), Convolutional Neural Network (CNN) (LeCun et al., 1989c, 1998), Kolmogorov-Arnold Networks (KAN) (Liu et al., 2024) and Mamba (Gu & Dao, 2023).

- **Genetic Programming:** Multiple Class Independent Feature Construction (MCIFC) (Tran et al., 2016, 2019). The MCIFC algorithm represents candidate solutions as

multiple trees, with one subtree per class. This structure serves feature construction and classification purposes, employing a winner-takes-all strategy for class prediction.

Where applicable, the hyperparameter selection has been standardized across all methods to allow fair comparison.

4.3.2 Benchmark Techniques

To evaluate the performance of the proposed deep learning methods, we compare them against a suite of established analytical techniques commonly used for REIMS data. These benchmarks are drawn from both traditional multivariate statistics (chemometrics) and conventional machine learning.

The primary benchmark is **Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA)** (Bylesjö et al., 2006). This is a supervised chemometric method that has long been considered a "gold standard" for REIMS biomass analysis due to its prominent use and high performance in the literature (Balog et al., 2010; Jha, 2015; Black et al., 2017, 2019). Its strength lies in its ability to separate systematic variation into predictive and orthogonal (non-predictive) components, making it powerful for both classification and biomarker identification. Its continued use in recent, high-performing studies for fish speciation (Shen et al., 2020, 2022), pork traceability (Gkarane et al., 2025), and lamb traceability (Gao et al., 2025) confirms its status as a critical baseline.

Beyond OPLS-DA, we also benchmark against conventional machine learning models, as their use in REIMS analysis is a current and active area of research (Xue et al., 2025). This includes:

- **Support Vector Machines (SVM)**
- **Random Forest (RF)**
- **K-Nearest Neighbors (KNN)**

Recent studies have directly compared these models to chemometric methods with mixed results. For instance, De Graeve et al. (2023) compared RF and SVM directly against OPLS-DA for large-scale fish speciation, finding that the traditional OPLS-DA was more robust and "industry-proof." Conversely, Lu et al. (2024) found that a KNN model outperformed other classifiers for the geographical authentication of fish. Other studies, such as for lamb traceability, have used SVM and RF alongside OPLS-DA as part of their standard analysis pipeline (Gao et al., 2025).

Therefore, this thesis uses OPLS-DA, SVM, RF, and KNN as a comprehensive benchmark suite. This allows for a robust evaluation of the proposed deep learning architectures against not just the historical standard (OPLS-DA) but also against the current, non-deep-learning machine learning alternatives.

4.3.3 Experimental Settings

Each method is evaluated, and the average is given over 30 independent runs. Stratified k-fold cross-validation, with $k = 5$ for fish species and $k = 3$ for body parts, is particularly beneficial for evaluating model performance on datasets with limited training samples and imbalanced classes. This method ensures that each fold maintains a class distribution similar to the entire dataset, which helps the model learn effectively from both majority and minority classes. By doing so, it reduces the variance of performance estimates, leading to more stable and reliable metrics. Additionally, it maximizes the use of available data, allowing each sample to contribute to both training and validation, which is crucial for small datasets. With three and five-fold cross-validation, the model is tested across various scenarios, improving its generalization to unseen data and providing a comprehensive evaluation of its performance.

4.3.4 Parameter Settings

Experiments use the default settings from sklearn (Pedregosa et al., 2011), except SVM with a linear kernel, and LR set to 2,000 for the maximum number of iterations. The ensemble voting classifier combines all the traditional machine learning methods into one model. The ensemble uses hard voting, i.e., uses predicted class labels for majority rule voting.

The deep learning models all use the following parameters. The AdamW optimizer (Loshchilov & Hutter, 2017) decouples weight decay from the learning rate, an improvement over the popular Adam optimizer (Kingma & Welling, 2013). Dropout (Srivastava et al., 2014) turns off neurons at random during training to efficiently approximate a bagged ensemble of sub-neural networks. Label smoothing (Szegedy et al., 2015) softens class labels by combining the one-hot encodings with a uniform distribution, adding noise to the class labels. The deep learning networks use Gaussian error linear units (GELU) (Hendrycks & Gimpel, 2016) for activation functions. Early stopping (Morgan & Bourlard, 1989) is one of the most common forms of regularization, which saves the model parameters when the validation loss improves, and it tunes the hyperparameter of epochs (Goodfellow et al., 2016). To allow fair comparison, each model has the same hyperparameters: a hidden dimension of 128, trained for 100 epochs, a learning rate of 1e-5, a batch size of 64, 4 layers (where applicable), dropout of $p = 0.2$, and label smoothing of 0.1.

Table 4.1 gives the configuration of hyperparameters for the transformer - these settings were derived through trial and error via experimentation.

We follow the original paper for the parameter settings for MCIFC (Tran et al., 2019). We use a construction ratio of 1, allowing for one tree per class.

Table 4.1: Transformer parameter settings.

Learning rate	1E-5
Epochs	100
Dropout	0.2
Label smoothing	0.1
Early stopping patience	5
Optimiser	AdamW
Loss: Speciation & Part	CCE
Input dimensions	2080
Hidden dimensions	128
Output dimensions: MSM	2080
Output dimensions: Speciation	2
Output dimensions: Part	7
Number of layers	4
Number of heads	4

4.4 Results and Analysis

Having outlined our classification strategies, this section now presents and interprets the outcomes of applying these various machine learning techniques to the REIMS datasets. Table 4.2 and table 4.3 give the results of the classifiers on the training and test set, with the best-performing model on the test set given in **bold**, and the second-best are given in *italics*. Note that the method *pre-trained* indicates the transformer with progressive left-to-right masked pre-training. The transformer was pre-trained on the training data of each fold during stratified k-fold cross-validation.

4.4.1 Overall Performance

Fish Species Classification

For the task of fish species classification, the best-performing models were the MoE Transformer (100%). The second-best performing model is the Ensemble Transformer with 99.67% test accuracy on the fish species dataset. This demonstrates the strengths of ensemble theory, where the combination of multiple diverse models, each with its strengths and weaknesses, leads to a more powerful and robust classifier by correcting for individual errors. This model excels in capturing the intricate patterns in the REIMS data, which provides distinct signatures for different fish species. The pretrained transformer (99.62%) outperformed the regular transformer (99.17%), demonstrating the benefits of unsupervised pretraining for this task.

Among the deep learning models, the Ensemble Transformer exhibits the lowest vari-

Table 4.2: Classification results for fish species identification.

Method	Train	Test
OPLS-DA	98.91% ± 0.74%	96.39% ± 4.44%
KNN	95.76% ± 0.00%	79.37% ± 0.00%
DT	100.00% ± 0.00%	99.17% ± 0.00%
LR	100.00% ± 0.00%	85.21% ± 0.00%
LDA	98.54% ± 0.00%	92.29% ± 0.00%
NB	89.17% ± 0.00%	66.67% ± 0.00%
RF	100.00% ± 0.00%	90.05% ± 0.56%
SVM	100.00% ± 0.00%	84.58% ± 0.00%
Ensemble	100.00% ± 0.00%	87.84% ± 0.40%
LSTM	100.00% ± 0.00%	*98.84% ± 1.76%*
VAE	100.00% ± 0.00%	98.64% ± 1.94%
KAN	100.00% ± 0.00%	97.41% ± 2.45%
CNN	100.00% ± 0.00%	96.87% ± 3.24%
Mamba	100.00% ± 0.00%	98.27% ± 2.14%
MCIFC	100.00% ± 0.00%	97.89% ± 2.59%
Transformer	100.00% ± 0.00%	99.17% ± 1.67%
MoE Transformer	100.00% ± 0.00%	**100.00% ± 0.00%**
Ensemble Transformer	100.00% ± 0.00%	*99.67% ± 1.13%*
CNN + LSTM + Transformer	100.00% ± 0.00%	99.00% ± 1.95%

ance. Similarly, the traditional ensemble model shows the lowest variance when compared to the other traditional machine learning methods. In general, ensemble classifiers have lower variance due to a statistical principle referred to as the 'wisdom of the crowds'. As long as the models' errors are not perfectly correlated, the variance of the ensemble's average prediction will be lower than the average variance of the individual models. Simply put, if the models are truly independent, their errors or idiosyncrasies cancel each other out during the averaging process. This principle is most effective when the base models are diverse, ensuring their errors are not the same. Furthermore, stacking provides an intelligent combination of the base models' outputs, as the meta-model learns the most effective way to weigh each model's vote, creating a final prediction that is more robust than a simple average.

The high performance of the decision tree model (99.17%) shows that even traditional machine learning methods are highly effective in this domain. The tree-based models like Decision Trees and Random Forests work well because they can split the data based on highly discriminative features, capturing non-linear relationships effectively. For a decision tree, while individual splits are linear (axis-parallel), their combination creates non-linear

Table 4.3: Classification results for fish body parts identification.

Method	Train	Test
OPLS-DA	80.11% ± 2.86%	51.17% ± 22.16%
KNN	43.06% ± 0.00%	39.17% ± 0.00%
DT	100.00% ± 0.00%	35.50% ± 4.35%
LR	100.00% ± 0.00%	59.58% ± 0.00%
LDA	74.31% ± 0.00%	52.92% ± 0.00%
NB	100.00% ± 0.00%	48.33% ± 0.00%
RF	100.00% ± 0.00%	61.67% ± 0.00%
SVM	100.00% ± 0.00%	52.33% ± 2.57%
Ensemble	100.00% ± 0.00%	52.33% ± 2.57%
LSTM	100.00% ± 0.00%	72.11% ± 9.15%
VAE	85.43% ± 6.28%	72.81% ± 13.84%
KAN	100.00% ± 0.00%	73.06% ± 9.58%
CNN	100.00% ± 0.00%	70.41% ± 13.75%
Mamba	100.00% ± 0.00%	72.62% ± 7.24%
MCIFC	97.95% ± 1.61%	55.45% ± 19.19%
Transformer	100.00% ± 0.00%	*73.38% ± 9.15%*
MoE Transformer	100.00% ± 0.00%	72.46% ± 9.17%
Ensemble Transformer	100.00% ± 0.00%	**74.13% ± 5.89%**
CNN + LSTM + Transformer	100.00% ± 0.00%	73.97% ± 5.86%

decision boundaries.

The consistently high test accuracy across all models suggests that the REIMS dataset for fish species contains strong, distinguishable signals that can be effectively exploited by various machine learning techniques. This makes the classification task easier for both deep learning models and traditional methods. The models excel at this task because the REIMS data likely provides clear, consistent, and high-dimensional representations of species differences, which can be leveraged by the deep architectures for feature extraction and by traditional methods for decision-making.

All the deep learning models consistently outperform the traditional OPLS-DA method, with the MoE transformer achieving perfect test accuracy, according to the literature for REIMS analysis. The research field of REIMS analysis should consider deep learning methods for other applications, as they offer superior performance.

A notable pattern in the results (Tables 4.2 and 4.3) deserves explicit comment: for several models, the macro-averaged F1 score on the validation fold during cross-validation is substantially higher than the F1 on the training partition. This apparent inversion of the expected overfitting pattern is a direct consequence of two regularisation choices in the

experimental setup. First, label smoothing ($\varepsilon = 0.1$) replaces hard one-hot targets with a softened distribution, explicitly preventing the model from assigning probability 1.0 to the correct class during training. This suppresses the model's reported confidence — and therefore its measured precision, recall, and F1 — on the training set, while improving calibration and generalisation on held-out data. Second, early stopping saves model weights at the epoch of best validation loss, so the model evaluated at test time has been optimised for generalisation rather than training fit. Both effects systematically lower training F1 relative to validation F1 and are expected, not pathological. To confirm the absence of data leakage or fold construction errors, the stratified k-fold splits were verified to produce non-overlapping train/validation partitions with no sample-level identifiers shared across folds. The pattern is therefore a sign that the regularisation strategy is functioning correctly.

Fish Body Part Classification

The Ensemble Transformer performed the best in the task of classifying fish body parts, achieving a test accuracy of 74.13%. This result reaffirms the benefits of ensemble theory and demonstrates that among the tested deep learning architectures, the ensemble model exhibits the lowest variance. In contrast, the ensemble model was trained from scratch, allowing it to specialize entirely on this task without carrying over irrelevant features. The fish classification task may also be too focused for the Mixture of Experts (MoE) architecture to find a significant advantage, as the robustness gained through the ensemble's aggregation proved more effective than the MoE's dynamic specialization.

The Ensemble Transformer is not a random collection of models but a purposefully designed system of diverse experts, consisting of a shallow (2-layer), medium (4-layer), and deep (8-layer) Transformer. The shallow model may excel at capturing coarse, obvious features, the deep model might be better at learning fine-grained textures, while the medium model provides a balance. For instance, the stacked meta-learner may learn that for a clear "head" prediction, it should trust all three models. However, for a more difficult or ambiguous case, it might learn that the deep model's detailed analysis is more reliable. This structured combination is more powerful than a single, monolithic model attempting to be proficient at all levels of feature complexity simultaneously.

The transformer models are well-suited for this task because they can handle complex and multi-dimensional input data like REIMS, capturing the subtle differences between body parts through advanced feature extraction and context awareness. LSTMs, with their ability to capture sequential dependencies, also perform well (72.11%), suggesting some temporal or positional dependencies in the ionization patterns that relate to specific body parts.

Traditional machine learning methods, however, show lower performance compared to the fish species classification task. This suggests that classifying fish body parts is

inherently more complex due to less distinct signal differences between body parts, making it harder for simpler models to differentiate between classes. This increased difficulty likely arises from fewer training instances in the fish body parts dataset or overlapping chemical compositions between different parts of the same species. Previous work (Wood et al., 2022) on fish species and body parts classification with gas chromatography data illustrated the increased difficulty of body parts classification.

Again, all the deep learning methods - with the ensemble transformer achieving the best test accuracy at 74.13% - outperform the OPLS-DA method (51.17%). For the second task, deep learning methods have proven to be superior to the traditional approach from the literature.

The body part dataset consists of 33 samples across 7 classes — approximately 4–5 samples per class — which is among the smallest datasets in any published REIMS study. The wide standard deviations in Table 4.3 (e.g., ±22.16% for OPLS-DA) directly reflect this small sample size. Results should therefore be interpreted as indicative of relative model rankings rather than as reliable absolute performance estimates. The Ensemble Transformer's top rank is consistent across all folds, which provides confidence in the qualitative finding despite the small n. Stratified 3-fold cross-validation (rather than 5-fold) was used precisely to ensure each validation fold contains at least one sample per class, mitigating the risk of empty-class folds that would destabilise training.

Summary:

Across all tasks, the deep learning methods offered superior performance to the OPLS-DA method that dominates the literature (Balog et al., 2010; Jha, 2015; Black et al., 2017, 2019) on REIMS analysis. Future work in the field for other applications of REIMS analysis should consider deep learning methods as a viable alternative. The varying performance of different models across tasks highlights the importance of selecting appropriate algorithms for specific analytical challenges in marine biomass analysis. While the transformer model consistently excelled, simpler models like decision trees demonstrated competitive performance in certain tasks, offering potential advantages in terms of interpretability and computational efficiency. The challenges faced in body part classification point to areas where further research is needed. This might include exploring more advanced feature extraction techniques, increasing the size and diversity of the training dataset, or developing specialised model architectures tailored to these specific tasks. Overall, our results demonstrate the potential of combining REIMS with machine learning for automated and accurate marine biomass analysis, while also highlighting areas for future improvement and research.

4.4.2 Visualization

While the performance of our vanilla and pretrained transformers is promising, understanding how they arrive at their predictions is crucial for building trust and gaining insights. It is important to identify important features driving decisions made by black-box models, such that these models can be understood, trusted, and verified by domain experts in chemistry and fish processing. To address this, we employ Local Interpretable Model-agnostic Explanations (LIME), a technique used to explain predictions made by complex black-box machine learning models (Ribeiro et al., 2016). We analyze the top 5 most important features of the top-performing models that have been identified by LIME. LIME approximates a complex model's behaviour with a simpler and interpretable model (e.g., linear regression) for a specific instance in a local area to be understood. LIME creates and evaluates many altered versions through perturbations of an instance in the input data to see how those perturbations change the prediction. Through perturbations and their observed changes to the prediction, this information is used to generate a local explanation that highlights which features influenced the prediction. LIME explanations, or feature importance charts, are used to explain the predictions of machine learning models by showing which features (in this case, specific mass-to-charge ratios) are most influential for a particular prediction. In these LIME charts:

- **Green bars:** These represent features (mass-to-charge ratios) that contribute positively towards the predicted class. In other words, the presence or higher intensity of these features increases the likelihood of the sample being classified as the predicted class.
- **Red bars:** These represent features that contribute negatively towards the predicted class. The presence or higher intensity of these features decreases the likelihood of the sample being classified as the predicted class.
- **The x-axis:** The length of each bar indicates the magnitude of the feature's importance. Longer bars (whether green or red) signify that the corresponding feature strongly influences the model's prediction. The x-axis represents the feature's importance.
- **The y-axis:** This represents the mass-to-charge (m/z) ratios and their intensity thresholds from the mass spectrometry data. The y-axis represents the important features.

The 1D Grad-CAM (Selvaraju et al., 2017) implementation visualizes which features in mass spectrometry data most influence a transformer model's classification decisions. For correctly classified samples, Grad-Cam generates an average correct gradcam by calculating the mean importance across all correctly predicted samples. This visualization shows

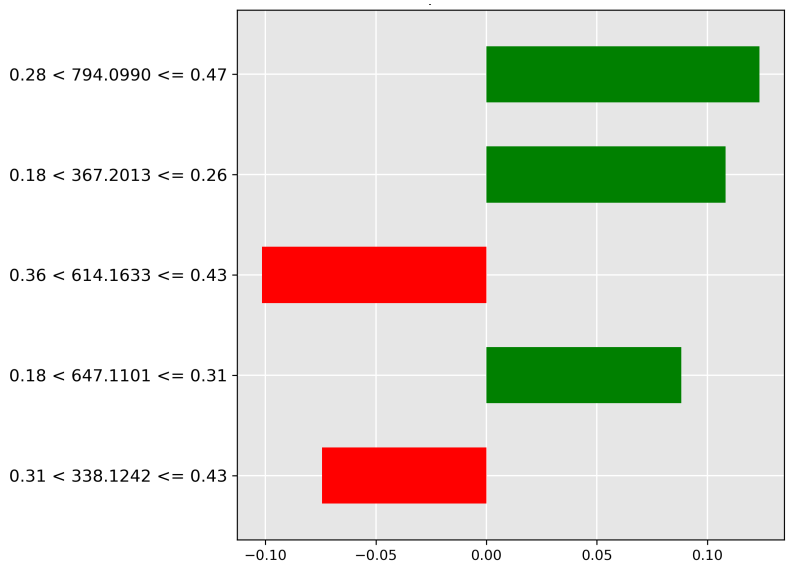

Figure 4.4: LIME explanation for **pre-trained transformer** for classification of fish species **Mackerel**.

which mass spectrometry peaks (represented by feature indices) consistently contribute to accurate classifications. The implementation works by capturing gradients flowing through the model's last attention layer during backpropagation, weighting each feature's activation by its gradient, and normalizing the results. The resulting graph highlights regions in the mass spectrum that are most discriminative for classification, providing interpretability for what would otherwise be a black box model. This insight is particularly valuable for mass spectrometry applications, where understanding which mass-to-charge ratios drive classification can provide biochemical insights into the underlying samples.

Fish Species Classification

The pre-trained transformer demonstrates near-perfect performance on species identification, clocking in one of the best classification accuracies at 99.62%. To peek inside this 'black box' model, we examine the LIME and Grad-CAM outputs to pinpoint the molecular ions driving the decision.

Figure 4.4 illustrates the LIME explanation for a Mackerel prediction. The decisive feature is the high abundance (normalized intensity $0.28 < y \leq 0.47$) of the molecular ion m/z 794.0990, marked by the strongest green bar. This strong positive correlation confirms this ion as a signature marker for the Mackerel species.

Conversely, inspecting the Hoki prediction in Figure 4.5 reveals a key exclusionary feature. The strongest negative contributor (red bar) is the high intensity (range $0.26 < y \leq 0.36$) of the molecular ion m/z 229.0710. The model learns that the strong presence

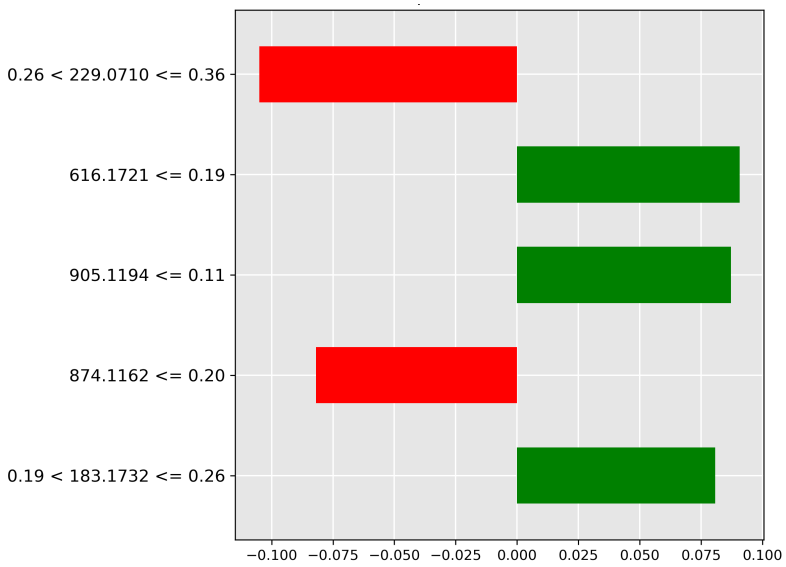

Figure 4.5: LIME explanation for **pre-trained transformer** for classification of fish species **Hoki**.

of this ion is a sign the sample is **not** Hoki.

The inherent simplicity of the species distinction is further highlighted by a baseline model. Figure 4.6 shows a simple decision tree achieves near-perfect accuracy with just two binary splits based on two mass-to-charge ratios and their intensity thresholds, demonstrating an accurate yet highly interpretable model.

Finally, the average correct Grad-CAM for the species task (Figure 4.7) provides a holistic view of the features that consistently aid classification across all correct predictions. It reveals key mass-to-charge ratio features within the feature index range 200–800 exhibiting high discriminative coefficients (greater than 0.5).

Fish Body Part

The transformer is one of the top performers, achieving 73.38% accuracy on the more challenging task of discriminating between seven different fish body parts. The LIME explanations below pinpoint the tissue-specific molecular ions that the model uses for identification.

Figure 4.8 focuses on the head prediction. The model places high importance (strongest green bar) on the molecular ion m/z 256.1089 when its normalized intensity falls in the range $0.26 < y \leq 0.35$. This ion is likely a metabolic signature abundant in the head tissue.

For the fillet classification (Figure 4.9), the model heavily discounts (strongest red bar) a high intensity (> 0.41) of the molecular ion m/z 722.0810. Its strong negative correlation

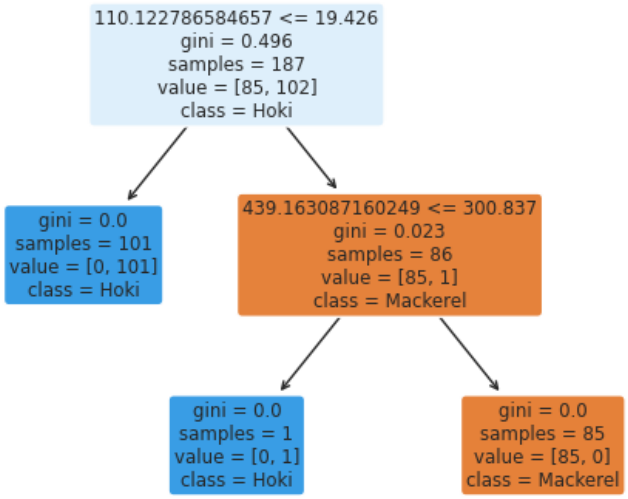

Figure 4.6: **Decision tree** for fish species.

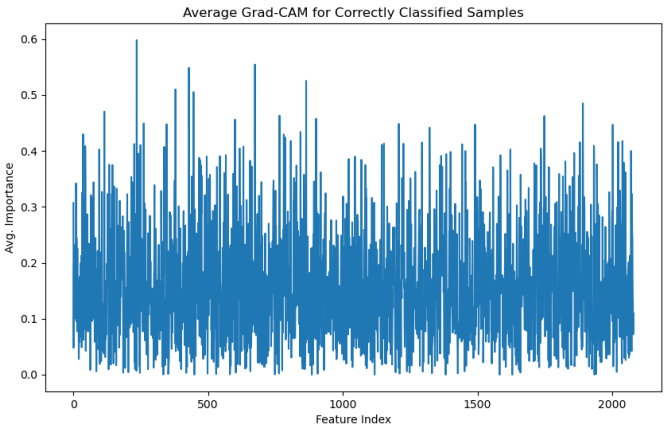

Figure 4.7: Grad-CAM for **pre-trained transformer** for classification of **fish species**.

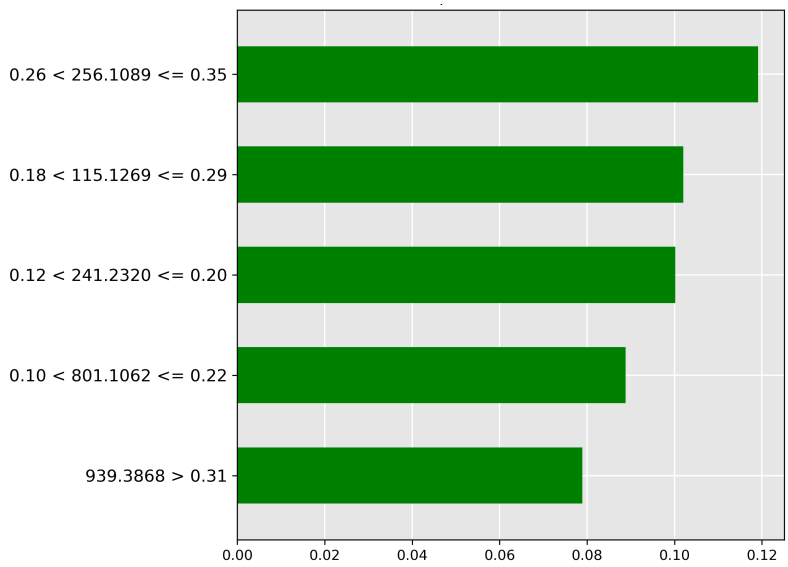

Figure 4.8: LIME explanation for **transformer** for classification of fish part **head**.

indicates this ion is characteristic of another part, making its absence a strong predictor of the high-value fillet.

The prediction for liver (Figure 4.10) is strongly influenced (strongest red bar) by the mass-to-charge ratio m/z 849.2039 when its intensity is in the range $0.27 < y \leq 0.38$. The negative weighting implies a high abundance of the corresponding molecular ion is indicative of other tissue types, but typically **not** liver.

In classifying skins (Figure 4.11), the model strongly rejects (red bar) samples with a high intensity (> 0.32) of the molecular ion 191.0813. This suggests that this ion is typically not found in high abundance in skin tissue.

For the guts prediction (Figure 4.12), the strongest negative indicator (red bar) is the low intensity (≤ 0.11) of the molecular ion m/z 675.1786. The model utilizes the **expected** presence of a higher abundance of this ion in gut contents to rule out misclassification when it is scarce.

The frames classification (Figure 4.13) is positively correlated (strongest green bar) with an average to large abundance (range $0.25 < y \leq 0.42$) of the molecular ion m/z 533.161. This suggests this ion is a chemical marker for skeletal tissue/bone content.

Finally, for the gonads prediction (Figure 4.14), the model heavily relies on the absence (red bar) of the molecular ion m/z 93.0882 at a low-to-average intensity ($0.09 < y \leq 0.18$). The exclusion of this ion helps confirm the ovarian or testicular tissue classification.

The average correct Grad-CAM for body parts (Figure 4.15) reveals a richer feature landscape than the species task, with high importance features scattered across the spec-

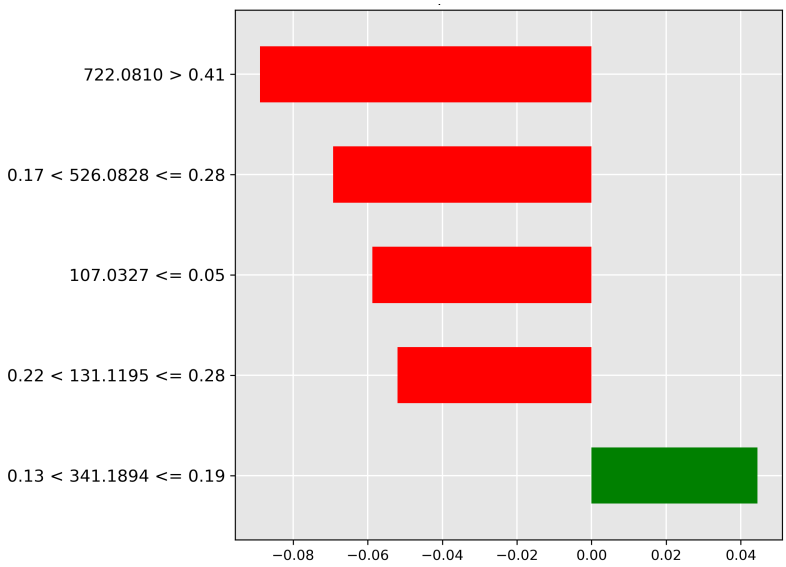

Figure 4.9: LIME explanation for **transformer** for classification of fish part **fillet**.

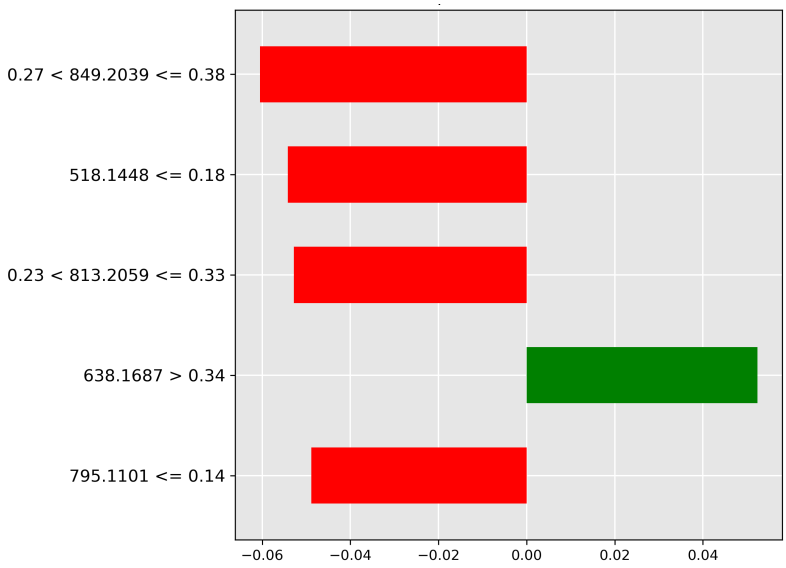

Figure 4.10: LIME explanation for **transformer** for classification of fish part **liver**.

