A rapid machine-learning approach for detecting fish species and body parts

using rapid evaporative ionisation mass spectrometry

Jesse Wood[1] Bach Hoai Nguyen[1] Bing Xue[1] Mengjie Zhang[1] Daniel Killeen[2]

[1]School of Engineering and Computer Science — Te Kura Mātai Pūkaha, Pūrorohiko
Victoria University of Wellington — Te Herenga Waka
[2]New Zealand Institute for Plant and Food Research Limited, Nelson, New Zealand

Island Bay, Wellington, New Zealand

GP [1, 2, 3] inspired by reproductive behaviour of animals

ECRG, VUW, New Zealand

Topics

1. Catfishing
2. Fish Oil
3. Mass Spectrometry
4. Classification
5. Transformer
6. Intepretable

Have you been catfished? [4]

Daily Mail AUSTRALIA

Home | U.K. | U.S. | News | World News | Sport | TV&Showbiz | Femail AU | Health | Scien

Monkeypox | Covid-19 | Russia-Ukraine War | Breaking News | Sydney | Melbourne | Brisbane | New Zealand |

Popular restaurant accused of serving cheap Vietnamese catfish to customers who thought they were getting Australian dory

- **A Melbourne restaurant has been accused of serving catfish to customers**
- **Hunky Dory has allegedly been selling frozen fillets of basa as dory**
- **Owner Greg Robotis has denied allegations he is misleading customers**
- **The City of Port Phillip is investigating Hunky Dory's Port Melbourne store**

By HARRY PEARL FOR DAILY MAIL AUSTRALIA
PUBLISHED: 14:31 AEDT, 27 May 2016 | **UPDATED:** 16:08 AEDT, 27 May 2016

Share 47 shares 9 View comments

A Melbourne restaurant has been accused of serving a Vietnamese catfish to customers who believe they are ordering Dory.

A whistleblower has alleged that Hunky Dory outlets have been selling frozen fillets of basa, a species of catfish native to the Mekong basin, as fish-of-the-day dory, **The Age** reports.

Owner Greg Robotis has denied the claims and said inexperienced staff may have been calling the fish the wrong name.

Aussies! No surprises there...

Catfishing [4], Mislabelling [5], and Quality Assurance [6]

