Wood *et al.*
Intelligent Marine Technology and Systems (2025) 3:16
https://doi.org/10.1007/s44295-025-00066-3

imts Intelligent Marine Technology and Systems

RESEARCH PAPER Open Access

Hook, line, and spectra: machine learning for fish species identification and body part classification using rapid evaporative ionization mass spectrometry

Jesse Wood[1*], Bach Nguyen[1], Bing Xue[1], Mengjie Zhang[1] and Daniel Killeen[2]

Abstract

Marine biomass composition analysis traditionally requires time-consuming processes and domain expertise. This study demonstrates the effectiveness of rapid evaporative ionization mass spectrometry (REIMS) combined with advanced machine learning (ML) techniques for accurate marine biomass composition determination. Using fish species and body parts as model systems representing diverse biochemical profiles, we investigate various ML methods, including unsupervised pretraining strategies for transformers. The deep learning approaches consistently outperformed traditional machine learning across all tasks. For fish species classification, the pretrained transformer achieved 99.62% accuracy, and for fish body parts classification, the transformer achieved 84.06% accuracy. We further explored the explainability of the best-performing and predominantly black box models using local interpretable model-agnostic explanations and gradient-weighted class activation mapping to identify the important features driving the decisions behind each of the best performing classifiers. REIMS analysis with ML can be an accurate and potentially explainable technique for automated marine biomass composition analysis. Thus, REIMS analysis with ML has potential applications in quality control, product optimization, and food safety monitoring in marine-based industries.

Keywords AI applications, Explainable AI, Machine learning, Marine biomass, Mass spectrometry, Multidisciplinary AI

1 Introduction

The fish processing industry forms a critical component of the global seafood supply chain, transforming raw marine biomass into consumer products through multiple stages. This process typically involves species sorting, cleaning, filleting, packaging, and quality control at various checkpoints. Each stage presents unique challenges that can benefit from artificial intelligence (AI) and machine learning (ML) solutions. The traditional fish processing workflow begins with the arrival of the catch, in which workers must rapidly sort different species–a task prone to human error, particularly with similar-looking fish. Then, the catch moves through the cleaning and filleting stations, where different body parts are separated for various products–from premium fillets to processed fish meal. Quality control occurs throughout, checking for freshness, proper handling, and accurate labeling. Finally, products are packaged and prepared for distribution. Several critical challenges that exist within this workflow are as follows:

*Correspondence:
Jesse Wood
jesse.wood@ecs.vuw.ac.nz
[1] Engineering and Computer Science, Victoria University of Wellington, Wellington 6012, New Zealand
[2] Seafood Technologies, Plant and Food Research New Zealand, Nelson 7010, New Zealand

(1) Quality control: Mislabeling and fraud remain persistent issues in the seafood industry (Black et al. 2017), with economic and food safety implications. Studies have shown significant rates of species substitution in various markets (Pardo et al. 2016).
(2) Product optimization: Different fish species and fish body parts have varying commercial values and uses, with some parts commanding premium prices in specific markets. The accurate classification of these parts ensures optimal resource utilization and maximizes economic value across the supply chain (Ghaly et al. 2013).
(3) Safety monitoring: The accurate tracking of processed species volumes is essential for regulatory compliance and stock management (Pauly and Zeller 2016).

These areas provide opportunities where automated analysis can significantly improve fish processing. Specifically, we explore the application of ML to rapid evaporative ionization mass spectrometry (REIMS) data across two critical classification tasks, i.e., fish species identification and body part classification. REIMS technology, combined with ML algorithms, aims to provide a promising solution for real-time, accurate analysis during processing operations. Our focus on these specific tasks is driven by their direct impacts on industry pain points:

(1) Species classification helps combat fraud and ensures proper resource management. (2) Body part identification helps optimize processing efficiency and product value. (3) Accurate species counting supports regulatory compliance and sustainability efforts.

This study demonstrates how ML techniques applied to REIMS data can enhance the efficiency and accuracy of these critical fish processing operations while supporting broader goals of sustainability and food security in the seafood industry.

REIMS marine biomass analysis faces several challenges, such as the time-consuming manual 'offline' analysis, costly domain expertise required, high dimensionality (Köppen 2000), few training samples, and the need for automated 'online' inference. However, 'online' inference in the domain of chemistry and fish processing, should not be confused with 'online' learning from ML. The rapid nature of REIMS necessitates equally rapid inference of its results, as traditional analytical chemistry techniques, which take several hours, are slow (Jha 2015). Furthermore, current analytical methods for REIMS data often require domain expertise in chemistry and fish processing, which does not match the speed of REIMS. Traditionally, samples would be sent away for 'offline' analysis by domain experts in chemistry. Thus, we aim to develop methods capable of automated inference for 'online' analysis on the production line of a fish processing factory. REIMS also produces high-dimensional data, with this particular dataset having 2080 mass-to-charge ratios as features but with limited training instances because of the time-consuming and expensive task of sample preparation. In addition, industry applications require fast, accurate, and interpretable models that can be verified and troubleshooted in real-world scenarios.

To address the aforementioned challenges, this study proposes several innovative ML approaches that provide automated inference, eliminating the need for domain expertise in chemistry and fish processing. To handle the high dimensionality of REIMS data, this study utilizes deep learning (Vaswani et al. 2017; Devlin et al. 2018) and evolutionary computation (Tran et al. 2016, 2019) that can address complex feature interactions in mass spectra with limited training instances. Techniques such as bidirectional encoder representations (BERT) (Devlin et al. 2018) and attention mechanisms (Vaswani et al. 2017) can capture complex, nonlinear relationships between features in high-dimensional data. To mitigate the limited number of training samples, we implement the unsupervised pertaining approach, which involves training the model on a large amount of unlabeled data before fine-tuning it on the limited labeled dataset. The model learns general features and patterns from the unlabeled data, which can then be transferred to the specific task at hand, significantly improving performance when labeled data are scarce. Finally, we employ local interpretable model-agnostic explanations (LIME) (McCann and Lowe 2012) and gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al. 2017) to provide interpretable outputs that identify important features and quantify their impact, making our models more accessible to domain experts in chemistry and fish processing.

The main contributions of the paper are as follows:

(1) Real-time marine biomass analysis: This study demonstrates the use of REIMS combined with advanced ML techniques to enable automated analysis of marine biomass, which represents a significant improvement over traditional, time-consuming methods.
(2) ML on sequential data: This study demonstrates that deep learning approaches, particularly transformers with and without progressive masking pretraining, consistently outperform traditional ML methods for analyzing sequential REIMS data.
(3) Feature importance: The important mass-to-charge ratios for the best performing models are identified to enhance domain knowledge in fish processing and chemistry.

2 Related works

Building upon the foundation laid in the 'Introduction', this section delves deeper into the existing body of research on marine biomass analysis, exploring traditional methods and recent advancements in REIMS technology, while highlighting the gaps and challenges that our proposed approach aims to address. We also provide the necessary background on deep learning required.

2.1 Marine biomass

Mislabeling is a significant issue in the global seafood industry, with a meta-analysis of genomic profiling methods finding an average mislabeling rate of 30% globally (Pardo et al. 2016). ML methods using REIMS data provide a promising solution to this problem by enabling more accurate fish species classification. For example, in 2016, a restaurant in Melbourne was accused of serving catfish instead of dory (Pearl 2016), highlighting the need for better species detection techniques. REIMS technology, which works on raw and cooked biomass, can combat fraud by ensuring species authenticity. Approximately 40% of a fish is edible fillet, whereas the remaining 60% can be repurposed into products such as fertilizers or high-value pharmaceutical-grade omega-3 concentrates. Fish oil, rich in omega-3 polyunsaturated fatty acids (Simopoulos 2011), is nutritionally essential but increasingly scarce in Western diets (FAO 2020). REIMS-based ML methods in fish processing also help identify high-value parts for repurposing into valuable products, contributing to the increasing consumer demand for omega-3 supplements made from diverse marine biomass (Panse and Phalke 2016).