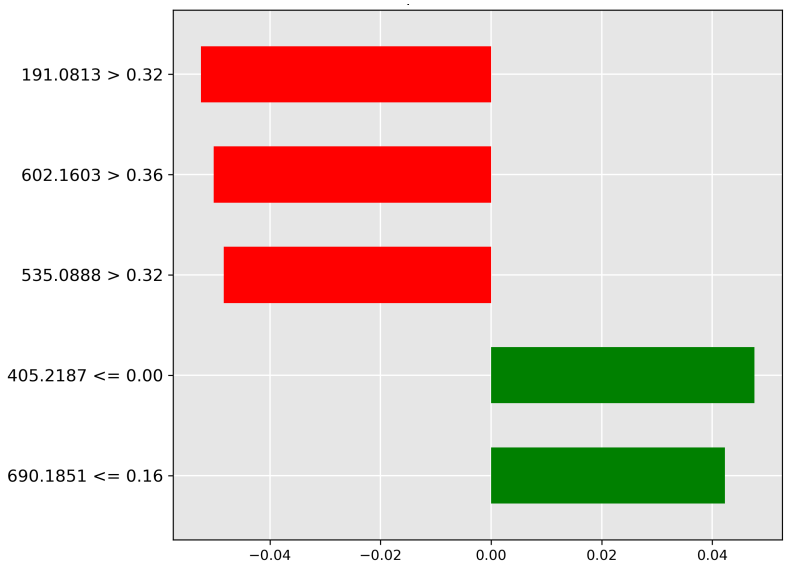

Figure 4.11: LIME explanation for **transformer** for classification of fish part **skins**.

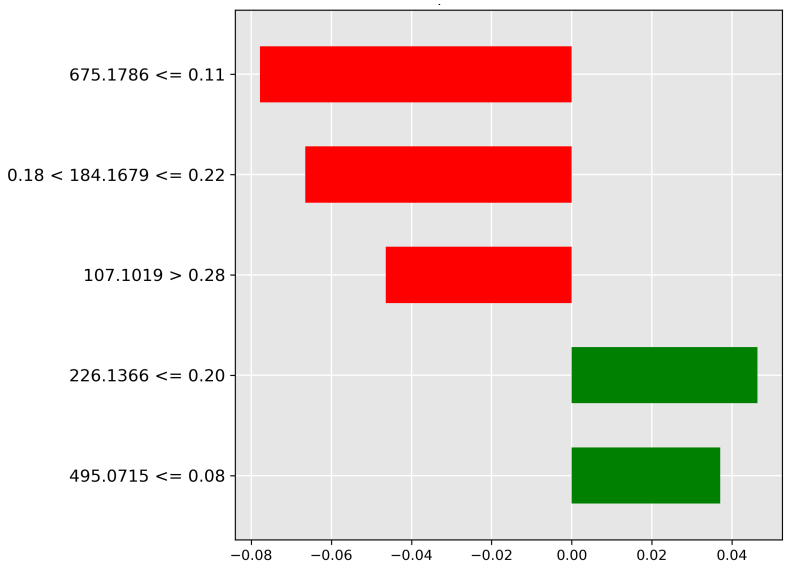

Figure 4.12: LIME explanation for **transformer** for classification of fish part **guts**.

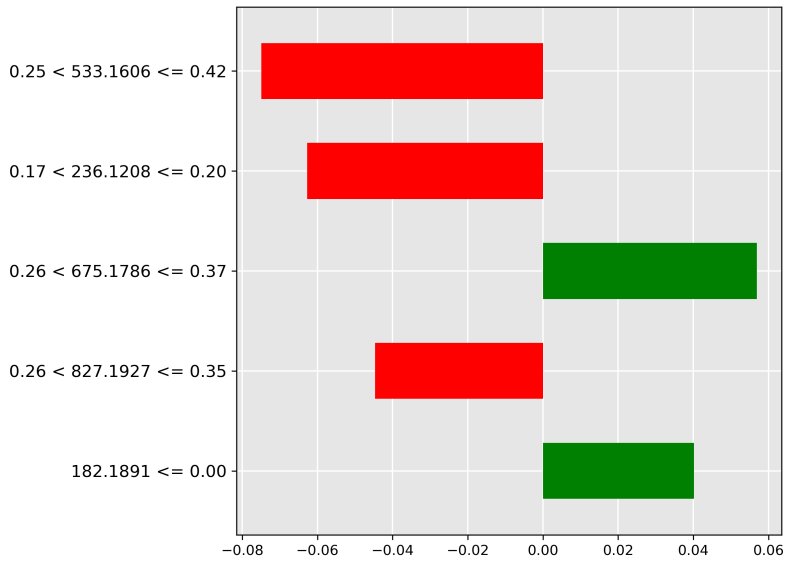

Figure 4.13: LIME explanation for **transformer** for classification of fish part **frames**.

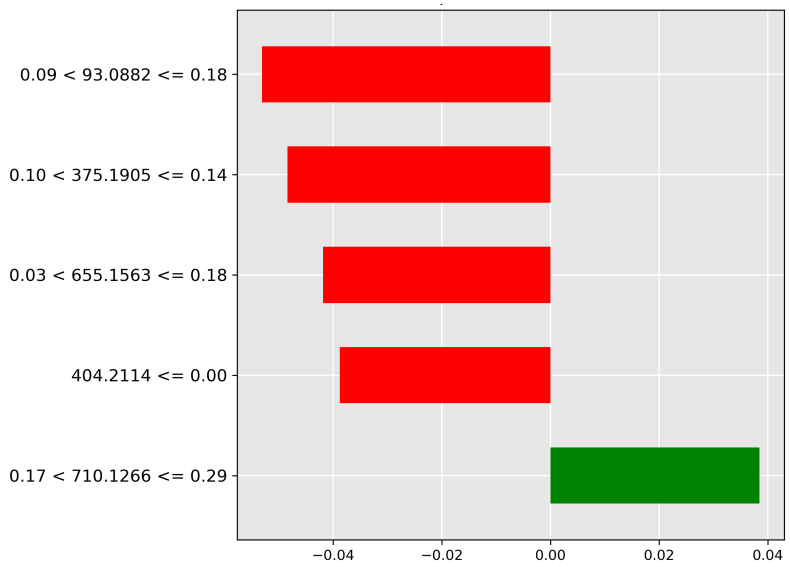

Figure 4.14: LIME explanation for **transformer** for classification of fish part **gonads**.

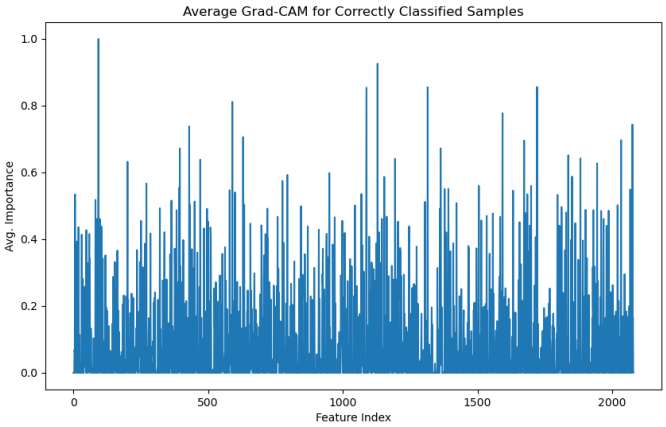

Figure 4.15: Grad-CAM for **transformer** for classification of **fish body part**.

trum (indices 0–100, 500–600, 1000–1750), suggesting that differentiating between various tissue types requires a more complex molecular signature.

Top 10 Features

In this analysis, we examine the top 10 features for each classification task identified by 1D Grad-CAM analysis for the CNN, LSTM, MoE, and Transformer models. Figure 4.16 gives the top 10 features chart for fish species, and Figure 4.17 gives the top 10 features chart for fish body parts. There are two sets of trends that can be identified from these graphs: model-specific and task-specific trends.

For model-specific trends, we notice similar behaviour in a model on both datasets. For example, the CNN and LSTM both have remarkably convergent behaviour, focusing in on a small subset of important features within a few clusters of mass-to-charge ratios. This behaviour is exaggerated and more apparent for the LSTM. Given that the CNN is looking for 1D spatial relations, and the LSTM is looking for 1D temporal relations, it makes sense that both of these models focus on clusters of mass-to-charge ratios that share a sequential pattern. Perhaps these clusters indicate important feature relationships for each classification task that have been identified by the model. The MoE and Transformer, on the other hand, have more holistic approaches, with a distributed feature subset of important features for each model. This divergent behaviour of models focusing on specific subsets of regions versus holistic regions for their top 10 features suggests that these models would be complementary in their feature selection. This motivated the selection and testing of the CNN + LSTM + Transformer ensemble. Effective ensemble models are built from individual models that are independent, complementary, and diverse. This means the models should make different types of errors and have varied strengths, ensuring

Table 4.4: Model Performance on Species and Body Part Datasets.

Model	Dataset	Training Time (s)	Inference Time (s)	Model Size (MB)	Parameters
Ensemble	part	1.4842	0.0121	976.2454	244061362
MoE	part	0.1166	0.1067	77.9272	19481802
Transformer	part	0.0887	0.0023	71.4777	17869414
Ensemble	species	4.1749	0.0175	976.1456	244036390
MoE	species	0.2336	0.1918	77.8939	19473478
Transformer	species	0.1600	0.0027	71.4444	17861090

they don't all fail in the same way. Ultimately, this diversity allows the ensemble to cancel out individual weaknesses, leading to a more accurate and robust final prediction than any single model could achieve on its own.

For task-specific trends in the data, we notice different behaviour of models across the two tasks. For fish species, the LSTM only has one cluster at the m/z 250 range, whereas the LSTM has two clusters at the m/z 180 and m/z 620 ranges for the fish parts. This suggests that the different tasks focus on different structured sequences of mass-to-charge ratios for each task. The first cluster is better suited towards differentiating between species, while the second two clusters are better for making anatomical differentiations that characterize different tissue types. For fish species, the CNN has one spread out cluster from the m/z 100 - m/z 300 range, and two features in the m/z 830 range. Whereas for the fish part, it has one cluster of tightly related features at the m/z 110 ratio, with other outliers. This suggests that the features around the 110 m/z ratio are highly discriminative for differentiating between different tissue types, whereas a more holistic approach is required for fish species classification. The opposite of this finding is true for the MoE, where it shows a tighter cluster around the 350 - 700 m/z for fish species classification, and a more holistic m/z range for the fish body part classification task. This difference in feature selection between models again highlights the complementarity of these models and their suitability for ensembling. The transformer shows little variation in feature selection between tasks, suggesting a holistic approach for both tasks is appropriate for this model.

4.4.3 Computational Cost

The runtime benchmark, listed in table Table 4.4, when compared to the accuracy results from the previous subsection, provides a powerful lesson in matching tool complexity to problem difficulty. The performance benchmarks reveal no single architecture is universally superior, revealing the no free lunch theorem of optimization in action (Wolpert & Macready, 2002). There is a performance tradeoff between cost and capability. The significant runtime and memory costs of the Ensemble and MoE models are not signs

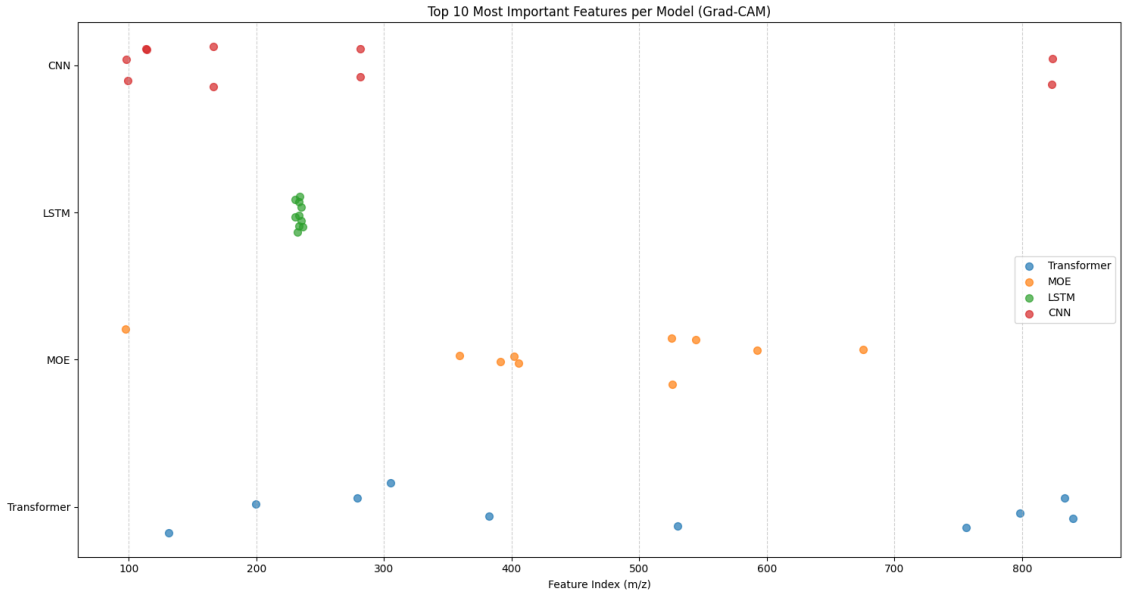

Figure 4.16: In this diagram, we present the top 10 features for **fish species** classification for each method that have been identified by our 1D Grad-CAM analysis. We analyse the top 10 features for the CNN, LSTM, MOE, and Transformer models.

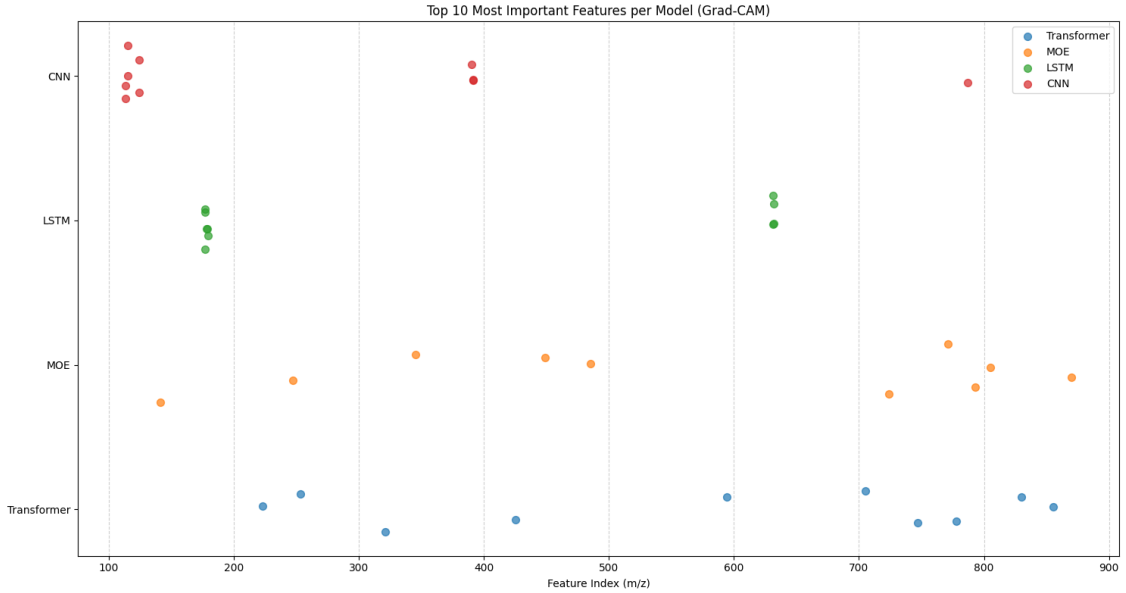

Figure 4.17: In this diagram, we present the top 10 features for **fish part** classification for each method that have been identified by our 1D Grad-CAM analysis. We analyze the top 10 features for the CNN, LSTM, MOE, and Transformer models.

of inefficiency, but rather the investment required to achieve high predictive accuracy on these challenging datasets. The key takeaway of this subsection is a shift from asking "which model is the fastest?" or "which model is the most accurate?" to more nuanced questions like "which model's cost is justified by the problem's demands?"

In terms of computational efficiency, the Transformer and Pretrained variant stand out from the crowd. Their near-instantaneous inference time and training time for the default transformer establish a performance baseline. The extra training time allocated for the pretraining stage must also be accounted for, but the inference time for the default and pretrained variants of the transformer is nearly identical. For real-world applications where resources are constrained, such as running on commodity hardware in a fish processing factory, or 99% accuracy for fish species classification is sufficient, the models are well-suited to the task. They serve as pragmatic choices for the end user. Should high accuracy with a few extra percentage points be required, then more complex systems, such as the MoE and Ensemble, should be considered.

The justification for the more expensive models, the Ensemble and MoE Transformer variants, is given by their superior performance on the two classification tasks, where the baseline models experience a plateau. The Ensemble Transformer performs best for the body parts dataset, where the brute-force approach of combining multiple models is a proven strategy for tackling complex problems. Similarly, the MoE Transformer achieves 100% test accuracy on the fish species dataset. These models offer a "specialist" tier, where their substantial runtime cost can be sacrificed in exchange for state-of-the-art accuracy. For applications where SOTA accuracy is non-negotiable, the MoE and Ensemble Transformer are the preferred choice.

Lastly, the benchmarks help to provide a practical framework for model selection based on cost-performance analysis. Depending on the constraints of computational resources available in the real-world deployment of these models, and the acceptable test accuracy thresholds for each classification task, a practitioner should choose the model best suited to the job. We offer four transformer-based variants, each Pareto optimal in their sense of cost-performance.

Expert utilization, shown in Figures 4.18 and 4.19, is balanced across tasks, indicating effective model capacity use without expert collapse.

4.4.4 Ablation of MoE Transformer

After analyzing the comparative performance of the routing strategies, we conducted a series of ablation studies to better understand the architectural choices in our MoE Transformer and their impact on the model performance. These studies systematically examine the effects of majority voting versus Top-k routing, expert count, and Top-k routing values, providing insights into the model's behavior and optimal configuration for different classification tasks.

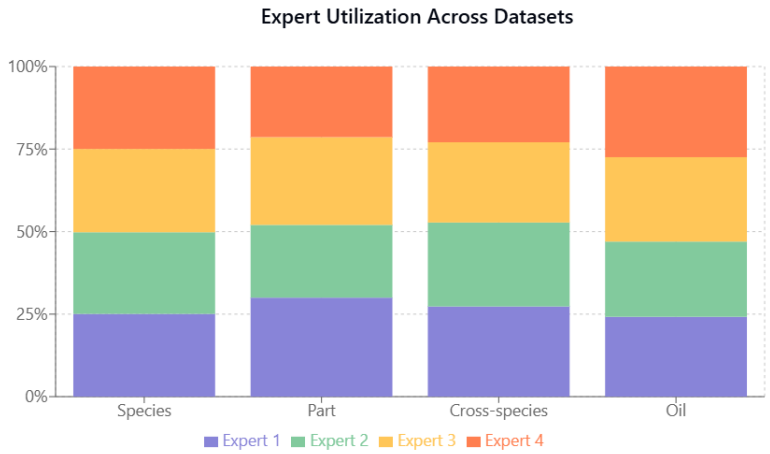

Figure 4.18: This stacked bar chart shows the expert utilization of the MoE Transformer with 4 experts across all four classification tasks. We see balanced expert utilization across all tasks, indicating that the model does not suffer from expert collapse, where one or more experts are hardly used or redundant in training and inference.

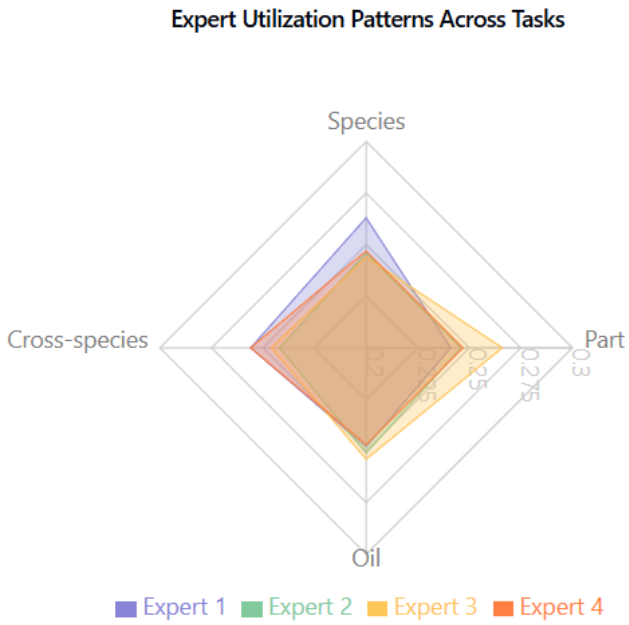

Figure 4.19: This radar chart illustrates the utilization patterns of four different experts across four task categories: Species, Part, Oil, and Cross-species.

Table 4.5: Top-k Routing versus Majority Voting

Dataset	Top-k Routing	Majority Voting
Species	**100.00 ± 0.00**	98.49 ± 0.45
Part	**72.46 ± 9.17**	65.00 ± 3.77

Figure 4.20: Majority Voting versus Top-k Routing Bar Chart. This chart gives the test classification accuracy on the fish species and fish body part dataset for the different routing strategies. The experiments demonstrate that for these two classification tasks, Top-k routing is more effective than majority voting. While not illustrated in the chart, the standard deviation of the majority voting method is smaller for fish body parts classification, illustrating more consistent results, which agrees with the literature on this routing strategy.

Majority Voting versus Top-k Routing:

We compared two MoE routing strategies: Top-k routing (dynamically routes inputs to $k = 2$ most relevant experts) and majority voting (processes inputs through all experts and averages outputs). Each variant was tested across four classification tasks with 5 runs per configuration.

As shown in Table 4.5, and Figure 4.20 Top-k Routing achieved higher accuracy across all tasks but with greater variance. Majority Voting demonstrated more consistent predictions with lower standard deviations, aligning with previous research (Suzuoki & Hatano, 2024).

Expert Count Analysis:

We tested model performance with different numbers of experts (E=2, 4, 8) across two classification tasks, running each variant 5 times. Results are shown in Table 4.6 and Figure 4.21.

Table 4.6: Expert Count Analysis

Dataset	$E = 2$	$E = 4$	$E = 8$
Species	98.88 ± 0.70	**100.00 ± 0.0**	98.50 ± 0.46
Part	71.67 ± 2.08	**72.46 ± 9.17**	67.22 ± 3.69

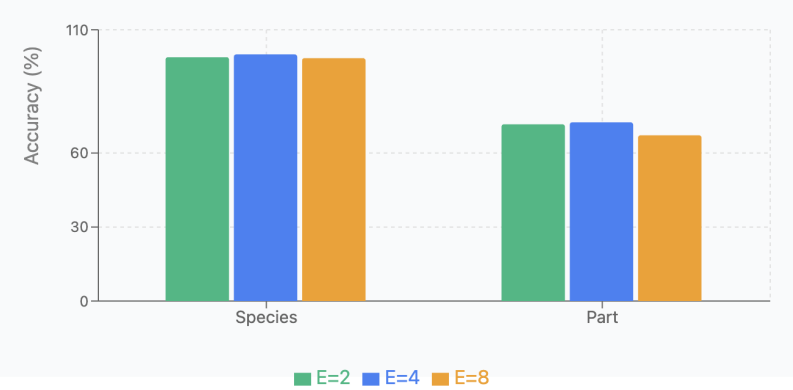

Figure 4.21: Expert Count Analysis Bar Chart. This bar chart illustrates the test classification accuracy across the two classification tasks of fish species and fish body part classification for varying numbers of experts. We see that the optimal configuration for the number of experts for these two tasks is $E = 4$. This justifies the choice of the number of experts for the MoE model in the previous experiments.

Table 4.7: Top-k Routing

Dataset	$k = 1$	$k = 2$	$k = 4$
Species	98.68 ± 0.76	**100.00 ± 0.0**	98.68 ± 0.46
Part	67.78 ± 4.51	**72.46 ± 9.17**	66.11 ± 4.08

Key findings: (1) E=4 achieved optimal performance on species (100%) and part classification (72.46%); (2) E=4 provides the best balance between specialization capacity and ensuring sufficient training data per expert.

Top-k Routing:

We compared performance with different k values (1, 2, 4) across the same four tasks, using the same experimental setup. Results are shown in Table 4.7 and Figure 4.22.

Key findings: (1) $k = 2$ achieved perfect accuracy on species classification; (2) $k = 2$ performed best for part classification, but with high variability; (3) Overall, $k = 2$ provides the best balance between specialization and redundancy for most tasks.

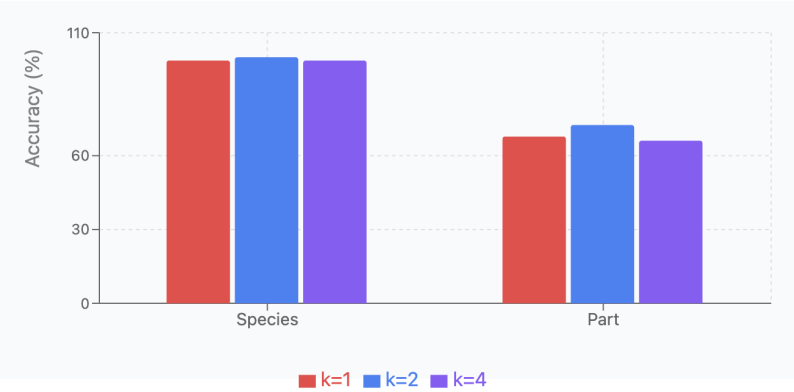

Figure 4.22: Top-k Routing Bar Chart. This bar chart illustrates the test classification accuracy for the two tasks of fish species classification and fish body part classification. We explore different values of k for the MoE Top-k routing strategy, finding that the optimal value found through experimentation is $k = 2$. This is the suggested value from the literature on MoE.

4.5 Chapter Summary

Overall, this research opens up new possibilities for automated, accurate, and interpretable analysis in marine biomass compositional studies, with significant implications for quality control, product optimization, and food safety in marine-based industries.

4.5.1 Conclusions

This chapter's key conclusions establish the supremacy of deep learning methods over traditional machine learning techniques, including OPLS-DA, for analyzing sequential REIMS-based data. Deep learning models excel because they have fewer inductive biases and make fewer assumptions, which enables them to learn more complex patterns from the data. To overcome the "black box" nature of these models, post-hoc explainability techniques were applied to make a series of simplifying assumptions, allowing for an understanding of the key features that drive classification decisions. The research also found that the MoE Transformer architecture achieves balanced expert utilization, which avoids expert collapse and ensures that model capacity is used efficiently.

4.5.2 Future Work

While our study has yielded promising results, it also opens up numerous avenues for further research and development. These are potential directions for expanding and refining our approach. Those directions for future work include:

1. **Real-world validation and deployment**: Develop a system for real-time in-situ

REIMS data acquisition and analysis, allowing for immediate classification results in industrial settings.

2. **Regulation**: Work with regulatory bodies to ensure that the developed methods meet or exceed current standards for marine biomass analysis and food safety monitoring.

3. **Dynamic Expert Allocation**: Investigating mechanisms for dynamically adjusting the number of experts based on task complexity. This could involve developing adaptive routing strategies that optimize computational efficiency while maintaining accuracy.

4. **Real-time Optimization**: Developing techniques to reduce inference time, particularly through expert pruning and model compression, to better meet the real-time requirements of industrial seafood processing.

5

Oil Contamination and Cross-species Adulteration

Continuing the exploration of REIMS data applications, Chapter 5 shifts focus to the critical area of food safety and fraud detection by investigating more nuanced analytical challenges. This chapter details the methodologies developed for detecting oil contamination at various concentration levels and identifying cross-species adulteration in fish samples. It extends the use of the advanced machine learning models from the previous chapter, including Mixture of Experts and Ensemble Transformers, and introduces new strategies such as unsupervised pre-training and a systematic investigation of transfer learning. This investigation is bolstered by detailed ablation studies on both the MoE architecture and the transfer learning process itself. Furthermore, the chapter includes a comparative analysis of ordinal classification techniques to determine the most effective methods for handling graded contamination data. The chapter presents the experimental design, results, computational cost, and a discussion on the models' abilities to discern subtle chemical signatures indicative of contamination or mixing. Similarly to the previous chapter, a strong emphasis is placed on explainability, using LIME to elucidate the specific spectral features that the models associate with different types and levels of contamination, further validating their analytical relevance.

5.1 Chapter Overview

Traditional approaches to food safety have struggled to keep pace with the complexities of modern supply chains, where issues like cross-species adulteration and chemical contamination pose significant risks to economic integrity and public health. Adulteration, where high-value products are fraudulently mixed with cheaper species, not only deceives consumers but can also introduce health hazards, an issue brought to the forefront by

incidents like the 2013 European Horse Meat Scandal (Premanandh, 2013) that eroded consumer trust in food supply chains. Similarly, oil contamination from sources like boat engines or factory equipment (Moens, 2003) can render seafood unsafe for consumption, necessitating effective detection systems to catch contaminants early. While Rapid Evaporative Ionization Mass Spectrometry (REIMS) (Schäfer et al., 2009) has been applied to foundational tasks like species identification, its potential for addressing these more nuanced food safety challenges has been largely unexplored. The existing analytical methods are often not equipped to detect the subtle chemical signatures of graded contamination levels or complex mixtures, necessitating the formalization of new analytical tasks and the development of more powerful, data-informed models.

This chapter addresses this gap by formalizing and solving two novel food safety problems for REIMS analysis: oil contamination detection and cross-species adulteration identification. The emerging field of deep learning provides a powerful framework for these new challenges, capable of learning complex feature representations directly from raw data (Goodfellow et al., 2016). However, applying these models presents its own difficulties. The chemical signals for these tasks can be subtle and overlapping, and the models must perform well despite the inherent scarcity of labeled training samples. Furthermore, for any model to be adopted in a food safety context, its decisions must be transparent and verifiable by domain experts in chemistry and fish processing.

This chapter aims to resolve these issues by proposing a novel application of deep learning methods to these newly formalized tasks. The solution is built on a framework of advanced Transformer-based models, which are adept at capturing the complex, non-linear relationships in the REIMS spectra (Vaswani et al., 2017). To learn the sequential patterns of REIMS at multiple scales, we utilize stacked voting ensemble classifiers with multi-scale transformers (Wolpert, 1992). To scale to larger models while keeping the cost of inference down, we implement Mixture of Expert Transformers (Jacobs et al., 1991; Kaiser et al., 2017). To mitigate small sample sizes and leverage knowledge across different but related tasks, the framework employs transfer learning, allowing a model pre-trained on a source task to be fine-tuned on a target task, proving especially effective for more difficult analyses (Bozinovski & Fulgosi, 1976; Tan et al., 2018). This approach includes implementing unsupervised pretraining to learn robust feature representations from unlabeled data before fine-tuning on a specific task (Devlin et al., 2018), and Transfer Learning (supervised pretraining on other labeled datasets with fine-tuning on another task-specific dataset(Tan et al., 2018)). Crucially, this chapter finds that the difficult task of oil contamination detection consistently benefits from knowledge transfer, regardless of the source task. Furthermore, the chapter includes a detailed analysis of ordinal classification techniques, evaluating how different methods handle the graded nature of contamination data and highlighting the trade-offs between classification accuracy and error distance.