Nutrition Facts

6 servings per container

Serving size **4-5 ounces(187g)**

Amount per serving

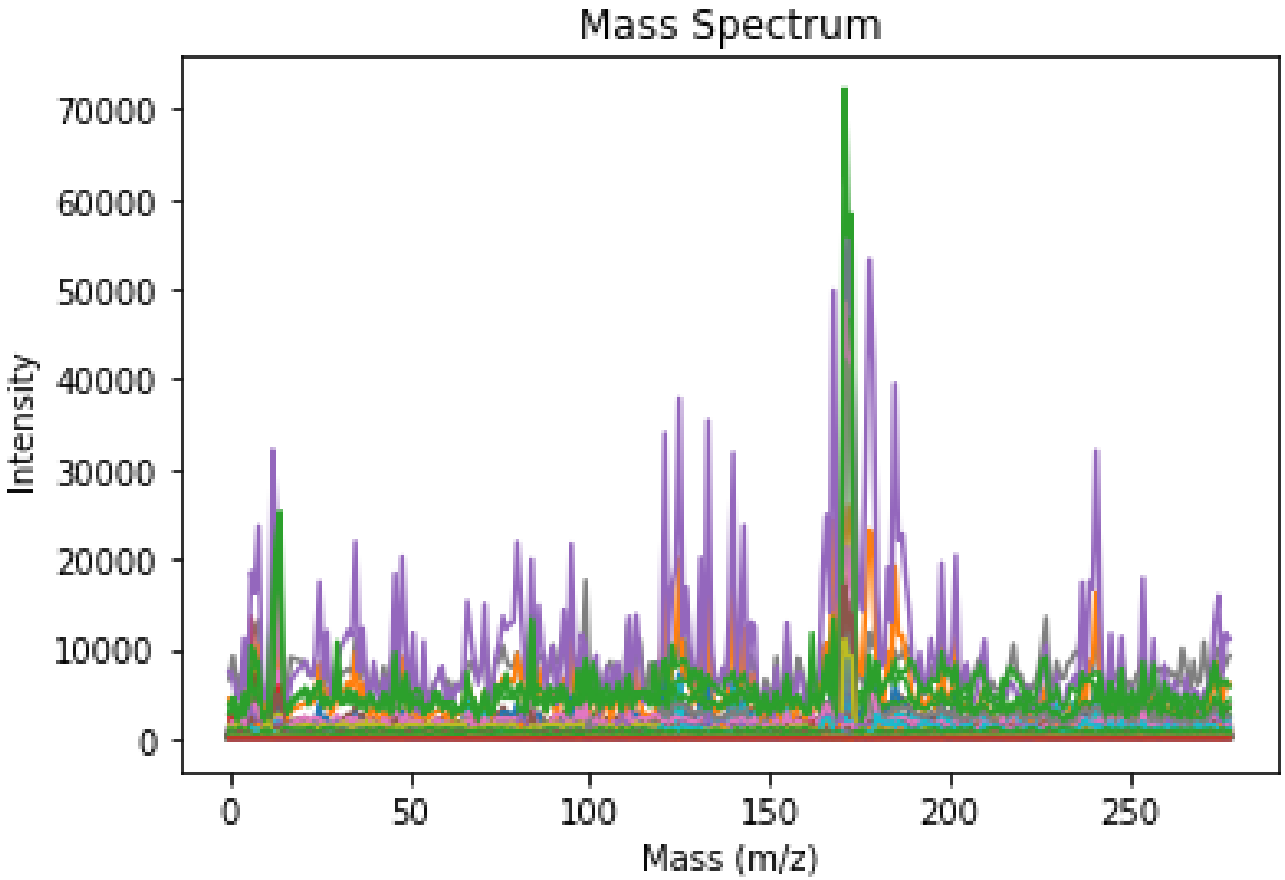
Calories **200**

	% Daily Value*
Total Fat 5g	**6%**
Saturated Fat 0.5g	**3%**
Trans Fat 0g	
Cholesterol 80mg	**27%**
Sodium 610mg	**27%**
Total Carbohydrate 10g	**4%**
Dietary Fiber 0g	**0%**
Total Sugars 3g	
Includes 0g Added Sugars	**0%**
Protein 27g	
Vitamin D 2mcg	10%
Calcium 79mg	6%
Iron 3mg	15%
Potassium 519mg	10%

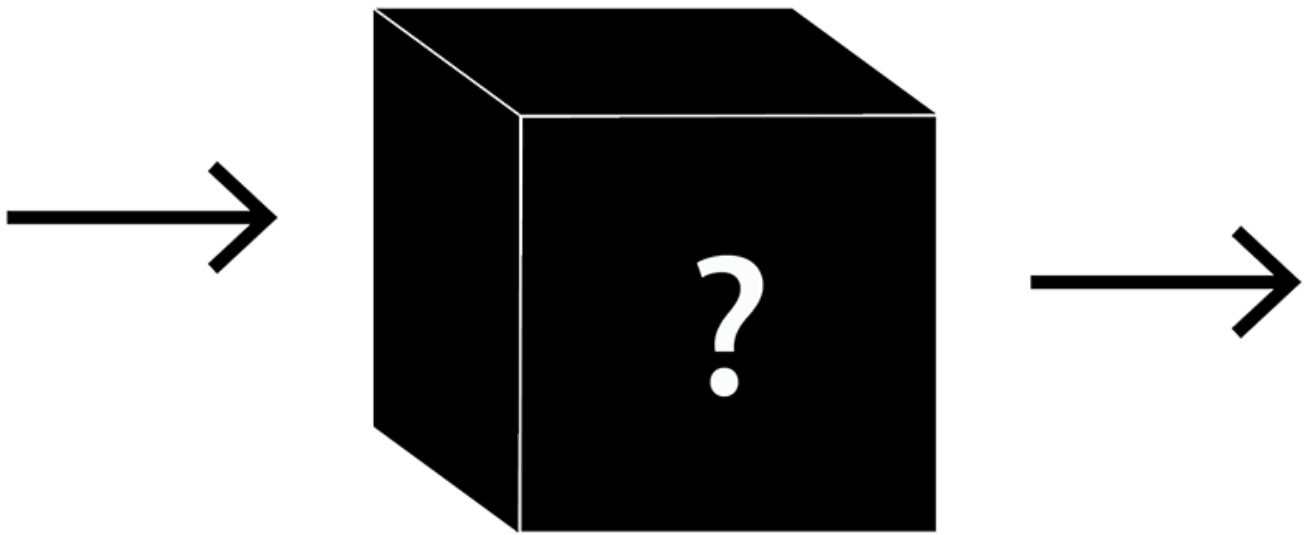
*The % Daily Value tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

Fish oil is brain food! [7, 8]

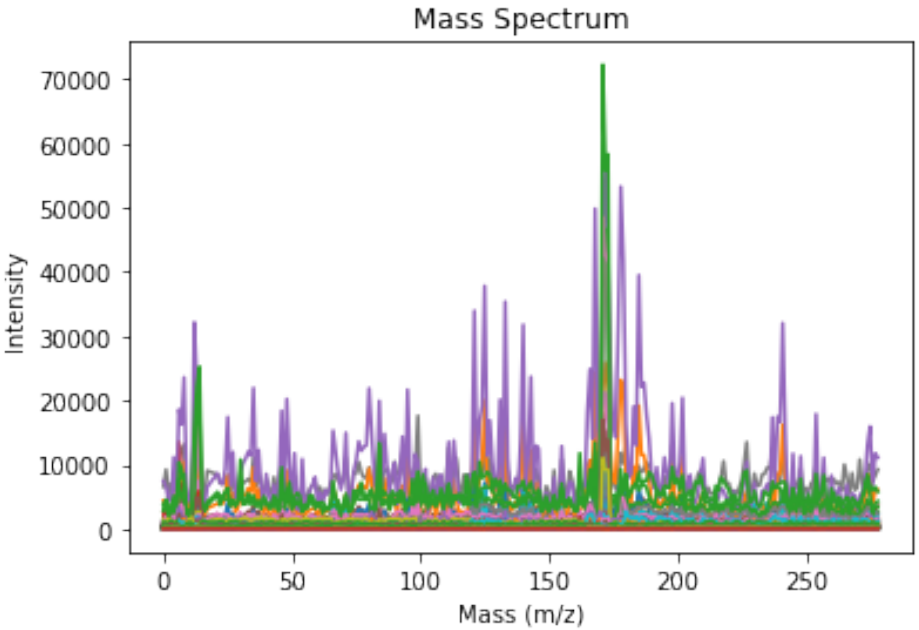
Fish oil analyzed with Mass Spectrometry! [6]