2.2 REIMS

Traditional approaches for analyzing marine biomass composition have long been the cornerstone of research and quality control in the seafood industry. These methods include gas chromatography-mass spectrometry (Wood et al. 2022), nuclear magnetic resonance spectroscopy (Bettjeman et al. 2018), and genomic profiling (Pardo et al. 2016). Although these techniques have been proven valuable, they often come with significant drawbacks. These techniques are typically time-consuming, requiring extensive sample preparation and analysis time. In addition, they are labor-intensive, demanding skilled technicians to operate complex equipment and interpret results. Most importantly, these methods necessitate substantial domain expertise, limiting their accessibility and scalability in real-world applications. These limitations have spurred the search for rapid techniques for marine biomass analysis in fish processing.

In recent years, REIMS has emerged as a promising technique for the rapid and accurate analysis of biological samples, addressing many of the limitations of traditional methods. Since its introduction in the original paper by Balog et al. (2010), REIMS has demonstrated its versatility and effectiveness across various applications. For instance, REIMS has been successfully employed to detect horse offal mixed with beef mince at concentrations as low as 1%–5%, showcasing its potential in addressing food fraud (Black et al. 2017). In the realm of seafood, REIMS has been applied to binary classification tasks for detecting fish species and catch methods, further illustrating its utility in combating fish fraud (Black et al. 2019). Historically, REIMS biomass analysis has primarily relied on orthogonal partial least squares discriminant analysis (OPLS-DA) (Balakrishnama and Ganapathiraju 1998; Bylesjö et al. 2006; Boccard and Rutledge 2013) with principal component analysis (PCA) for dimensionality reduction (Abdi and Williams 2010). However, this PCA-OPLS-DA approach has limitations, particularly in its reliance on outlier thresholding for adulteration detection, which requires manually defined hyperparameters set by domain experts in chemistry. This work proposes automated methods with learnable parameters that do not require domain expertise in chemistry to be configured. In addition, this work proposes deep learning and evolutionary computation methods from ML that outperform the traditional OPLS-DA approach.

2.3 ML for REIMS

Deep learning models were selected for REIMS marine biomass analysis because of their capability to handle complex, high-dimensional data with sequential or structured dependencies, which are inherent in REIMS data. Transformers (Vaswani et al. 2017; Devlin et al. 2018), known for their powerful self-attention mechanisms, excel at weighing the importance of different features in sequential data, making them well-suited for identifying patterns in REIMS spectra. Because REIMS data, similar to sequences in language, consist of ordered data points (mass-to-charge ratios) with varying degrees of importance, the attention mechanism of the transformer enables it to dynamically focus on critical parts of the spectrum for classification or prediction. Long short-term memory (LSTM) networks (Hochreiter and Schmidhuber 1997), a type of recurrent neural network, are also effective for REIMS data because they capture long-term dependencies in sequential data, which is crucial for REIMS analysis, as spectral data might contain dependencies across distant mass-to-charge values. The capability of LSTM to store and retrieve information over long sequences enhances its performance in such tasks, especially when the signal patterns may not be immediately adjacent. Variational autoencoders (VAEs) (Kingma

and Welling 2013) provide an effective approach to handling the complexity and variability of REIMS data by learning a compressed, latent representation of the spectral information. VAEs can also reconstruct these data, making them ideal for different tasks such as species and body part classification, where they can model and detect small anomalies or deviations in the spectral data. Kolmogorov-Arnold networks (KAN) (Liu et al. 2024) are highly efficient at approximating complex functions, which is essential in REIMS data analysis, where subtle differences in spectra can indicate different classes of species and body parts. The capability of KAN to improve function approximation makes it especially powerful for handling nonlinear patterns in mass spectrometry data, which traditional models may struggle to capture. Convolutional neural networks (CNN) (LeCun 1989; LeCun et al. 1989a, 1998b), although primarily used in image processing, are highly effective for REIMS data because of the spatial connectivity in mass spectra. Similar to neighboring pixels in images that share spatial relationships, neighboring mass-to-charge ratios in REIMS data also exhibit dependencies. CNN can exploit this structure to identify patterns in one-dimensional (1D) data, treating mass spectra similarly to 1D images. Finally, Mamba (Gu and Dao 2023), a state-space model, provides an efficient alternative to transformers for sequential data processing. Mamba is designed for high-performance handling of complex time series data, making it an excellent fit for REIMS analysis, where computational efficiency and the capability to model sequential dependencies are essential for automated or large-scale biomass analysis.

3 Methods

With the background established, this discusses the heart of our analytical approach, i.e., the classification methods that extract meaningful insights from the REIMS spectra.

3.1 Deep learning and evolutionary computation

The intricate nature and high dimensionality of REIMS data demand advanced models capable of managing sequential relationships and complex spectral features. We explore various architectures tailored to these demands. Transformers (Vaswani et al. 2017; Devlin et al. 2018) utilize self-attention mechanisms to prioritize different mass-to-charge ratios within the spectra, akin to their effectiveness in processing sequential data for language tasks, which makes them adept at pinpointing key regions in REIMS spectra for classification purposes. LSTM networks (Hochreiter and Schmidhuber 1997) are proficient at capturing extended dependencies across mass-to-charge values, which is vital for identifying intricate spectral patterns that may extend over broad data ranges. Their capacity to retain and leverage information across long sequences is essential for detecting faint contamination signals. VAEs (Kingma and Welling 2013) generate compact representations of REIMS spectra, proving useful for spotting anomalies and cross-species contamination by modeling and reconstructing intricate spectral distributions. We also investigate KAN (Liu et al. 2024), which can approximate complex functions in spectral data that could reveal species distinctions or contamination levels. CNN (LeCun et al. 1998) can explore local dependencies within the spectra, treating them as 1D signals with significant neighboring connections. Finally, the Mamba architecture (Gu and Dao 2023) provides an efficient state-space method for processing sequential REIMS data, striking a balance between computational efficiency and modeling power, which is critical for real-time analysis.

Genetic programming (Koza 1994) is an evolutionary computation method that solves a given problem by iteratively evolving a population of solutions, often represented by trees. Genetic programming has been used for feature construction (Tran et al. 2016, 2019), which can potentially enhance mass spectrometry data classification by automatically generating and evolving complex features from raw spectral data; thus, the constructed high-level features can better capture intricate patterns and relationships that simpler methods or original features might miss. This approach can improve model performance by tailoring features to the specific characteristics of mass spectrometry datasets, such as peak intensities and mass-to-charge ratios, leading to more accurate identification of compounds or contaminants.

3.2 Transformer

3.2.1 Architecture

The transformer model, originally proposed in the seminal paper by Vaswani et al. (2017), revolutionized natural language processing and other tasks involving sequential data by relying entirely on self-attention mechanisms instead of recurrent or convolutional layers. Our transformer model consists of an encoder-only structure, in which encoders are stacked as layers. Each encoder layer comprises multi-head self-attention and position-wise feed-forward layers. We implement a transformer without positional embeddings (Wang et al. 2024).