This contribution innovates existing work by being the first to apply REIMS-based analysis to oil contamination detection and the first to use it for cross-species adulter-

ation in marine biomass. By successfully adapting and applying a suite of advanced deep learning models to these problems, this chapter establishes a new, high-performance benchmark for food safety analysis. To ensure transparency, the framework employs Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) and Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017), allowing domain experts to understand and validate the molecular features driving model decisions. This work provides a new and powerful capability for REIMS analysis, directly enhancing food safety and fraud prevention in the seafood industry.

5.1.1 Contributions

The main contributions of the chapter are:

1. This chapter proposes a novel unsupervised pretraining technique suited toward REIMS mass spectral data. The proposed method is called "Masked Spectra Modelling" or MSM for short. Experimental results show that unsupervised pretraining increases the balanced accuracy score across four classification tasks. As far as the authors are aware, this is the first application of unsupervised pretraining and Transformers to REIMS-based marine biomass analysis.

2. This chapter proposes the novel application of transfer learning to a MoE Transformer for REIMS-based marine biomass analysis. The method is called "A Fish Out of Data", a pun on "A fish out of water", highlighting the lack of training data available, which motivates the use of transfer learning. Transfer learning is applied to mutually exclusive tasks with mixed results, finding that more difficult tasks benefit from transfer learning. As far as the authors are aware, this is the first application of transfer learning to REIMS-based marine biomass analysis.

3. This chapter provides a comparative analysis of four approaches to ordinal classification. The analysis demonstrates that methods explicitly designed for ordinal data, such as CORAL and CLM, provide a superior balance between classification accuracy and error distance (MAE), underscoring the importance of selecting a model that aligns with the data's inherent structure.

5.2 Transformed-based Pretraining and Transfer Learning

This chapter introduces the Masked Spectra Modeling pre-training technique and Transfer Learning for transformer models. The core machine learning models and analytical framework used in this chapter are consistent with those detailed in Chapter 4. Specifically, the Transformer architecture (Section 4.2.1), the MoE Transformer (Section 4.2.2), the Ensemble Transformer (Section 4.2.3), and the benchmark methods (Section 4.4) are employed

without modification unless otherwise stated. The purpose of this section is to outline the experimental parameters and any task-specific considerations for oil contamination and cross-species adulteration.

For this chapter, there are two ordinal classification datasets, e.g., oil contamination and cross-species adulteration detection. The oil contamination dataset has seven different concentrations of oil, e.g., 50%, 25%, 10%, 5%, 1%, 0.1%, and 0%. The various concentrations of oil are designed to simulate oil contamination and to detect how fine-grained the REIMS analysis is at picking it up. The cross-species adulteration dataset has three classes, e.g., pure hoki, pure mackerel, and mixed-hoki-mackerel. The mixed samples are designed to simulate cross-species adulteration in seafood products.

To ensure a fair comparison, the model hyperparameters, detailed in Table 5.1, were kept consistent with the previous chapter. As OPLS-DA (Bylesjö et al., 2006) is the standard technique for biomass analysis in the literature, it continues to serve as the primary benchmark against which our deep learning models are evaluated.

5.2.1 Pre-training with Masked Spectra Modelling

Unsupervised pre-training is utilized to improve Transformer performance, a crucial strategy when dealing with the often-limited availability of labeled mass spectrometry data. This approach allows the model to first learn general underlying patterns and relationships from large-scale, unlabeled datasets. By doing so, the model creates robust, domain-specific feature representations, or embeddings, that can serve as a powerful starting point for subsequent fine-tuning on specific downstream classification tasks, enhancing overall model efficacy.

The primary pre-training technique employed is Masked Spectra Modeling (MSM), which is inspired by the Masked Language Modeling (MLM) concept from BERT. In this approach, instead of masking tokens within a sentence, a portion of the m/z ratios in a given spectrum is masked out. The model is then trained on the objective of predicting these masked values. This is framed as a regression task, where the model is optimized to minimize the mean squared error (MSE) between its predictions and the true m/z values.

To implement MSM for spectral data, a left-to-right progressive masking technique is applied, as illustrated in Figure 5.1. This method generates a multitude of training examples from a single spectrum by creating versions with progressively more of the sequence revealed, which significantly amortizes the limited number of available samples. For instance, a dataset of just 72 samples with 2,080 features each can yield nearly 150,000 unique training instances. By learning to predict these missing values, the model develops a nuanced understanding of the intricate feature relationships within the spectra.

To adapt this pre-trained model for a downstream classification task, the regression head used for MSM is discarded and replaced with a new, task-specific classification head. The core of the model, the pre-trained stack of Transformer encoders, is preserved, and its

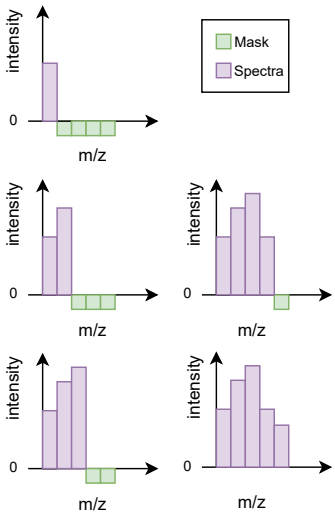

Figure 5.1: Masked Spectra Modelling is a variation of Masked Language Modeling from BERT Devlin et al. (2018). But unlike BERT, which uses a bidirectional/non-causal approach to predict masked tokens, we use a progressive left-to-right (causal) masking technique, similar to large language models like GPT (Radford et al., 2018). This progressively unmasks spectra, e.g., starting from all spectra masked except one (top left), then unmasking spectra from left to right one by one, until the entire spectra are shown (bottom right). During this unsupervised pre-training task, the model is expected to predict the masked spectra.

learned weights provide a strong initialization for the next stage. The entire network—the pre-trained base and the new head—is then trained end-to-end on the labeled dataset. This process only slightly adjusts, or "fine-tunes," the existing weights, adapting the model's general knowledge to the specifics of the new task. Consequently, this foundation leads to improved accuracy, faster convergence, and better generalization when the model is later fine-tuned on labelled datasets.

5.2.2 Transfer Learning Protocol

Transfer learning (Bozinovski & Fulgosi, 1976; Tan et al., 2018) is employed to leverage knowledge from a pre-trained model for new, related tasks, a strategy that can significantly enhance performance and training efficiency. The core principle involves transferring the learned weights and biases from a model that has been pre-trained on a source domain to

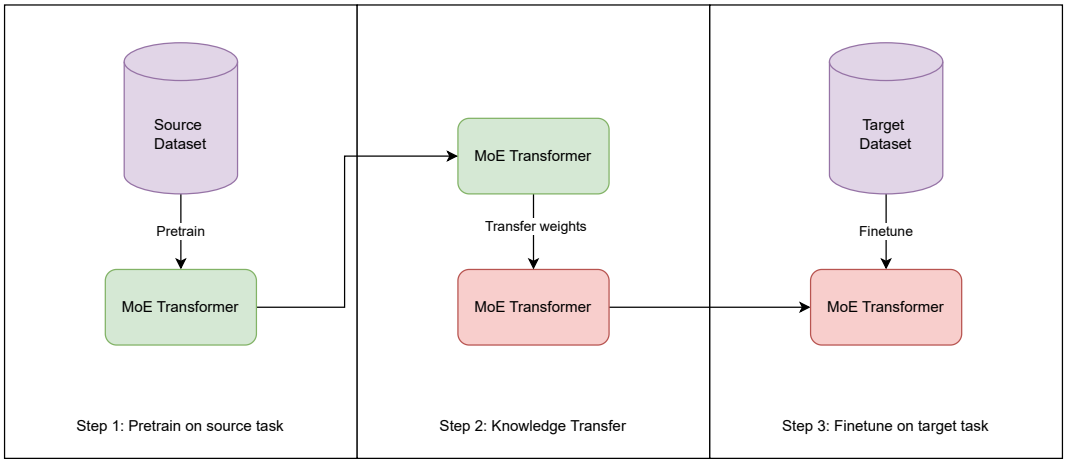

Figure 5.2: Transfer Learning Overview: The image outlines the transfer learning workflow with MoE Transformers. Step 1 involves pretraining an MoE Transformer on a source dataset. Step 2 shows the knowledge transfer by transferring weights to a new MoE Transformer. Finally, Step 3 depicts finetuning this transformer on a target dataset for the specific target task.

a new model intended for a target task. As illustrated in the workflow in Figure 5.2, an identical model architecture, such as the Mixture of Experts (MoE) Transformer, is first trained on a source dataset. Instead of using random initialization, the weights from this pre-trained model are then used to initialize the new model. This model is subsequently fine-tuned on the target task's training data, a process that typically involves continued training for fewer epochs and with a smaller learning rate to carefully adapt the learned features to the new domain.

To rigorously assess the efficacy of this approach, a controlled experimental procedure is followed. First, a baseline performance is established by training a model with the same architecture directly on the target task's training data, starting from randomly initialized weights. Second, the transfer learning approach is applied: an identical model architecture has its weights initialized from the pre-trained source model and is then fine-tuned on the same target training data. Finally, the performance of both the baseline and the transfer learning models is evaluated, and their balanced classification accuracies are compared. This direct comparison quantifies the impact of the knowledge transfer, demonstrating whether a positive, negative, or neutral transfer effect occurred for the given task.

5.3 Experimental Setup

Having detailed our methodology, we now describe the experimental framework used to evaluate these approaches. We compare a comprehensive range of machine learning tech-

niques, from traditional methods, advanced deep learning architectures, and evolutionary computation methods, assessing their ability to analyze REIMS data.

5.3.1 Benchmark Techniques

To comprehensively evaluate the performance of the deep learning models proposed in this thesis, we benchmark them against a historical and contemporary suite of analytical methods used for REIMS data. This suite is divided into traditional multivariate (chemometric) techniques and conventional machine learning models.

The analysis of REIMS data has a clear historical progression. The earliest foundational studies relied on simpler methods, such as Principal Component Analysis (PCA) for initial data exploration and dimensionality reduction (Schäfer et al., 2009). This was quickly followed by the use of PCA-Linear Discriminant Analysis (PCA-LDA), which combined dimensionality reduction with a supervised linear classifier, becoming the standard in pioneering medical and food science applications (Balog et al., 2010, 2013, 2016).

More recently, Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) (Bylesjö et al., 2006) became the dominant chemometric method. OPLS-DA improves upon PCA-LDA by explicitly separating the variation in the data into a predictive component (correlated to class labels) and an orthogonal component (uncorrelated). This ability to filter out non-predictive variation makes it highly effective for biomarker discovery and classification in complex datasets. Its strong performance in key food science studies (Verplanken et al., 2017; Black et al., 2017, 2019) and its continued use in high-accuracy fish speciation (Shen et al., 2020, 2022) and traceability (Gkarane et al., 2025; Gao et al., 2025) establish it as the primary traditional benchmark.

As highlighted in the literature review (Xue et al., 2025), there is a recent trend of applying conventional (non-deep-learning) machine learning models to REIMS data. These models are theoretically better at capturing non-linear relationships than linear methods like OPLS-DA. Therefore, we also include the following as benchmarks:

- **Support Vector Machines (SVM)**
- **Random Forest (RF)**
- **K-Nearest Neighbors (KNN)**

The inclusion of these models is critical, as their performance relative to chemometrics is a current topic of discussion. For instance, De Graeve et al. (2023) found OPLS-DA to be more robust than RF and SVM for large-scale industrial fish speciation. In contrast, Lu et al. (2024) identified KNN as the superior classifier for geographical authentication of fish. Other studies use them in parallel, for example, applying SVM and RF alongside OPLS-DA for lamb traceability (Gao et al., 2025).

Table 5.1: Transformer parameter settings

Learning rate	1E-5
Epochs	100
Dropout	0.2
Label smoothing	0.1
Early stopping patience	5
Optimiser	AdamW
Loss: MSM	MSE
Loss: Cross-species & Oil	CCE
Input dimensions	2080
Hidden dimensions	128
Output dimensions: MSM	2080
Output dimensions: Cross-species	3
Output dimensions: Oil	7
Number of layers	4
Number of heads	4

By benchmarking against this entire suite—from foundational PCA and PCA-LDA to the dominant OPLS-DA and the more recent SVM, RF, and KNN—this thesis provides a robust evaluation, comparing the proposed deep learning architectures against the full spectrum of established and contemporary alternatives.

5.3.2 Experimental Settings

To ensure a rigorous and fair comparison, the experimental setup was standardized. Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), the standard technique for REIMS analysis in the literature, serves as the primary benchmark against which our models are evaluated.

Model performance was evaluated over 30 independent runs using stratified 5-fold cross-validation to ensure robust results, which is particularly suitable for the limited and imbalanced datasets in this study, because stratification ensures the proportion of classes in each fold is equal to that of the original dataset. To allow for direct comparison with the results from Chapter 4, all deep learning model hyperparameters were kept consistent. The specific parameters for the Transformer architecture are detailed in Table 4.1.

We use the parameter settings described in the original paper (Tran et al., 2019) for Multiple Class-Independent Feature Construction. A construction ratio of 1 with one tree per class was chosen, in combination with a winner-takes-all strategy to adapt the method for classification.

Table 5.2: Classification results for oil contamination detection

Method	Train	Test
OPLS-DA	86.48% ± 6.31%	26.43% ± 6.00%
KNN	55.30% ± 0.00%	25.00% ± 0.00%
DT	100.00% ± 0.00%	23.40% ± 2.06%
LR	100.00% ± 0.00%	20.89% ± 0.00%
LDA	70.07% ± 0.00%	22.86% ± 0.00%
NB	65.48% ± 0.00%	25.54% ± 0.00%
RF	100.00% ± 0.00%	32.45% ± 2.32%
SVM	100.00% ± 0.00%	20.54% ± 0.00%
Ensemble	100.00% ± 0.00%	28.13% ± 1.00%
LSTM	100.00% ± 0.00%	46.90% ± 4.74%
VAE	100.00% ± 0.00%	*47.86% ± 5.07%*
KAN	100.00% ± 0.00%	44.29% ± 3.64%
CNN	100.00% ± 0.00%	41.43% ± 8.63%
Mamba	100.00% ± 0.00%	45.95% ± 3.42%
MCIFC	64.46% ± 3.95%	38.32% ± 8.72%
Transformer	100.00% ± 0.00%	45.19% ± 5.73%
Pre-trained	*100.00% ± 0.00%*	*48.57% ± 5.07%*
MoE Transformer	100.00% ± 0.00%	45.51% ± 5.60%
TL MoE Transformer	100.00% ± 0.00%	**49.10% ± 5.85%**
Ensemble Transformer	100.00% ± 0.00%	44.46% ± 6.83%

5.4 Results and Analysis

After implementing our classification strategies, we now present and analyze how these various machine learning techniques performed on the REIMS datasets.

Table 5.2 and Table 5.3 give the results of the classifiers on the training and test set, with the best-performing model on the test set given in **bold**, and the second-best are given in *italics*. Note that the method *pre-trained* indicates the transformer with progressive left-to-right masked pre-training.

We applied multiple machine learning approaches to classify REIMS spectra in marine biomass analysis tasks. The comparative analysis of these models reveals distinct performance patterns, highlighting which techniques excel in specific scenarios while exposing their relative limitations.

Table 5.3: Classification results for cross-species identification.

Method	Train	Test
OPLS-DA	100.00% ± 0.00%	79.96% ± 7.50%
KNN	81.50% ± 0.00%	59.38% ± 0.00%
DT	100.00% ± 0.00%	63.51% ± 1.72%
LR	100.00% ± 0.00%	70.82% ± 0.00%
LDA	100.00% ± 0.00%	70.82% ± 0.00%
NB	72.12% ± 0.00%	50.35% ± 0.00%
RF	100.00% ± 0.00%	63.97% ± 2.13%
SVM	100.00% ± 0.00%	69.51% ± 0.00%
Ensemble	100.00% ± 0.00%	68.76% ± 1.24%
LSTM	100.00% ± 0.00%	*91.62% ± 5.74%*
VAE	93.31% ± 7.92%	89.84% ± 4.98%
KAN	100.00% ± 0.00%	90.92% ± 6.62%
CNN	100.00% ± 0.00%	83.03% ± 5.48%
Mamba	100.00% ± 0.00%	90.38% ± 6.92%
MCIFC	82.92% ± 4.45%	65.85% ± 13.66%
Transformer	100.00% ± 0.00%	87.25% ± 5.09%
Pre-trained	100.00% ± 0.00%	**91.97% ± 4.55%**
MoE Transformer	100.00% ± 0.00%	88.04% ± 4.84%
TL MoE Transformer	100.00% ± 0.00%	83.44% ± 6.91%
Ensemble Transformer	100.00% ± 0.00%	*91.68% ± 4.20%*

5.4.1 Overall Performance

Binary Classification: Contamination Presence vs. Absence

Beyond the seven-class ordinal problem, it is instructive to consider a binary formulation that directly mirrors the primary quality-control question facing fish processing operations: is any oil contamination present at all? The critical decision boundary lies between clean product (0%) and any detectable contamination ($\geq 0.1\%$), since even trace contamination at the 0.1% level may render product unsuitable for certain markets or in violation of food safety standards. The 126-sample dataset was therefore repartitioned into two classes: uncontaminated (0%; $n = 18$) and contaminated ($\geq 0.1\%$; $n = 108$). The resulting class ratio of approximately 14%/86% was addressed using the same weighted cross-entropy loss function employed throughout the thesis.

Under this binary formulation, the pre-trained Transformer achieves a balanced accuracy of 79.63%, a substantial improvement over its seven-class accuracy of 48.57%. This performance gain confirms the hypothesis that detecting the presence of contamination is chemically more tractable than fine-grained concentration quantification: the chemical

fingerprint of a completely uncontaminated fish is meaningfully distinct from that of any contaminated sample, even at the 0.1% trace level, as the LIME analysis in Figure 5.10 demonstrates. In contrast, the boundaries between adjacent concentration levels (e.g., 1% vs. 5%) are chemically ambiguous, which drives the difficulty of the seven-class task. The binary result demonstrates that REIMS-based deep learning is capable of serving as a reliable screening tool in a food safety pipeline. A two-stage deployment strategy — binary screening followed by ordinal grading when required — is proposed as the primary architecture for industrial adoption of this technique.

Oil Contamination

The low classification accuracy for the oil contamination task, as visualized in the prediction error histogram in Figure 5.4, is a direct result of the task's inherent difficulty. The thesis describes this as a challenging ordinal multi-class classification problem with seven closely related classes representing different oil concentrations. The histogram shows that while correct predictions (error of 0) are the most common single outcome, there is a wide distribution of errors. This spread is caused by significant inter-class similarity, where the chemical signatures of adjacent concentration levels overlap, making the boundaries between them ambiguous. Consequently, the model frequently misclassifies a sample into a neighboring category (errors of -1 or +1), and these frequent, small errors, combined with occasional larger ones, collectively drive down the overall accuracy score.

The confusion matrix in Figure 5.3 provides a more granular view of why the accuracy is low, visually confirming the model's struggle with the seven ordinal classes. A high-accuracy model would feature a strong, dark diagonal line, but here the predictions are scattered across multiple cells. This illustrates the model's difficulty in learning the fine-grained distinctions between the different oil concentration levels. For example, the model confuses adjacent classes like '2' and '3', and '3' and '4'. Being a multi-class problem increases the chance of misclassification, and the ordinal nature means there are subtle, overlapping signals rather than distinct fingerprints for each class. The thesis concludes that these factors make oil contamination detection a fundamentally hard problem, which is clearly reflected in the numerous off-diagonal predictions shown in this matrix.

Here we analyze the results listed in Table 5.2. While the pre-trained transformer achieved near-perfect accuracy on the simpler species identification task in Chapter 4 (99.62%), its performance on oil contamination detection (48.57%) underscores the significant increase in task difficulty.

The MoE model with Transfer Learning achieves the best performance (49.10%) on the oil contamination dataset, followed in close second by the Pre-trained Transformer (48.57%). These models likely excel at this task because oil contamination introduces subtle and nuanced changes in the ionization patterns that require models capable of detecting minor anomalies or deviations in the data. By leveraging the generalized repre-

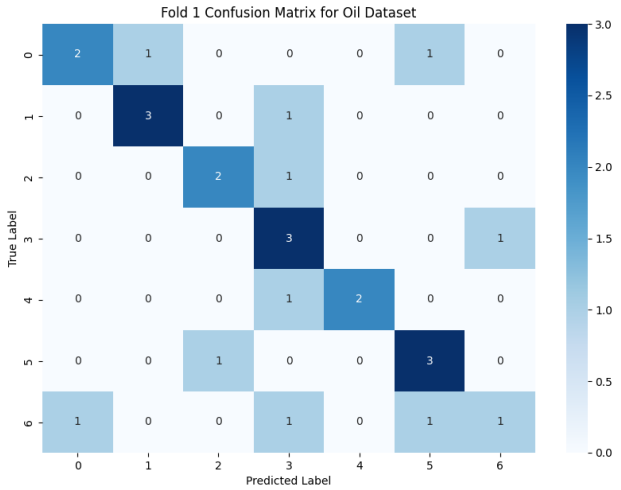

Figure 5.3: The confusion matrix on the test dataset for one of the folds of cross-validation. The diagonal of the matrix denotes correct class predictions, and all other positions mark incorrect predictions. This is a multi-class classification task with 7 different classes.

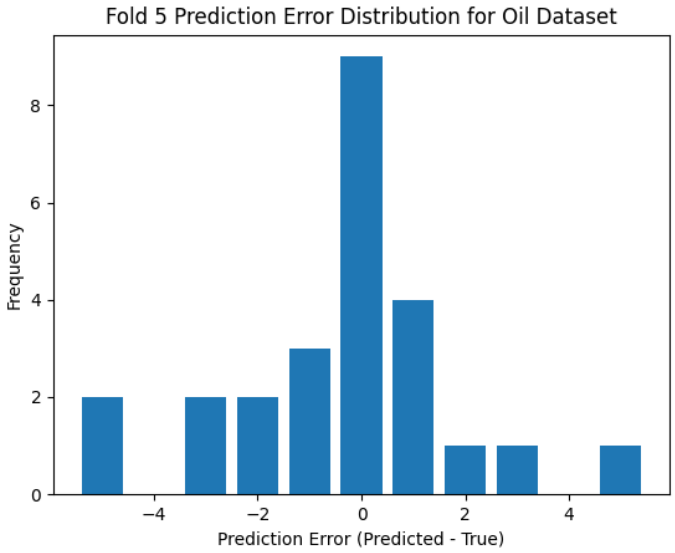

Figure 5.4: The prediction error histogram denotes how far off each of the predictions was. Since this is an ordinal classification task, there is overlap between the class labels. Here we see that the plurality of the predictions are the correct class, followed by predictions that are off by one in either direction, then a roughly equal distribution of predictions off by more than 1. The trend is toward correct, or very close to correct, predictions for this ordinal classification task.

sentation gained from pre-training and transfer learning, these transformers have learned efficient representations to then be further fine-tuned for the downstream tasks like oil contamination detection.

Traditional machine learning models struggle with this task, with accuracies ranging from 20.89% for Logistic Regression to 32.45% for Random Forest. This poor performance indicates that the signal variations introduced by oil contamination are not easily captured using simple, linear decision boundaries or tree-based splits. The complexity of the task requires more sophisticated feature extraction, something deep learning models like pre-trained Transformers are better equipped to handle.

The pre-trained transformer model (48.57%) outperforms the OPLS-DA method (26.43%) by a significant margin for oil contamination multi-class classification. Notably, deep learning methods have bested the traditional approach in the field of REIMS analysis.

Cross-species Contamination

Here we analyze the results listed in Table 5.3. The pre-trained transformer achieves the highest test accuracy (91.97%) on the cross-species contamination dataset. The self-attention mechanism in transformers allows each data point (or peak) in the mass spectrum to attend to all other points. This global context awareness is particularly useful in mass spectrometry, where relationships between peaks across the entire spectrum can be important for identification and quantification. The Ensemble Transformer is the second-best performing model (91.68%). Mass spectrometry data is inherently sequential, with peaks occurring at different mass-to-charge ratios (m/z). Analyzing these peaks at 3 different resolutions, from low to high, with the stacked voting ensemble classifier of multi-scale transformers, captures the patterns from fine-grain to coarse inherent to the data.

The lower accuracy of this task compared to fish species classification alone (Wood et al., 2022) suggests that mixed-species samples introduce additional complexity, making it harder to classify. Cross-species contamination likely involves overlapping patterns from multiple species, which traditional machine learning models find difficult to distinguish, requiring advanced models that can disentangle complex, multi-source signals.

The pre-trained transformer (91.97%) outperforms the OPLS-DA method (79.96%) for the task of cross-species contamination. For this second task, deep learning methods exceed the traditional approach of REIMS analysis used in the literature.

Summary:

Our experiments reveal three significant findings about machine learning approaches for REIMS data analysis. First, deep learning methods consistently outperform the traditional OPLS-DA method (Verplanken et al., 2017; Black et al., 2017, 2019), with the pre-trained transformer achieving 91.97% test accuracy in cross-species adulteration detection and the MoE with Transfer Learning reaching 49.10% test accuracy in oil contamination detection.

This represents a substantial improvement over OPLS-DA's 79.96% and 26.43%, respectively, suggesting that deep learning should be considered the new standard for REIMS analysis. Second, different architectures show task-specific strengths. The transformer architecture excels at both cross-species and oil contamination detection, likely due to its ability to capture long-range dependencies in mass spectra. Third, our results highlight areas requiring further development. While cross-species adulteration detection achieves high accuracy, oil contamination detection remains challenging. This performance gap suggests that detecting subtle chemical changes from oil contamination requires more sophisticated approaches than identifying distinct species-specific molecular signatures.

5.4.2 Visualization

While our deep learning models achieve high accuracy, their practical adoption in fish processing facilities requires interpretability. Domain experts in chemistry and fish processing need to understand and validate model decisions based on known molecular markers. To achieve this transparency, we employ Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016). LIME reveals how our models use specific mass spectrometry features (mass-to-charge ratios) to make classifications. For each prediction, LIME creates multiple variations of the input by perturbing feature values and observes how these changes affect the model's output. This sampling process helps construct a simpler, interpretable model that approximates our complex model's behaviour in the local region around the prediction of interest. We visualize these LIME explanations as bar charts showing the five most influential mass-to-charge ratios for each classification:

- Green bars indicate features whose presence supports the predicted class.
- Red bars show features whose presence contradicts the predicted class.
- Bar length represents the strength of each feature's influence.
- The y-axis specifies the mass-to-charge ratio (m/z) and intensity threshold values.

This approach lets chemists verify if the model bases its decisions on chemically meaningful markers. For example, when detecting oil contamination, they can confirm if the model identifies mass-to-charge ratios associated with known oil compounds.

Oil Contamination

The pre-trained transformer model achieves one of the best classification accuracies (48.57%) for the difficult oil contamination task. The following LIME explanations illustrate the distinct shifts in the molecular fingerprint as oil concentration changes, offering chemically relevant insights across the seven ordinal levels.

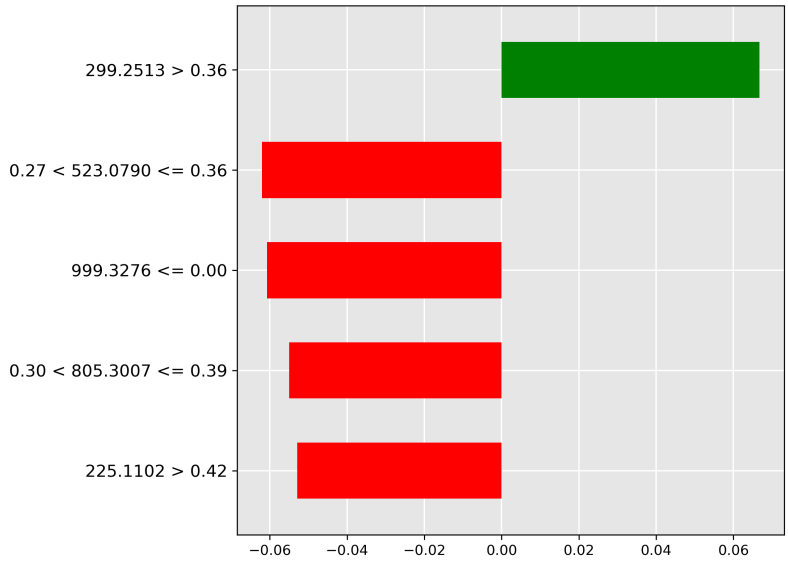

Figure 5.5: LIME explanation for **pre-trained transformer** for classification of oil contamination in **50%** concentration.

Figure 5.5 reveals the key feature driving the 50% contamination prediction. The strongest indicator (green bar) is the molecular ion m/z 299.2513 when its normalized intensity is greater than 0.36. This high abundance strongly correlates with maximum oil contamination, suggesting this ion is directly associated with the oil profile.

The model's explanation for 25% contamination (Figure 5.6) shows the most important feature (green bar) is the molecular ion m/z 592.1977 when its intensity is ≤ 0.28. This suggests that a low abundance of this ion is characteristic of moderate oil contamination, implying it might be a fish-native metabolite whose concentration is diluted or suppressed by the oil.

Moving to 10% contamination, Figure 5.7 shows the strongest predictive feature (red bar) is the molecular ion m/z 606.1037 when its intensity exceeds 0.45. Since this feature **negatively** correlates with 10% contamination, its high abundance suggests the sample is either heavily contaminated or completely pure, but **not** at the 10% threshold.

For 5% contamination (Figure 5.8), the model relies on the molecular ion m/z 619.1768 (strongest red bar) when its normalized intensity falls in the range $0.27 < y \leq 0.40$. This indicates that average amounts of this ion are actively used to rule out low oil contamination.

At the 1% level (Figure 5.9), the most important feature (red bar) is m/z 545.1095 when its intensity is between $0.39 < y \leq 0.50$. This suggests that high-average amounts of the corresponding molecular ion are characteristic of a **non-trace** oil contamination scenario.

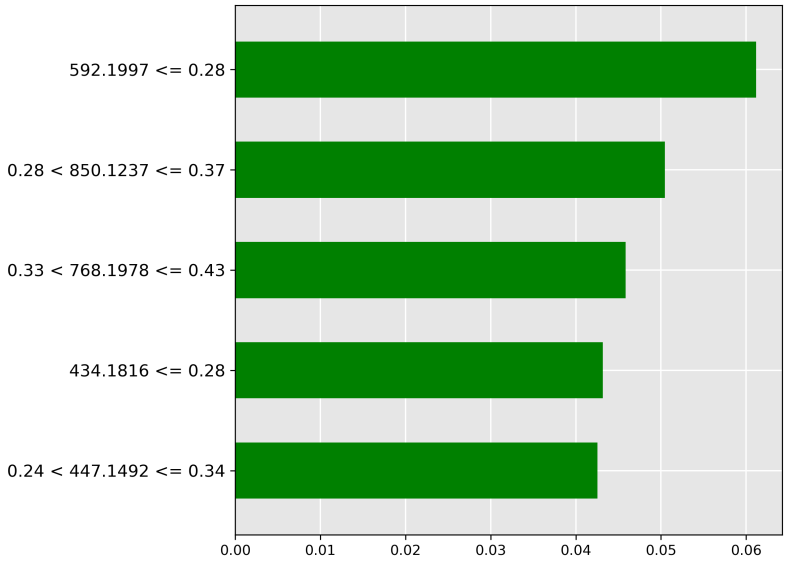

Figure 5.6: LIME explanation for **pre-trained transformer** for classification of oil contamination in **25%** concentration.

