A rapid machine-learning approach for detecting fish species and body parts using rapid evaporative ionisation mass spectrometry

Abstract. Marine biomass compositional analysis is traditionally time-consuming and requires domain expertise. Rapid evaporative ionisation mass spectrometry (REIMS) with automated machine learning may be able to accelerate this process without compromising quality. This study demonstrates REIMS's effectiveness in rapid marine biomass composition determination, using fish species and body parts as model systems that represent diverse biochemical profiles. Using decision trees, genetic programming and novel unsupervised pre-training strategies for transformers, the research achieved 99.58% accuracy in fish speciation and 63.33% accuracy for fish body parts classification, two model systems that capture marine biomass with different compositions. REIMS analysis with machine learning proves to be a fast, accurate, and interpretable technique for real-time marine biomass compositional analysis, with potential applications in marine-based industry quality control and product optimization.

Keywords: AI applications · explainable AI · machine learning · marine biomass · mass spectrometry · multidisciplinary AI

1 Introduction

Marine biomass analysis for compositional classification is traditionally time-consuming and requires domain expertise [4, 5, 9, 20, 37]. Rapid evaporative ionisation mass spectrometry (REIMS) with automated machine learning offers the potential to accelerate this process without compromising quality. Rapid evaporative ionisation mass spectrometry (REIMS) offers a promising alternative, providing high-resolution chemometric profiling in *real time*. This technology [4,5,20] is faster than slower genomic profiling [26], spectroscopy [3], and gas chromatography [35] methods. This study demonstrates REIMS's effectiveness in rapid marine biomass composition determination [26, 37], using fish species and body parts [35], as a model system representing diverse biomass compositions.

Our objective is to illustrate that REIMS can be used for rapid profiling of marine biomass compositions, such as high lipid, high protein, or high omega-3 content. We utilize two datasets derived from fish samples: (1) fish speciation and (2) fish body parts. These datasets serve as proxies for diverse marine biomass compositions, allowing us to test the capability of REIMS in distinguishing between different biochemical profiles.

This research opens avenues for efficient biomass composition analysis in marine-based industries, potentially supporting optimized processing, product

development, and resource utilization. By automating these analyses, our methods aim to maintain or exceed current quality standards while accelerating the throughput and efficiency of marine biomass compositional assessment.

The rapid evaporative ionisation mass spectrometry datasets introduce four challenges.

- **Time-consuming manual analysis**: Current REIMS methods [4, 5, 9] require domain expertise and manual tuning. This paper automates analysis for rapid inference in fish processing applications.
- **High-dimensionality**: The REIMS datasets have 1023 features. Deep learning [10, 34], evolutionary computation [23, 32, 33], and ensembles [15] are suitable for this high-dimensional data.
- **Few training instances**: Limited samples - 108 instances for fish speciation and 30 instances for fish body part classification - are addressed using unsupervised pre-training [10] and data augmentation [29].
- **Jargon**: Introduces machine learning terminology to the traditional chemistry and statistics-dominated field [4, 5, 20], promoting interdisciplinary collaboration.

2 Datasets

Having established the context and challenges of marine biomass analysis, we now turn to the specific datasets used in this study to address these issues.

(a) Mackerel (b) Hoki

Fig. 1: Mackerel (left) Hoki (right) fish species

This study leverages datasets from New Zealand Plant and Food Research, part of the Cyber-Marine project aimed at developing a flexible factory that uses rapid evaporative ionisation mass spectrometry (REIMS) for quality assurance in marine biomass analysis. The goal of Cyber-Marine's flexible factory is to maximise value from seafood resources. The datasets include samples analysed with REIMS using an electrosurgical knife, evaporated into an aerosol, and then that aerosol is directed into a mass spectrometer for ionisation and mass-to-charge ratio analysis. Each sample undergoes multiple 3-5 second incisions to generate chemometric models within the mass range m/z 77.04 - 474.1631,

facilitating rapid analysis to match with the operational speed of ambient mass spectrometry devices [20].