Fish oil analysis can't be blackbox! [9, 10]

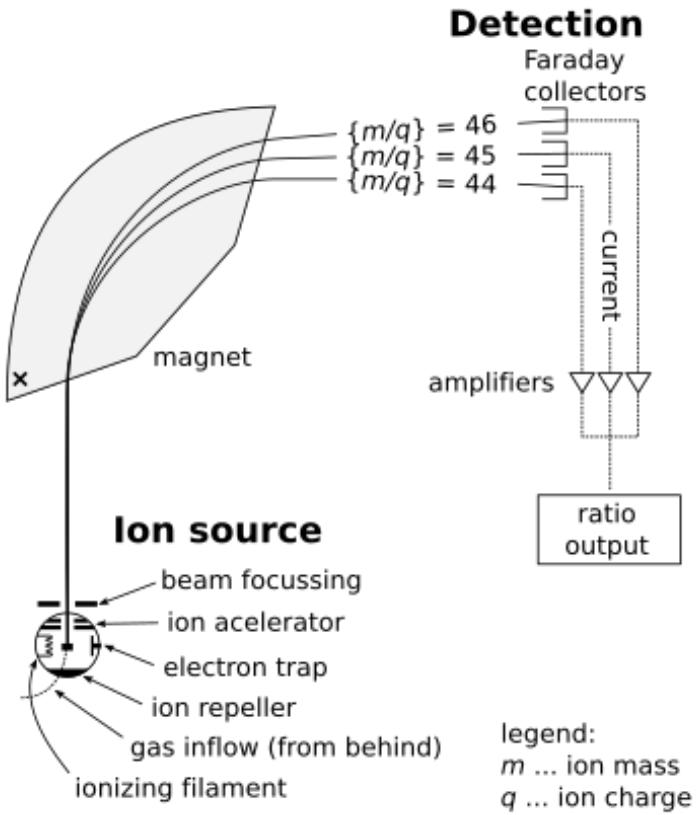


Mass Spectrometry [11, 6, 12] ≈ Chemical Fingerprint



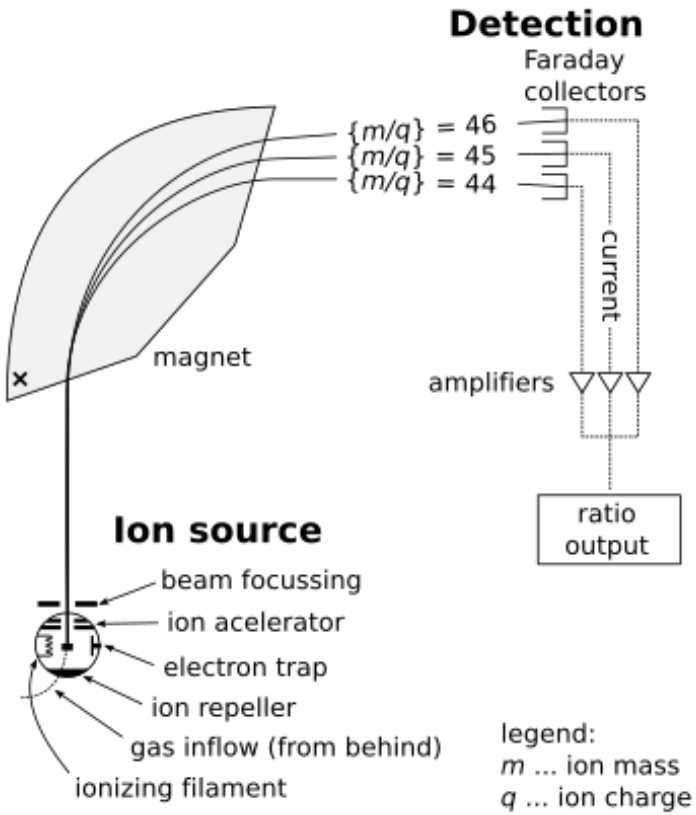
Mass Spectrometry: Steps

1. Laser Pen
2. Vacuum
3. Electromagnetic Field (EMF)
4. Detector



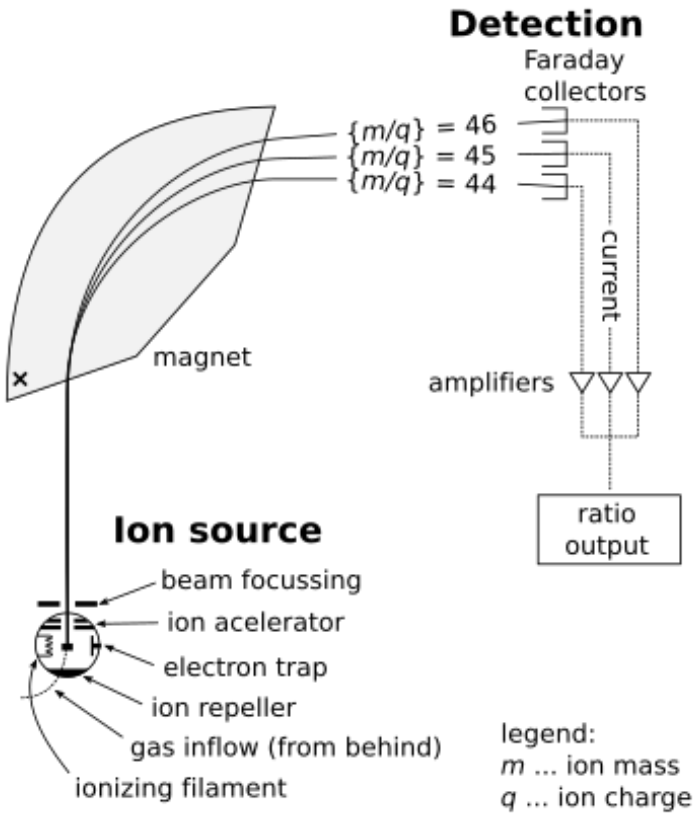
Mass Spectrometry: Steps

1. Laser Pen
2. Vacuum
3. Electromagnetic Field (EMF)
4. Detector



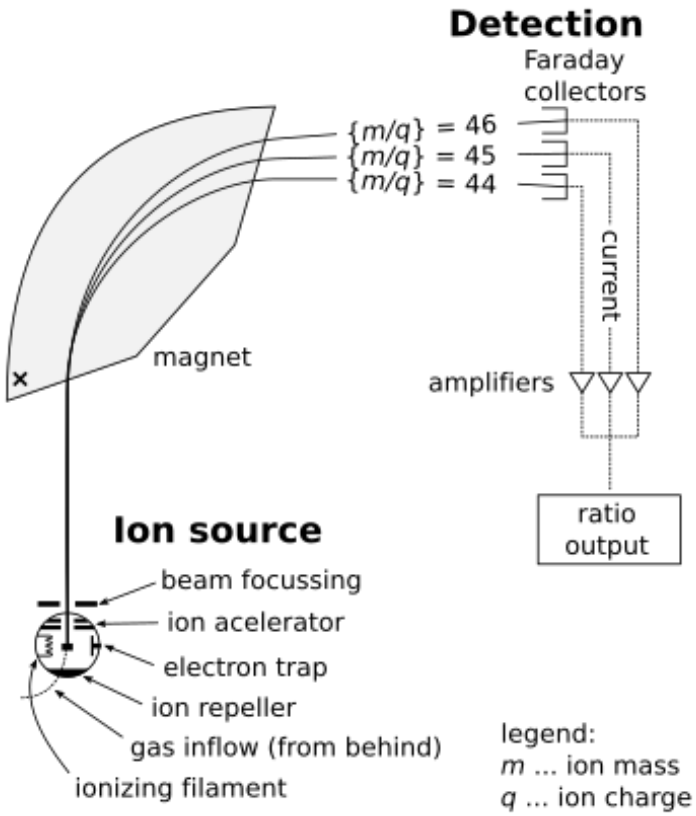
Mass Spectrometry: Steps

1. Laser Pen
2. Vacuum