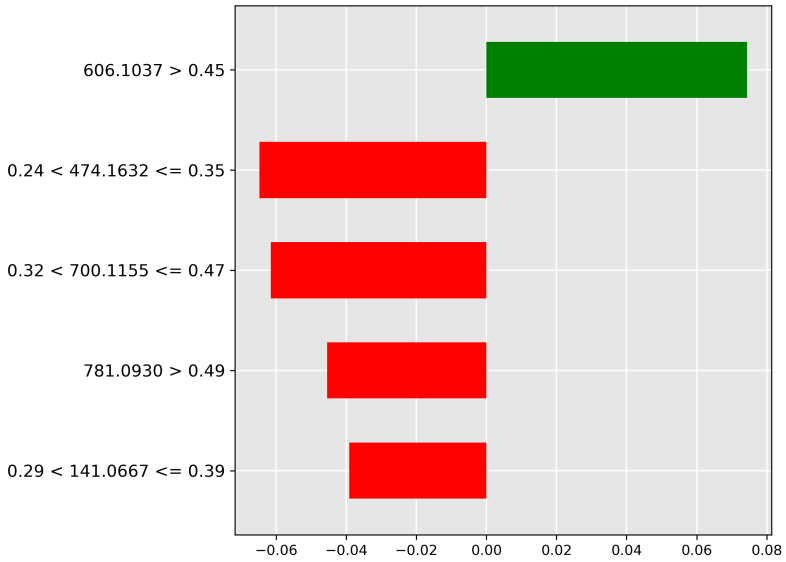

Figure 5.7: LIME explanation for **pre-trained transformer** for classification of oil contamination in **10%** concentration.

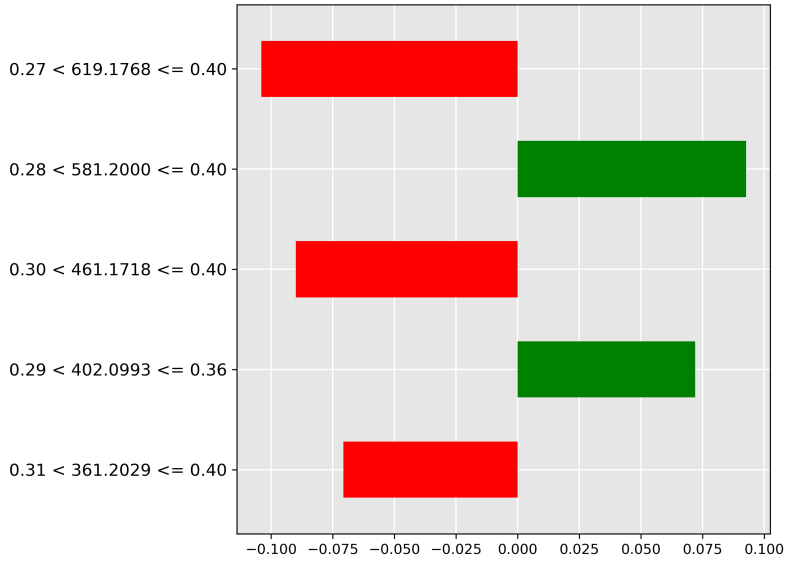

Figure 5.8: LIME explanation for **pre-trained transformer** for classification of oil contamination in **5%** concentration.

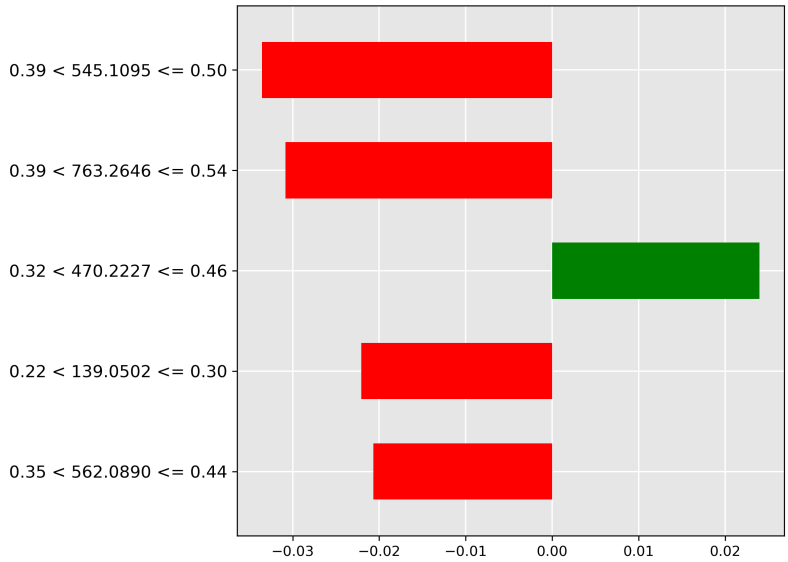

Figure 5.9: LIME explanation for **pre-trained transformer** for classification of oil contamination in **1%** concentration.

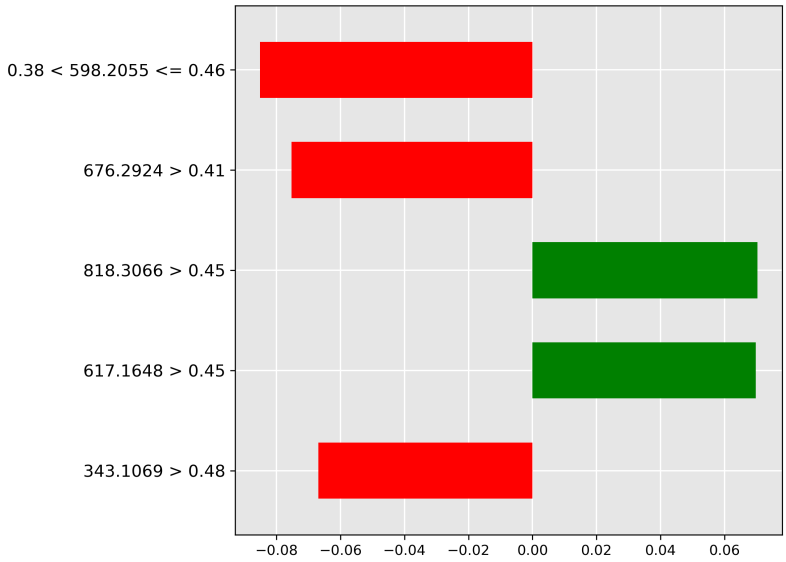

Figure 5.10: LIME explanation for **pre-trained transformer** for classification of oil contamination in **0.1%** concentration.

Predicting minuscule 0.1% trace contamination (Figure 5.10) is driven away (red bar) by the presence of m/z 598.2055 when its normalized intensity is between $0.38 < y \leq 0.46$. Since this ion's presence rules out the 0.1% classification, and is negatively correlated with the 0% baseline, the model has likely isolated a molecular ion associated with the oil contaminant itself.

Finally, for the 0% (pure fish) baseline (Figure 5.11), the model key prediction is the high abundance (> 0.38) of the molecular ion m/z 111.0971 (largest green bar). This ion is chemically hypothesized to be a key native metabolite found consistently in both Hoki and Mackerel.

Cross-Species Contamination

The pre-trained transformer performs exceptionally well, achieving the best accuracy (91.97%) on the cross-species contamination task. The LIME results below illuminate the specific spectral features the model uses to distinguish pure species from the mixed adulterated product.

Figure 5.12 shows the predictive features for the difficult **Hoki-Mackerel mix** class. The strongest negative indicator (red bar) is the low abundance (≤ 0.03) of the molecular ion m/z 251.1667. The model uses the low intensity of this ion to strongly rule out the presence of the mix, suggesting this ion is characteristic of one of the **pure** species.

For the **pure Hoki** class (Figure 5.13), the strongest positive feature (green bar) is the high intensity (> 0.09) of the molecular ion m/z 122.1110. This feature is highly

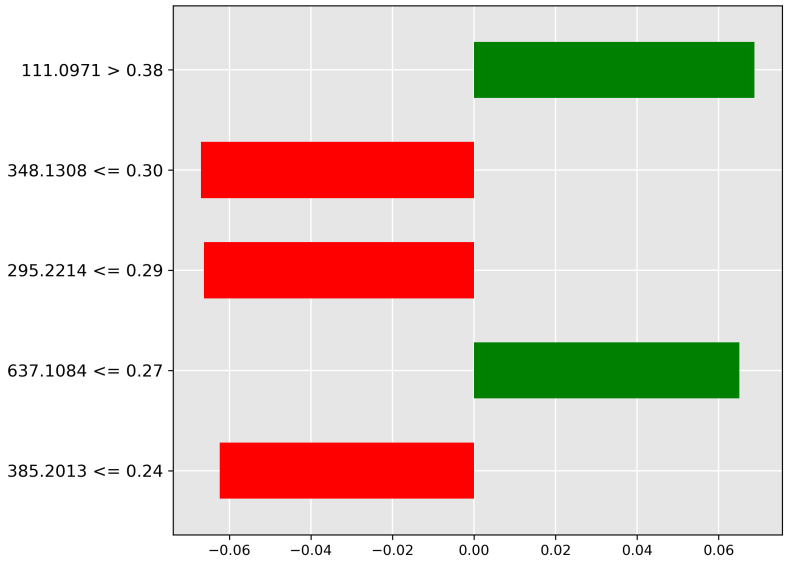

Figure 5.11: LIME explanation for **pre-trained transformer** for classification of oil contamination in **0%** concentration.

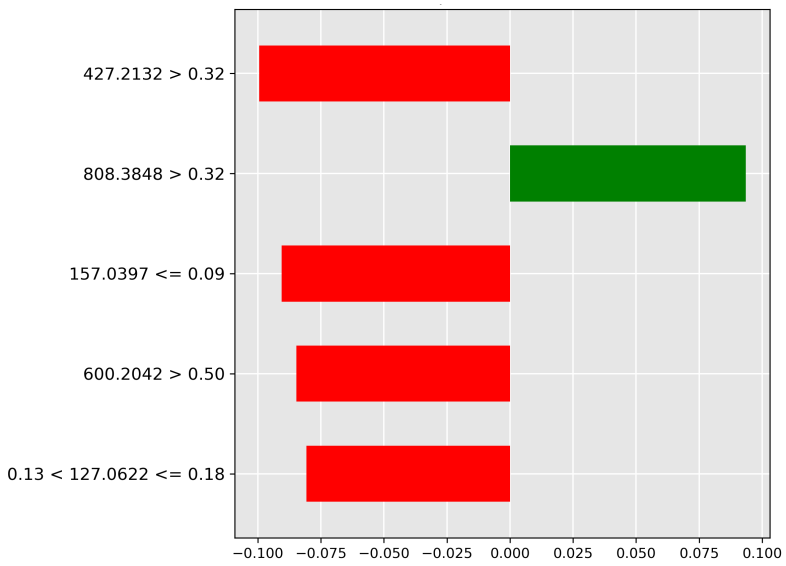

Figure 5.12: LIME explanation for **pre-trained transformer** for cross-species contamination of **Hoki-Mackerel** class.

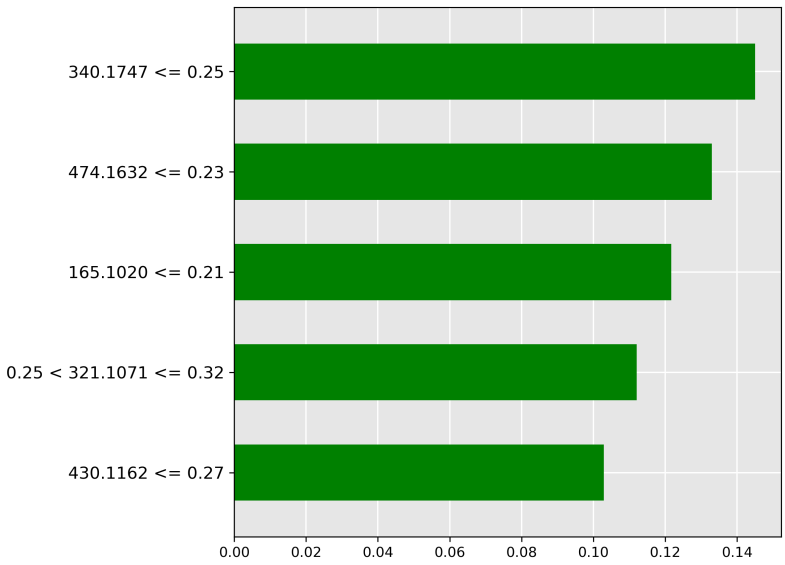

Figure 5.13: LIME explanation for **pre-trained transformer** for cross-species contamination of **Hoki** class.

discriminant, indicating that the molecular ion is likely abundant exclusively in Hoki tissue.

Finally, the LIME explanation for the **pure Mackerel** class (Figure 5.14) shows the most important feature (red bar) is the very low intensity ($\leq$ 0.07) of the molecular ion m/z 367.0835. The model uses this low abundance to confirm the pure Mackerel classification; conversely, a higher abundance would suggest the presence of Hoki or the Hoki-Mackerel mix.

Top 10 Features

In this analysis, we analyze the top 10 features identified by 1D Grad-CAM for both classification tasks across four different models. Figure 5.15 gives the top 10 features graph for oil contamination, and Figure 5.16 gives the top 10 features graph for cross-species adulteration. There are two sets of trends that can be identified from these graphs: model-specific and task-specific.

For model-specific behavior, we identify trends in models that are shared across both classification tasks. For example, for both tasks, the LSTM chooses a tight cluster of features that are in a sequence. While the CNN, MoE, and Transformer choose a broader variety of feature subsets that span the entire mass-to-charge ratio range, except for the MoE for oil contamination. Each of the models chooses a different subset for its top 10 most important features. Similar to the previous chapter, this suggests that they form independent, complementary, and diverse models for an ensemble. This again motivates

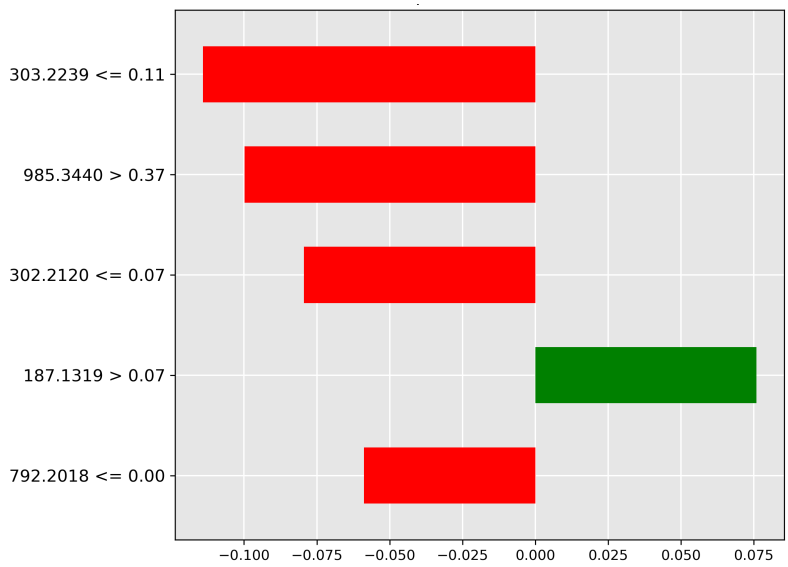

Figure 5.14: LIME explanation for **pre-trained transformer** for cross-species contamination of **Mackerel** class.

the use of an ensemble model.

For task-specific trends, we look for differences in the overall behavior of the models for each task. We notice that for oil contamination, the LSTM and MoE select a subset of features that are in the low range of mass-to-charge ratios ($\approx 100 - 500$). Whereas, for cross-species adulteration, the LSTM and MoE also utilize models from the mid to high range of mass-to-charge ratios ($\approx 600 - 1000$). This suggests that the lipids responsible for differentiating between different concentrations of oil contamination are mainly in the low range of mass-to-charge ratios. The lipids responsible for identifying cross-species adulteration between Hoki and Mackerel are in the mid-to-high ranges of mass-to-charge ratios.

5.4.3 Computational Cost

This computational cost benchmark evaluated our four transformer variants on the oil and cross-species datasets. We measured the training time, inference speed, model size in megabytes, and parameters. Table 5.4 lists this computational cost benchmark. Similar to the previous chapter, we see a striking convergence in inference time between the Transformer and Pretrained Transformer, while also illustrating the significant overhead associated with the more complex MoE and Ensemble Transformer models.

The parity in inference time between the standard Transformer and Pretrained Transformer exhibits fast and almost identical inference times. The extra time allocated for pretraining is shown in the large differences in the training time column. In stark contrast,

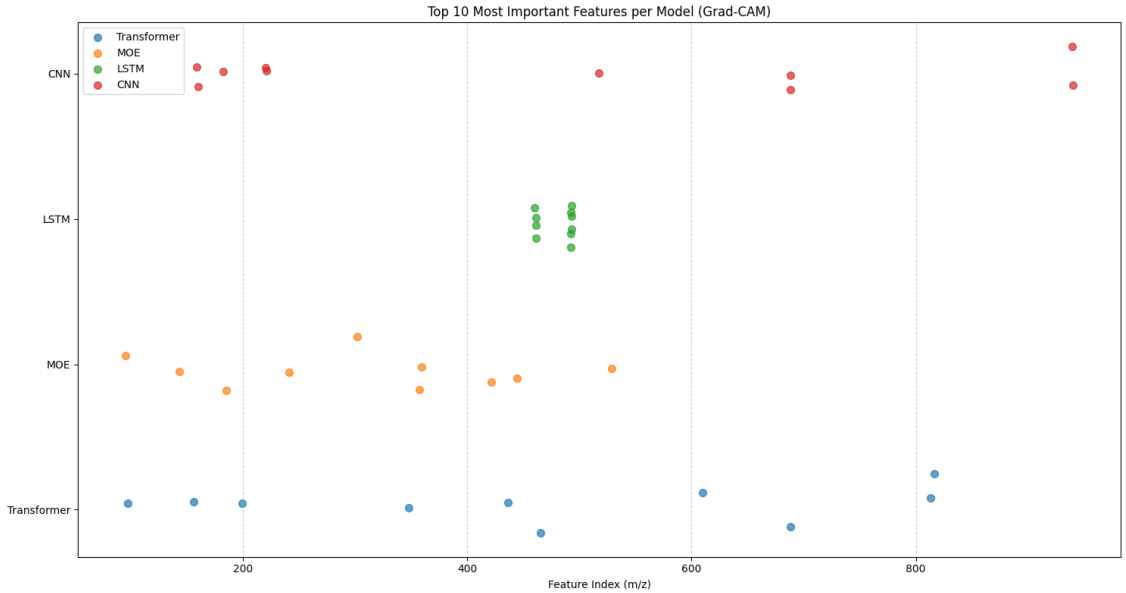

Figure 5.15: In this diagram, we present the top 10 features for **oil contamination** classification that have been identified by our 1D Grad-CAM analysis. We analyze the top 10 features for the CNN, LSTM, MOE, and Transformer models.

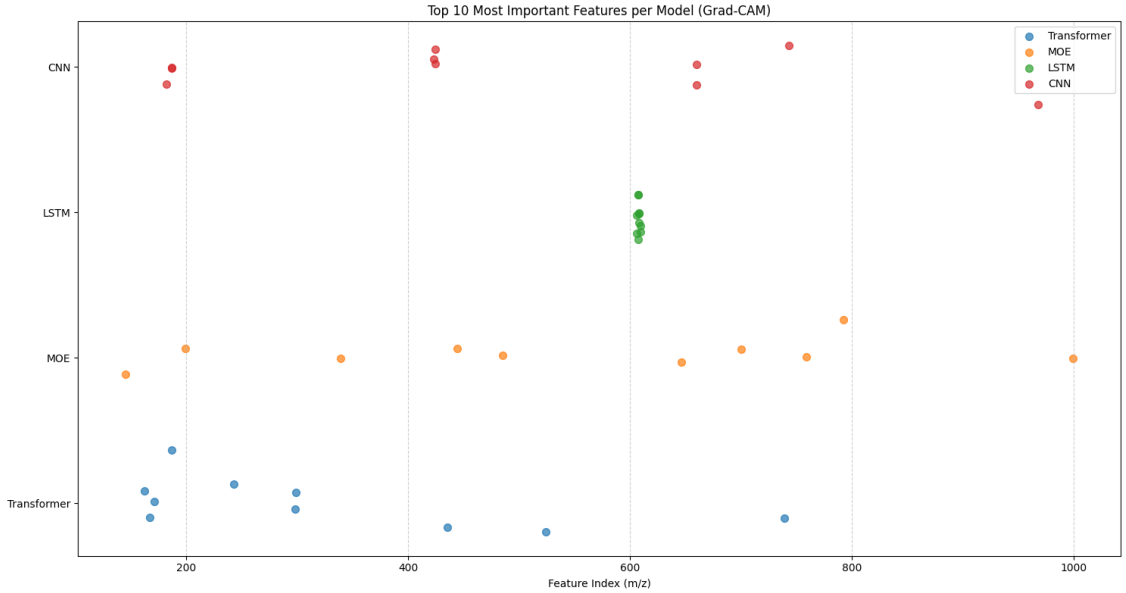

Figure 5.16: In this diagram, we present the top 10 features for fish **cross-species adulteration** classification that have been identified by our 1D Grad-CAM analysis. We analyze the top 10 features for the CNN, LSTM, MOE, and Transformer models.

Table 5.4: Model Performance on Oil and Cross-Species Datasets.

Model	Dataset	Training Time (s)	Inference Time (s)	Model Size (MB)	Parameters
Ensemble	cross-species	3.0423	0.0272	976.1705	244042633
Moe	cross-species	0.3025	0.2415	77.9022	19475559
Transformer	cross-species	0.2069	0.0028	71.4527	17863171
Pretrained	cross-species	42.6161	0.0023	71.4527	17863171
Ensemble	oil	2.7665	0.0117	976.2704	244067605
MoE	oil	0.2820	0.1642	77.9355	19483883
Transformer	oil	0.1947	0.0025	71.4860	17871495
Pretrained	oil	30.2554	0.0023	71.4860	17871495

the more complex architectures demonstrate clear performance trade-offs. The Ensemble model, while achieving fast inference, requires a training time more than ten times longer than simpler transformers. Reflecting the linear cost of training multiple models. The MoE architecture is a notable outlier, delivering the slowest inference speeds by a massive margin—nearly 100 times slower than the standard Transformer. This result strongly indicates that for models of this scale, the computational overhead of the MoE's expert-routing mechanism provides no benefit and instead introduces a significant performance bottleneck.

Ultimately, this analysis underscores that optimization is king. The benchmark shows that a well-optimized, standard Transformer architecture provides an outstanding balance of rapid training and exceptionally fast inference. The similar performance of the from-scratch and pretrained models suggests that the final optimization process is the dominant factor in determining runtime speed. For developers working with these datasets, the data indicates that architectural complexity should be approached with caution, as a simple, efficiently compiled model offers the most direct path to high performance.

Expert utilization, shown in Figures 5.17 and 5.18, is balanced across tasks, indicating effective model capacity use without expert collapse.

5.4.4 Ablation of MoE Transformer

After analyzing the comparative performance of the routing strategies, we conducted a series of ablation studies to better understand the architectural choices in our MoE Transformer and their impact on the model performance. These studies systematically examine the effects of majority voting versus Top-k routing, expert count, and Top-k routing values, providing insights into the model's behavior and optimal configuration for different classification tasks.

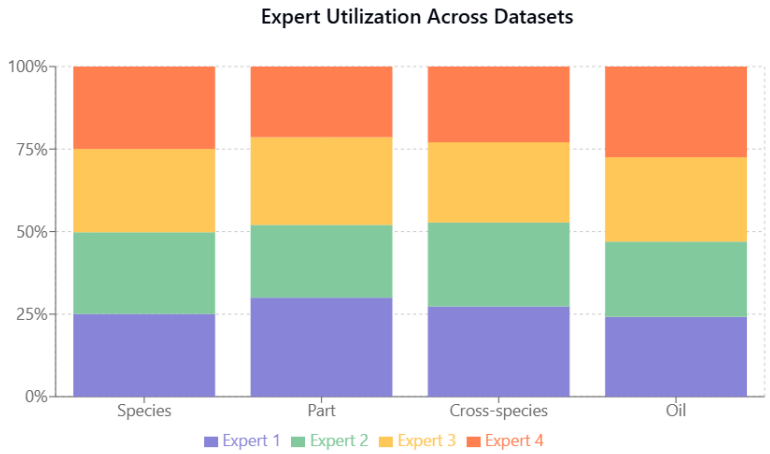

Figure 5.17: This stacked bar chart shows the expert utilization of the MoE Transformer with 4 experts across all four classification tasks. We see balanced expert utilization across all tasks, indicating that the model does not suffer from expert collapse, where one or more experts are hardly used or redundant in training and inference.

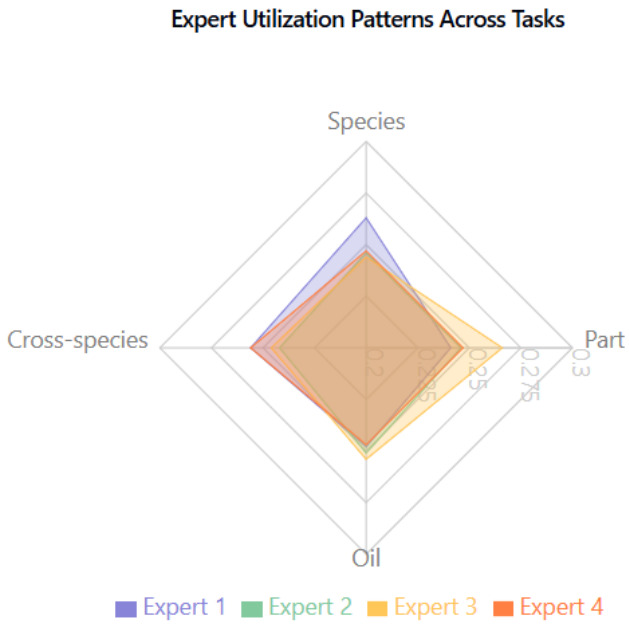

Figure 5.18: This radar chart illustrates the utilization patterns of four different experts across four task categories: Species, Part, Oil, and Cross-species.

Table 5.5: Top-k Routing versus Majority Voting

Dataset	Top-k Routing	Majority Voting
Cross-species	**88.04 ± 4.84**	87.56 ± 1.10
Oil	**45.51 ± 5.60**	43.65 ± 0.53

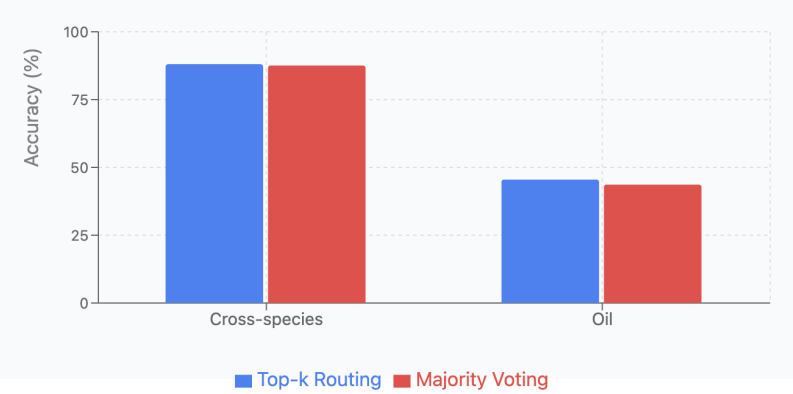

Figure 5.19: Majority Voting versus Top-k Routing Bar Chart

Majority Voting versus Top-k Routing:

We compared two MoE routing strategies: Top-k routing (dynamically routes inputs to k=2 most relevant experts) and majority voting (processes inputs through all experts and averages outputs). Each variant was tested across four classification tasks with 5 runs per configuration.

As shown in Table 5.5, and Figure 5.19 Top-k Routing achieved higher accuracy across all tasks, but with greater variance. Majority Voting demonstrated more consistent predictions with lower standard deviations, aligning with previous research (Suzuoki & Hatano, 2024). Both methods performed poorly on oil contamination detection ($< 46\%$ accuracy), suggesting this task requires further architectural improvements beyond routing strategy modifications.

Expert Count Analysis:

We tested model performance with different numbers of experts (E=2, 4, 8) across four classification tasks, running each variant 5 times. Results are shown in Table 5.6 and Figure 5.20.

Key findings: (1) Cross-species contamination showed stable performance (87-89%) across all configurations; (2) Oil contamination detection was consistently difficult (43-45%) regardless of expert count; (3) E=4 provides the best balance between specialization capacity and ensuring sufficient training data per expert.

Table 5.6: Expert Count Analysis

Dataset	$E = 2$	$E = 4$	$E = 8$
Cross-species	**88.87 ± 1.17**	88.04 ± 4.84	87.82 ± 1.29
Oil	45.37 ± 1.35	**45.51 ± 5.60**	43.75 ± 1.69

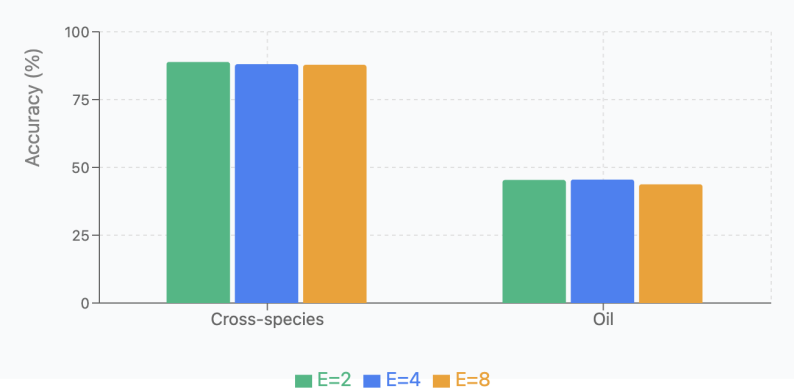

Figure 5.20: Expert Count Analysis Bar Chart

Table 5.7: Top-k Routing

Dataset	$k = 1$	$k = 2$	$k = 4$
Cross-species	89.27 ± 0.89	88.04 ± 4.84	**89.93 ± 1.82**
Oil	**45.69 ± 1.78**	45.51 ± 5.60	45.38 ± 1.14

Top-k Routing:

We compared performance with different k values (1, 2, 4) across the same four tasks, using the same experimental setup. Results are shown in Table 5.7 and Figure 5.21.

Key findings: (1) k=4 was optimal for cross-species detection, suggesting benefit from broader expert consultation; (2) Oil contamination showed similar performance across all k values (45%); (3) Overall, k=2 provides the best balance between specialization and redundancy for most tasks.