Using different fish tissues and species as a proxy for different fish compositions, the datasets comprise two main classification tasks: fish speciation and fish body parts classification. The fish speciation task includes 106 samples divided into Hoki and Mackerel classes, representing significant species in the New Zealand seafood industry [8]. The fish body parts classification involves 30 samples categorised into six classes: fillet, heads, livers, skins, guts, and frames, presenting a multi-class classification challenge. Both REMS datasets - fish speciation and fish body parts - are normalised between $x \in [0, 1]$ and stratified into training (80%), validation (10%), and test (10%) sets, with balanced class distributions to ensure robust model training and evaluation.

3 The Classification Approaches

With a clear understanding of the REIMS datasets, we can now explore the various machine learning classifiers employed to analyse this complex spectral data. Classification in machine learning involves categorising input data into predefined classes based on their features. The goal is to learn a mapping from these features to class labels, allowing accurate predictions for new instances.

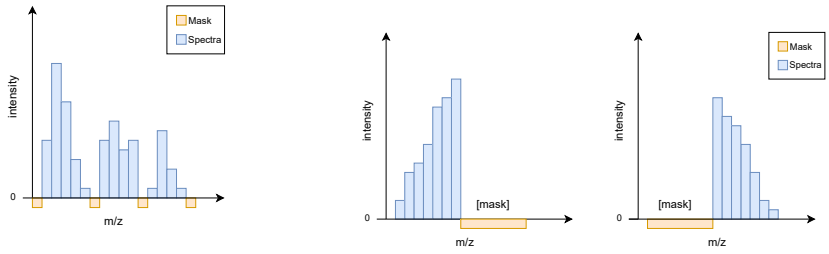
3.1 Commonly Used Classification Algorithms

In this paper, we apply classification algorithms to marine biomass analysis for the REIMS datasets. We apply seven classifiers to the tasks of fish speciation and body parts classification: random forest (RF) [19], k-nearest neighbour (KNN) [11], decision trees (DT) [6], naive Bayes (NB) [14], logistic regression (LR) [22], support vector machines (SVM) [7], linear discriminant analysis (LDA) [2]. With an ensemble voting classifier [15] combining them all. They use default settings from sklearn [27], except SVM with a linear kernel and LR set to 2,000 max iterations, these exceptions were found experimentally with trial and error. The ensemble voting classifier uses hard voting. Additionally, more advanced classification methods of transformers and genetic programming are detailed below.

3.2 Transformer

While traditional classifiers offer a baseline for performance, we now delve into the application of transformers, a more advanced deep learning approach that has shown promise in handling high-dimensional data with complex spectral patterns.

This paper introduces two novel unsupervised pre-training methods for mass spectrometry data, illustrated in fig. 2, which were inspired by techniques used in the BERT model [10]. These methods allow the model to learn general patterns from larger, unlabeled mass spectrometry datasets, creating useful embeddings



(a) MSM (b) NSP

Fig. 2: Masked spectra modelling (left) Next spectra prediction (right)

that improve performance on smaller, fine-tuned datasets for specific downstream tasks.

1. **Masked Spectra Modelling (MSM)**: adapts masked language modelling to mass spectrometry. MSM randomly masks mass-to-charge ratios in spectra and predicts the missing values, evaluated as a regression task using mean square error (MSE) as the loss function.
2. **Next Spectra Prediction (NSP)**: is inspired by next sentence prediction, NSP splits spectra in half and predicts whether two halves belong to the same spectrum, evaluated as a pair-wise comparison using categorical cross entropy (CCE) as the loss function.

This paper adapts the transformer architecture, introduced by Vaswani et al. [34] for machine translation, to marine biomass analysis. The architecture features stacked encoder-decoder layers with residual connections [17]. The study explores various weight initialization methods - Xavier [12], Kaiming [16], and orthogonal [28]) - and compares pre-norm [36] and post-norm [1] layer normalization approaches in the context of marine biomass analysis.

Table 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: NSP & Speciation	CCE
Input dimensions	1023	Hidden dimensions	128
Output dimensions: MSM	1023	Output dimensions: NSP & Speciation	2
Output dimensions: Part	6	Number of layers	3
Number of heads	3		

Table 1 outlines the transformer configuration used. The transformer model in this paper employed the AdamW optimiser [24], this improves the Adam [21]

by decoupling weight decay from the learning rate. Dropout [30] approximates a bagged ensemble of neural networks efficiently. For regularisation, label smoothing [31] softens class label targets by combining one-hot encodings with a uniform distribution. The transformer network utilises the Gaussian error linear unit (GELU) activations [18]. Data augmentation inflates the number of training instances, it duplicates each instance five times and injects noise [29], effectively expanding the training set five-fold. Early stopping [25] saves model parameters whenever validation loss improves, effectively tuning the hyperparameter of epochs [13].