3. Electromagnetic Field (EMF)
4. Detector



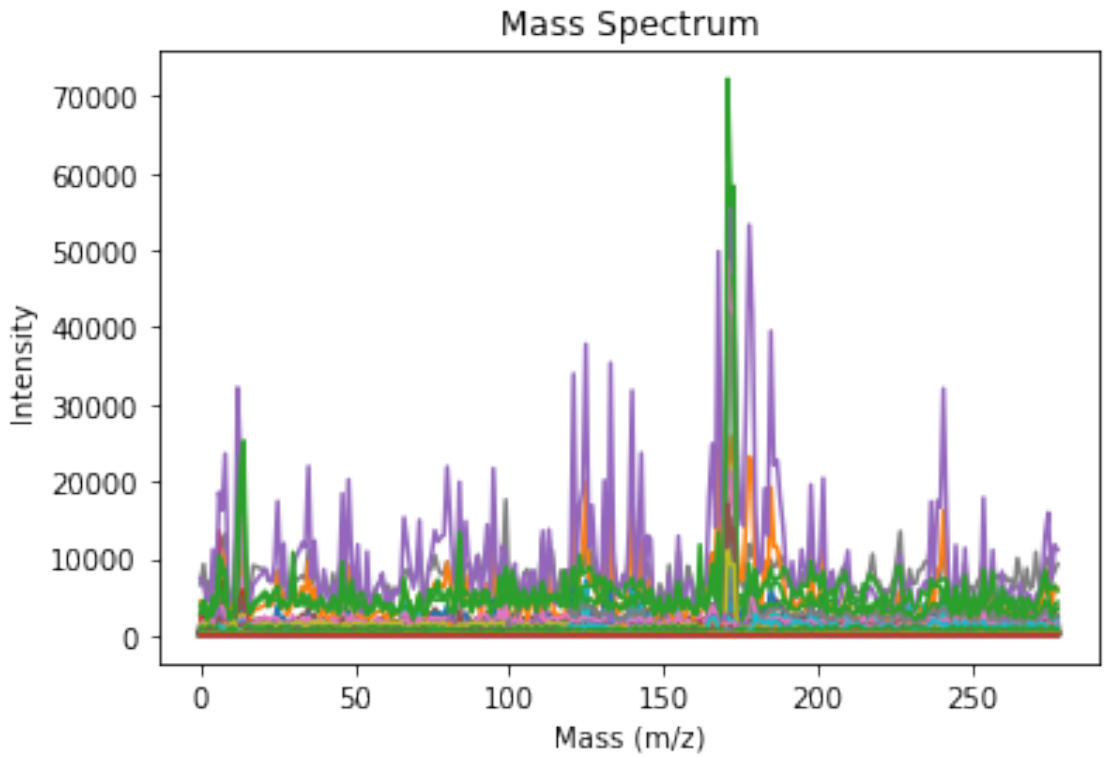
Mass Spectrometry: Steps

1. Laser Pen
2. Vacuum
3. Electromagnetic Field (EMF)
4. Detector



Mass Spectrometry: Steps

1. Laser Pen
2. Vacuum
3. Electromagnetic Field (EMF)
4. Detector



Classification: Datasets

Dataset

Species 

Parts 

Classification: Methods

Dataset	**Method**
Species  Parts 	RF [13]
	KNN [14]
	DT [15]
	NB [16]
	LR [17]
	SVM [18]
	LDA [19]
	Ensemble [20]
	Transformer [21, 22]
	MCIFC [2, 3]

Classification: Fish Species