In the architecture used in this work (Fig. 1), the encoder blocks are equipped with residual connections (He et al. 2016), allowing gradients to flow efficiently during backpropagation. These residual connections act as 'gradient superhighways', enabling deeper models without the risk of vanishing or exploding gradients, thus allowing better training stability.

A notable aspect of the transformer architecture used in this work is the choice of pre-norm layer normalization

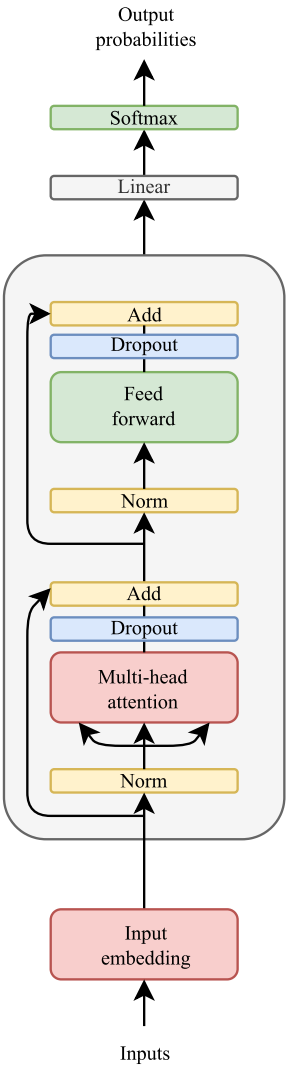

Fig. 1 Transformer architecture

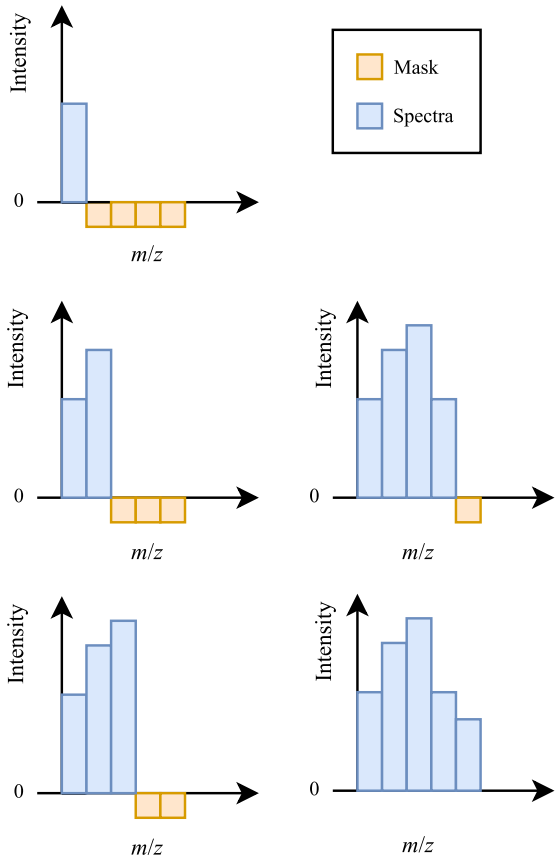

Fig. 2 Masked language modeling, where *m/z* stands for the mass-to-charge ratio

(Ba et al. 2016; Xiong et al. 2020), where layer normalization is applied before the multi-head self-attention and position-wise feed-forward layers. This approach contrasts with post-norm layer normalization (used after the attention block), as it stabilizes training and improves the convergence of deep transformers by ensuring more consistent gradients across layers. By normalizing before the main components of each layer, the pre-norm structure helps maintain better gradient flow across the network, contributing to more effective training of the encoder layers.

3.2.2 Progressive masking

Figure 2 illustrates the concept of progressive masking in pretraining transformer models. At the bottom right, we see the original mass spectra. On the top left, we see the first mask, which applies a mask to all spectra except the first one. From that, we see masks that slowly shrink down until we reach the original spectra. These patterns illustrate how the masking process evolved, starting with masking just one spectrum and progressively unmasking more spectra in the sample. Mask 1 shows only the first spectra, with the rest masked. Mask 2 reveals two spectra, masking the remainder. Mask 3 unmasks one more spectrum, showing three spectra. The final mask shows all of the spectra except the final one. This progressive masking technique creates multiple training examples from a single spectrum, effectively increasing the amount and diversity of training data for the transformer model. In this work, we apply left-to-right progressive masking to REIMS data. Instead of sentences in natural language processing, we are masking mass spectra, and pretraining has the model predict the masked spectra, which amortized the limited number of training samples by creating 2080 masked spectra per instance to train from, resulting in a training set of 2080 features $\times$ 72 samples = 149760 instances.

3.2.3 Pretrained transformers

Pretraining is an extension of transformers that enables them to be pretrained on a general task, and transfer the pretrained weights to a transformer model to be fine-tuned on a downstream task. This study adopts unsupervised pretraining inspired by BERT (Devlin et al. 2018) to improve the performance of transformer models on mass spectrometry tasks. Unsupervised pretraining has significant benefits, particularly for models working with limited labeled data. By training on large-scale, unlabeled datasets, the model learns general patterns that capture the underlying structure of the data, resulting in useful embeddings that can be fine-tuned for specific downstream tasks with small-scale, labeled datasets. This approach mitigates the need for extensive labeled data while still providing high-quality results.

This approach is an adaptation of the masked language modeling (MLM) task used in BERT to handle mass spectrometry data. In MLM, tokens in a sentence are progressively masked, and the model is trained to predict these masked tokens. Analogously, in masked spectra modeling (MSM), mass-to-charge ratios in spectra are progressively masked, and the model learns to predict the missing values, which is framed as a regression task, where the loss function is the mean squared error. By learning to predict missing mass-to-charge ratios, the model develops a robust understanding of the relationships between features in the spectra, making it well-suited for downstream tasks. We use left-to-right progressive masking to amortize the limited number of training instances.

By pretraining on this task, the transformer network learns valuable domain-specific representations. When fine-tuned on small-scale, labeled datasets, the model can leverage these pretrained weights, resulting in improved accuracy, faster convergence, and better generalization. This approach is particularly advantageous in some fields, such as mass spectrometry, where labeled data are limited, but large amounts of unlabeled data are readily available.

4 Experimental setup

Having outlined our various ML 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.

4.1 Comparison methods

This study evaluates a diverse range of ML techniques to classify the REIMS spectra:

(1) Benchmark 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.
(2) Traditional machine learning methods: Random forest (RF) (Ho 1995), K-nearest neighbors (KNN) (Fix and Hodges 1989), decision trees (DT) (Breiman 2017), naive Bayes (NB) (Hand and Yu 2001), logistic regression (LR) (Kleinbaum et al. 2002), support vector machines (SVM) (Cortes and Vapnik 1995), and linear discriminant analysis (LDA) (Balakrishnama and Ganapathiraju 1998).
(3) Ensemble method (Hansen and Salamon 1990): A combination of the aforementioned 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.
(4) Deep neural networks: Transformer (Vaswani et al. 2017; Devlin et al. 2018), LSTM (Hochreiter and Schmidhuber 1997), VAE (Kingma and Welling 2013), CNN (LeCun 1989; LeCun et al. 1989a,b 1998), KAN (Liu et al. 2024), and Mamba (Gu and Dao 2023).
(5) Genetic programming: Multiple class independent feature construction (MCIFC) (Tran et al. 2016, 2019) 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.