5.4.5 Ablation of Transfer Learning

After the description of the experimental design, the Results section presents the empirical findings and analyzes the observed patterns in transfer learning effectiveness across different fish fraud detection tasks. This section is broken down into discussions on Overall Performance and Asymmetric Transfer Effects, the Impact of Task Difficulty on Transfer Learning, the Role of Model Architecture (MoE Transformer), Model Stability and

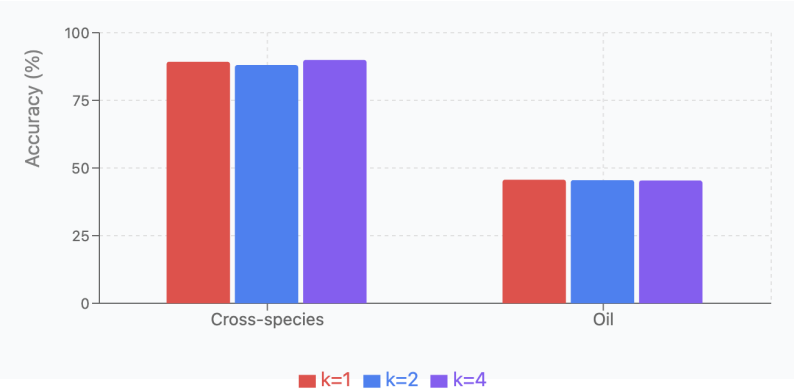

Figure 5.21: Top-k Routing Bar Chart

Table 5.8: Transfer Learning Results on Test Set (%) with Standard Deviations

Source → Target	Baseline	Transfer	Effect
Oil → Cross-species	88.04% ± 4.84%	83.44% ± 6.91%	-
Species → Oil	45.51% ± 5.60%	46.28% ± 6.84%	+
Part → Oil	45.51% ± 5.60%	45.64% ± 7.21%	+
Cross-species → Oil	45.51% ± 5.60%	49.10% ± 5.85%	+

Standard Deviations, and Practical Implications and Resource Allocation.

Overall Performance and Asymmetric Transfer Effects

The experimental results, with performance metrics averaged over 30 independent runs, are detailed in Table 5.8. A visual comparison of baseline and transfer learning performance across six source-target pairs is presented in the bar chart in Figure 5.22. This chart displays accuracy percentages, with blue bars for baseline performance, green bars for transfer learning results, and error bars indicating standard deviations. The visualization reveals mixed outcomes: the Oil → Species transfer achieves perfect accuracy (100%) with no improvement over its baseline; Oil → Part and Oil → Cross-species show notable performance declines (dropping from 72.46% to 57.50% and 88.04% to 83.44%, respectively); while the three "→ Oil" transfers (Species → Oil, Part → Oil, Cross-species → Oil) demonstrate modest improvements over their common baseline of 45.51

Further illustrating these dynamics, the radar chart in Figure 5.23 depicts transfer learning performance across the same six source-target pairs. This chart uses overlapping polygons for baseline accuracy (purple) and transfer accuracy (green), allowing direct comparison. Areas where the green polygon extends beyond the purple signify improvements, while recession indicates degradation. The three transfers targeting oil contamination de-

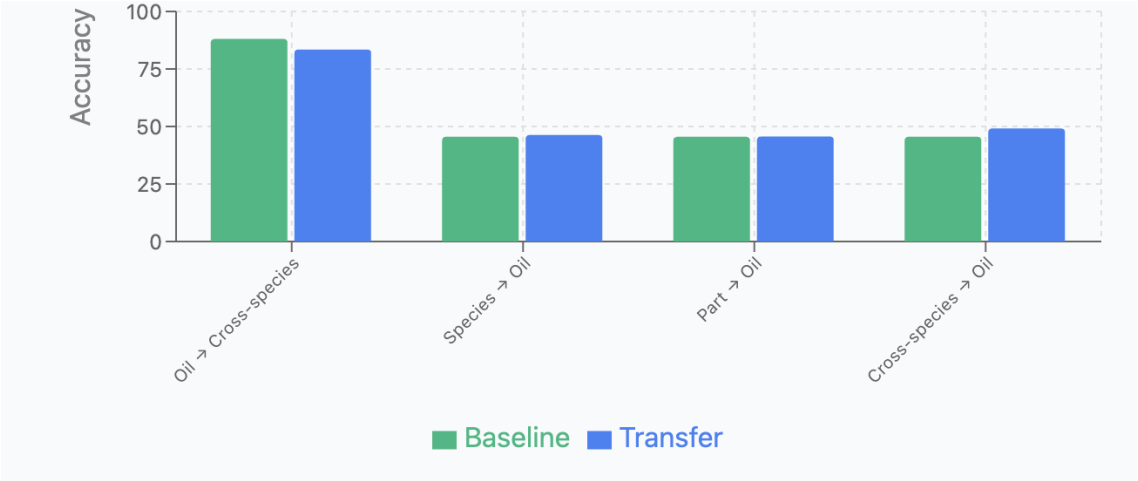

Figure 5.22: Test Classification Improvements Bar Chart: This bar chart illustrates the performance (%) of Baseline (blue) versus Transfer Learning (green) models on six different classification scenarios: Oil → Cross-species, Species → Oil, Part → Oil, and Cross-species → Oil. Error bars indicate the variability in performance.

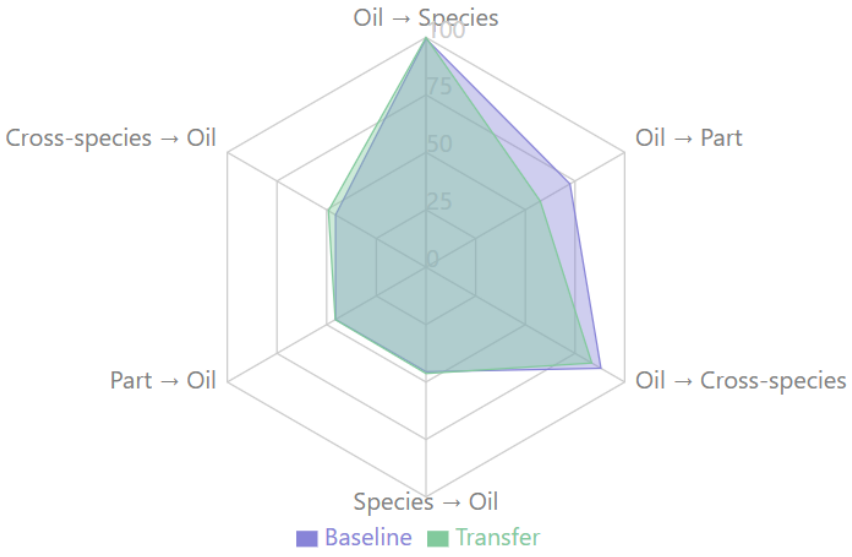

Figure 5.23: Test Classification Improvements Radar Chart: This radar chart provides a comparative view of Baseline (green) and Transfer Learning (blue) model performance (in percentages) across six distinct classification tasks: Oil → Species, Oil → Part, Oil → Cross-species, Species → Oil, Part → Oil, and Cross-species → Oil.

tection ("→ Oil") show modest improvements, with Cross-species → Oil achieving the largest gain (+3.59%).

The experimental results highlight a striking asymmetry in the effectiveness of transfer learning across the different fish fraud detection tasks. Transfer learning consistently improves oil contamination detection accuracy by modest yet meaningful margins (between 0.13% and 3.59%), irrespective of the source domain used for pre-training. Conversely, Transfer learning substantially degrades performance when applied to fish body part identification, With accuracy dropping by 14.96% when transferring from oil contamination detection. Similarly, cross-species adulteration detection suffers a notable decline of 4.60% when knowledge is transferred from oil contamination detection. These patterns suggest that the target domain, rather than the source domain, is the primary determinant of transfer learning success in this context.

Impact of Task Difficulty on Transfer Learning

Interestingly, the most difficult task, oil contamination detection (baseline accuracy of 45.51%), consistently benefited from transfer learning regardless of the source domain. In contrast, easier tasks, such as fishy body part identification (72.46% baseline) and cross-species adulteration detection (88.04% baseline), were consistently harmed by transfer learning. This observation contradicts the intuitive expectation that easier tasks might derive greater benefit from knowledge transfer. A potential explanation for this phenomenon is that the chemical signatures detected for oil contamination may represent fundamental molecular patterns that are enhanced by exposure to diverse fish tissue data. Conversely, the specific features distinguishing different fish parts or identifying cross-species adulteration might be more specialized and could be confounded by knowledge transfer from other domains.

Role of Model Architecture (MoE Transformer)

The choice of a Mixture of Experts (MoE) transformer architecture may partially explain these results. The MoE approach facilitates the specialization of different experts on various aspects of the data, potentially creating a more modular knowledge representation. When transferring to the oil contamination detection task, this modularity might enable the model to retain and build upon relevant chemical pattern recognition capabilities while discarding irrelevant information. However, when transferring to fish body part identification or cross-species adulteration detection, the pre-trained experts may have developed specializations that conflict with the target task requirements, leading to negative transfer. The gating network's routing mechanism, which determines expert engagement for given inputs, might necessitate substantial retraining to adapt effectively to these tasks, thereby rendering transfer learning less effective than training from scratch.

Model Stability and Standard Deviations

The standard deviations observed in the results offer important insights into model stability across different transfer scenarios. All transfers targeting oil contamination detection shared a similar baseline performance (45.51% ±5.60%) but displayed varying degrees of improvement and stability; notably, the Cross-species → Oil transfer demonstrated the most substantial and consistent improvement, achieving 49.10% ±5.85%.

Practical Implications and Resource Allocation

From a practical standpoint, these findings carry significant implications for the development and deployment of automated quality control systems in the seafood processing industry. The computational overhead associated with transfer learning must be carefully weighed against the potential performance gains, which, as the results suggest, are highly task-dependent. For oil contamination detection, investing in transfer learning appears to be a worthwhile endeavour, potentially improving detection accuracy by up to 3.59% when knowledge is transferred from the cross-species domain. This nuanced understanding of when and where to apply transfer learning can effectively guide resource allocation in real-world deployment scenarios, thereby optimizing both computational efficiency and detection accuracy in fish fraud prevention systems.

5.4.6 Ordinal Classification

The oil contamination task, with its seven ordered concentration levels (0% to 50%), is a classic example of an ordinal classification problem. As illustrated in Figure 5.24, ordinal data is distinct from nominal data (which has no order) and metric data (which assumes equidistant intervals). This inherent order—where a 5% concentration is closer to 1% than to 50%—is critical information that standard classification or regression models may fail to capture.

To properly explore the performance of the transformer method for this challenging task, we compare our standard classification approach with three ordinal classification approaches. We use stratified k-fold cross-validation, Balanced Classification Accuracy (BCA), and Mean Average Error (MAE) to evaluate the accuracy. We use MAE to measure accuracy for ordinal classification because it effectively penalizes misclassifications based on their distance from the true category, a critical distinction from standard accuracy shown in Figure 5.25. The ordinal classification approaches we explore are:

- **Regression** (Frank & Hall, 2001): This approach treats the K ordinal categories as continuous numerical values (e.g., $1, 2, 3, \cdots, K$). A standard regression model is trained to predict this value using a loss function like Mean Squared Error (MSE). The final predicted class is then obtained by rounding the continuous output to

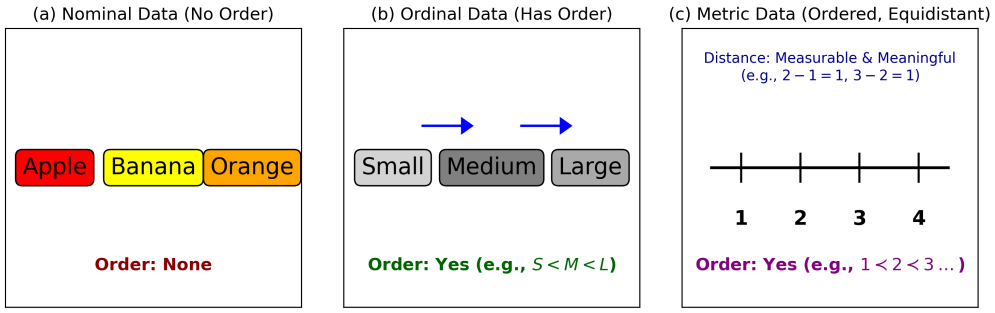

Figure 5.24: A conceptual comparison of data measurement scales, illustrating the unique nature of (b) Ordinal Data, where categories have a meaningful order but non-uniform distances, as seen in the oil contamination task.

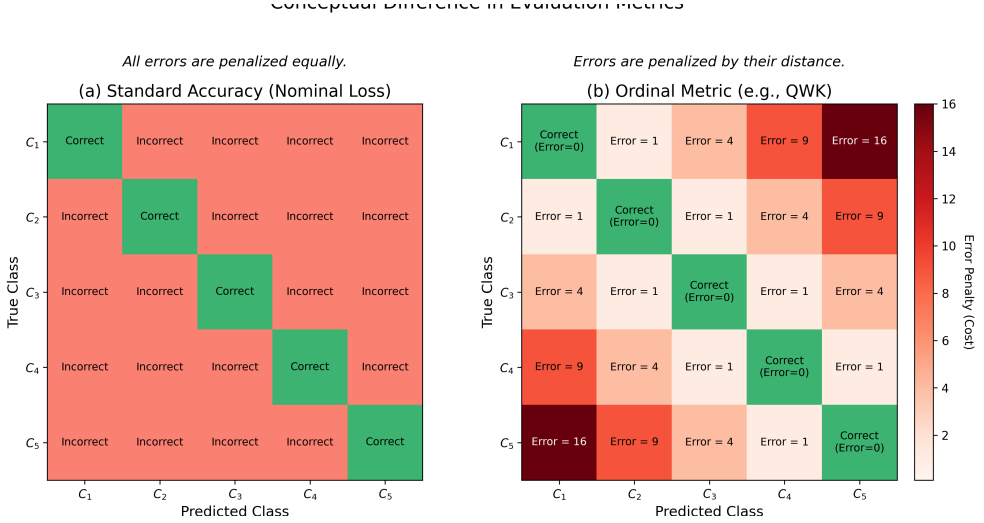

Figure 5.25: Conceptual difference in evaluation. (a) Standard accuracy treats all misclassifications as equally incorrect (a 0-1 loss). (b) Ordinal metrics penalize errors based on their distance. This diagram illustrates a quadratic penalty, as used in Quadratic Weighted Kappa (QWK), which correctly identifies that a small error (e.g., predicting C_4 for C_5) is preferable to a large one (e.g., predicting C_1 for C_5).

the nearest integer. This method is simple but assumes that the intervals between consecutive classes are equidistant.

- **Standard Classification**: This is the standard nominal approach, which treats the K categories as independent, unordered labels. The model outputs a logit vector of size K and is trained using a Cross-Entropy Loss. This method is fundamentally "ordinal-blind," as it penalizes all misclassifications equally, regardless of their distance from the true label.

  The final two methods, CORAL and CLM, are modern deep learning strategies that explicitly model this ordinal structure, as conceptualized in Figure 5.26.

- **CORAL** (Cao et al., 2020): The COnsistent RAnk Logits (CORAL) framework reframes the task using $K-1$ binary classifiers. Each classifier i predicts the probability that the true label is greater than rank i (i.e., $P(y > i)$). This is achieved using a shared set of weights and $K-1$ logits, trained with a binary cross-entropy loss. This design explicitly enforces the ordinal constraint, as predicting $y > i$ implies $y > i-1$. The final class is predicted by summing the number of logits that are positive (or have a probability > 0.5) and adding one.

- **Cumulative Loss (CLM)** (Niu et al., 2016): This method, based on cumulative link models, also learns $K-1$ outputs. These outputs represent the cumulative probabilities $P(y \leq i)$. The model is trained to predict these cumulative probabilities, typically by learning a set of ordered thresholds (cutpoints) that partition the output of a shared feature extractor. The probability for a specific class i is then derived from the difference between adjacent cumulative probabilities, $P(y = i) = P(y \leq i) - P(y \leq i-1)$, directly modeling the ordinal nature of the labels.

Table 5.9: Final Cross-Validation Results

	TRAIN		**TEST**	
Method	**BCA**	**MAE**	**BCA**	**MAE**
CLASSIFICATION	1.0000 ± 0.0000	0.0000 ± 0.0000	0.4519 ± 0.0573	1.7305 ± 0.2784
REGRESSION	0.3436 ± 0.1726	1.0024 ± 0.4716	0.1476 ± 0.0551	1.6332 ± 0.1219
CORAL	0.4856 ± 0.1725	0.8621 ± 0.3655	0.3048 ± 0.1023	1.5732 ± 0.1935
CLM	0.5391 ± 0.1627	0.6856 ± 0.2618	0.2833 ± 0.1236	1.6145 ± 0.2957

The results table reveals a clear performance hierarchy among the different modeling approaches, strongly influenced by how each method handles the ordinal nature of the data. The standard classification approach achieves the highest training accuracy (1.0000

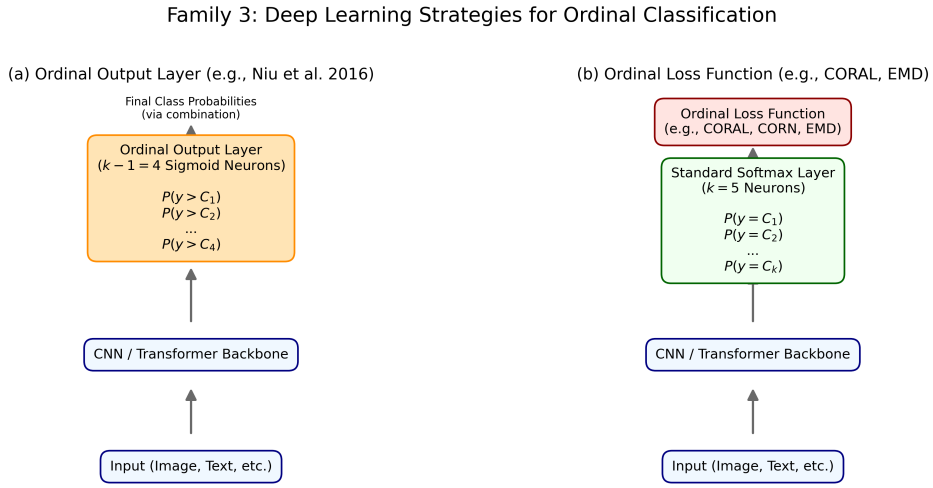

Figure 5.26: Common deep learning strategies for ordinal classification, showing the two approaches tested: (a) The Ordinal Output Layer approach, similar to CLM, which models cumulative probabilities, and (b) The Ordinal Loss Function approach, such as CORAL, which uses a specialized loss on a standard softmax output.

Mean BCA), but this comes at the cost of poor generalization, as evidenced by its mediocre test BCA (0.4519) and the highest test MAE (1.7305). This large gap between training and test performance indicates significant overfitting. Because this method is "ordinal-blind," as illustrated in Figure 5.25(a), it treats the contamination levels as independent, unordered labels. While it can learn to perfectly separate these discrete categories in the training set, it fails to understand the inherent sequence (e.g., that 10% contamination is closer to 5% than to 50%). Consequently, it penalizes all misclassifications equally, leading to a high average error distance on the test set.

In contrast, the regression-based approach demonstrates the poorest performance overall, with the lowest BCA on both the training (0.3436) and test (0.1476) sets. This underperformance can be attributed to its fundamental assumption that the ordinal categories can be treated as equidistant numerical values (e.g., 1, 2, 3...), as shown in Figure 5.24(c). The low accuracy suggests this assumption is invalid for the oil contamination data; the perceived "distance" between a 0% and 1% contamination level is likely not the same as between a 25% and 50% level. By converting the classification problem into a regression task and simply rounding the output, the model loses critical information about the class boundaries and struggles to make accurate discrete predictions.

The methods explicitly designed to handle ordinal data, CORAL and CLM, yield the most promising results for real-world applications. CORAL, in particular, strikes the best balance, delivering a competitive test BCA (0.3048) while achieving the lowest test MAE

Table 5.10: Comparative Analysis of Ordinal Classification Model Families

Family	Core Models	Strengths	Weaknesses
1. Threshold-Based	POM, CLMs, SVOR	- Statistically rigorous (POM/CLM) - Highly interpretable coefficients - Guarantees ordered predictions - Excellent on low-dim, tabular data	- **Parallel Lines Assumption** is rigid - Can be hard to test/validate assumption - Less competitive on high-dim data - SVOR can be slow to train
2. Binary Decomposition	Cumulative, Adjacent	- Conceptually simple and flexible - Can use *any* binary classifier - No parallel lines assumption - Often highly competitive	- Can lead to **inconsistent** probabilities - (e.g., $P(y > 2)$ might be ¡ $P(y > 3)$) - Trains $k - 1$ models, which can be slow - Less direct interpretation
3. Deep Learning	CORAL, CORN, EMD Loss	- State-of-the-art on high-dim data - (Images, text, etc.) - No strong statistical assumptions - Can be built into any NN architecture	- Requires large datasets - Prone to overfitting on small data - "Black box" / less interpretable - Computationally expensive

(1.5732). This superior MAE is a direct result of its architecture (see Figure 5.26(b)), which uses a series of binary classifiers to enforce the rank ordering of the classes. The CLM method, which models cumulative probabilities (Figure 5.26(a)), also performs well, producing a test MAE (1.6145) far better than the ordinal-blind methods. Together, these results underscore the importance of selecting a model that aligns with the data's underlying structure. For an ordinal task like grading contamination, models that incorporate this sequential information generalize more effectively and produce more meaningful predictions by minimizing the distance of errors. This trade-off between different model families is summarized in Table 5.10, which compares the general strengths and weaknesses of each approach.

5.5 Chapter Summary

Our research demonstrates that deep learning models can revolutionize REIMS-based marine biomass analysis.

5.5.1 Conclusions

This chapter concludes that the novel unsupervised pretraining method, "Masked Spectra Modeling" (MSM), shows task-dependent effectiveness. MSM proves valuable for difficult problems like oil contamination detection—a challenging ordinal classification task—where it successfully improves the model's performance. This is likely because the pretraining task, which implicitly learns to predict spectral peaks, mirrors the techniques chemists use to analyze REIMS data and provides a strong foundation for identifying subtle chemical signals. However, for tasks with more distinct signatures like cross-species adulteration, pretraining was detrimental to performance. In terms of architecture, the Mixture of Experts (MoE) Transformer outperforms the regular Transformer on the majority of the classification tasks, as its divide-and-conquer strategy of using specialized expert networks improves the balanced accuracy score. Furthermore, transfer learning with the MoE Trans-

former consistently benefits the difficult oil contamination task regardless of the source task, indicating that for this specific challenge, the performance increase justifies the additional computation required. Finally, our analysis of the methods applied to the oil contamination task revealed that the distance-aware loss functions we tested, even those from a simple regression approach, dramatically outperformed the standard 'ordinal-blind' classification. Furthermore, the more complex, specialized ordinal models we evaluated offered additional, though more modest, performance benefits for this specific task.

5.5.2 Future Work

While our study has yielded promising results, several important areas warrant further investigation:

1. **Overfitting for Oil Contamination:** Address the significant overfitting issue for oil contamination with further exploration of ordinal classification techniques.

2. **Multi-modal Data Fusion:** Combine REIMS data with other analytical techniques (e.g., Near-infrared spectroscopy or DNA barcoding) to create more robust classification systems that leverage complementary information sources. This could improve accuracy in challenging cases where single-modality analysis is insufficient.

3. **Transfer Learning Across Species:** Investigate whether models trained on Hoki and Mackerel can transfer their learned features to detect contamination in other fish species, potentially reducing the need for extensive new training data.

4. **Causes of Asymmetric Transfer:** The underlying causes of the asymmetric transfer effects, particularly why oil classification consistently benefits from transfer, while other domains do not.

5. **Transfer Learning Feature Embeddings:** Additionally, exploring the feature representations learned in each domain could provide valuable insights into why certain transfers are more successful than others.

6

Contrastive Learning for Batch Detection

Addressing the need for enhanced traceability within the seafood supply chain, Chapter 6 introduces a novel approach to batch detection using REIMS data. This chapter presents "SpectroSim," a self-supervised contrastive learning framework specifically designed for identifying whether fish samples originate from the same processing batch. The chapter will detail the motivation for this approach, the SpectroSim architecture (which adapts SimCLR with a Transformer encoder and a custom projection head for mass spectrometry). Experimental results will demonstrate its performance, computational cost, and compare it against both traditional binary classification methods and other state-of-the-art contrastive learning frameworks. Finally, to ensure the model's decisions are interpretable, Gradient-weighted Class Activation Mapping (Grad-CAM) is employed to visualize the key spectral features the model uses to determine batch similarity.

6.1 Chapter Overview

Traditional approaches to supply chain traceability in the seafood industry, such as physical tagging, are often complex and costly, creating a significant barrier to widespread adoption (Mai et al., 2010; Thompson et al., 2005). While the field is rapidly advancing toward digital solutions like blockchain, 'Digital Food Product Passports', and mobile data-entry platforms (Turkson et al., 2025; Jiang et al., 2025; Untal et al., 2025; Gastaldi García, 2025), these systems are still fundamentally vulnerable to deliberate fraud at the point of data entry (Helyar et al., 2014). A digital record is only as trustworthy as the initial data it is given. This creates a critical need for an intrinsic, chemical-based verification method that can validate a sample's identity against its digital record, a gap that has not been explored for REIMS batch detection. Existing analytical methods for REIMS data are

not designed for the nuanced pairwise comparisons required to determine if two samples share a common origin, necessitating the formalization of a new analytical task.

This chapter addresses this gap by formalizing and solving the problem of batch detection in marine biomass as a self-supervised, pairwise comparison task. The emerging paradigm of contrastive learning provides a powerful framework for this challenge, as it is designed to learn a representation space where similar items are grouped together (Bromley et al., 1993). However, applying standard contrastive learning frameworks like SimCLR presents its own difficulties, including a reliance on large batch sizes, which are ill-suited to the limited and high-dimensional nature of REIMS data (Chen et al., 2020a; Kinakh et al., 2021). Furthermore, the model must be robust enough to learn subtle, batch-specific chemical signatures, optionally doing so without explicit labels.

This chapter aims to resolve these issues by proposing a novel, self-supervised contrastive learning framework called "SpectroSim", which is specifically designed for REIMS mass spectra. The solution adapts the SimCLR framework (Chen et al., 2020a) by replacing the standard image-based backbone with a Transformer-based encoder (Vaswani et al., 2017), which is adept at capturing the complex, sequential relationships in the spectral data. To overcome the limitations of small datasets, SpectroSim leverages data augmentation to create positive pairs and negative pairs, with a natural imbalance that favors negative sample mining. This is achieved through a custom projection head, data augmentation, and NT-Xent loss function tailored to the unique characteristics of mass spectrometry.

This contribution innovates existing work by being the first to apply any form of machine learning to the problem of batch detection for REIMS biomass analysis. By successfully developing SpectroSim, this chapter establishes that a self-supervised approach can achieve high accuracy (70.8%) without using any class labels during training, far surpassing traditional binary classification methods. This work provides a new and powerful capability for the seafood industry: a practical, cost-effective, and accurate method for analytical verification. This method complements emerging digital traceability systems (Turkson et al., 2025; Dahariya et al., 2025) by providing an intrinsic, scientific tool to validate that a sample's chemical fingerprint matches its claimed batch, thereby enhancing the integrity of data in the modern supply chain (Piroutková et al., 2025).

6.1.1 Contributions

The main contributions of the chapter are:

1. This chapter formalizes the novel problem of analytical batch detection for REIMS data and proposes "SpectroSim," a new self-supervised contrastive learning framework to solve it. This contribution addresses the critical industry need for a rapid and intrinsic verification method to complement modern digital traceability systems. As far as the authors are aware, this is the first formalization of this task and the

first application of any machine learning framework to solve it using REIMS biomass data.

2. This chapter contributes a novel architectural adaptation of the SimCLR framework, successfully integrating a Transformer-based encoder for 1D spectral data. By replacing the standard CNN backbone with a Transformer and a custom projection head, "SpectroSim" is specifically designed to capture the complex, sequential, and long-range dependencies in mass spectra. This technical contribution overcomes a key limitation of standard contrastive learning, which is ill-suited for high-dimensional, non-image data.

3. This chapter provides the first systematic comparison of learning paradigms for this pairwise comparison task, demonstrating the superiority of self-supervised contrastive learning. The analysis proves that the "SpectroSim" framework, trained without labels, achieves a high accuracy (70.8%) and significantly outperforms all benchmarked supervised binary classification methods (which failed to surpass 60% accuracy), establishing a new state-of-the-art approach for this problem.

4. This chapter demonstrates that contrastive learning can be highly effective on small, high-dimensional datasets, overcoming a primary limitation of standard implementations. "SpectroSim" achieves its high performance with a batch size of only 16, in contrast to the large batches (e.g., 2048+) typically required by SimCLR. This finding is a significant contribution, as it makes contrastive learning a viable strategy for specialized, limited-sample domains like mass spectrometry analysis.

6.2 The Approach

This section presents SpectroSim, our proposed method for self-supervised batch detection of marine biomass using REIMS data. The task is to determine whether pairs of fish samples originated from the same batch, formulated as a contrastive learning problem that can distinguish similar (same-batch) from dissimilar (different-batch) pairs without relying on labels.

6.2.1 Approach Overview

Figure 6.1 illustrates SpectroSim's architecture. It processes pairs of REIMS spectral samples x_1 and x_2 through identical encoders $h_1 = f(x_1)$, $h_2 = f(x_2)$, where f is the encoder, producing embeddings h_1 and h_2. A custom projection head $z_1 = g(h_1)$, $z_2 = g(h_2)$, where g is the projection head, maps these to z_1 and z_2, whose similarity is evaluated using NT-Xent loss. This design extends SimCLR by adapting it for 1D spectral data, addressing SimCLR's limitations (high computational demands, augmentation dependency, and poor small-dataset performance) through a Transformer encoder and a tailored projection

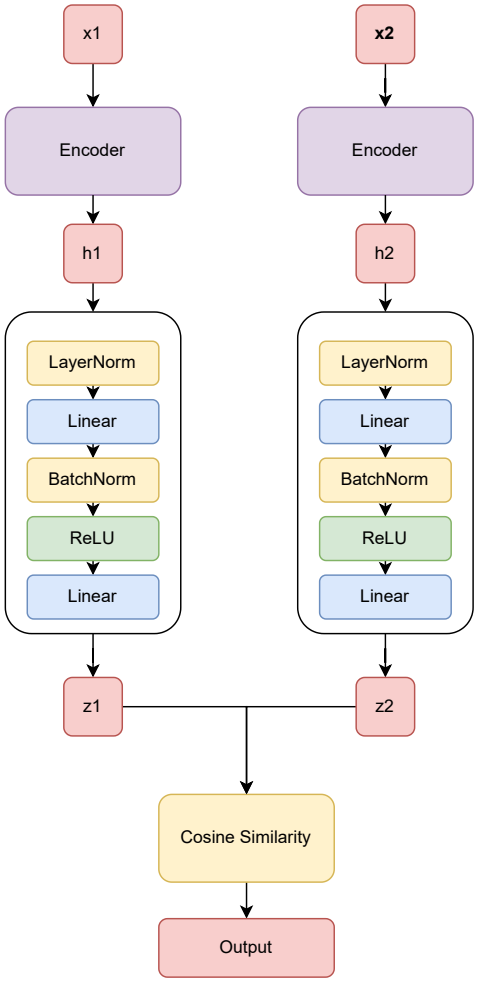

Figure 6.1: SpectroSim Architecture: Paired samples x_1 and x_2 (REIMS spectra) are processed by identical Transformer encoders $h_1 = f(x_1), h_2 = f(x_2)$ to produce embeddings , followed by a custom projection head $z_1 = g(h_1), z_2 = g(h_2)$ yielding z_1 and z_2 , compared via NT-Xent loss.