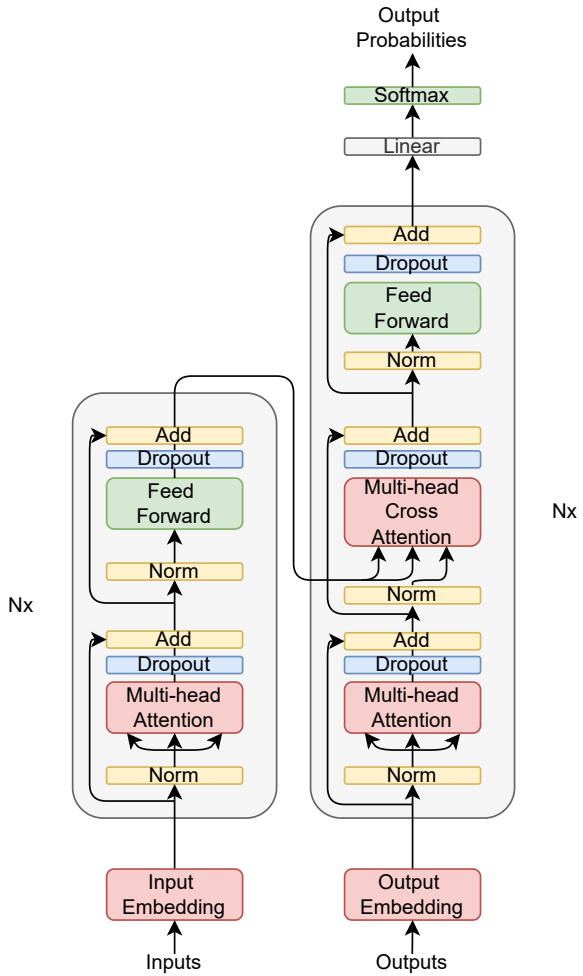
Dataset	Method	**Train**	**Test**
Species 	RF [13]	100.0% ± 0.00%	95.88% ± 4.47%
	KNN [14]	93.24% ± 2.43%	83.69% ± 6.91%
	DT [15]	*100.0% ± 0.00%*	*99.13% ± 1.72%*
	NB [16]	100.0% ± 0.00%	87.97% ± 9.57%
	LR [17]	100.0% ± 0.00%	96.72% ± 4.75%
	SVM [18]	100.0% ± 0.00%	95.97% ± 5.06%
	LDA [19]	98.67% ± 0.77%	96.47% ± 3.67%
	Ensemble [20]	100.0% ± 0.00%	98.16% ± 3.00%
	Transformer [21, 22]	**100.0% ± 0.00%**	**99.58% ± 1.31%**
	MCIFC [2, 3]	99.97% ± 0.15%	94.72% ± 10.25%