4.2 Benchmark technique

To evaluate the performance of the proposed methods, OPLS-DA (Bylesjö et al. 2006) is used as a benchmark to compare new approaches with the existing methods for REIMS analysis. OPLS-DA is the standard technique for biomass analysis using REIMS, as supported by prominent use in the literature (Balog et al. 2010; Jha 2015; Black et al. 2017, 2019). Therefore, we use OPLS-DA as a benchmark for comparative performance, showing the contrast between traditional and new approaches for biomass analysis with REIMS. OPLS-DA is considered an ML technique, specifically a supervised dimensionality reduction and classification method. OPLS-DA falls into the category of linear supervised ML algorithms, similar to partial least squares discriminant analysis and linear discriminant analysis. However, the primary strength of OPLS-DA lies in its capability to separate systematic variation into predictive and orthogonal (non-predictive) components, which makes it particularly useful for

classification and biomarker identification in some fields, such as metabolomics and chemometrics.

4.3 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 the majority and minority classes. By doing so, this method reduces the variance of performance estimates, leading to more stable and reliable metrics. In addition, this method maximizes the use of available data, allowing each sample to contribute to training and validation, which is crucial for small datasets. With threefold and fivefold cross-validations, the model is tested across various scenarios, improving its generalization to unseen data and providing a comprehensive evaluation of its performance.

4.4 Datasets

This study utilizes datasets provided by New Zealand Plant and Food Research as part of Cyber-Marine (Plant and Food Research 2020). REIMS can be used to optimize the value obtained from seafood resources. The dataset consists of mass spectrometry samples collected using REIMS, where an electrosurgical knife is used to create an aerosol from the samples. This aerosol is then directed into a mass spectrometer, where ionization occurs, enabling mass-to-charge ratio analysis. Each sample undergoes multiple incisions lasting 3–5 s, providing detailed chemometric data in the mass range of m/z 77.04–999.32. Figure 3 shows the two wild-caught fish species, i.e., hoki and mackerel, that are the subject of this study. These are two important fish species in New Zealand's seafood industry, especially given that New Zealand's largest fishery is hoki (Ministry for Primary Industries 2024).

For illustrative purposes, the different fish body parts, which are shared across both species of fish, are shown in Fig. 4.

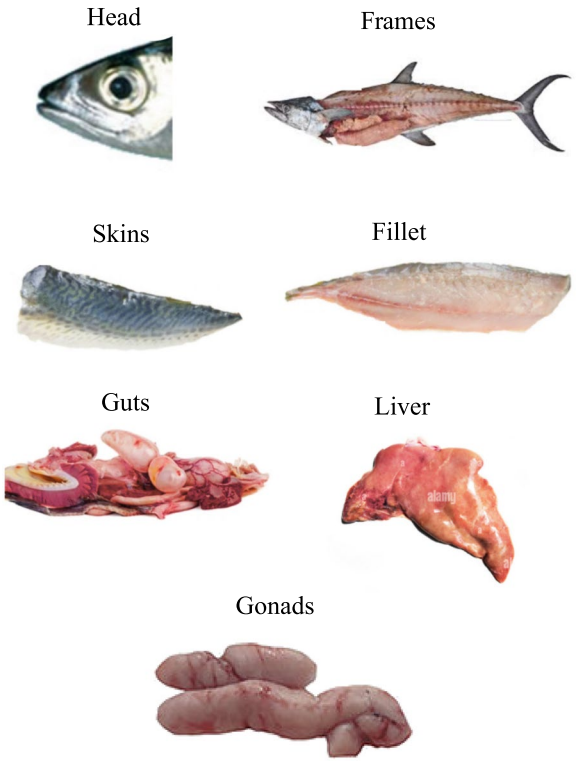

Fig. 4 Fish body parts

The dataset used consists of REIMS spectra collected from two fish species and seven body parts. In particular, we will have two different datasets corresponding to two different tasks:

(1) Species classification: The task is to distinguish between two species of fish (i.e., hoki and mackerel) based on 2080 features derived from REIMS analysis. This classification is crucial for food authentication and quality control in the seafood industry, helping prevent species substitution fraud and ensure accurate product labeling. We focus on pure (i.e., non-contaminated/unmixed) samples to establish a reliable baseline for species identification. The dataset contains 106 samples, with a relatively balanced distribution of 44.44% hoki and 55.56%

(a) Mackerel (b) Hoki

Fig. 3 Mackerel (**a**) and hoki (**b**) fish species

mackerel. These proportions reflect the natural availability of samples while maintaining sufficient representation for both species to train a robust classifier.

(2) Body parts classification: This multi-class classification task aims to identify seven distinct fish parts (i.e., fillets, head, livers, skins, gonads, guts, and frames) using REIMS data. The classification supports process automation by enabling automated sorting and processing in seafood production lines while helping maximize the value of each fish part, such as using fillets for premium products and frames for fish meal. Furthermore, precise classification ensures proper tracking and documentation of different fish components throughout the supply chain. The dataset contains 33 samples with a distribution of 16.66% each for fillets, heads, livers, skins, and guts and 8.33% each for gonads and frames. The relatively small sample size per class is attributed to the limited number of annotated samples for each class of body part.

The REIMS spectra were normalized to be within $x \in [0, 1]$, fitted to the training set of each fold. Let $X = \{x_1, x_2, \cdots, x_n\}$ be a dataset containing n elements. The normalized value x_i' for each element x_i is obtained as follows:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where $x_{\min}$ is the minimum value in the dataset X, $x_{\max}$ is the maximum value in the dataset X.

4.5 Parameter settings

Experiments use the default settings from sklearn (Pedregosa et al. 2011), except for SVM with a linear kernel, and LR set to 2000 for the maximum number of iterations. The ensemble voting classifier combines all of the traditional ML methods into one model. The ensemble uses hard voting, i.e., the predicted class labels for majority rule voting.

The deep learning models all use the following parameters: The AdamW optimizer (Loshchilov and Hutter 2017) decouples weight decay from the learning rate, an improvement over the popular Adam optimizer (Kingma and Ba 2014). Dropout (Srivastava et al. 2014) turns off neurons at random during training to efficiently approximate a bagged ensemble of subneural networks. Label smoothing (Szegedy et al. 2016) 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 (Hendrycks and Gimpel 2016) as the activation functions. Early stopping (Morgan and Bourlard 1989) is one of the most common forms of regularization that saves the model parameters when the validation loss is improved and tunes the hyperparameter of epochs (Goodfellow et al. 2016). To ensure a fair comparison, each model has the same hyperparameters, i.e., 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 1 presents 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.

5 Results and discussions

Having outlined our classification strategies, this section now presents and interprets the outcomes of applying these various ML techniques to the REIMS datasets. Tables 2 and 3 list the results of the classifiers on the training and test sets, with the best performing model on the test set rendered in bold, and the second best are rendered in italics. Notably, the method 'pretrained' indicates the transformer with progressive left-to-right masked pretraining. The transformer was pretrained on the training data of each fold during stratified k-fold cross-validation.