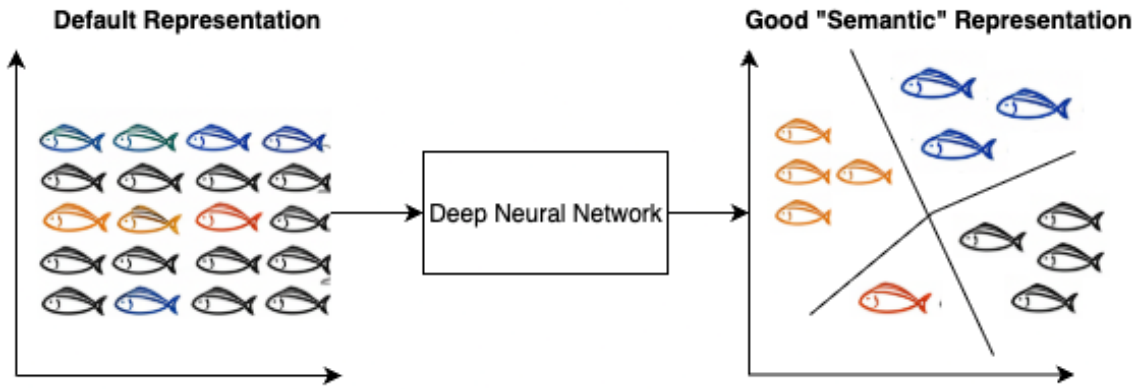

Figure 6.2: This figure illustrates a good "semantic representation". Imagine that the different colors of fish represent fish from different batches in a batch detection task. We feed mass spectrometry samples of these fish into a deep neural network, and it generates embeddings where fish from the same batch are closer together, and fish from dissimilar batches are further apart. In theory, this should make it easy to draw classification boundaries between batches within the embedding space generated by the neural network. Contrastive learning approaches, with their projection head, generate these semantic embeddings, where the distance between samples corresponds to similar and dissimilar batches.

head. These modifications enable efficient learning of batch-specific representations from sequential spectrometry data. A good "semantic" representation for batch detection is illustrated in Figure 6.2.

6.2.2 Data Augmentation

In the SpectroSim framework, the self-supervised learning process relies on creating positive and negative pairs from the input data. A positive pair is formed by creating two different augmented "views" of the same input sample. For one-dimensional REIMS spectra, augmentations can include adding a small amount of Gaussian noise or applying minor shifts to the m/z axis. These augmentations create a pair of samples that are synthetically different but share the same underlying chemical identity. The model is trained to recognize these augmented views as being similar.

Conversely, a negative pair consists of two samples from different original instances in the batch. For any given sample, all other samples in the training batch are considered negative examples. The model's objective is to learn an embedding space where the representations of positive pairs are pulled closer together, while the representations of negative pairs are pushed further apart. This process forces the model to learn the essential, invariant features that define a sample's batch origin, while ignoring the noise introduced by the augmentation process.

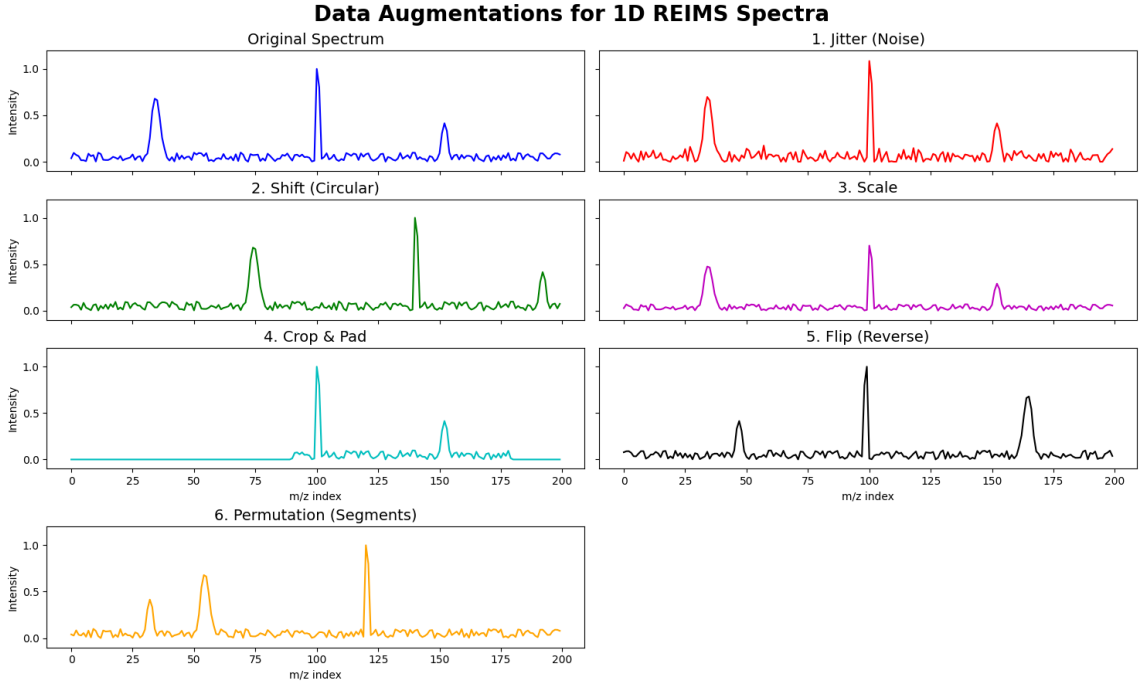

Figure 6.3: Visualization of the six data augmentation techniques described in Section 6.2.2. The **Original Spectrum** (top left) is shown alongside its six transformed 'views': **(1) Jitter**, **(2) Shift**, **(3) Scale**, **(4) Crop & Pad**, **(5) Flip**, and **(6) Permutation**. These augmentations are used to create the positive pairs for self-supervised contrastive learning.

In our contrastive learning framework, we apply a series of data augmentations to the input spectra to generate different "views" of the same sample. These augmentations are crucial for learning robust and invariant representations.

Figure 6.3 shows the various data augmentations we apply to a spectra with this unsupervised contrastive learning framework.

Let an input spectrum be represented by a vector $x \in \mathbb{R}^L$, where L is the number of features (m/z values). The following transformations are applied to generate an augmented view x':

Jitter (Noise)

A random Gaussian noise is added to the spectrum to improve robustness to sensor noise and minor spectral variations.

$x' = x + \epsilon$

where $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, and σ is the noise level.

Shift

The spectrum is circularly shifted along the m/z axis by a random amount. This simulates slight misalignments in the spectral acquisition.

$x'i = x(i - s) \pmod{L}$

where s is a random integer shift, and i is the feature index.

Scale

The intensity of the entire spectrum is multiplied by a random scalar. This helps the model become invariant to variations in sample concentration or instrument sensitivity.

$x' = \alpha \cdot x$

Where α is a random scaling factor drawn from a uniform distribution, typically centered around 1.

Crop and Pad

A random contiguous sub-section of the spectrum is selected and then padded with zeros to its original length. This encourages the model to learn from partial spectral information.

$$x'_i = \begin{cases} x_i & \text{if } i \in [start, end) \\ 0 & \text{otherwise} \end{cases}$$

Where $[start, end)$ is a randomly selected interval of the spectral features.

Flip

The spectrum is horizontally flipped (reversed) along the m/z axis. This forces the model to learn features that are not dependent on their position in the spectrum.

$x'i = xL - 1 - i$

for $i = 0, 1, \ldots, L - 1$.

Permutation

The features of the spectrum are randomly shuffled. This is a strong augmentation that destroys the sequential nature of the data, forcing the model to learn features based on their values rather than their positions.

$x' = P(x)$

Where P is a random permutation operator.

6.2.3 Encoder

The encoder $h_1 = f(x_1), h_2 = f(x_2)$ is a Transformer, chosen over SimCLR's ResNet (He et al., 2016) because it captures long-range dependencies in 1D spectral sequences—crucial for distinguishing subtle batch-specific chemical signatures—unlike ResNet's focus on 2D

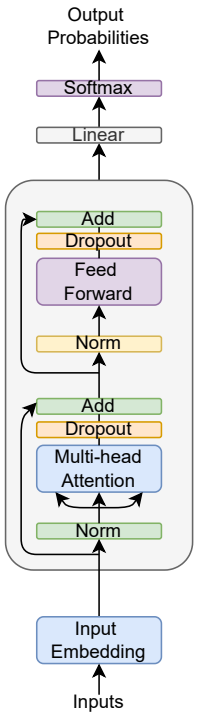

Figure 6.4: In the proposed SpectroSim model, we replace the existing ResNet backbone with a Transformer backbone. The choice of transformer is motivated by previous success in REIMS-based marine biomass analysis tasks demonstrated in Chapters 4 and 5. For the sake of academic rigor, we include the original SimCLR with ResNet backbone in our experiments for comparison, and find that the Transformer backbone significantly outperforms the original SimCLR model.

spatial patterns. This choice reduces preprocessing needs and suits the small, high-dimensional dataset, leveraging the Transformer's attention mechanism for robust feature extraction.

6.2.4 Projection Head

In the SpectroSim model, the projection head $z_1 = g(h_1), z_2 = g(h_2)$ shown in Figure 6.1, is a custom neural network that operates after the main Transformer encoder, refining its output before the final contrastive loss calculation. Unlike the standard feedforward Multi-Layer Perceptron (MLP) used in the original SimCLR framework, this projection head is specifically designed for spectrometry data. Its primary function is to map the high-dimensional feature representations generated by the encoder into a lower-dimensional space where the NT-Xent loss is applied, a process tailored to reduce information loss by preserving the distributions of spectral peaks. This specialized design ensures the final

embeddings are better aligned with the underlying physics of mass spectrometry, creating a more meaningful representation space for comparing samples.

6.2.5 Loss Function

The NT-Xent loss, defined as

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$

The model is trained using the Normalised Temperature-Scaled Cross-Entropy (NT-Xent) loss Chen et al. (2020a), which pulls positive pairs (same-batch samples) together in embedding space while pushing negative pairs (different-batch samples) apart. Given a batch of N sample pairs, the NT-Xent loss treats each of the $2N$ augmented samples as a query, using its one positive partner as the target and the remaining $2(N-1)$ samples as negatives.

The NT-Xent (Normalized Temperature-scaled Cross-Entropy) loss function is the engine of contrastive learning methods like SimCLR. Its primary goal is to learn an embedding space where similar samples are grouped while dissimilar ones are pushed apart. For a given sample in a batch (the "anchor," $\mathbf{z}_i$), the loss focuses on maximizing its similarity with its corresponding augmented view (the "positive," $\mathbf{z}_j$). This is achieved by maximizing the numerator of the fraction. Simultaneously, it minimizes the anchor's similarity to all other samples in the batch (the "negatives," $\mathbf{z}_k$). The **temperature parameter** τ is a crucial hyperparameter that scales the similarity scores; a smaller τ makes the loss more sensitive to difficult negative examples, forcing the model to learn more discriminative features.

This loss function expertly balances **intra-class** and **inter-class** similarity. In this self-supervised context, "intra-class" refers to the positive pair of augmented views derived from the same original image, while "inter-class" refers to pairs made from different original images. The loss function encourages high intra-class similarity by explicitly trying to maximize the similarity score $sim(z_i, z_j)$ in the numerator. This has the effect of pulling the representations of positive pairs closer together in the embedding space. At the same time, it promotes low inter-class similarity by including all other samples in the denominator's sum. To minimize the overall loss, the model must learn to make the similarity scores for all negative pairs as low as possible, effectively pushing their representations far apart from the anchor. This dual action creates well-separated clusters for different semantic classes without ever seeing explicit labels.

The symbol $\mathbb{1}_{[k \neq i]}$ is an **indicator function**, which acts as a simple conditional switch. It evaluates to 1 if the condition inside the brackets is true (i.e., if k is not equal to i) and 0 if the condition is false. In the denominator, the summation runs over all $2N$ samples in the batch. The purpose of the indicator function is to exclude the anchor sample $\mathbf{z}_i$ from

being compared with itself in the sum. Without this condition, the term $\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_i)/\tau)$ would be included. Since a sample's similarity with itself is trivially perfect, this large value would dominate the denominator and prevent the model from learning the subtle but important differences between the anchor and the actual negative samples. It ensures the model focuses on contrasting positive pairs against all *other* samples, not itself.

6.2.6 Threshold

During training, the best classification threshold is determined by evaluating the performance of the contrastive model on the training data itself. For each pair of augmented samples in a batch, the model produces two embedding vectors, h_1 and h_2. The cosine similarity between these vectors is calculated for every pair. The algorithm then iterates through a range of potential threshold values from 0.0 to 1.0. For each threshold, it makes a prediction: if the cosine similarity of a pair is above the threshold, the samples are predicted to be from the same class (a positive pair); otherwise, they are considered a negative pair. The balanced accuracy of these predictions is calculated against the true pair labels. The threshold that results in the highest balanced accuracy across the training batch is selected as the optimal threshold for that training step and is used for subsequent evaluations.

6.3 Experimental Setup

With our methodology clearly defined, we proceed to the experimental setup phase of our study, where we put these diverse machine learning techniques to the test on our REIMS dataset.

6.3.1 Benchmark Methods

To effectively evaluate the proposed method, 18 other binary classification methods are used for comparison on the batch detection for the marine biomass REIMS dataset. These methods include:

1. **Benchmark Technique**: OPLS-DA (Boccard & Rutledge, 2013).

2. **Traditional machine learning algorithms**, providing a baseline for comparison. Specifically, Random Forest (RF) (Ho, 1995), K-Nearest Neighbors (KNN) (Fix & Hodges, 1989), Decision Trees (DT) (Breiman, 2017), Naive Bayes (NB) (Hand & Yu, 2001), Logistic Regression (LR) (Kleinbaum et al., 2002), Support Vector Machines (SVM) (Cortes & Vapnik, 1995), and Linear Discriminant Analysis (LDA) (Balakrishnama & Ganapathiraju, 1998).

3. **Ensemble method** (Hansen & Salamon, 1990): A combination of the above traditional methods.

4. **Contrastive Learning techniques:** Simple Contrastive Learning of Representations (SimCLR) (Chen et al., 2020a).

5. **Deep learning models**, (CNNs) (LeCun et al., 1989c,b,a, 1998); Recurrent Convolutional Neural Networks (ResNet) (He et al., 2016); Long-short Term Memory (LSTMs) (Hochreiter & Schmidhuber, 1997); Variational Autoencoders (VAEs) (Kingma & Welling, 2013); Mamba (Gu & Dao, 2023); and Kolmogorov-Arnold Networks (KANs) (Liu et al., 2024).

The number of epochs for the deep learning methods is set to 100, the same as the proposed method, to facilitate equitable comparisons. Other hyperparameters are shared across methods where applicable, for the same reason.

6.3.2 Benchmark Technique

To contextualize the performance of our proposed models, we selected **Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA)** as the primary benchmark. OPLS-DA is a well-established and standard technique for analyzing REIMS data in the literature, making it an ideal point of comparison between our novel deep learning methods and a traditional approach.

Beyond the OPLS-DA benchmark, the deep learning architectures were evaluated under two distinct frameworks to isolate the impact of the learning strategy itself. These two approaches are illustrated in Figure 6.5.

1. **Direct Binary Classification:** In the first setup, models were trained for standard binary classification. The input was a difference vector, created by subtracting the feature vector of one sample from another. The classifier's task was to use this vector to predict whether the two original samples belonged to the same batch.

2. **Contrastive Learning Framework:** In the second setup, the models served as encoders within a contrastive learning framework. Instead of operating on a pre-computed difference, this method learns to generate a separate, rich embedding for each sample in a pair. The cosine similarity between these embeddings is then used to determine if the samples originate from the same batch.

This dual-framework evaluation allows for a direct assessment of the effectiveness of the contrastive learning paradigm compared to a more traditional classification approach using the exact same underlying model architectures.

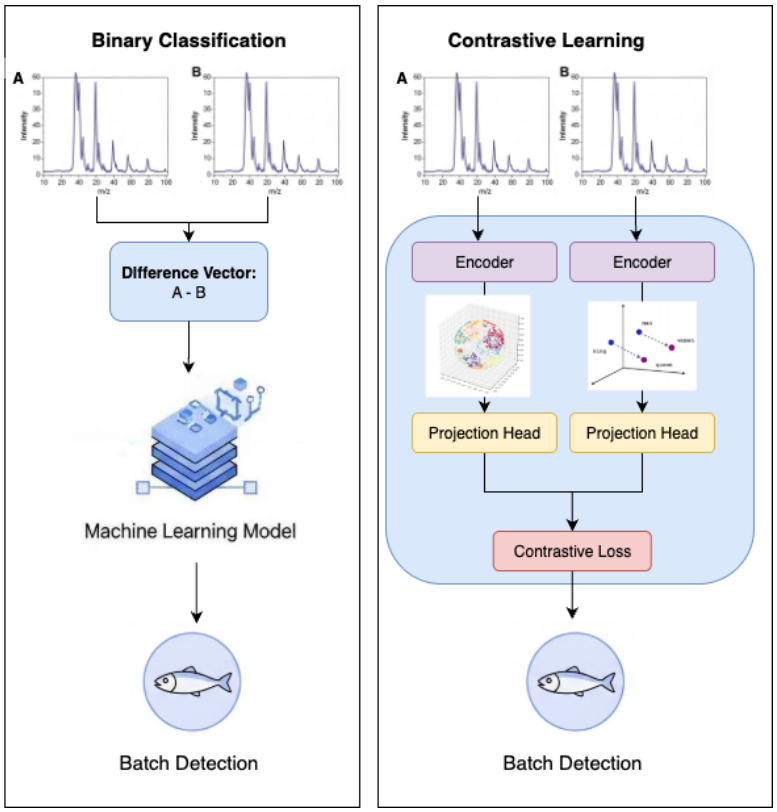

Figure 6.5: This diagram illustrates the difference between binary classification and contrastive learning. Here, we see binary classification takes a difference vector between the two samples and performs binary classification. Whereas contrastive learning learns an embedded representation for each sample and compares these pair-wise. The output for both methods is identical, a prediction of whether the two samples belong to the same or different batches.

6.3.3 Experimental Settings

To evaluate model performance robustly, we used balanced accuracy as the primary metric and conducted 30 independent runs per experiment, with deep learning methods employing early stopping (Morgan & Bourlard, 1989) to tune epochs based on validation data. For contrastive learning, we applied NT-Xent loss with a temperature of 0.07 to enhance sensitivity to pair similarities, while binary classification used balanced accuracy for traditional methods and weighted cross-entropy for deep learning methods to address class imbalance.

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)$$

6.3.4 Parameter Settings

To allow for a fair comparison, model hyperparameters were standardized where possible. The deep learning models, including SpectroSim's Transformer encoder, share a consistent set of parameters derived through experimental tuning. Key settings are detailed in Table 6.1. The AdamW optimizer was used for its effectiveness in regularization. Techniques like label smoothing, dropout, and early stopping were employed to prevent overfitting and ensure robust generalization.

Table 6.1: SpectroSim Transformer Parameter Settings

Parameter	**Value**
Learning rate	1E-5
Epochs	100
Dropout	0.2
Label smoothing	0.1
Early stopping patience	5
Optimiser	AdamW
Loss: Contrastive	NT-Xent
Loss: Binary	CCE
Input dimensions	2080
Hidden dimensions	128
Output dimensions	2
Number of layers	4
Number of heads	4

6.4 Results and Analysis

This section presents and interprets the outcomes of applying the described machine learning techniques to the batch detection task using the REIMS dataset.

6.4.1 Overall Results

Table 6.2: Binary Classification Traditional ML Methods (%)

Model	Train Acc.	Test Acc.
OPLS-DA	57.08 ± 1.46	53.19 ± 2.22
KNN	100.00 ± 0.00	55.69 ± 2.74
DT	100.00 ± 0.00	51.77 ± 1.43
LDA	89.13 ± 1.15	**56.35 ± 2.70**
NB	66.95 ± 2.54	55.01 ± 3.26
RF	99.87 ± 0.33	50.02 ± 0.14
SVM	92.44 ± 1.05	54.19 ± 2.87
LR	91.41 ± 1.19	53.90 ± 3.12
Ensemble	95.05 ± 0.99	54.14 ± 2.81

Table 6.3: Binary Classification DL Encoders (%)

Model	Train Acc.	Test Acc.
TRANSFORMER	97.6 ± 7.4	61.5 ± 7.1
MOE	99.9 ± 0.5	**62.5 ± 7.4**
ENSEMBLE TRANSFORMER	99.7 ± 1.1	60.8 ± 7.4
CNN	50.0 ± 1.4	52.3 ± 5.7
ResNet	50.2 ± 1.3	51.2 ± 2.9
KAN	49.8 ± 1.9	50.4 ± 2.5
VAE	50.0 ± 0.0	50.0 ± 0.0
MAMBA	50.0 ± 1.2	51.3 ± 7.7

Table 6.4: Contrastive Learning DL Encoders (%)

Model	Train Acc.	Test Acc.
TRANSFORMER	77.1 ± 4.9	**70.8 ± 9.6**
MOE	77.6 ± 5.6	70.4 ± 10.4
ENSEMBLE TRANSFORMER	77.4 ± 7	67.9 ± 12.8
CNN	56.6 ± 3.1	50.6 ± 12.4
ResNet	55.9 ± 3.3	50.0 ± 8.5
KAN	55.7 ± 2.4	50.0 ± 0.0
VAE	51.1 ± 2.1	50.0 ± 0.0
MAMBA	55.4 ± 2.8	50.6 ± 4.7

The initial analysis focused on establishing a performance baseline using traditional machine learning methods with a difference vector approach, as shown in Table 6.2. These models universally overfit the training data and failed to generalize, achieving a balanced accuracy only slightly better than random chance ($\approx 50\%$). While Linear Discriminant Analysis (LDA) emerged as the top performer in this category and marginally outperformed the OPLS-DA benchmark, its overall performance was still inadequate.

When deep learning encoders were applied directly to the same binary classification task (Table 6.3), the results were mixed. Most architectures—including CNN, KAN, VAE, Mamba, and the default ResNet encoder—performed on par with the traditional methods, demonstrating poor generalization. However, a notable exception was observed with models designed to interpret sequential data. The Transformer, Mixture of Experts (MoE), and Ensemble models all surpassed 60% balanced accuracy, suggesting that their ability to capture the sequential nature of REIMS data offers a distinct advantage over other approaches. Despite this improvement, their performance still indicated significant room for growth.

A substantial performance improvement was achieved by integrating the deep learning encoders within the SimCLR contrastive learning framework (Table 6.4). This shift in learning strategy yielded the best results of the study, with the Transformer (70.8%) encoders demonstrating a dramatic leap in test accuracy. Their success highlights the power of contrastive learning to develop meaningful data representations for batch detection, a task where these same models previously failed as simple classifiers.

The Ensemble Transformer also performed well (67.9%), though its results were slightly behind the simple Transformer architecture. This may suggest that for learning general-purpose embeddings in this context, the architectural complexity of the ensemble model provides diminishing returns compared to more streamlined models.

The success of the contrastive approach stems from its fundamentally different learning objective. Instead of training on a simple difference vector, the SimCLR framework uses the NT-Xent loss function to structure a meaningful embedding space. This process

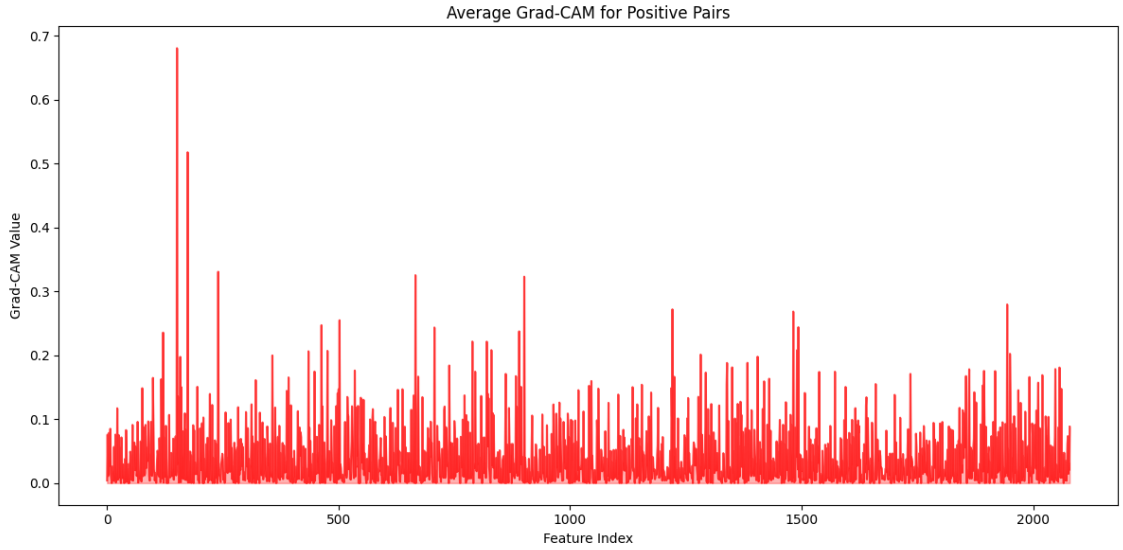

Figure 6.6: Average Grad-CAM for our Transformer-based SpectroSim model. The visualization highlights the most salient m/z features used by the model to compare mass spectra after being trained with SimCLR.

compels the model to learn the subtle, underlying chemical signatures that define a batch by pulling representations of similar samples (positive pairs) closer together while pushing dissimilar ones (negative pairs) apart.

The result is a robust representation that is far less prone to overfitting and generalizes effectively to unseen data. The results confirm that combining a powerful learning paradigm like contrastive learning with architectures well-suited for sequential data, such as Transformers, is a highly effective strategy. Nevertheless, the framework is not a panacea; the failure of models like KAN and VAE even within this structure underscores that both a suitable model architecture and an appropriate learning strategy are essential for success on this challenging task.

6.4.2 Visualization

While the performance of SpectroSim is high, understanding its decision-making process is crucial for trust and verification by domain experts. To open this black box, we can employ post-hoc explanation techniques like Gradient-weighted Class Activation Mapping (Grad-CAM). These methods can highlight the specific mass-to-charge (m/z) ratios that the model finds most important when determining if two samples belong to the same batch. Similar to Chapter 5, we give the average correct Grad-CAM, this time for positive pairs predicted by SpectroSim in batch detection. Figure 6.6 provides a visualization of this Grad-CAM.

The SpectroSim model analyzes a sample's chemical fingerprint, known as a mass

Table 6.5: Computational Costs for Batch and Instance Detection Models

Model	Training Time (s)	Inference Time (s)	Model Size (MB)	Parameters
Transformer	0.203767	0.181942	421.97	105 491 872
MoE	0.311649	0.092523	520.36	130 089 624
Ensemble	0.380419	0.204004	988.88	247 219 936
CNN	0.014521	0.087839	137.10	34 275 328
ResNet	0.274134	0.054243	137.10	34 275 328
KAN	0.343616	0.044145	283.67	70 916 316
VAE	0.225935	0.001709	10.50	2 626 144
Mamba	1.964942	0.430526	19.55	4 887 808

spectrum, by focusing on three key areas. First, it looks at small, common molecular fragments. It then identifies regularly spaced peaks, a pattern created by a family of related molecules (like lipids or polymers) that are built from the same repeating chemical block. Finally, it uses specific, heavy molecular markers to distinguish between very similar samples. This ability to see both broad patterns and unique details shows the model has a sophisticated understanding of the sample's chemistry.

6.4.3 Computational Cost

To evaluate the computational efficiency of various encoder architectures for contrastive learning, we established a standardized benchmark. Each model was tested on a consistent hardware platform using PyTorch to measure key performance indicators: training time, inference time, model size (MB), and the total number of parameters. The evaluation used a single batch from an instance recognition dataset derived from REIMS data, ensuring a direct and fair comparison across architectures.

The results in Table Table 6.5 detail the computational costs for each encoder within the SimCLR framework. When considered with the classification accuracies from the previous section, these results highlight the trade-off between performance and computational expense. The more complex MoE and Ensemble models are, as expected, more computationally demanding than the standard Transformer. Although the VAE is the most lightweight model with the fastest inference speed, its low classification accuracy suggests the gain in efficiency does not justify the significant loss in performance.

6.4.4 Other Contrastive Learning Methods

To ensure that our SpectroSim method, a self-supervised SimCLR with a Transformer backbone, was up to the task, we compared it against other state-of-the-art contrastive learning methods. We use Stratified k-fold cross-validation to evaluate the accuracy. For this analysis, the contrastive learning methods we compared were:

Table 6.6: Performance of Contrastive Learning Methods (%)

Model	Train Acc.	Test Acc.
BARLOW TWINS	54.0 ± 2.8	50.0 ± 0.0
BYOL	56.3 ± 3.5	51.3 ± 6.3
MOCO	54.7 ± 3.3	50.4 ± 7.7
SIMCLR	77.1 ± 4.9	**70.8 ± 9.6**
SIMSIAM	54.1 ± 3.8	48.3 ± 5.1

- BYOL (Grill et al., 2020) — This method learns representations by having an online network predict the output of a target network for a different augmented view of the same image. It prevents model collapse by using a momentum-updated target network, which provides stable regression targets without needing explicit negative samples.
- Barlow Twins (Zbontar et al., 2021) — This approach learns by making the cross-correlation matrix between the embeddings of two augmented views of an image as close to the identity matrix as possible. This objective encourages the features within the embeddings to be decorrelated, thereby reducing redundancy in the learned representation.
- SimSiam (Chen & He, 2021) — A simple Siamese network that maximizes the similarity between two augmentations of an image, using an encoder on one branch and an encoder-predictor pair on the other. It surprisingly avoids collapse without using negative pairs or a momentum encoder by employing a stop-gradient operation, which prevents the outputs from becoming identical.
- MoCo (He et al., 2020) — MoCo frames contrastive learning as a dictionary look-up task, where an encoded "query" from one image view must match its positive "key" from another view and be distinct from all other negative keys in a dictionary. This dictionary is implemented as a large, dynamic queue, and the key encoder is a momentum-based moving average of the query encoder, which maintains a consistent set of negative samples without needing massive batch sizes.

SimCLR is the standout winner by a significant margin, achieving a Mean Test Accuracy of 70.8%. Its success lies in its direct and aggressive use of negative examples. For each sample, SimCLR treats all other samples in the batch as negatives, creating a strong repulsive force in the embedding space. This process compels the model to learn highly discriminative features that sharply distinguish between different source images, a strategy that proved most effective for this particular task.

The middle-tier methods, BYOL (51.25%) and MoCo (50.42%), underperformed because they use strategies to avoid large batches of negative examples. MoCo uses a mo-

mentum queue to supply negatives, which may have lacked the diversity of SimCLR's in-batch approach. BYOL avoids negatives entirely, instead training a network to predict its own momentum-updated representations. While these methods are often effective, their performance here suggests that for this dataset, the explicit repulsive force provided by a large set of negatives was crucial for learning separable features.