3.3 Mutli-tree Genetic Programming

Complementing the deep learning approach of transformers, we next examine the use of multi-tree genetic programming, which offers a different paradigm for feature construction and classification. We employ multiple class independent feature construction (MCIFC) [33] in a novel application for marine biomass analysis using REIMS. Algorithm 1 gives the pseudocode for MCIFC.

Algorithm 1 GP-based multiple feature construction

Input : train_set, m;
Output : Best set of m trees;
Initialise a population of GP individuals. Each individual is an array of m trees;
best_inds ← the best e individuals;
while Maximum generation is not reached **do**
  for $i = 1$ to Population Size **do**
    $transf\_train$ ← Calculate constructed features of individual i on train_set;
    $fitness$ ← Apply fitness function on $transf\_train$;
    Update best_inds the best e individuals from elitism and offspring combined;
  end for
  Select parent individuals using tournament selection for breeding;
  Create new individuals from selected parents using crossover or mutation;
  Place new individuals into population for next generation;
end while
Return best individual in best_inds;

MCIFC represents candidate solutions as multiple trees, with one subtree per class (construction ratio of 1). This approach serves both feature construction and classification, using a winner-takes-all strategy for class prediction. MCIFC uses a two-tree representation when applied to the fish species classification task since there are two classes, i.e. Hoki and Mackerel. Additionally, MCIFC uses a six-tree representation when applied to the fish body parts classification task since there are six classes, i.e. fillet, heads, livers, skins, guts, and frames.

The genetic operators in MCIFC are crossover (80% probability) and mutation (20% probability) - both adapted from conventional genetic programming - and Tournament selection with a size of 7 is used for parent selection. Crossover

operates only between trees of the same class, while mutation randomly alters one subtree.

The fitness evaluation combines accuracy with a distance regularisation term:

$$\alpha \cdot \text{balanced\_accuracy} + (1 - \alpha)(\beta \cdot \text{inter} + (1 - \beta) \cdot (1 - \text{intra}))$$

where $\alpha = 0.8$ - prioritising accuracy and $\beta = 0.5$ - balancing inter and intra-class distances. The interclass and intraclass distances are calculated as:

$$\text{inter} = \frac{1}{|S|} \sum_{i,j} d(i,j) \quad \forall \quad i \neq j \quad \text{and} \quad \text{class}(i) \neq \text{class}(j)$$

$$\text{intra} = \frac{1}{|S|} \sum_{i,j} d(i,j) \quad \forall \quad i \neq j \quad \text{and} \quad \text{class}(i) = \text{class}(j)$$

where $d(i,j)$ is the Euclidean distance between points i and j, and $|S|$ is the total number of pairs. This approach aims to maximise interclass distances while minimising intraclass distances, promoting better class separation in the constructed feature space.

Table 2 outlines the parameter settings of the MCIFC method. The construction ratio is the number of trees per class.

Table 2: MCFIC parameter settings

Function Set	$+, -, \times, \cos, \sin, \tan, -1*$	Terminal Set	$x_1, x_2, ..., x_n$
Maximum Tree Depth	6	Population size	1 * 1023 (= 1× #features)
Initial Population	Ramped Half-and-Half	Generations	400
Crossover	0.8	Mutation	0.2
Elitism	0.1	Selection	Tournament
Tournament Size	7	Construction ratio	1
Fitness weighting α	0.8	Distance weighting β	0.5

4 Results and Discussions

With our methodologies established, we now present the results of applying these various machine learning techniques to the REIMS datasets and elaborate in the discussion related to those results.

4.1 Classification Accuracy

Table 3 gives the results of the classifiers, giving the average over 30 independent runs, with the best-performing model on the test set given in **bold**.