Table 1 Transformer parameter settings

Parameter	Setting
Learning rate	1e-5
Epochs	100
Dropout	0.2
Label smoothing	0.1
Early stopping patience	5
Optimiser	AdamW
Loss: MSM	MSE
Loss: speciation and 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

Table 2 Classification results of fish species identification

Method	Train (%)	Test (%)
OPLS-DA	98.91 ± 0.74	96.39 ± 4.44
KNN	95.76 ± 0	79.37 ± 0
DT	100.00 ± 0	99.17 ± 0
LR	100.00 ± 0	85.21 ± 0
LDA	98.54 ± 0	92.29 ± 0
NB	89.17 ± 0	66.67 ± 0
RF	100.00 ± 0	90.05 ± 0
SVM	100.00 ± 0	84.58 ± 0
Ensemble	100.00 ± 0	87.84 ± 0.40
Transformer	100.00 ± 0	*99.17 ± 1.67*
Pretrained	100.00 ± 0	**99.62 ± 1.15**
LSTM	100.00 ± 0	98.84 ± 1.76
VAE	100.00 ± 0	98.64 ± 1.94
KAN	100.00 ± 0	97.41 ± 2.45
CNN	100.00 ± 0	96.87 ± 3.24
Mamba	100.00 ± 0	98.27 ± 2.14
MCIFC	100.00 ± 0	97.89 ± 2.59

Table 3 Classification results of fish body part identification

Method	Train (%)	Test (%)
OPLS-DA	80.11 ± 2.86	51.17 ± 22.16
KNN	43.06 ± 0	39.17 ± 0
DT	100.00 ± 0	35.50 ± 4.35
LR	100.00 ± 0	59.58 ± 0
LDA	74.31 ± 0	52.92 ± 0
NB	100.00 ± 0	48.33 ± 0
RF	100.00 ± 0	61.67 ± 0
SVM	100.00 ± 0	52.33 ± 2.57
Ensemble	100.00 ± 0	52.33 ± 2.57
Transformer	100.00 ± 0	**84.06 ± 6.42**
Pretrained	100.00 ± 0	*83.94 ± 7.12*
LSTM	100.00 ± 0	82.11 ± 9.15
VAE	85.43 ± 6.28	74.81 ± 13.84
KAN	100.00 ± 0	73.06 ± 9.58
CNN	100.00 ± 0	70.41 ± 13.75
Mamba	100.00 ± 0	80.67 ± 8.73
MCIFC	97.95 ± 1.61	55.45 ± 19.19

5.1 Fish species classification

For the fish species classification task, the best performing model was the pretrained transformer (99.62%). This model excels in capturing the intricate patterns in the REIMS data, which provide distinct signatures for different fish species. The high performance of the DT model (99.17%) shows that even traditional ML methods are highly effective in this domain. Tree-based models, such as DT and RF, work well because they can split the data based on highly discriminative features, capturing non-linear relationships effectively. For a DT, although individual splits are linear (axis-parallel), their combination creates non-linear decision boundaries.

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

All of the deep learning models consistently outperform the traditional OPLS-DA method–with the pretrained transformer achieving 96.39% test accuracy–in the literature for REIMS analysis. The research field of REIMS analysis should consider deep learning methods for other applications, as they exhibit superior performance.

5.2 Fish body part

The transformer without pretraining performed the best in the fish body parts classification task, achieving a test accuracy of 84.06%. These models are well-suited for this task because they can handle complex and multidimensional input data, such as REIMS, capturing the subtle differences between body parts through advanced feature extraction and context awareness. LSTMs, with their capability to capture sequential dependencies, also perform well (82.11%), indicating some temporal or positional dependencies in the ionization patterns that relate to specific body parts.

However, traditional ML methods show lower performance in the fish body part task, compared to the fish species task. This finding indicates that the fish body parts classification task is inherently more complex because of less distinct signal differences between body parts, making it harder for simpler models to differentiate between classes. This increased difficulty 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 part classification with gas chromatography data illustrated the increased difficulty of body part classification.

Again, all of the deep learning methods–with the transformer achieving the best test accuracy of 84.06%–outperform the OPLS-DA method (51.17%). For the second task, deep learning methods have been proven to be superior to the traditional approach in the literature.

5.3 Summary

Across all tasks, the deep learning methods exhibited 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 the appropriate algorithms for specific analytical challenges in marine biomass analysis. Although the transformer model consistently excelled, simpler models, such as DT, exhibited competitive performance in certain tasks, with potential advantages in terms of interpretability and computational efficiency. The challenges faced in body part classification point to areas where further research is needed, which might include exploring more advanced feature extraction techniques, increasing the size and diversity of the training dataset, or developing specialized model architectures tailored to these specific tasks. Overall, our results indicate the potential of combining REIMS with ML for automated and accurate marine biomass analysis while also highlighting areas for future improvement and research.

6 Further analysis on feature importance

Although the performance of our simple and pretrained transformers is promising, understanding how they arrive at their predictions is crucial for building trust and gaining insights. The important features driving the decisions made by black box models need to be identified so that these models can be understood, trusted, and verified by domain experts in chemistry and fish processing. To address this challenge, we employ LIME, a technique used to explain predictions made by complex black box ML models (Ribeiro et al. 2016). We analyze the top 5 most important features of the best performing models that have been identified by LIME. LIME approximates a complex model 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 the features that influenced the prediction. LIME explanations, or feature importance charts, are used to explain the predictions of ML models by identifying the features (in this case, specific mass-to-charge ratios) that are most influential in a particular prediction. In these LIME charts:

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

The 1D gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al. 2017) implementation visualizes the features in the mass spectrometry data that most influence the classification decisions of a transformer model. For correctly classified samples, Grad-CAM generates an 'average correct Grad-CAM' by calculating the mean importance across all correctly predicted samples. This visualization shows the mass spectrometry peaks (represented by feature indices) that consistently contribute to accurate classifications. The implementation works by capturing gradients flowing through the final attention layer of the model during backpropagation, weighting the activation of each feature by its gradient, and normalizing the results. The resulting graph highlights regions in the mass spectra 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 the mass-to-charge ratios that drive classification can provide biochemical insights into the underlying samples.

6.1 Fish species classification

The pretrained transformer achieves the best classification accuracy (99.62%) for fish species classification. Figure 5 shows the LIME chart for the pretrained transformer for the mackerel fish species. The most important feature, denoted by the strongest green bar, is detected when the mass-to-charge ratio (m/z) of 794.0990 is within the normalized intensity range of $0.28 < y \leq 0.47$,

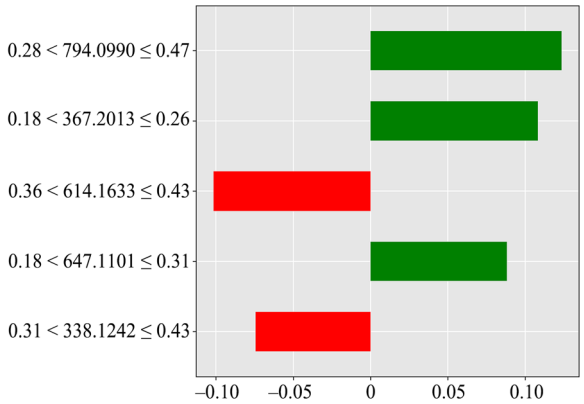

Fig. 5 LIME explanation for the pretrained transformer model using in mackerel species classification

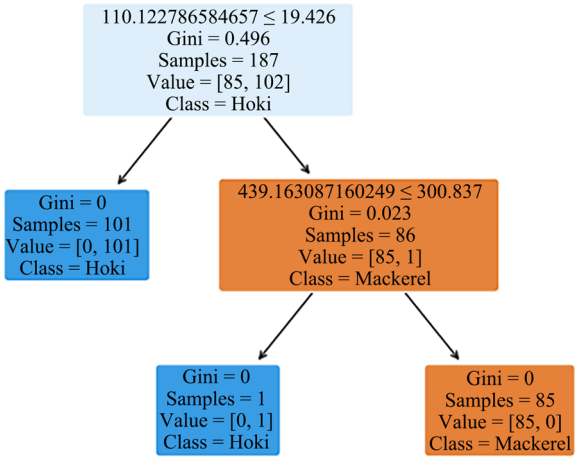

Fig. 7 Decision tree for fish species

indicating that large amounts of this molecule are present in the mackerel fish species.