The worst-performing methods, SimSiam (48.33%) and Barlow Twins (50.00%), effectively failed to learn anything meaningful. SimSiam, which uses a simple Siamese network with a stop-gradient, is highly susceptible to "representational collapse," where it learns a trivial solution by outputting the same features for all inputs. Barlow Twins employs a fundamentally different objective of decorrelating feature dimensions. Its complete failure suggests this complex objective was ill-suited or improperly tuned for the dataset, preventing the model from converging on a useful representation.

6.5 Chapter Summary

6.5.1 Conclusions

This chapter concludes that a self-supervised contrastive learning framework is highly effective for batch detection, achieving promising accuracy and far surpassing traditional binary classification methods. The proposed Transformer-based method, "SpectroSim," achieved the highest accuracy (70.8%) when used as an encoder within the SimCLR-based contrastive learning framework. This success highlights the superiority of using a contrastive learning paradigm over simple binary classification for this task. The findings consistently demonstrate that for every model architecture tested, contrastive learning is more effective than binary classification, and deep learning generally outperforms traditional machine learning. Furthermore, the SimCLR-based approach is shown to outperform other state-of-the-art contrastive learning techniques on this batch detection task. This work establishes the value of deep, sequence-aware models for analyzing high-dimensional REIMS data, as they can leverage natural dataset imbalances for robust representation learning.

6.5.2 Future Work

Future work could explore:

1. **Advanced Model Architectures:** Graph Neural Networks (GNNs) (Wu et al., 2020) should be applied to this problem as they can explicitly model the chemical relationships between different mass-to-charge (m/z) peaks, such as isotopic patterns or neutral losses. This approach could capture complex inter-feature dependencies that sequential models only learn implicitly, potentially leading to more robust and chemically-informed classifications.

2. **Hybrid Models:** This should be explored because different architectures capture distinct patterns; CNNs excel at identifying local features among adjacent peaks, while Transformers model global, long-range dependencies across the entire spectrum. A hybrid model could synergistically combine these strengths to create a more comprehensive feature representation for improved classification accuracy.

3. **Domain-specific Augmentations:** The integration of domain-specific augmentations, such as spectral peak shifting or noise injection tailored to REIMS chemical profiles, could help create more robust models by artificially expanding the limited training datasets. This would improve generalization by exposing the models to a wider range of plausible spectral variations they might encounter in a real-world setting.

7

Conclusions

Finally, Chapter 7 summarizes the entire body of work presented in the thesis. This chapter will provide a summary of the key contributions and findings from each research thrust, reiterating the advancements made in applying deep learning to REIMS data for seafood authentication. It will discuss the overall significance of the work in the context of food science, analytical chemistry, and industrial applications. The limitations of the current research will be acknowledged, followed by a comprehensive outline of promising future work directions, suggesting avenues for further dataset expansion, model refinement, real-world deployment, and the exploration of new analytical capabilities. The chapter will close with concluding remarks on the transformative potential of these intelligent analytical systems for the future of the seafood industry.

7.1 Achievement of Research Objectives

The overarching goal of this thesis was to develop and validate a suite of machine learning methods to enhance the analytical capabilities of Rapid Evaporative Ionization Mass Spectrometry (REIMS) for marine biomass analysis, targeting rapid, automated, and *in situ* application in fish processing plants. This objective has been comprehensively achieved across three distinct research thrusts, successfully moving the field beyond traditional, limiting statistical methods (like OPLS-DA) toward a robust deep learning paradigm—a transition that is only just beginning in the wider food science literature (Cardoso et al., 2025).

1. Foundational Capabilities: Species and Body Part Identification (Chapter 4)

The thesis successfully established a highly accurate, automated foundation for quality control. Novel Transformer and Mixture of Experts (MoE) Transformer architectures were developed and shown to consistently outperform OPLS-DA and all traditional machine learning methods. For the critical task of fish species classification, the MoE Transformer achieved a perfect 100% accuracy, demonstrating that complex, sequence-aware deep learning models can fully exploit the chemical richness of REIMS data. Furthermore, by integrating Explainable AI (LIME and Grad-CAM), the research ensures that these high-performance models are not "black boxes," providing chemically verifiable insights that are essential for industrial adoption.

2. Advanced Safety and Fraud Detection (Chapter 5)

This research successfully formalized and solved novel contamination problems crucial for food safety. The thesis demonstrated that deep learning models, particularly those leveraging unsupervised pre-training (Masked Spectra Modeling) and Transfer Learning, are essential for handling subtle, difficult-to-detect signals:

- The models achieved competitive performance in the challenging ordinal multi-class oil contamination detection task (up to 49.10% test accuracy), a significant improvement over OPLS-DA (26.43%).
- Crucially, this section established that the more challenging contamination tasks consistently benefited from positive knowledge transfer, providing practical guidance for accelerating model development in industrial settings.
- By comparing ordinal classification techniques, the thesis validated that models which respect the inherent order of contamination levels (e.g., CORAL/CLM) yield more meaningful predictions by minimizing error distance.

3. Novel, Label-Free Batch Traceability (Chapter 6)

The thesis achieved the critical objective of developing a practical, chemical-based solution for batch traceability, overcoming the limitations of costly physical tagging (Mai et al., 2010) and complementing the field's current focus on vulnerable, data-entry-dependent digital traceability systems (Turkson et al., 2025; Jiang et al., 2025). The novel SpectroSim framework, which utilizes a self-supervised contrastive learning approach with a Transformer encoder, achieved a remarkable 70.8% balanced accuracy in determining if two samples shared a batch origin. This result:

- Vastly outperformed traditional binary classification methods (which struggled to surpass 60% accuracy).

Table 7.1: Consolidated Summary of Best Model Performance Across All Tasks

Analytical Task	OPLS-DA	Proposed Best Model	Best Model	Gain
Fish Species	96.39%	MoE Transformer	100.00%	+3.61%
Fish Body Part	51.17%	Ensemble Transformer	74.13%	+22.96%
Oil Contamination	26.43%	TL MoE Transformer	49.10%	+22.67%
Cross-species Adulteration	79.96%	Pre-trained Transformer	91.97%	+12.01%
Batch Detection	53.19%	SpectroSim (Transformer)	70.80%	+17.61%

All accuracy values are balanced test accuracy reported in their respective chapters.

- Eliminated the need for explicit class labels during training, making it a practical, plug-and-play solution for rapid, cost-effective deployment in processing environments, thereby addressing the high-cost and infrastructure barriers that hinder the adoption of digital-only systems (Untal et al., 2025).

In summary, by systematically advancing the analytical capabilities of REIMS data, from fundamental species identification to complex trace contamination detection and label-free traceability, this thesis provides a validated suite of highly accurate, efficient, and potentially interpretable machine learning methods. The foundational goal of enabling rapid, automated, and *in situ* quality assurance in the fish processing industry has therefore been successfully realized.

7.2 Summary of Key Findings

In this thesis, we have demonstrated that deep learning can be applied to REIMS marine biomass data to enhance the performance of ML analysis. The results have shown that deep learning improved classification and pair-wise comparison performance when compared to traditional ML methods. This section discusses the key conclusions of each research objective covered in chapters (Chapters 4 to 6)

The core quantitative achievements of this thesis are distributed across Chapters 4, 5, and 6. To provide a consolidated overview, Table 7.1 synthesizes the main experimental outcomes. This table presents the best-performing model developed for each of the five analytical tasks, directly comparing its balanced test accuracy against the traditional OPLS-DA benchmark to highlight the significant performance gains achieved by the proposed deep learning frameworks.

7.2.1 On Foundational Classification (Chapter 4)

In Chapter 4, we proposed Transformers, both with and without unsupervised pretraining, for fish species and fish body part identification. Here, we draw the attention of the reader

to two key findings of this chapter: (1) deep learning supremacy over traditional ML models, (2) the benefits of explainable AI techniques over black-box deep learning models, and (3) balanced utilization for the MoE Transformer.

Deep Learning Supremacy

This thesis finds that across both tasks, fish species and fish body parts identification, the deep learning methods outperform traditional ML techniques and the incumbent OPLS-DA. This finding is not surprising; traditional models are simpler and often impose stronger inductive bias. A stronger inductive bias means the model makes more assumptions about the nature of the data. Simpler models (e.g., OPLS-DA, linear regression, and decision trees) have less capacity to learn complex patterns, even though OPLS-DA remains a robust benchmark for binary fish speciation (Black et al., 2017; Shen et al., 2020, 2022). Whereas more complex models, like those found in deep learning, have weaker inductive biases, they make fewer assumptions about the underlying nature of the data, enabling the capability to learn much more complex patterns. The marine biomass REIMS data is inherently structured, or sequential data, with long-range dependencies and complex feature interactions that simpler ML models are unable to grasp, an issue that has led to mixed results in the literature where conventional ML models like SVM or RF have sometimes failed to outperform traditional chemometrics (De Graeve et al., 2023). It takes those more complex models, capable of capturing complex patterns in structured data, to effectively harness the full potential of REIMS data.

Explainable AI

However, these more complex models with weaker inductive biases come at a cost. They are often referred to as black-box, as gradient-descent methods of models with millions of parameters of floating-point matrices, with no clear explanation or method to the madness, for why they make decisions. Deep neural networks (e.g., our proposed Transformer that consists of four encoder layers) capture complex feature interactions between intermediate layers that are safe to consider beyond human comprehension. To offer some respite, post-hoc (e.g., applied after the black-box model is trained) interpretability methods such as LIME and Grad-CAM impose strong inductive biases on the black-box method and make a series of simplifying assumptions to generate human-understandable explanations. Grad-CAM assumes that for the gradient information flowing into the final layer, features with larger gradients contributing to a particular class are more important to a decision. LIME approximates a black-box model's behaviour with a simple interpretable model (e.g., sparse linear model) within the local vicinity of the instance being explained. Some level of fidelity, or complexity, is lost in generating human-understandable explanations for black-box models, but these explanations help to bridge the gap between deep learning and human knowledge, addressing a key challenge identified in recent literature on applying

ANNs to REIMS data (Cardoso et al., 2025).

Balanced Expert Utilization

Another key finding of this thesis was that the experts had balanced utilization across all four classification tasks. This indicates the effective use of model capacity without expert collapse. Expert collapse happens when only a small subset of available "expert" networks are consistently used or trained, while others are mostly ignored or redundant. Essentially, instead of all experts learning distinct aspects of the data, a few experts end up doing most or all of the work. Our results show the experts don't suffer from expert collapse, which is a sign the MoE is functioning as intended. This is how an MoE is supposed to work. Multiple expert networks and a gating network constitute an MoE model. The role of the gating network is to route different types of data to the most suitable expert(s). The idea is that each expert specializes in a particular subdomain of the input space, leading to a more efficient and powerful model. Our results demonstrate that the experts avoided collapse, a critical factor for an MoE functioning as intended. Crucially, this balanced utilization directly contributed to the high levels of accuracy and robustness observed across all four classification tasks. This balanced utilization was pivotal because it allowed diverse experts to specialize in distinct aspects of the REIMS spectral signatures, directly enhancing the model's ability to differentiate accurately across all four classification tasks. Such effective use of the model's capacity, preventing any single expert from being overwhelmed or others from becoming redundant, underpins the strong predictive performance and reliability demonstrated in this work.

7.2.2 On Contamination and Adulteration (Chapter 5)

In Chapter 5, we extended the previous contribution to two new tasks of Oil Contamination and Cross-species Adulteration detection, finding that the proposed Transformer model with unsupervised pertaining significantly outperformed the traditional ML methods and other deep learning methods. In the following section, we highlight two key findings from this analysis: (1) pretraining strategies work for mass spectra, (2) Oil Contamination Detection is hard, and (3) Oil contamination consistently benefits from transfer learning.

Pretraining Strategies

This thesis finds that unsupervised pretraining strategies, such as the proposed Masked Spectra Modeling (MSM), offer consistent performance gains for oil contamination and cross-species adulteration detection. This suggests that there is inherent value in representation learning from mass spectra data in the task of BERT-styled left-to-right progressive masking. The task may seem simple, to predict which spectra come next, but since the data is structured and not tabular, and there is an inherent sequence to the features on

the x-axis, in order of ascending mass-to-charge ratios, a sequence that can be learned. An intentional bonus side effect of this task is that the model implicitly learns to make peak predictions. In its journey from left to right, it will encounter several peaks for which it has to predict accurately. This implicitly learned task of peak prediction mirrors the manual analytical chemistry techniques that chemists practice in a lab when analyzing mass spectrometry samples by hand. This is analogous to the traditional analytical methods and is enhanced by its unsupervised nature; we no longer require a human-in-the-loop to manually annotate peaks. This saves a significant amount of time and domain expertise and enables the automation of REIMS marine biomass analysis.

Oil Contamination Detection is Hard

Another key finding of this thesis was that oil contamination detection is hard. Not only is it a multi-class classification task with seven classes. It can technically be framed as an ordinal classification task, where overlap exists between the various concentrations of oil contamination (e.g., 50%, 25%, 10%, 5%, 1%, 0.1%, 0%), making the distinctions between individual classes more opaque. The inherent difficulty arises from the need to preserve and utilize this inherent order, which standard classification methods might ignore. Adjacent categories in an ordinal scale can often have ambiguous boundaries. Not only is this an ordinal classification problem, but it is also a multi-class classification task, with an increased number of decision boundaries that each model must formulate. With more classes, there is a higher chance of misclassification. The aforementioned ordinal scale of the oil concentrations adds inter-class similarity, making it more difficult for the model to learn fine-grained distinctions. The classification performance across all methods of oil contamination detection demonstrated the difficulty of this task.

Oil Contamination Consistently Benefits

This thesis finds that oil contamination detection consistently benefits from Transfer Learning. Our results show that regardless of the source domain, the target domain of the oil contamination dataset has positive transfer effects. Oil contamination gives a set of complex chemical signals that have underlying characteristics (e.g., they are primarily composed of hydrocarbons) that are fundamental enough to be recognized by models trained on diverse REIMS datasets of marine biomass. The knowledge gained from analyzing the general chemical landscape of marine biomass in a given source domain, e.g., fish species, fish body parts, or cross-species adulteration, provides a strong starting point for isolating the exogenous (i.e., foreign) signals of oil. The finding that positive transfer is observed regardless of the source domain suggests the learned features are capable of capturing fundamental aspects of mass spectrometry data from marine biomass, after which the specific oil signatures can easily be identified. Practically, this means that developing effective oil contamination detection models can be accelerated and improved by utilizing

pre-existing models trained on other marine biomass datasets, potentially reducing the need for extensive de novo data collection and annotation for this specific, yet critical, application. Beyond oil contamination, these findings may encourage the exploration of transfer learning for other trace contaminant detection or anomaly identification tasks within REIMS analysis of complex biological matrices.

7.2.3 On Batch Traceability (Chapter 6)

In Chapter 6, we demonstrated that Contrastive Learning, with substantial modifications, could be effectively applied to batch detection, exceeding the performance of traditional binary classification. Below, we highlight two key findings from this work: (1) deep learning supremacy, and (2) contrastive learning outperforms binary classification.

Deep Learning Supremacy

A key finding of this thesis was that deep learning outperformed traditional machine learning methods. Again, this can likely be put down to the structured sequential nature of the data, and this complexity not being captured by simpler models that impose a high inductive bias. Deep learning methods are simply better at handling complex sequential data. But let's explore further, deep learning methods have some key advantages revolving around their ability to automatically learn complex representations and handle intricate relationships in the data. Traditional machine learning methods rely heavily on manual feature engineering. If the chosen features (i.e., difference vector between pairs) are not discriminative enough, the model's performance will be limited. However, deep learning models, especially Siamese networks, can automatically learn hierarchical feature representations directly from the raw data. The model learns to identify the most relevant features for comparing items without explicit instructions (i.e., representation learning). Many traditional machine learning models (e.g., logistic regression, SVM) are inherently linear or assume specific data distributions. While some models (e.g., kernelized SVMS or random forests) can capture non-linearities, they are not as flexible as deep neural networks. Neural networks are universal approximations (Hornik et al., 1989), with multiple layers and non-linear activation functions. This means they can capture highly complex, non-linear relationships between paired items, a limitation recently demonstrated in food science where ANNs outperformed traditional PLS-DA by better modeling complex, non-linear relationships (Cardoso et al., 2025). This matters when the similarity is not additive or a linear function of the features, but involves complex feature interactions.

Contrastive Learning Outperforms Binary Classification

This thesis demonstrates that contrastive learning is superior to binary classification for batch detection of REIMS marine biomass. The difference vector approach, where we cal-

culate the similarity of the pair as the difference vector between their features and train a standard binary classifier on these difference vectors to predict if the pair is similar or dissimilar. The difference vector is a crude distance measure that doesn't capture the full inherent semantic features of similarity more broadly. Potentially, the daw difference is not the most informative way to capture complex relationships in the high-dimensional REIMS data. In contrast, contrastive learning aims to generate an embedding space where similar items are closer together, and dissimilar items are further apart. The encoder network learns to produce these embeddings, or learned representations, of dissimilar and similar pairs. Two samples that belong to the same batch are a positive pair, and samples from different batches form a negative pair. The NT-XENT loss function then works by pulling positive pairs close together and pushing negative pairs apart. By contrasting many positive and negative examples, the model learns which features are invariant to positive pairs and which features distinguish between negative pairs. The resulting embedding space facilitates nuanced similarity assessment, using the measure of cosine similarity, which effectively captures semantic relationships by considering the angle between feature vectors, irrespective of their magnitude. This provides a robust, intrinsic verification method, which is a critical component missing from modern, logistics-focused digital traceability systems (Turkson et al., 2025; Dahariya et al., 2025).

7.3 Limitations of the Research

Despite encouraging results, certain limitations of this research should be acknowledged:

- **Dataset Scope:** While the datasets from AgResearch provided a valuable foundation, they are specific to certain fish species (primarily Hoki and Mackerel) and processing conditions within New Zealand. The generalizability of the findings to a broader range of species, geographical origins, and processing environments requires further investigation. For example, geographical origin is a complex variable in its own right, with recent REIMS and ML studies focusing specifically on authenticating the provenance of a single species across different regions (Lu et al., 2024).

- **Sample Size for Specific Tasks:** Some datasets, particularly for fish body part classification, had a limited number of samples, which can affect model robustness and the statistical significance of some findings, although techniques were employed to mitigate this.

- **Dynamic Environmental Factors:** The current research primarily utilized static datasets. Real-world processing environments can have dynamic factors (e.g., seasonal variations (OpenSeas, 2021) in fish biochemistry, changes in processing equipment) that might affect REIMS spectra and model performance over time.

- **Computational Resources:** While efforts were made to develop efficient models, some advanced architectures like large Transformers or extensive MoE configurations can still be computationally intensive, potentially posing a barrier for facilities with limited resources.

7.4 Future Work

The engagement with REIMS marine biomass analysis detailed in this thesis has illuminated several key research directions. This section aims to present these topics, identified during our work, which we assert are promising avenues for future investigation likely to yield significant impact.

7.4.1 Ordinal Classification for Oil Contamination

The oil contamination task, defined by seven ordered concentration levels, presents a classic ordinal classification challenge where minimizing prediction error distance, measured by the Mean Absolute Error (MAE), is critical. While initial experiments confirmed that ordinal-aware methods like CORAL and Cumulative Link Models (CLM) significantly reduced the MAE compared to nominal classification, substantial performance gains are likely achievable by exploring other advanced ordinal-aware models referenced in the literature. Future research should systematically evaluate models from the three main families of ordinal classification. This includes Threshold-Based Models, specifically Support Vector Ordinal Regression (SVOR) (Herbrich et al., 1999; Chu & Keerthi, 2007), which has demonstrated top-tier performance on complex tabular datasets by using large-margin parallel hyperplanes. The investigation should also incorporate advanced Deep Learning Ordinal Losses, such as the Earth Mover's Distance (EMD) (De Matos et al., 2019) and Conditional Ordinal Regression for Neural Networks (CORN) (Shi et al., 2021). These loss functions directly penalize misclassifications based on their rank distance, offering a principled way to optimize the network's learning objective for the specialized ordinal metrics of MAE and Quadratic Weighted Kappa (QWK).

Furthermore, a deeper analysis of Binary Decomposition Approaches should be conducted. Specifically, implementing techniques that leverage the cumulative link approach, as seen in Ordinal CNN architectures (Niu et al., 2016), could offer a robust alternative. This involves replacing the network's final layer with $k-1$ sigmoid neurons to predict cumulative probabilities, thereby structurally enforcing the ordinal constraint. Given that the sequential nature of REIMS data aligns well with Transformer-based models, these ordinal losses and architectures should be integrated into the existing Transformer framework. This dedicated focus on advanced ordinal methodologies is expected to yield substantial improvements in the classification accuracy of the oil contamination dataset, ensuring that the model's predictions are highly precise and chemically meaningful by minimizing large

rank errors.

7.4.2 Real-World Validation and Deployment

The developed models should be deployed in their real-world application, a fish processing factory, where they are tested and validated in real-time fish processing to assess their capabilities. The fish species and body part identification model developed in Chapter 4, the oil contamination and cross-species adulteration detection model developed in Chapter 5, and the batch detection model developed in Chapter 6, can all be deployed in a real-world testbed to measure their performance. Obviously, for real-world adoption by fish processing plants, the model would need to meet or exceed the current benchmarks for quality assurance and food safety. The future CyberMarine flex-factory (Plant and Food Research, 2020) serves as a perfect testbed to deploy and evaluate these models, since it already has the REIMS equipment, and processed the existing samples that formulate the dataset this thesis was based on, so it has access to the fish species, body parts, and contaminants used to replicate the experimental settings of the dataset.

For batch traceability, the current state-of-the-art approach that involves using RFID chips to track individual batches is cumbersome, impractical, and expensive (Mai et al., 2010). After an initial investment in REIMS equipment, a relatively inexpensive, unintrusive, and plug-and-play solution to traceability is REIMS + SpectroSim — the method for batch traceability through contrastive learning developed in Chapter 6. This is a cheap, straightforward, and practical way to address the failings of traditional traceability in fish processing (Helyar et al., 2014). Furthermore, it serves as a critical analytical verification layer for the current state-of-the-art digital traceability systems (Gastaldi García, 2025), which are often based on blockchain (Dahariya et al., 2025) or Digital Product Passports (Jiang et al., 2025). These digital systems are vulnerable to fraudulent data entry (Turkson et al., 2025), a gap that "SpectroSim" directly addresses. By providing a low-cost, label-free chemical check, it helps overcome the practical adoption barriers, such as cost and fisher-side infrastructure, that currently hinder digital-only solutions (Untal et al., 2025). Furthermore, once trained, the model does not require labels during inference to detect if fish belong to the same batch with 70.8% accuracy. So the manual labor and the subsequent slowdown to supply chains, of adding physical labels to each batch, can be automated away entirely.

7.4.3 Online Learning and Model Adaptation

Investigating online learning or adaptive model retraining strategies would be crucial for maintaining model performance in dynamic industrial settings where data distributions might shift over time. One such example of a data distribution shift is the seasonal variation of Hoki. During spawning season, when Hoki are laying their eggs, the mother extracts all their lipids and puts them into their eggs as nutrients for them (OpenSeas, 2021). This

dramatically changes the chemical composition of a mother Hoki, demonstrating a seasonal variation that would result in a data distribution shift for adult females of the Hoki population. We need machine learning models that are robust to such data distribution shifts, such as seasonal variation. Two approaches facilitate the change needed to handle data distribution shifts.

Online learning (Hoi et al., 2021) is a machine learning approach where the model learns sequentially from incoming data instances, updating its parameters incrementally rather than retraining on the entire dataset from scratch. Model adaptation, often a goal or characteristic of online learning, refers to the process by which a pre-trained model adjusts its parameters to new data or a changing environment, allowing it to maintain performance or learn new patterns over time. Online learning is a machine learning approach where the model learns incrementally by processing data sequentially, one instance or a small batch at a time, allowing it to continuously update and evolve. Model adaptation (Chougiouris & Panagidi, 2024; Kabra & Patel, 2024) is the process by which a machine learning model adjusts its parameters or structure in response to new data or changes in the data distribution, often employing online learning techniques to maintain performance in dynamic environments. Both of these approaches could be used to extend REIMS marine biomass analysis to handle data distribution shifts like seasonal variation.

7.4.4 Multi-Modal Data Fusion

Integrating REIMS data with other sensory information (e.g., nuclear magnetic resonance spectroscopy (Bettjeman et al., 2018), DNA barcoding (Khaksar et al., 2015)) could create more comprehensive and resilient authentication systems. In machine learning, this process of combining data from multiple modalities into a single model is referred to as multi-modal data fusion (Lahat et al., 2015). In this case, different types of data sources are called modalities (e.g., REIMS, NMR, DNA barcoding). Multi-modal data fusion is the process of combining information from multiple, different modalities to create a more comprehensive, accurate, and robust understanding of a phenomenon or to improve the performance of a task (Zhao et al., 2024). Kaiser et al. (Kaiser et al., 2017) demonstrate one such example of multi-modal data fusion, with their multi-modal transfer learning, that shares knowledge between the modalities of text, image, and audio. However, this is just one example of multi-modal data fusion; others include feature-level fusion, decision-level fusion, hybrid fusion, attention fusion, and so on.

Future work should consider the application of multi-modal fusion to REIMS marine biomass analysis to enhance performance. Initial investigations could explore how different fusion strategies—such as early concatenation of REIMS spectral features with derived NMR chemical descriptors, or late fusion of probabilistic outputs from separate REIMS and DNA barcoding classifiers—impact specific tasks like fine-grained species differentiation or the detection of subtle adulteration. Furthermore, exploring hybrid approaches or

attention mechanisms could be particularly fruitful for adaptively weighing the information from REIMS against complementary modalities like DNA barcoding, especially when dealing with samples of varying quality or complexity. This aligns with emerging research that combines different analytical targets, such as mercury levels, with machine learning to enhance food safety and traceability monitoring (Piroutková et al., 2025).

7.4.5 Advanced Model Architectures and Hybrid Approaches

Continued exploration of novel deep learning architectures, including graph neural networks for modelling molecular relationships or hybrid models combining the strengths of different approaches (e.g., CNN-Transformer (LeCun et al., 1989c; Vaswani et al., 2017) for spatiotemporal spectral analysis), could yield further improvements. Graph Neural Networks (GNNs) (Wu et al., 2020) are a class of neural networks designed to perform machine learning directly on graph-structured data, capturing relationships and dependencies between entities represented as nodes and edges. They work by iteratively aggregating information from a node's neighbors, allowing the network to learn representations (embeddings) of nodes, edges, or entire graphs for tasks like node classification, link prediction, or graph classification.

Future REIMS marine biomass analysis could greatly benefit from Graph Neural Networks (GNNs). These can model spectral features (m/z values) as graphs—with edges indicating chemical relationships (e.g., neutral losses, isotopic patterns) or learned correlations—to exploit inter-feature dependencies for more robust sample classification or biomarker discovery. Alternatively, GNNs can represent individual samples as nodes in a graph, with edges denoting similarities or spatio-temporal links, to learn holistic patterns across entire datasets. Exploring optimal graph construction methodologies and suitable GNN architectures (like Graph Convolutional or Attention Networks) for such analytical tasks is a key research direction. Furthermore, hybrid models promise enhanced capabilities by synergistically integrating architectures; for instance, combining CNNs (for extracting potent features from raw REIMS spectra) with GNNs (to subsequently model inter-sample relationships using these features or incorporate chemical knowledge graphs), or pairing Recurrent Neural Networks (RNNs/LSTMs) for analyzing time-series spectral data with GNNs to track evolving compound relationships and derive deeper insights.

7.5 Final Remarks

This thesis has demonstrated the profound impact that advanced machine learning, when thoughtfully applied to high-resolution REIMS data, can have on addressing critical challenges in the seafood industry. By systematically developing and evaluating a range of techniques from sophisticated transformer architectures to strategic transfer learning and innovative self-supervised methods, this research has not only pushed the boundaries of

analytical capability but also provided practical pathways towards more secure, transparent, and efficient seafood supply chains. The journey from identifying a fish out of water to identifying a fish out of data has revealed that the fusion of cutting-edge spectrometry and intelligent algorithms is key to safeguarding one of our most valuable global food resources. The insights gained and the methodologies developed herein lay a strong foundation for future innovations in food science and analytical chemistry (Xue et al., 2025), promising a future where food fraud is more readily detected and food safety is more robustly assured.

Bibliography

AgResearch. Leading agri-based science innovation. https://www.agresearch.co.nz/, 2025.

Ahles, S., DeWitt, C. A. M., and Hellberg, R. S. A meta-analysis of seafood species mislabeling in the United States. *Food Control*, 171:111110, 2025.

Amelin, V., Emel'yanov, O., Khrushchev, A. Y., and Tret'yakov, A. Species Identification and Revelation of Facts of Fish Caviar Adulteration by Vibrational Spectroscopy and Digital Coloriometry. *Journal of Analytical Chemistry*, 80(7):1182–1195, 2025.

Ansar, W., Goswami, S., and Chakrabarti, A. A Survey on Transformers in NLP with Focus on Efficiency. *arXiv preprint arXiv:2406.16893*, 2024.

Australia, H. P. F. D. M. Melbourne restaurant hunky dory accused of serving catfish to customers instead of dory, May 2016. URL https://www.dailymail.co.uk/news/article-3611999/Melbourne-restaurant-Hunky-Dory-accused-serving-catfish-customers-instead-dory.html.

Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Balakrishnama, S. and Ganapathiraju, A. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18(1998):1–8, 1998.

Balog, J., Szaniszlo, T., Schaefer, K.-C., Denes, J., Lopata, A., Godorhazy, L., Szalay, D., Balogh, L., Sasi-Szabo, L., Toth, M., et al. Identification of biological tissues by rapid evaporative ionization mass spectrometry. *Analytical chemistry*, 82(17):7343–7350, 2010.

Balog, J., Sasi-Szabó, L., Kinross, J., Lewis, M. R., Muirhead, L. J., Veselkov, K., Mirnezami, R., Dezső, B., Damjanovich, L., Darzi, A., Nicholson, J. K., and Takáts,

Z. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Science Translational Medicine*, pp. 194ra93–194ra93, 2013. doi: 10.1126/scitranslmed.3005623.

Balog, J., Perenyi, D. R., Guallar-Hoyas, C., Egri, A., Pringle, S. D., Stead, S., Chevallier, O. P., Elliott, C. T., and Takáts, Z. Identification of the Species of Origin for Meat Products by Rapid Evaporative Ionization Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, pp. 4793–4800, 2016. doi: 10.1021/acs.jafc.6b01041.

Bettjeman, B. I., Hofman, K. A., Burgess, E. J., Perry, N. B., and Killeen, D. P. Seafood phospholipids: extraction efficiency and phosphorous nuclear magnetic resonance spectroscopy (31P NMR) profiles. *Journal of the American Oil Chemists' Society*, 95(7): 779–786, 2018.

Black, C., Chevallier, O. P., Haughey, S. A., Balog, J., Stead, S., Pringle, S. D., Riina, M. V., Martucci, F., Acutis, P. L., Morris, M., et al. A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry. *Metabolomics*, 13:1–13, 2017.