Table 3 reveals insights into model performance on REIMS datasets for fish speciation and part classification. For speciation, the transformer model excelled

Table 3: Fish speciation and fish part classification results

	Fish speciation		Fish part	
Method	Train	Test	Train	Test
RF	100.0% ± 0.00%	95.88% ± 4.47%	100.0% ± 0.00%	40.00% ± 15.27%
KNN	93.24% ± 2.43%	83.69% ± 06.91%	42.88% ± 5.37%	31.66% ± 14.49%
DT	100.0% ± 0.00%	99.13% ± 1.72%	100.0% ± 0.00%	27.22% ± 13.25%
NB	93.40% ± 6.99%	87.97% ± 9.57%	100.0% ± 0.00%	45.00% ± 15.60%
LR	100.0% ± 0.00%	96.72% ± 4.75%	100.0% ± 0.00%	56.66% ± 15.27%
SVM	100.00 ± 0.00%	95.97% ± 5.06%	100.0% ± 0.00%	56.11% ± 14.58%
LDA	98.67% ± 0.77%	96.47% ± 3.67%	75.61% ± 3.20%	45.55% ± 16.06%
Ensemble	100.0% ± 0.00%	98.16% ± 3.00%	100.0% ± 0.00%	51.66% ± 15.72%
Transformer	**100.0% ± 0.00%**	**99.58% ± 1.31%**	**100.0% ± 0.0000**	**63.33% ± 24.59%**
MCIFC	99.97% ± 0.15%	94.72% ± 10.25%	97.93% ± 1.59%	55.83% ± 18.97%

with 99.58% test accuracy, leveraging its attention mechanism to capture complex relationships between mass-to-charge ratios. The transformer's attention mechanism in REIMS analysis allows each mass-to-charge ratio to interact with all others, learning which combinations best identify specific molecular structures. This enables recognition of complex patterns characteristic of fish species or body parts by weighting m/z ratio interactions. The mechanism's adaptability focuses on the most relevant spectral features, potentially uncovering subtle patterns missed by simpler models or human experts. The decision tree followed closely at 99.13% test accuracy, suggesting clear hierarchical decision boundaries. The ensemble model achieved 98.16% test accuracy, benefiting from diverse model aggregation. MCIFC showed signs of overfitting with 94.72% accuracy, failing to generalize to unseen data.

Fish part classification proved more challenging, with overall lower accuracies indicating subtler distinguishing features. The transformer led with 63.33% test accuracy, still outperforming other models despite the task's complexity. Logistic regression and SVM followed at 56.66% and 56.11% test accuracy, respectively. Their simpler structures may have helped avoid overfitting on this complex task with limited data. Surprisingly, complex models like random forest and decision tree underperformed, likely due to overfitting. This performance pattern suggests that for this task, the ability to find linear separations was more valuable than capturing intricate non-linear relationships. Furthermore, this was highlighted by the linear kernel being the best kernel for the SVM.

4.2 Further Analysis on Transformer

Building on these initial results, we conduct additional analyses to gain deeper insights into the performance and characteristics of our models. Sensitivity analysis explored various transformer architectures, including weight initialization strategies, layer normalization techniques, and label smoothing regularization. We compared these architectures on the fish speciation task, chosen due to the transformer's near-perfect classification accuracy on the test set as shown in earlier results.

The default weight initialization with pre-training was most effective, with alternative strategies - Xavier, Kaiming, and orthogonal - showing no improvement. Ablation studies on the fish speciation dataset revealed that Pre-LN transformers outperformed Post-LN variants, achieving higher accuracy (99.16% ± 1.66%) and faster convergence (15 epochs vs. 50+ epochs). Label smoothing further improved accuracy to 99.58% ± 1.31%, compared to 99.16% ± 1.66% without it.

4.3 Analysis on Interpretability

While performance metrics provide quantitative measures of success, we now turn our attention to the interpretability of our models, a crucial aspect for practical application in marine biomass analysis.

Decision Tree To begin our exploration of model interpretability, we first examine the structure and decision-making process of the decision tree classifier, which offers clear insights into the key features driving classification.