Figure 6 shows the LIME chart for the pretrained transformer for the hoki fish species. The most important feature, denoted by the strongest red bar, is when the mass-to-charge ratio (m/z) of 229.0710 is within the normalized intensity range of $0.26 < y \leq 0.36$, indicating that large amounts of this molecule are present in a sample does not belong to the hoki fish species.

Figure 7 illustrates the DT with near-perfect accuracy, showing how a simple model with two splits, can classify fish species, giving both a highly accurate and interpretable model. The figure shows the two mass-to-charge ratios and their intensity thresholds on which they based their decision boundaries.

Figure 8 shows the average correct Grad-CAM for the body part classification task, highlighting the important features in the mass spectrograph that contribute to correct classifications, averaged across all of the classes. From the figure, we identify important features with high discriminative capability in the feature index range of 200–800 for the mass-to-charge ratios, with coefficients greater than 0.5.

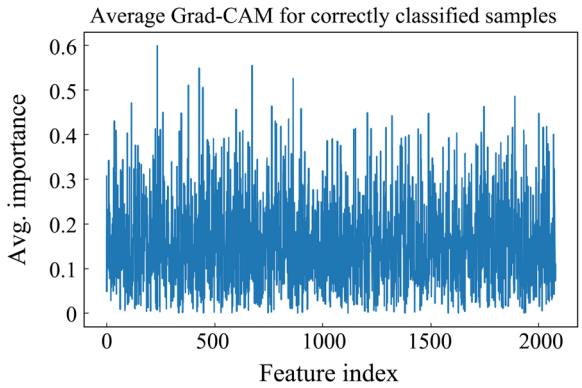

Fig. 8 Grad-CAM for the pretrained transformer model in fish species classification

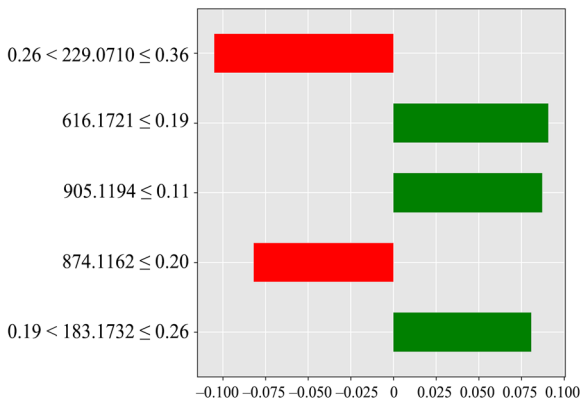

Fig. 6 LIME explanation for the pretrained transformer model used in hoki species classification

6.2 Fish body part

The transformer performs the best (83.94%) on the fish parts dataset. Figure 9 shows the LIME chart for the transformer for the fish part classification of fish heads. The most important feature, denoted by the strongest green bar, is detected when the mass-to-charge ratio (m/z) of 256.1089 is within the normalized intensity range of $0.26 < y \leq 0.35$, indicating that large amounts of this molecule are likely present in fish heads.

Figure 10 shows the LIME chart for the transformer for the fish body part classification of fish fillets. The most important feature, denoted by the strongest red bar, is detected when the mass-to-charge ratio (m/z) of 722.0810 is greater than the normalized intensity of 0.41, indicating that large amounts of this molecule are not expected in fish fillets.

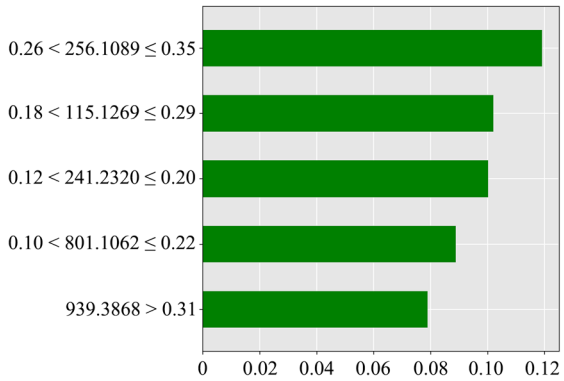

Fig. 9 LIME explanation for the transformer model used in fish head classification

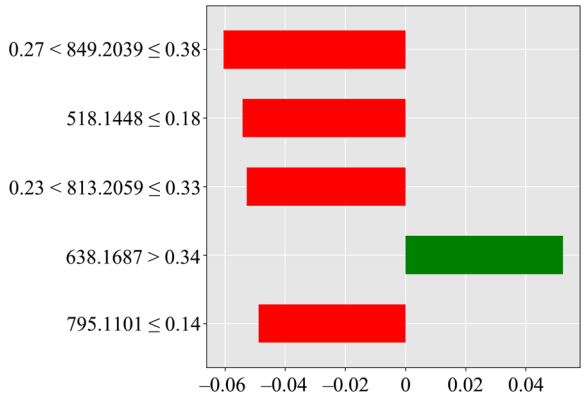

Fig. 11 LIME explanation for the transformer model used in fish livers classification

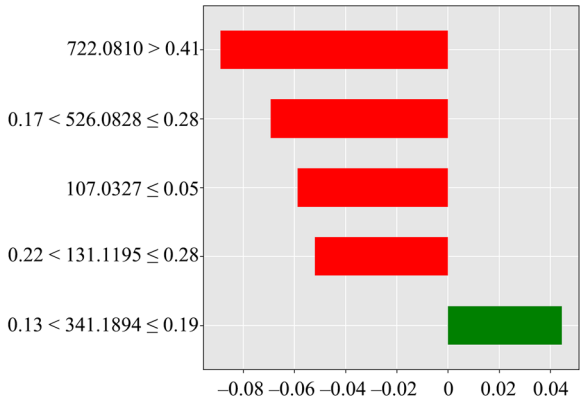

Fig. 10 LIME explanation for the transformer model used in fish fillets classification

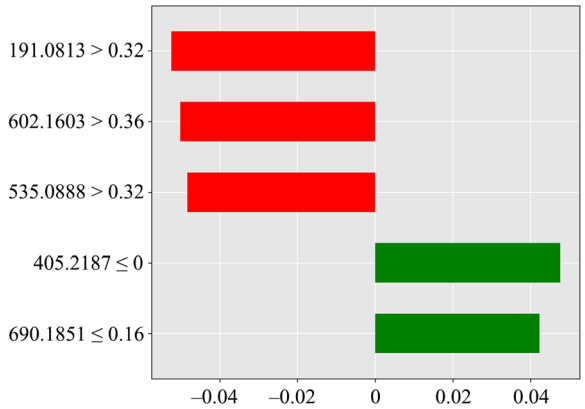

Fig. 12 LIME explanation for the transformer model used in fish skins classification

Figure 11 shows the LIME chart for the transformer for the fish body part classification of fish livers. The most important feature, denoted by the strongest red bar, is detected when the mass-to-charge ratio (m/z) of 849.2039 is within the normalized intensity range of $0.27 < y \leq 0.38$, indicating that large amounts of this molecule are not likely to be found in fish liver.

Figure 12 shows the LIME chart for the transformer for the fish body part classification of fish skins. The most important feature, denoted by the strongest red bar, is detected when the mass-to-charge ratio (m/z) of 191.0813 is greater than the normalized intensity of 0.32, indicating that large amounts of this molecule are not usually found in fish skins.