Classification: Fish Body Parts

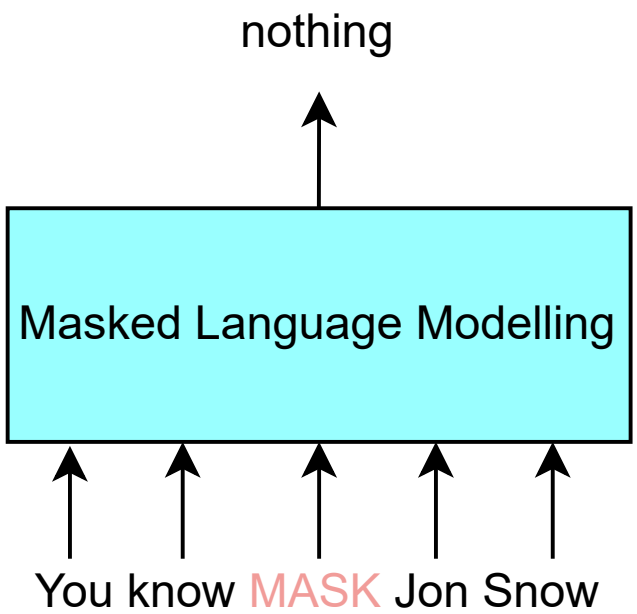
Dataset	Method	**Train**	**Test**
Parts 	RF [13]	100.0% ± 0.00%	40.00% ± 15.27%
	KNN [14]	42.88% ± 5.37%	31.66% ± 14.49%
	DT [15]	100.0% ± 0.00%	27.22% ± 13.25%
	NB [16]	100.0% ± 0.00%	45.00% ± 15.60%
	LR [17]	*100.0% ± 0.00%*	*56.66% ± 15.27%*
	SVM [18]	100.0%± 0.00%	56.11% ± 14.58%
	LDA [19]	75.61% ± 3.20%	45.55% ± 16.06%
	Ensemble [20]	100.0% ± 0.00%	51.66% ± 15.72%
	Transformer [21, 22]	**100.0% ± 0.00%**	**63.33% ± 24.59%**
	MCIFC [2, 3]	97.93% ± 1.59%	55.83% ± 18.97%

Classification: Avoid Catfishing [4] & Mislabelling [5]

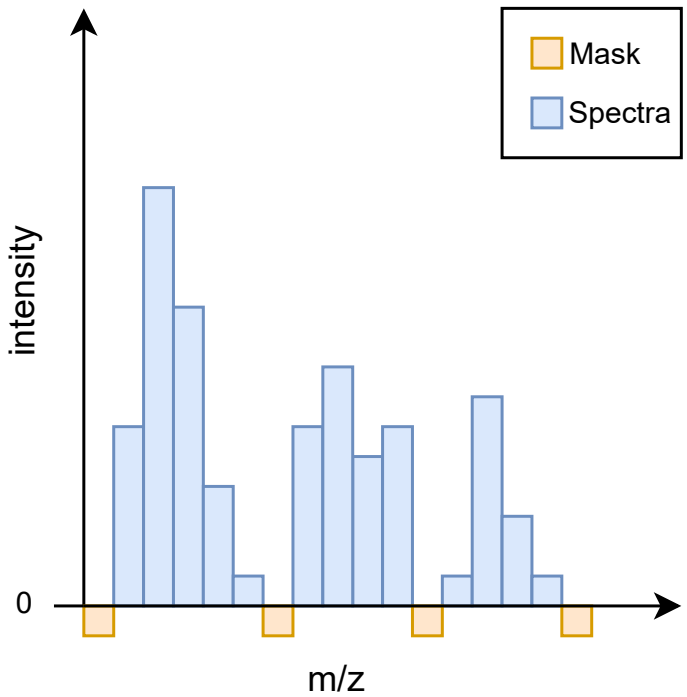
Transformer Architecture [21]



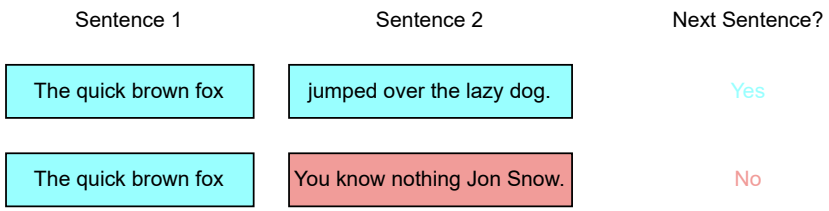
Pre-Training: Masked Spectra Modelling [22]



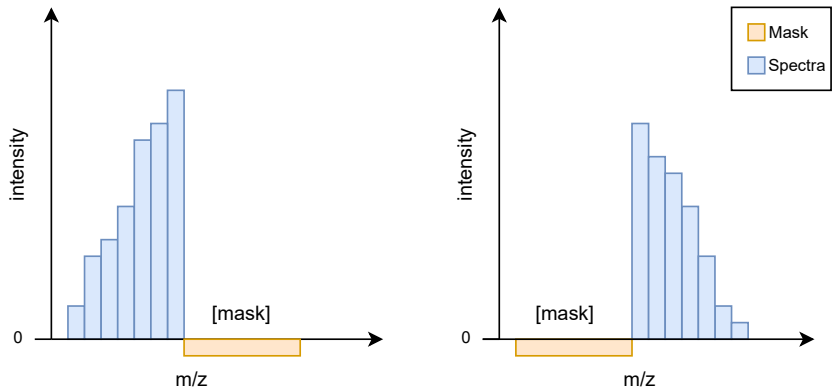
Pre-Training: Masked Spectra Modelling [22]