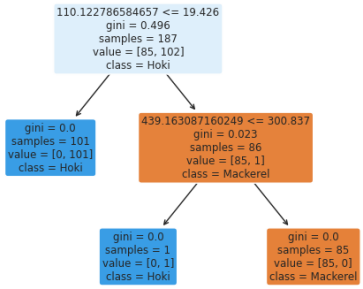
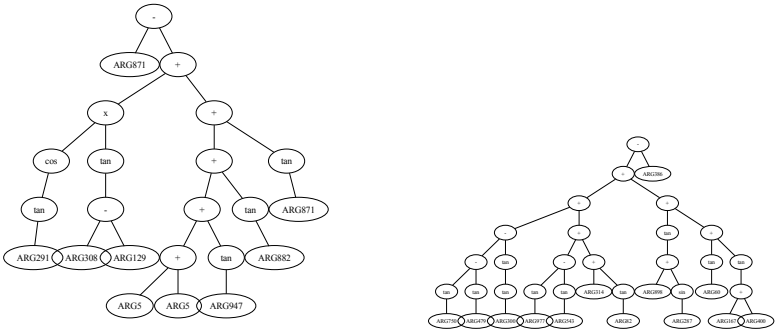


Fig. 3: Fish speciation: decision tree

Figure 3 gives the decision tree. This tree splits data when key mass-to-charge ratios exceed a threshold. For when 110.1228 m/z exceeds intensity 19.426, and later when 439.1631 m/z exceeds intensity 300.837. The intensity threshold for 439.1631 m/z is much greater than 110.1228 m/z, suggesting a large abundance of that molecule in Mackerel.

Assigning compounds to high-resolution mass spectrometry (HRMS) is challenging. This is due to the enormous amounts of metabolites present in homogenised fish tissues. The feature at 110.1128 m/z is consistent with a diphenol group. The feature 439.1631 m/z is consistent with a fragmented phospholipid (1-Lauroyl-2-hydroxy-sn-glycero-3-phosphocholine). These compounds are known to vary between fish species and tissues [3]. This means there can be a dramatic variation in the molecules such as the lipid profiles, between samples of the same species or body part.

Genetic Programming: Trees Having examined the interpretability of decision trees, we now turn our attention to the genetic programming approach, which provides a different perspective on feature importance and classification logic.



(a) Mackerel (b) Hoki

Fig. 4: Hoki (left) Mackerel (right) genetic programming tree

For Mackerel a genetic programming (GP) tree with seven features in its terminal set achieves perfect training accuracy and 95% test accuracy. Notably feature 5, 81.0893 m/z, is included twice. This suggests this is an important feature highly correlated with Mackerel species prediction. More features are needed for Hoki than Mackerel, with thirteen features selected as terminal nodes, suggesting more fragment ions are present in Hoki than in Mackerel. Perhaps there is a larger abundance of molecules in Hoki than in Mackerel.

4.4 Further Discussion

Having examined both the performance and interpretability of our models, we now discuss the broader implications and potential applications of this research. Transformer models excel in distinguishing between REIMS datasets, demonstrating deep learning's ability to discern complex spectral patterns associated with different biochemical profiles. Decision trees and genetic programming offer interpretable insights into key mass-to-charge ratios driving classifications, while challenges in some datasets highlight the need for further refinement. Rapid biomass composition analysis can discriminate between biomass with different compositions, enabling optimal processing decisions to be made for every batch of marine biomass, whether that be whole fish, mixed species, or various mixtures of frames, heads and whole fish. While not the focus of this study, the same approach may be useful for discriminating different species of fish i.e. to help combat the nefarious practice of mislabelling fish [26] species/adulteration [4, 5, 20] of high-value fish e.g. Hoki and Cod, with lower-grade fish e.g. Rig.

5 Conclusions and Future Work

Drawing from our findings and discussion, we can now summarise the key outcomes and significance of this study, and avenues for future research.

5.1 Conclusions

This study showcases the effectiveness of machine learning techniques, particularly transformers, genetic programming, and decision trees applied to REIMS data for marine biomass compositional analysis. Key findings include:

- Transformers achieving high accuracy in distinguishing between different biomass compositions (99.58% in the fish speciation dataset and 63.33% in the body part dataset), demonstrating their potential for automated classification of diverse marine biomass compositions.
- Genetic programming and decision trees offered interpretable insights into mass spectrometry data, highlighting important mass-to-charge ratios linked to different biomass compositions.
- These results suggest that REIMS, coupled with advanced machine learning techniques, has significant potential for rapid and accurate classification of marine biomass based on compositional differences.