Figure 13 shows the LIME chart for the transformer for the fish body part classification of fish guts. The most important feature, denoted by the strongest red bar, is detected when the mass-to-charge ratio (m/z) 675.1786 is less than or equal to the normalized intensity of 0.11, indicating that small amounts of this molecule are not usually found in fish guts.

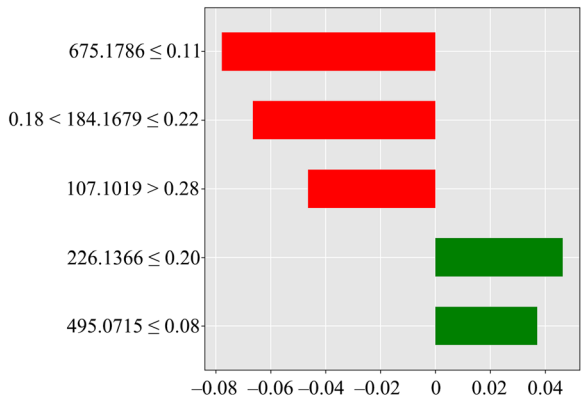

Fig. 13 LIME explanation for the transformer model using in fish guts classification

Figure 14 shows the LIME chart for the transformer for the fish body part classification of fish frames. The most important feature, denoted by the strongest green bar, is detected when the mass-to-charge ratio (m/z) of 533.161 is within the normalized intensity range of

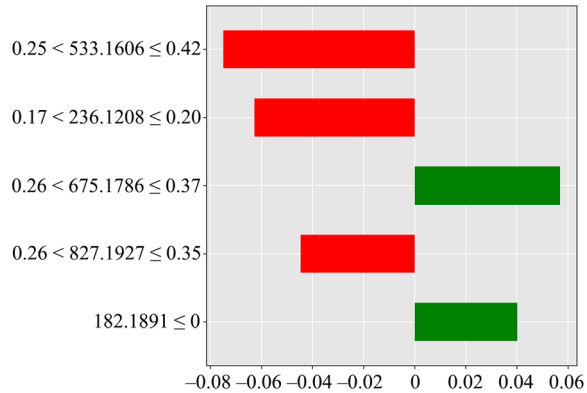

Fig. 14 LIME explanation for the transformer model used in fish frames classification

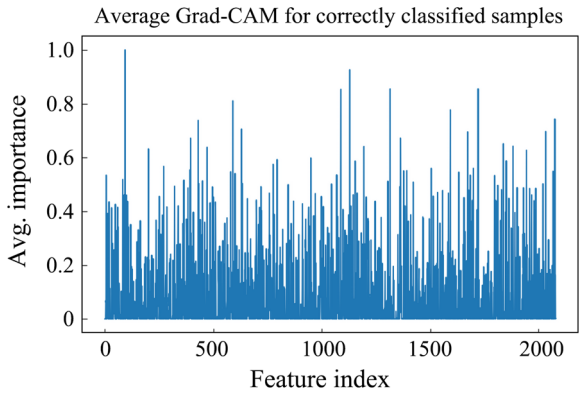

Fig. 16 Grad-Cam for the transformer model used in fish body part classification

$0.25 < y \leq 0.24$, indicating that average to large amounts of this chemical are not expected to be found in fish frames.

Figure 15 shows the LIME chart for the transformer for the fish body part classification of fish gonads. The most important feature, denoted by the strongest red bar, is detected when the mass-to-charge ratio (m/z) of 93.0882 is less than or equal to the normalized intensity threshold of $0.09 < y \leq 0.18$. Biochemically, this finding indicates that the m/z of 93.0882 might correspond to a compound not found in fish gonads.

Figure 16 shows the average correct Grad-CAM for the body part classification task, highlighting the key features in the mass spectrograph that contribute to accurate classifications, averaged across all of the classes. From the figure, we identify one important feature in the feature index range of 0–100, another in the feature index range of 500–600, and four more in the feature index range of 1000–1750 for the mass-to-charge ratios, with coefficients greater than 0.8. This Grad-CAM identifies more important features, with a higher average importance, for fish body parts than for fish species.

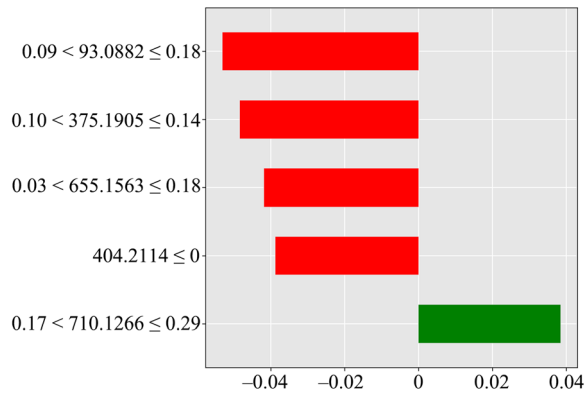

Fig. 15 LIME explanation for the transformer model used in fish gonads classification

7 Conclusions and future work

The results of these classification tasks indicate that deep learning models, particularly transformer, are well-suited for handling the complex, high-dimensional data generated by REIMS data. These models consistently outperform traditional ML methods, especially for tasks involving subtle or overlapping signal differences, such as body part detection. The pretrained transformer outperforms the regular transformer on fish species classification, indicating that pretraining captures meaningful embeddings that improve the performance of downstream classification tasks. Although traditional models, such as DT, show excellent performance in simpler tasks, such as fish species classification, their performance decreased significantly in more challenging tasks, highlighting the need for advanced feature extraction and representation learning that deep learning models provide. The overall strong performance across the board indicates that REIMS data provide rich, discriminative information, particularly for fish species classification. However, body part identification requires more sophisticated modeling approaches, where deep learning excels because of its capability to capture complex patterns and subtle signal deviations.

The application of explainable AI techniques, i.e., LIME and Grad-CAM, provided valuable insights into the decision-making processes of our models. These explanations revealed specific mass-to-charge ratios that strongly influence classifications, enhancing our understanding of the biochemical markers associated with different fish species and body parts. For instance, the LIME analysis of fish speciation highlighted distinct spectral regions that differentiate mackerel from hoki. The Grad-CAM analysis generated class activation maps that quantitatively highlighted the ion peaks in the m/z range at feature index 200–800 for fish species, which exhibited significance coefficients > 0.5 across all correctly classified

samples, confirming these spectral features as statistically significant biomarkers rather than artifacts of the classification algorithm.

This interpretability not only increases confidence in the predictions of the model but also opens up possibilities for new scientific insights into the biochemical composition of marine biomass, demonstrating that our approach can provide accurate classifications and meaningful, chemically relevant explanations for those classifications.

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

Although 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. These directions for future work include (1) developing a system for real-time REIMS data acquisition and analysis, allowing for immediate classification results in industrial settings, and (2) working with regulatory bodies to ensure that the developed methods meet or exceed current standards for marine biomass analysis and food safety monitoring.

Acknowledgements
This work was supported by the MBIE Fund on Research Program under contract C11X2001. The authors would like to thank Sue Marshall from Plant and Food Research for her leadership on this project.

Additional information
Edited by: Lin Gao.

Authors' contributions
Jesse Wood conceived and implemented the methodology, conducted experiments, analyzed results, and wrote the manuscript. Bach Nguyen, Bing Xue, Mengjie Zhang and Daniel Killeen provided supervision, proofreading, and feedback on the research and manuscript, and they contributed equally to this work. All authors read and approved the final manuscript.