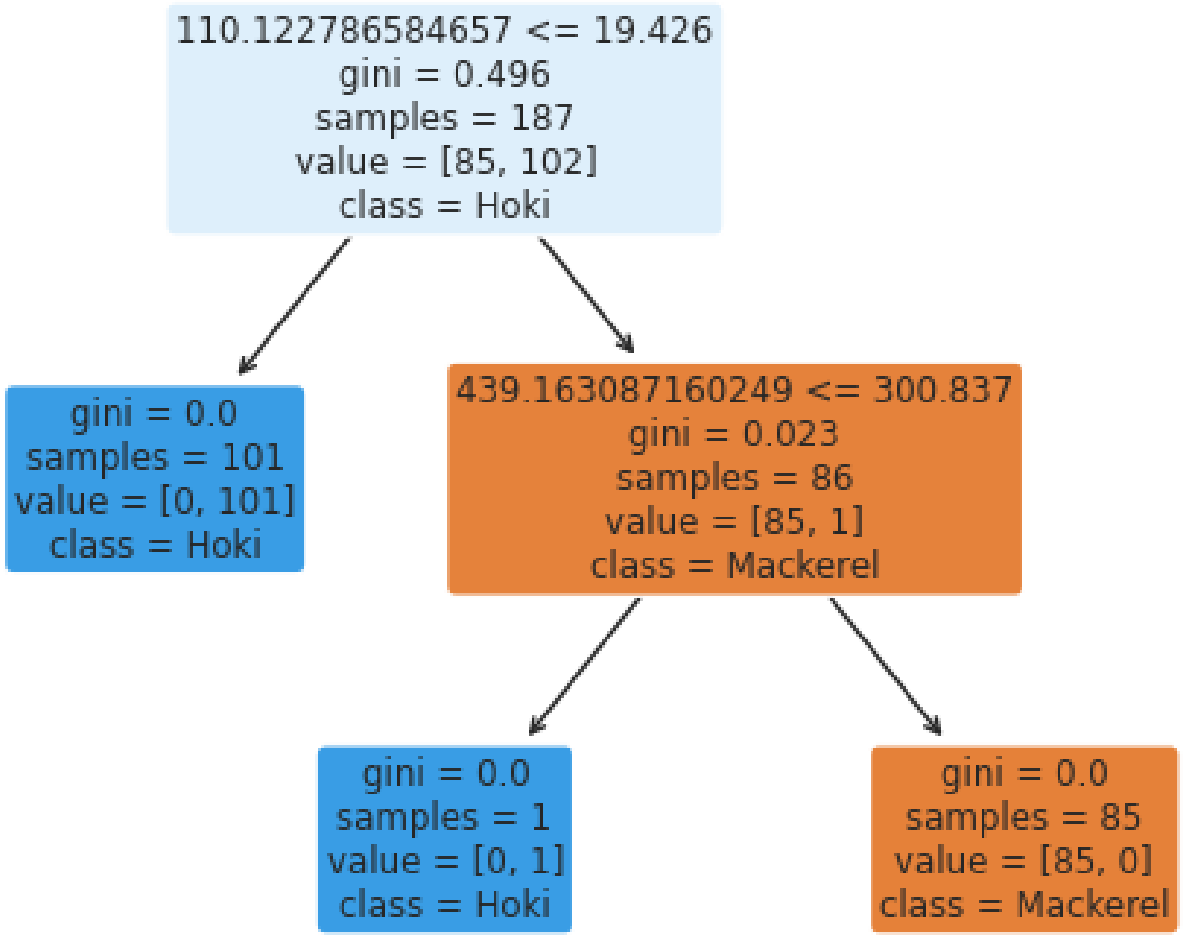
Pre-Training: Next Spectra Prediction [22]



Pre-Training: Next Spectra Prediction [22]