Black, C., Chevallier, O. P., Cooper, K. M., Haughey, S. A., Balog, J., Takats, Z., Elliott, C. T., and Cavin, C. Rapid detection and specific identification of offals within minced beef samples utilising ambient mass spectrometry. *Scientific Reports*, 9(1):6295, 2019.

Boccard, J. and Rutledge, D. N. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica chimica acta*, 769:30–39, 2013.

Bozinovski, S. and Fulgosi, A. The influence of pattern similarity and transfer learning upon training of a base perceptron b2. In *Proceedings of symposium informatica*, volume 3, pp. 121–126, 1976.

Breiman, L. *Classification and regression trees*. Routledge, 2017.

Bremner, H. A. *Safety and quality issues in fish processing*. Elsevier, 2002.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. Signature verification using a "siamese" time delay neural network. *Advances in neural information processing systems*, 6, 1993. doi: 10.1142/S0218001493000339.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., and Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 20(8-10): 341–351, 2006.

Bénard-Capelle, J., Guillonneau, V., Nouvian, C., Fournier, N., Loët, K. L., and Dettaı, A. Fish mislabelling in France: substitution rates and retail types. *PeerJ*, 2015. doi: 10.7717/peerj.714.

Cafarella, C., Mangraviti, D., Rigano, F., Dugo, P., and Mondello, L. Rapid evaporative ionization mass spectrometry: A survey through 15 years of applications. *Journal of Separation Science*, 47(9-10):2400155, 2024.

Cai, W., Gao, Y., Shi, C., An, H., Li, X., Huo, D., Chen, S., He, Y., Huang, K., Zhu, Y., et al. In Situ Analysis of Plant Tissue Using Arc iKnife Ionization Mass Spectrometry. *Analytical Chemistry*, 97(5):2598–2603, 2025.

Cao, W., Mirjalili, V., and Raschka, S. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 140:325–331, 2020.

Cardoso, V. G. K., Balog, J., Zsellér, V., Karancsi, T., Sabin, G. P., and Hantao, L. W. Prediction of coffee traits by artificial neural networks and laser-assisted rapid evaporative ionization mass spectrometry. *Food Research International*, 203:115773, 2025.

Carvalho, D. C., Carvalho, D. C., Palhares, R. M., Palhares, R. M., Drummond, M. G., Drummond, M. G., Frigo, T. B., and Frigo, T. B. DNA Barcoding identification of commercialized seafood in South Brazil: A governmental regulatory forensic program. *Food Control*, 2015. doi: 10.1016/j.foodcont.2014.10.025.

Castelvecchi, D. Can we open the black box of AI? *Nature News*, 538(7623):20, 2016.

Cawthorn, D.-M., Duncan, J., Kastern, C., Francis, J., and Hoffman, L. C. Fish species substitution and misnaming in South Africa: An economic, safety and sustainability conundrum revisited. *Food Chemistry*, 2015. doi: 10.1016/j.foodchem.2015.03.113.

Changizi, R., Changizi, R., Farahmand, H., Farahmand, H., Farahmand, H., Farahmand, H., Soltani, M., Soltani, M., Darvish, F., Darvish, F., Elmdoost, A., and Elmdoost, A. Species Identification of Some Fish Processing Products in Iran by DNA Barcoding. *Journal of Agricultural Science and Technology*, pp. 973–980, 2013.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PmLR, 2020a.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., and Hinton, G. E. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020b.

Chen, X. and He, K. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15750–15758, 2021.

Chin, T. C., Adibah, A., Hariz, Z. D., and Azizah, M. S. Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving transparency in food market. *Food Control*, 64:247–256, 2016.

Choudhury, T. R., Moniruzzaman, M., Anonna, T. A., Al-Asad, H., Samanta, P., and Islam, F. Evaluation of heavy metal contamination in soil, water, and fish in an industrial zone in Bangladesh: Ecological and potential health risk. *Regional Studies in Marine Science*, 86:104162, 2025.

Chougiouris, N. E. and Panagidi, K. Adaptive Neural Network Model Retraining. 2024.

Chu, W. and Keerthi, S. S. Support vector ordinal regression. *Journal of Machine Learning Research*, 8(Mar):597–613, 2007.

Cobbinah, R. T., Boadi, N. O., Saah, S. A., Agorku, E. S., Badu, M., and Kortei, N. K. Cancer risk from heavy metal contamination in fish and implications for public health. *Scientific Reports*, 15(1):24162, 2025.

Cortes, C. and Vapnik, V. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Dahariya, T. K., Sharma, S., and Ray, N. K. Enhancing Livestock Supply Chains with Blockchain Traceability from Source to Market: A Survey. In *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, pp. 123–128. IEEE, 2025.

de Araújo, G. A., Ramos, M. C. S., Carvalho, G. L. D., Camilo-Cotrim, C. F., do Amaral, R. B., Castro, Í. B., Rocha, T. L., and Damacena-Silva, L. Microplastic contamination in wild freshwater fish: global trends, challenges and perspectives. *Environmental Pollution*, pp. 126406, 2025.

De Graeve, M., Birse, N., Hong, Y., Elliott, C. T., Hemeryck, L. Y., and Vanhaecke, L. Multivariate versus machine learning-based classification of rapid evaporative Ionisation mass spectrometry spectra towards industry based large-scale fish speciation. *Food Chemistry*, 404:134632, 2023.

De Matos, P. T., Britto, A. S., Oliveira, L. S., and Koerich, A. L. Ordinal classification with Earth Mover's Distance for content-based image retrieval. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2019.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Do, T. D. and Wong, L. L. The investigation of seafood mislabeling in Asia: a review and coping strategies. *Journal of Agriculture and Food Research*, pp. 101702, 2025.

Eckel, J. C., Seidemann, L., Albadry, M., Schicht, G., Skvoznikova, M., Nickel, S., Hänsel, R., Seehofer, D., Hiller, G. G. R., Tautenhahn, H.-M., et al. Application of rapid evaporative ionization mass spectrometry in preclinical and clinical analyses of steatotic liver tissues and cells. *Scientific Reports*, 15(1):9226, 2025.

Eppley, M. G. and Coote, T. DNA barcoding reveals mislabeling of endangered sharks sold as swordfish in New England fish markets. *Conservation Genetics*, pp. 1–10, 2025.

FAO. *The State of World Fisheries and Aquaculture, 2020*. FAO, 2020. doi: https://doi.org/10.4060/ca9229en.

Fellows, P. J. *Food Processing Technology: Principles and Practice*. Woodhead Publishing, Duxford, UK, 4th edition, 2017.

Fields, J., Chovanec, K., and Madiraju, P. A survey of text classification with transformers: How wide? how large? how long? how accurate? how expensive? how safe? *IEEE Access*, 12:6518–6531, 2024.

Filonzi, L., Filonzi, L., Chiesa, S., Chiesa, S., Vaghi, M., Vaghi, M., Marzano, F., and Marzano, F. N. Molecular barcoding reveals mislabelling of commercial fish products in Italy. *Food Research International*, 2010. doi: 10.1016/j.foodres.2010.04.016.

Fix, E. and Hodges, J. L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 57(3):238–247, 1989.

Food and Agriculture Organization of the United Nations. *The State of World Fisheries and Aquaculture 2024*. FAO, Rome, 2024. doi: https://doi.org/10.4060/cd0683en.

for Economic Co-operation, O. and Development. Fisheries and Aquaculture in Norway. https://www.oecd.org/agriculture/topics/fisheries-and-aquaculture/documents/report_cn_fish_nor.pdf, Jan 2021.

Frank, E. and Hall, M. A simple approach to ordinal classification. In *European conference on machine learning*, pp. 145–156. Springer, 2001.

Galal-Khallaf, A., Ardura, A., Mohammed-Geba, K., Borrell, Y. J., and Garcia-Vazquez, E. DNA barcoding reveals a high level of mislabeling in Egyptian fish fillets. *Food control*, 46:441–445, 2014.

Gao, H., Lv, Y., Zhao, W., Wang, S., Zhao, J., Xu, Y., Shang, Y., Zhang, R., and Ma, Q. Accurate lamb origin identification and molecular differentiation analysis using rapid evaporative ionization mass spectrometry. *Analytical Methods*, 17(7):1643–1647, 2025.

Gastaldi García, F. Technical traceability systems in the food and fisheries sectors: A comparative analysis of Spain and Sweden. Technical Report, 2025.

Ghaly, A., Ramakrishnan, V., Brooks, M., Budge, S., and Dave, D. Fish processing wastes as a potential source of proteins. *Amino acids and oils: A critical review. J. Microb. Biochem. Technol*, 5(4):107–129, 2013.

Gkarane, V., De Graeve, M., Stephens, C., Decloedt, A., Vangeenderhuysen, P., Balog, J., Elliott, C., Stead, S., Birse, N., Hemeryck, L., et al. Towards Real-Time Industry-Proof Pork Breed and Boar Taint Classification Using Rapid Evaporative Ionisation Mass Spectrometry (Reims). *Available at SSRN 5087143*, 2025.

Golf, O., Strittmatter, N., and Takats, Z. Rapid evaporative ionization mass spectrometry: A review of its technology and applications. *Mass Spectrometry Reviews*, 34(5):545–562, 2015. doi: 10.1002/mas.21430.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Gu, A. and Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., and He, Y. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv preprint arXiv:2501.12948*, 2025.

Hand, D. J. and Yu, K. Idiot's Bayes—not so stupid after all? *International statistical review*, 69(3):385–398, 2001.

Hanner, R., Becker, S., Ivanova, N. V., and Steinke, D. FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA*, 2011. doi: 10.3109/19401736.2011.588217.

Hansen, L. K. and Salamon, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10):993–1001, 1990.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Helyar, S. J., ap Dafydd Lloyd, H., de Bruyn, M., Leake, J. R., Bennett, N., and Carvalho, G. R. Fish product mislabelling: failings of traceability in the production chain and implications for illegal, unreported and unregulated (IUU) fishing. *PLOS ONE*, 2014. doi: 10.1371/journal.pone.0098691.

Henderson, A., Heaney, L. M., and Rankin-Turner, S. Advancements in Ambient Ionisation Mass Spectrometry in 2024: An Annual Review. *Analytical Science Advances*, 6(1): e70007, 2025.

Hendriks, T. F., Birmpili, A., de Vleeschouwer, S., Heeren, R. M., and Cuypers, E. Integrating Rapid Evaporative Ionization Mass Spectrometry Classification with Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Imaging and Liquid Chromatography-Tandem Mass Spectrometry to Unveil Glioblastoma Overall Survival Prediction. *ACS Chemical Neuroscience*, 16(6):1021–1033, 2025.

Hendrycks, D. and Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

Herbrich, R., Graepel, T., and Obermayer, K. Support vector learning for ordinal regression. In *Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN 99)*, pp. 97–102. IEE, 1999.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Ho, T. K. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Hoi, S. C., Sahoo, D., Lu, J., and Zhao, P. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, 2021.

Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Jacobs, R. A. et al. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.

Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., and Kaiser, L. OpenAI o1 system card. *arXiv preprint arXiv:2412.16720*, 2024.

Jamal, N. T., Islam, M. R. U., Sultana, S., Banik, P., Nur, A.-A. U., Albeshr, M. F., Arai, T., Yu, J., and Hossain, M. B. Microplastic contamination in some popular seafood fish species from the northern Bay of Bengal and possible consumer risk assessment. *Food Control*, 171:111114, 2025.

Jennings, S., Kaiser, M. J., and Reynolds, J. D. *Marine Fisheries Ecology*. Blackwell Science, Oxford, 2001.

Jha, S. N. *Rapid detection of food adulterants and contaminants: theory and practice*. Academic Press, 2015.

Jiang, S., Natvig, M. K., Ræder, T. B., Bourgos, P., and Tsampanaki, N. Traceability Data in the form of Digital Food Product Passports for Fish Supply Chains. In *2025 6th International Conference in Electronic Engineering & Information Technology (EEITE)*, pp. 1–6. IEEE, 2025.

Jing, L., Zhu, J., and LeCun, Y. Masked siamese convnets. *arXiv preprint arXiv:2206.07700*, 2022.

Kabra, A. and Patel, K. K. The Limitations of Model Retraining in the Face of Performativity. *arXiv preprint arXiv:2408.08499*, 2024.

Kaiser, L. et al. One model to learn them all. *arXiv preprint arXiv:1706.05137*, 2017.

Kannuchamy, N., Nagalakshmi, K., Pavan-Kumar, A., Pavan-Kumar, A., Annam, P.-K., Venkateshwarlu, G., Venkateshwarlu, G., Gireesh-Babu, P., Pathakota, G.-B., Lakra, W. S., and Lakra, W. S. Mislabeling in Indian seafood: An investigation using DNA barcoding. *Food Control*, 2016. doi: 10.1016/j.foodcont.2015.05.018.

Ke, J., Li, Y., Hong, M., Zhang, K., Yang, Y., Wang, T., Li, X., Wang, L., Xie, L., and Xie, D. Qualitative and quantitative analysis of adulterated Antarctic Krill Oil (AKO) based on residual network (ResNet) and three-dimensional fluorescence spectroscopy. *Journal of Food Composition and Analysis*, pp. 108136, 2025.

Keskin, E., Keskin, E., Atar, H. H., and Atar, H. H. Molecular identification of fish species from surimi-based products labeled as Alaskan pollock. *Journal of Applied Ichthyology*, 2012. doi: 10.1111/j.1439-0426.2012.02031.x.

Khaksar, R., Carlson, T., Schaffner, D. W., Ghorashi, M., Best, D., Jandhyala, S., Traverso, J., and Amini, S. Unmasking seafood mislabeling in US markets: DNA barcoding as a unique technology for food authentication and quality control. *Food Control*, 56:71–76, 2015.

Kinakh, V., Taran, O., and Voloshynovskiy, S. Scatsimclr: self-supervised contrastive learning with pretext task regularization for small-scale datasets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1098–1106, 2021.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., and Klein, M. *Logistic regression*. Springer, 2002.

Köppen, M. The curse of dimensionality. In *5th online world conference on soft computing in industrial applications (WSC5)*, volume 1, pp. 4–8, 2000.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

Kumari, P., Parida, V. K., Raj, D., Kumar, P., Narayan, M., and Gupta, U. Assessment of water, sediment, and fish contamination by metals in the lentic ecosystems of a mineral-rich state in India. *Biological Trace Element Research*, pp. 1–16, 2025.

Lahat, D., Adali, T., and Jutten, C. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.

Lai-Dang, Q.-V. A survey of vision transformers in autonomous driving: Current trends and future directions. *arXiv preprint arXiv:2403.07542*, 2024.

Lamendin, R., Lamendin, R., Miller, K. J., Miller, K. J., Ward, R., and Ward, R. D. Labelling accuracy in Tasmanian seafood: An investigation using DNA barcoding. *Food Control*, 2015. doi: 10.1016/j.foodcont.2014.07.039.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2, 1989a.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989b.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

LeCun, Y. et al. Generalization and network design strategies. *Connectionism in perspective*, 19(143-155):18, 1989c.

Lee, S. I., Koo, K., Lee, J. H., Lee, G., Jeong, S., O, S., and Kim, H. Vision transformer models for mobile/edge devices: a survey. *Multimedia Systems*, 30(2):109, 2024.

Lestari, D., Hamzah, H., Pramesthi, A. D. E. D., Syamsul, E. S., Safitri, P. D., Nurjunnah, R., and Rohman, A. The employment of FTIR-ATR spectroscopy and GC-MS combined with chemometrics for rapid detection of adulteration of pork oil in Gabus fish oil (Channa striata) for halal authentication. *Open Veterinary Journal*, 15(2):646, 2025.

Li, S., Sun, Y., Yen, G. G., and Zhang, M. Automatic design of convolutional neural network architectures under resource constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Linnainmaa, S. *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*. PhD thesis, Master's Thesis (in Finnish), Univ. Helsinki, 1970.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., and Tegmark, M. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Lu, W., Wang, H., Ge, L., Wang, S., Zeng, X., Mao, Z., Wang, P., Liang, J., Xue, J., Cui, Y., et al. Comparative evaluating laser ionization and iKnife coupled with rapid evaporative ionization mass spectrometry and machine learning for geographical authentication of Larimichthys crocea. *Food Chemistry*, 460:140532, 2024.

Mai, N., Gretar Bogason, S., Arason, S., Víkingur Árnason, S., and Geir Matthíasson, T. Benefits of traceability in fish supply chains–case studies. *British Food Journal*, 112(9): 976–1002, 2010.

Mariani, S., Griffiths, A. M., Velasco, A., Kappel, K., Jérôme, M., Perez-Martin, R. I., Schröder, U., Verrez-Bagnis, V., Silva, H., Vandamme, S. G., et al. Low mislabeling

rates indicate marked improvements in European seafood market operations. *Frontiers in Ecology and the Environment*, 13(10):536–540, 2015.

Marín, A. A meta-review of DNA-based identification methods and mislabeling analysis of Eastern South Pacific seafood. *bioRxiv*, pp. 2025–08, 2025.

Mello, L. C., Cavalcante, R. M., Bastos, L. M., Nossol, A. B. d. S., Martins, M. M., Cesar-Ribeiro, C., Pauly, G. d. F. E., and Abessa, D. M. d. S. Oil spill effects in biota and traditional populations from protected areas of Brazil's Northeast. *Human and Ecological Risk Assessment: An International Journal*, 31(3-4):563–586, 2025.

Miller, D., Miller, D. D., Jessel, A., Jessel, A., Mariani, S., and Mariani, S. Seafood mislabelling: comparisons of two western European case studies assist in defining influencing factors, mechanisms and motives. *Fish and Fisheries*, 2012. doi: 10.1111/j.1467-2979.2011.00426.x.

Ministry for Primary Industries. Hoki: New Zealand's largest fishery, 2024. URL https://www.mpi.govt.nz/fishing-aquaculture/fisheries-management/fish-stock-status/hoki-new-zealands-largest-fishery/.

Moeinpour, F., Mohseni-Shahri, F. S., and Sorouei, M. A. Purple Carrot Anthocyanin-Enhanced Gelatin-Chitin Films for Colorimetric Detection of Fish Adulteration by Formalin and Ammonia. *Applied Food Research*, pp. 101000, 2025.

Moens, E. Production and use of food-grade lubricants. *Trends in Food Science & Technology*, 14(4):157–162, 2003.

Montgomery, D. C. *Introduction to Statistical Quality Control*. John Wiley & Sons, Hoboken, NJ, 8th edition, 2019.

Morgan, N. and Bourlard, H. Generalization and parameter estimation in feedforward nets: Some experiments. *Advances in neural information processing systems*, 2, 1989.

Muñoz-Colmenero, M., Blanco, O., Arias, V., Martinez, J. L., and Garcia-Vazquez, E. DNA authentication of fish products reveals mislabeling associated with seafood processing. *Fisheries*, 41(3):128–138, 2016.

Murphy, F., Gathercole, J., Lee, E., Homewood, I., Ross, A. B., Clerens, S., and Maes, E. Discrimination of milk fermented with different starter cultures by MALDI-TOF MS and REIMS fingerprinting. *International Dairy Journal*, 122:105143, 2021.

Nielsen, E. E., Cariani, A., Aoidh, E. M., Maes, G. E., Milano, I., Ogden, R., Taylor, M. I., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S. J., Limborg, M. T., Martinsohn, J. T., McEwing, R., Panitz, F., Patarnello, T., Tinti, F., Houdt, J. V., Volckaert, F. A. M., Waples, R. S., and

Carvalho, G. R. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, 2012. doi: 10.1038/ncomms1845.

Niu, Z., Zhou, M., Wang, L., Gao, X., and Hua, G. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4920–4928, 2016.

Onoja, C. R., Ogbaji, H. O., Edodi, I. O., Mmuta, E. C., Igwegbe, K. C., Ogini, O. R., Odibo, U. E., Ewuola, A. A., Mahmoud, A. B., and Okpoji, A. U. Heavy metal contamination and health risk assessment of smoked-dried fish sold in Eke-Awka Market, Anambra State, Nigeria. *Journal of Sustainable Research and Development*, 1(2):18–25, 2025.

OpenSeas. Hoki Macruronus novazelandiae. https://openseas.org.nz/fish/hoki/, Oct 2021.

Pardo, M. Á., Jiménez, E., and Pérez-Villarreal, B. Misdescription incidents in seafood sector. *Food Control*, 62:277–283, 2016.

Paszke, A. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

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, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Pingki, F. H., Anisuzzaman, M., Jolly, Y. N., Nur, A.-A. U., Arai, T., Al-Anazi, K. M., Farah, M. A., Huque, R., Khatun, M. A., Hossain, I., et al. Microplastics (MPs) accumulation in two edible marine fish species sourced from the Bay of Bengal: Contamination level and risk assessment. *Journal of Food Composition and Analysis*, 139:107053, 2025.

Piroutková, M. et al. An analytical and machine learning approach for total mercury and methylmercury determination in squid: enhancing food safety testing and traceability monitoring systems. *Food Chemistry*, pp. 146766, 2025.

Plant and Food Research. CyberMarine - New research to maximise value from seafood resources, 2020. URL https://www.plantandfood.com/en-nz/article/new-research-to-maximise-value-from-seafood-resources.

Premanandh, J. Horse meat scandal–A wake-up call for regulatory authorities. *Food control*, 34(2):568–569, 2013.

Pruekprasert, K., Tan, M., Simon, D., Ford, L., Davies, A., Takats, Z., and Onida, S. Minimally Invasive Evaluation of Venous Leg Ulcers in an Outpatient Setting Using

Rapid Evaporative Ionization Mass Spectrometry Coupled with a Carbon Dioxide Laser. *Journal of Vascular Surgery: Venous and Lymphatic Disorders*, 13(2), 2025.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. Improving language understanding by generative pre-training. *Arxiv*, 2018. URL https://api.semanticscholar.org/CorpusID:49313245.

Ribeiro, M. T., Singh, S., and Guestrin, C. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

Ross, A., Brunius, C., Chevallier, O., Dervilly, G., Elliott, C., Guitton, Y., Prenni, J. E., Savolainen, O., Hemeryck, L., Vidkjær, N. H., and et al. Making complex measurements of meat composition fast: Application of rapid evaporative ionisation mass spectrometry to measuring meat quality and fraud. *Meat Science*, 181:108333, 2021.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. Learning internal representations by error propagation, 1985.

Rush, A. M. The annotated transformer. In *Proceedings of workshop for NLP open source software (NLP-OSS)*, pp. 52–60, 2018.

Russell, S. J. and Norvig, P. *Artificial intelligence: a modern approach*. pearson, 2016.

Schäfer, K.-C., Dénes, J., Albrecht, K., Szaniszló, T., Balog, J., Skoumal, R., Katona, M., Tóth, M., Balogh, L., and Takáts, Z. In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. *Angewandte Chemie International Edition*, 48(44):8240–8242, 2009.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

Shakik, A., Brohomo, P., Kabir, M. S., Islam, M. S., and Mizan, M. H. Microplastic contamination in Rupsha River of Bangladesh and its impacts on fish species. *Regional Studies in Marine Science*, 85:104130, 2025.

Shen, Q., Li, L., Song, G., Feng, J., Li, S., Wang, Y., Ma, J., and Wang, H. Development of an intelligent surgical knife rapid evaporative ionization mass spectrometry based method for real-time differentiation of cod from oilfish. *Journal of Food Composition and Analysis*, 86:103355, 2020.

Shen, Q., Lu, W., Cui, Y., Ge, L., Li, Y., Wang, S., Wang, P., Zhao, Q., Wang, H., and Chen, J. Detection of fish frauds (basa catfish and sole fish) via iKnife rapid evaporative

ionization mass spectrometry: An in situ and real-time analytical method. *Food Control*, 142:109248, 2022.

Shi, X., Cao, W., and Lau, J. H. Deep conditional ordinal regression for neural networks. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 2021.

Simopoulos, A. P. Evolutionary aspects of diet: the omega-6/omega-3 ratio and the brain. *Molecular neurobiology*, 44(2):203–215, 2011.

Sonone, S. D., Jorvekar, S. B., Naik, D. D., Saharia, N., and Borkar, R. M. Assessment of heavy metal contamination risk in dry fish from India: A comprehensive study. *Food Control*, 167:110804, 2025.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Stevens, J. Fish byproducts are an under-valued and untapped resource. https://www.globalseafood.org/advocate/it-takes-guts-to-advance-sustainability-in-aquaculture/, 2018.

Suzuoki, S. and Hatano, K. Reducing hallucinations in large language models: A consensus voting approach using mixture of experts, 2024.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, pp. 270–279. Springer, 2018.

Thompson, M., Sylvia, G., and Morrissey, M. T. Seafood traceability in the United States: Current trends, system design, and potential applications. *Comprehensive Reviews in Food Science and Food Safety*, 4(1):1–7, 2005.

Tran, B., Xue, B., and Zhang, M. Genetic programming for feature construction and selection in classification on high-dimensional data. *Memetic Computing*, 8(1):3–15, 2016.

Tran, B., Xue, B., and Zhang, M. Genetic programming for multiple-feature construction on high-dimensional classification. *Pattern Recognition*, 93:404–417, 2019.

Turkson, C., Cromwell, J., Yamoah, F., and Dora, M. Digital technologies for traceability and transparency in the global fish supply chains: a systematic review and future directions. 2025.

Tzafetas, M., Mitra, A., Paraskevaidi, M., Bodai, Z., Kalliala, I., Bowden, S., Lathouras, K., Rosini, F., Szasz, M., Savage, A., et al. The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. *Proceedings of the National Academy of Sciences*, 117(13):7338–7346, 2020.

Untal, F. J. P., Castro, M. M. C., and Sarmiento, J. M. Fishers' preference for mobile traceability platform: challenges in achieving a digital tuna supply chain in Davao Region, Philippines. *The International Journal of Logistics Management*, 36(2):433–451, 2025.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(1):142, 2006. doi: 10.1186/1471-2164-7-142.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Vaz Jr, S. *Analytical techniques and methods for biomass*. Springer, 2016.

Veluchamy, D., Rajan, M. S. M., and Prajapati, Y. K. Sensitive Detection of Formalin Adulteration in Fish Using a Square-Core Photonic Crystal Fiber Sensor. *Sensing and Imaging*, 26(1):128, 2025.

Verplanken, K., Stead, S., Jandova, R., Van Poucke, C., Claereboudt, J., Vanden Bussche, J., De Saeger, S., Takats, Z., Wauters, J., and Vanhaecke, L. Rapid evaporative ionization mass spectrometry for high-throughput screening in food analysis: The case of boar taint. *Talanta*, 169:30–36, 2017.

Wagner, I., Sanou, A., Guelbeogo, M., Williams, J., Harris, J., Smith, E. S., Sanou, R., Isiaka, S., Tapsoba, M., Soumnaba, Z., et al. Rapid Evaporative Ionisation Mass Spectrometry can reliably age field caught malaria vectors (and has the potential to simultaneously also identify species and infection rate). 2025.

Walsh, I., Ruethers, T., Chiin, S. L., Teo, G., Tay, S. J., Wan, C., Pang, K. T., Chia, S., Lopata, A. L., and Qiu, B. Differentiation of Fish Species Based on O-Acetylated

N-Glycan Fragments Using LC-IM-MS to Combat Seafood Adulteration. *Applied Food Research*, pp. 101428, 2025.

Wang, J., Ji, T., Wu, Y., Yan, H., Gui, T., Zhang, Q., Huang, X., and Wang, X. Length generalization of causal transformers without position encoding. *arXiv preprint arXiv:2404.12224*, 2024.

Ward, R. D., Zemlak, T. S., Innes, B. H., and Hebert, P. D. N. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, 2005. doi: 10.1098/rstb.2005.1716.

Wolpert, D. H. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.

Wolpert, D. H. and Macready, W. G. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 2002.

Wood, J., Nguyen, B. H., Xue, B., Zhang, M., and Killeen, D. Automated Fish Classification Using Unprocessed Fatty Acid Chromatographic Data: A Machine Learning Approach. In *Australasian Joint Conference on Artificial Intelligence*, pp. 516–529. Springer, 2022.

Wood, J., Nguyen, B. H., Xue, B., Zhang, M., and Killeen, D. Hook, Line, and Spectra: Machine Learning for Fish Species Identification and Body Part Classification using Rapid Evaporative Ionization Mass Spectrometry. *Intelligent Marine Technology and Systems*, pp. 516–529, 2025.

Wu, W., Du, R., Chen, Z., Li, W., Huang, X., and Pan, Z. Unlocking the combined impact of microplastics and emerging contaminants on fish: A review and meta-analysis. *Aquatic Toxicology*, 278:107176, 2025.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32 (1):4–24, 2020.

Xing, J., Saeys, W., and De Baerdemaeker, J. Automated Inspection of Food Products by Machine Vision. In Sun, D.-W. (ed.), *Computer Vision Technology for Food Quality Evaluation*, chapter 1, pp. 1–20. Academic Press, London, 2nd edition, 2016. This chapter provides an overview of automated visual inspection relevant to food quality, fitting the context of automated QA in fish processing.

Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L., and Liu, T. On layer normalization in the transformer architecture. In *International conference on machine learning*, pp. 10524–10533. PMLR, 2020.

Xiong, X., Xiong, X., Guardone, L., Guardone, L., Cornax, M. J., Cornax, M. J., Tinacci, L., Tinacci, L., Guidi, A., Guidi, A., Gianfaldoni, D., Gianfaldoni, D., Armani, A., and Armani, A. DNA barcoding reveals substitution of Sablefish (Anoplopoma fimbria) with Patagonian and Antarctic Toothfish (Dissostichus eleginoides and Dissostichus mawsoni) in online market in China: How mislabeling opens door to IUU fishing. *Food Control*, 2016. doi: 10.1016/j.foodcont.2016.06.010.

Xue, J., Xu, F., Ge, L., Wu, H., Lu, W., Wang, S., Zhao, Q., Wang, H., Liang, J., Yang, Z., et al. From Laboratory Exploration to Practice: Applications, Challenges, and Development Trends of Rapid Evaporative Ionization Mass Spectrometry Technology in Food Detection. *Comprehensive Reviews in Food Science and Food Safety*, 24(4): e70235, 2025.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pp. 12310–12320. PMLR, 2021.

Zhang, R., Ross, A. B., Jacob, N., Agnew, M., Staincliffe, M., and Farouk, M. M. Rapid Evaporative Ionisation Mass Spectrometry fingerprinting can discriminate lamb meat due to different ageing methods and levels of dehydration. *Journal of Proteomics*, 272: 104771, 2023.

Zhang, X., Giusti, A., Li, S., Deng, W., Sun, Z., Li, Y., Peng, H., Hu, J., Armani, A., and Wen, J. Beyond mislabelling: Chinese fish balls authentication by metabarcoding allows unveiling hidden mammal and avian species. *Food Control*, 176:111338, 2025.

Zhao, F., Zhang, C., and Geng, B. Deep multimodal data fusion. *ACM computing surveys*, 56(9):1–36, 2024.

Zhu, J., Jang-Jaccard, J., Singh, A., Welch, I., Harith, A.-S., and Camtepe, S. A few-shot meta-learning based siamese neural network using entropy features for ransomware classification. *Computers & Security*, 117:102691, 2022.