5.2 Future Work

Future work will focus on (1) enhancing model interpretability through advanced AI explainability techniques tailored for mass spectrometry data, such as SHAP values, LIME, and attention visualization; (2) validating model outputs through collaboration with marine biology, biochemistry, and food science experts, correlating REIMS-based classifications with traditional compositional analysis methods; Additionally, we aim to expand classifier capabilities to (3) detect oil and cross-species contamination; and (4) directly identify compositional attributes such as lipid, protein, and omega-3 fatty acid content.

References

1. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
2. Balakrishnama, S., Ganapathiraju, A.: Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing **18**(1998), 1–8 (1998)
3. Bettjeman, B.I., Hofman, K.A., Burgess, E.J., Perry, N.B., 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)
4. Black, C., Chevallier, O.P., Cooper, K.M., Haughey, S.A., Balog, J., Takats, Z., Elliott, C.T., Cavin, C.: Rapid detection and specific identification of offals within minced beef samples utilising ambient mass spectrometry. Scientific reports **9**(1), 1–9 (2019)

5. 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**(12), 1–13 (2017)
6. Breiman, L.: Classification and regression trees. Routledge (2017)
7. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)
8. Czechowski, P., de Lange, M., Heldsinger, M., Kardailsky, A., Rayment, W., Hepburn, C., Ladds, M., Knapp, M.: Comparison of traditional and molecular surveys of fish biodiversity in southern te wāhipounamu/fiordland (aotearoa/new zealand). Environmental DNA **6**(1), e514 (2024)
9. De Graeve, M., Birse, N., Hong, Y., Elliott, C.T., Hemeryck, L.Y., 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)
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
11. Fix, E., Hodges, J.L.: Discriminatory analysis. nonparametric discrimination: Consistency properties. International Statistical Review/Revue Internationale de Statistique **57**(3), 238–247 (1989)
12. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
13. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016), `http://www.deeplearningbook.org`
14. Hand, D.J., Yu, K.: Idiot's bayes—not so stupid after all? International statistical review **69**(3), 385–398 (2001)
15. Hansen, L.K., Salamon, P.: Neural network ensembles. IEEE transactions on pattern analysis and machine intelligence **12**(10), 993–1001 (1990)
16. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
18. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
19. Ho, T.K.: Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. vol. 1, pp. 278–282. IEEE (1995)
20. Jha, S.N.: Rapid detection of food adulterants and contaminants: theory and practice. Academic Press (2015)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: Logistic regression. Springer (2002)
23. Koza, J.R., et al.: Genetic programming II, vol. 17. MIT press Cambridge (1994)
24. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

25. Morgan, N., Bourlard, H.: Generalization and parameter estimation in feedforward nets: Some experiments. Advances in neural information processing systems **2** (1989)
26. Pardo, M.Á., Jiménez, E., Pérez-Villarreal, B.: Misdescription incidents in seafood sector. Food Control **62**, 277–283 (2016)
27. 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., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
28. Saxe, A.M., McClelland, J.L., Ganguli, S.: Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv preprint arXiv:1312.6120 (2013)
29. Sietsma, J., Dow, R.J.: Creating artificial neural networks that generalize. Neural networks **4**(1), 67–79 (1991)
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
31. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., 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)
32. Tran, B., Xue, B., Zhang, M.: Genetic programming for feature construction and selection in classification on high-dimensional data. Memetic Computing **8**(1), 3–15 (2016)
33. Tran, B., Xue, B., Zhang, M.: Genetic programming for multiple-feature construction on high-dimensional classification. Pattern Recognition **93**, 404–417 (2019)
34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
35. Wood, J., Nguyen, B.H., Xue, B., Zhang, M., 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)
36. Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L., Liu, T.: On layer normalization in the transformer architecture. In: International Conference on Machine Learning. pp. 10524–10533. PMLR (2020)
37. Zhang, L., Agrawala, M.: Adding conditional control to text-to-image diffusion models. arXiv preprint arXiv:2302.05543 (2023)