Data availability
The data used in this study is not publicly available due to confidentiality requirements.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests. Mengjie Zhang and Daniel Killeen are the Editorial Board Members, but they were not involved in the journal's review of or decision related to this manuscript.

Received: 19 March 2025 Revised: 17 April 2025 Accepted: 18 May 2025
Published online: 12 June 2025

References

Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev-Comput Stat 2(4):433–459
Ba JL, Kiros JR, Hinton GE (2016) Layer normalization. Preprint at arXiv:1607.06450
Balakrishnama S, Ganapathiraju A (1998) Linear discriminant analysis-a brief tutorial. Inst Signal Inf Proc 18(1998):1–8
Balog J, Szaniszlo T, Schaefer KC, Denes J, Lopata A, Godorhazy L et al (2010) Identification of biological tissues by rapid evaporative ionization mass spectrometry. Anal Chem 82(17):7343–7350
Bettjeman BI, Hofman KA, Burgess EJ, Perry NB, Killeen DP (2018) Seafood phospholipids: extraction efficiency and phosphorous nuclear magnetic resonance spectroscopy (^{31}P NMR) profiles. J Am Oil Chem Soc 95(7):779–786. https://doi.org/10.1002/aocs.12086
Black C, Chevallier OP, Cooper KM, Haughey SA, Balog J, Takats Z et al (2019) Rapid detection and specific identification of offals within minced beef samples utilising ambient mass spectrometry. Sci Rep 9(1):1–9
Black C, Chevallier OP, Haughey SA, Balog J, Stead S, Pringle SD et al (2017) A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry. Metabolomics 13(12):1–13
Boccard J, Rutledge DN (2013) A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock omics data fusion. Anal Chim Acta 769:30–39
Breiman L (2017) Classification and regression trees. Routledge
Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemometr 20(8–10):341–351
Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at arXiv:1810.04805
FAO (2020) The state of world fisheries and aquaculture, 2020. FAO, Rome. https://doi.org/10.4060/ca9229en
Fix E, Hodges JL (1989) Discriminatory analysis. Nonparametric discrimination: consistency properties. Int Stat Rev 57(3):238–247. https://doi.org/10.2307/1403797
Ghaly A, Ramakrishnan V, Brooks M, Budge S, Dave D (2013) Fish processing wastes as a potential source of proteins. J Microb Biochem Technol 5(4):107–129. https://doi.org/10.4172/1948-5948.1000110
Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press. Accessed 2 Jan 2025. http://www.deeplearningbook.org
Gu A, Dao T (2023) Mamba: linear-time sequence modeling with selective state spaces. Preprint at arXiv:2312.00752
Hand DJ, Yu K (2001) Idiot's bayes-not so stupid after all? Int Stat Rev 69(3):385–398. https://doi.org/10.1111/j.1751-5823.2001.tb00465.x
Hansen LK, Salamon P (1990) Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 12(10):993–1001
He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
Hendrycks D, Gimpel K (2016) Gaussian error linear units (GELUs). Preprint at arXiv:1606.08415
Ho TK (1995) Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. IEEE, pp 278–282. https://doi.org/10.1109/ICDAR.1995.598994
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Jha SN (2015) Rapid detection of food adulterants and contaminants: theory and practice. Academic Press
Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. Preprint at arXiv:1412.6980
Kingma DP, Welling M (2013) Auto-encoding variational bayes. Preprint at arXiv:1312.6114
Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M (2002) Logistic regression. Springer, New York
Köppen M (2000) The curse of dimensionality. In: 5th Online World Conference on Soft Computing in Industrial Applications (WSC5). pp 4–8
Koza JR (1994) Genetic programming II, vol 17. MIT Press, Cambridge
LeCun Y (1989) Generalization and network design strategies. Connectionism Perspect 19:143–155

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W et al (1989a) Handwritten digit recognition with a back-propagation network. In: Touretzky D (ed) Advances in Neural Information Processing Systems. Morgan Kaufmann, pp 1–9

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W et al (1989b) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–51. https://doi.org/10.1162/neco.1989.1.4.541

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

Liu ZM, Wang YX, Vaidya S, Ruehle F, Halverson J, Soljačić M et al (2024) KAN: Kolmogorov-Arnold networks. Preprint at arXiv:2404.19756

Loshchilov I, Hutter F (2017) Decoupled weight decay regularization. Preprint at arXiv:1711.05101

McCann S, Lowe DG (2012) Local naive bayes nearest neighbor for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 3650–3656

Ministry for Primary Industries (2024) Hoki: New Zealand's largest fishery. https://www.mpi.govt.nz/fishing-aquaculture/fisheries-management/fish-stock-status/hoki-new-zealands-largest-fishery/. Accessed 6 Jan 2025.

Morgan N, Bourlard H (1989) Generalization and parameter estimation in feedforward nets: some experiments. In: Proceedings of the 3rd International Conference on Neural Information Processing Systems. MIT Press, pp 630–637

Panse ML, Phalke SD (2016) World market of omega-3 fatty acids. In: Hegde M et al (eds) Omega-3 fatty acids. Springer, Cham, pp 79–88. https://doi.org/10.1007/978-3-319-40458-5_7

Pardo MÁ, Jiménez E, Pérez-Villarreal B (2016) Misdescription incidents in seafood sector. Food Control 62:277–283

Pauly D, Zeller D (2016) Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. Nat Commun 7(1):10244. https://doi.org/10.1038/ncomms10244

Pearl H (2016) Melbourne restaurant hunky dory accused of serving catfish to customers instead of dory. In: Daily Mail Australia, May 2016. https://www.dailymail.co.uk/news/article-3611999/Melbourne-restaurant-Hunky-Dory-accused-serving-catfish-customers-instead-dory.html. Accessed 4 Jan 2025.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Plant and Food Research (2020) New research to maximise value from seafood resources - plant & food research. https://www.plantandfood.com/en-nz/article/new-research-to-maximise-value-from-seafood-resources. Accessed 2 Jan 2025.

Ribeiro MT, Singh S, Guestrin C (2016) "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. ACM, pp 1135–1144. https://doi.org/10.1145/2939672.2939778

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 618–626. https://doi.org/10.1109/ICCV.2017.74

Simopoulos AP (2011) Evolutionary aspects of diet: the omega-6/omega-3 ratio and the brain. Mol Neurobiol 44(2):203–215. https://doi.org/10.1007/s12035-010-8162-0

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tran B, Xue B, Zhang MJ (2016) Genetic programming for feature construction and selection in classification on high-dimensional data. Memet Comput 8(1):3–15

Tran B, Xue B, Zhang MJ (2019) Genetic programming for multiple-feature construction on high-dimensional classification. Pattern Recognit 93:404–417

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al (2017) Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017). pp 1–11

Wang J, Ji T, Wu YB, Yan H, Gui T, Zhang Q et al (2024) Length generalization of causal transformers without position encoding. Preprint at arXiv:2404.12224

Wood J, Nguyen BH, Xue B, Zhang MJ, Killeen D (2022) Automated fish classification using unprocessed fatty acid chromatographic data: a machine learning approach. In: Aziz H (eds) AI 2022: Advances in Artificial Intelligence. Lecture notes in computer science, vol 13728. Springer, Cham, pp 516–529. https://doi.org/10.1007/978-3-031-22695-3_36

Xiong RB, Yang YC, He D, Zheng K, Zheng SX, Xing C et al (2020) On layer normalization in the transformer architecture. In: Proceedings of the 37th International Conference on Machine Learning. PMLR, pp 10524–10533

Publisher's Note