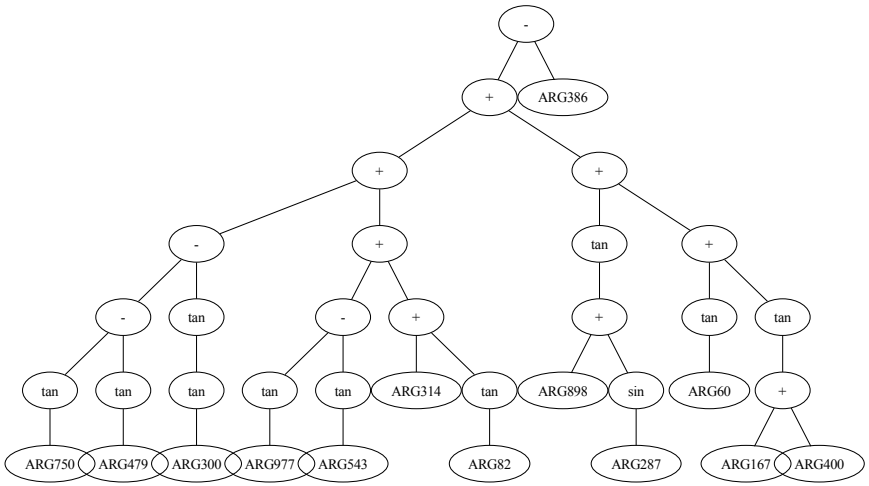


Decision Tree



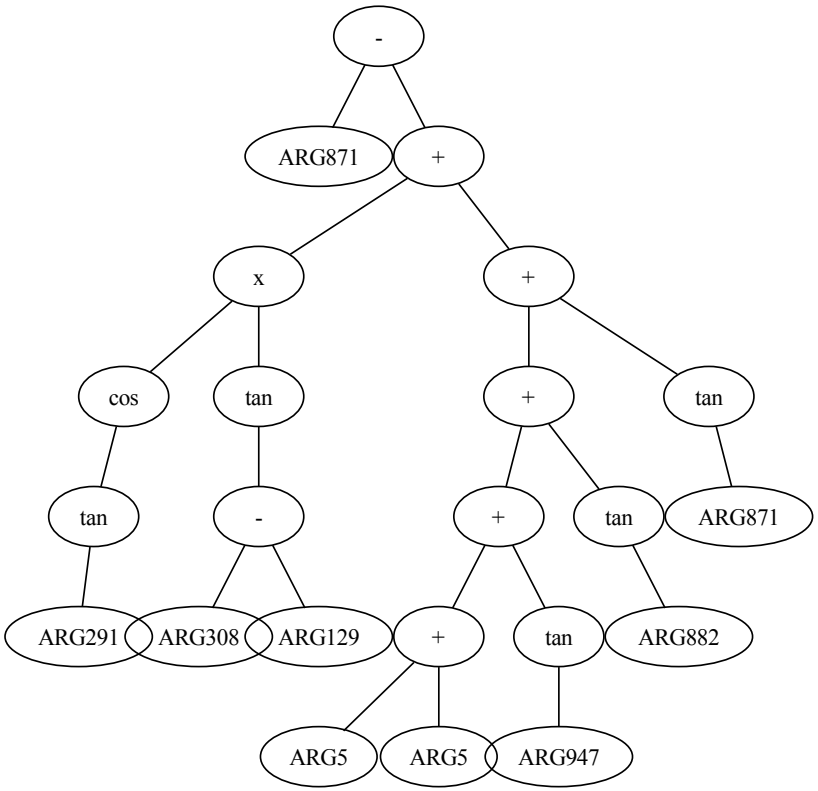
Decision tree for fish species

Genetic Programming Tree - Hoki



Genetic Programming Tree - Fish Species Hoki

Genetic Programming Tree - Mackerel



Genetic Programming Tree - Fish Species Hoki

TLDR;

Transformer can predict fish species with near-perfect accuracy, **DT** and **GP** provide **accurate**, **interpretable** and **efficient** models for **Rapid Evaporative Ionisation Mass Spectrometry**.

Download the slides & paper.

[1] J. R. Koza *et al.*, *Genetic programming II*. MIT press Cambridge, 1994, vol. 17.

[2] B. Tran, B. Xue, and M. Zhang, "Genetic programming for feature construction and selection in classification on high-dimensional data," *Memetic Computing*, vol. 8, no. 1, pp. 3–15, 2016.

[3] ——, "Genetic programming for multiple-feature construction on high-dimensional classification," *Pattern Recognition*, vol. 93, pp. 404–417, 2019.

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

[5] M. Á. Pardo, E. Jiménez, and B. Pérez-Villarreal, "Misdescription incidents in seafood sector," *Food Control*, vol. 62, pp. 277–283, 2016.

[6] C. Black, O. P. Chevallier, S. A. Haughey, J. Balog, S. Stead, S. D. Pringle, M. V. Riina, F. Martucci, P. L. Acutis, M. Morris *et al.*, "A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry," *Metabolomics*, vol. 13, no. 12, pp. 1–13, 2017.

[7] A. P. Simopoulos, "Evolutionary aspects of diet: the omega-6/omega-3 ratio and the brain," *Molecular neurobiology*, vol. 44, no. 2, pp. 203–215, 2011.

[8] M. L. Panse and S. D. Phalke, "World market of omega-3 fatty acids," *Omega-3 Fatty Acids*, pp. 79–88, 2016.

[9] K. Bi, D. Zhang, T. Qiu, and Y. Huang, "Gc-ms fingerprints profiling using machine learning models for food flavor prediction," *Processes*, vol. 8, no. 1, p. 23, 2020.

[10] D. D. Matyushin and A. K. Buryak, "Gas chromatographic retention index prediction using multimodal machine learning," *Ieee Access*, vol. 8, pp. 223 140–223 155, 2020.

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

[12] C. Black, O. P. Chevallier, K. M. Cooper, S. A. Haughey, J. Balog, Z. Takats, C. T. Elliott, and C. Cavin, "Rapid detection and specific identification of offals within minced beef samples utilising ambient mass spectrometry," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[13] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[14] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.

[15] W.-Y. Loh, "Classification and regression trees," *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[16] D. J. Hand and K. Yu, "Idiot's bayes—not so stupid after all?" *International statistical review*, vol. 69, no. 3, pp. 385–398, 2001.

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

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and information Processing*, vol. 18, no. 1998, pp. 1–8, 1998.

[20] H. M. Gomes, J. Montiel, S. M. Mastelini, B. Pfahringer, and A. Bifet, "On ensemble techniques for data stream regression," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

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