Automated Fish Classification

Using Unprocessed Fatty Acid Chromatographic Data

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

TLDR;

Linear SVM can accurately predict fish species, **PSO** makes that process 4 times faster, producing an **accurate**, **interpretable** and **efficient** model for **Gas Chromatography**.

Table of Contents

Introduction

Gas Chromatography is an analytical chemistry method that produces high-dimensional low-sample data.

This study compares classification and feature selection when applied to gas chromatography data from fish oil.

Classification to predict Fish Species and Body Parts, two datasets that share the same features.

Feature selection to reduce the dimensionality, improve computation efficiency, and (even) improve classification performance.

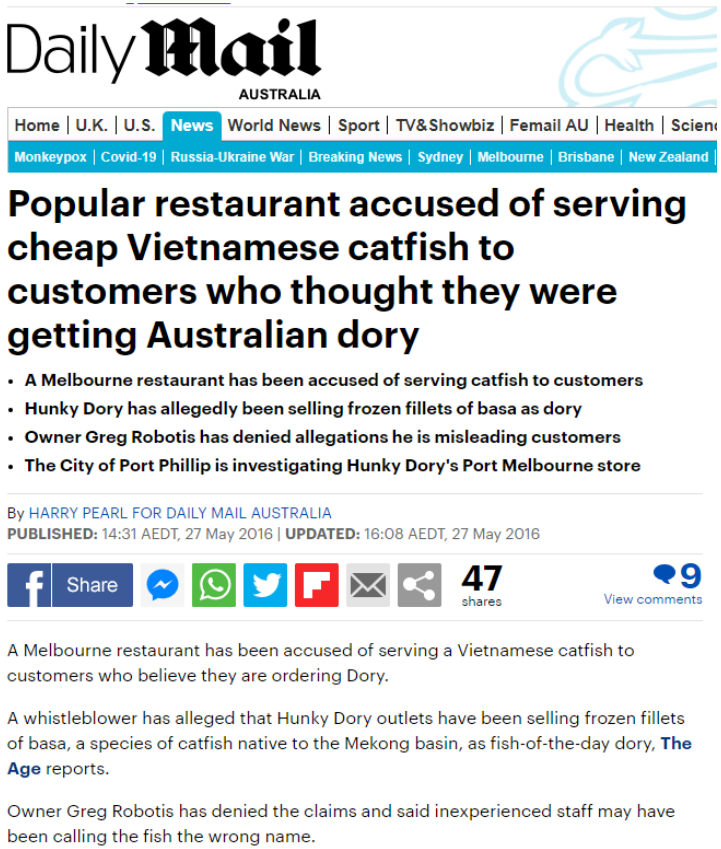
Background

Catfishing: when you order Australian Dory, you were served Vietnamese catfish[1]

For quality assurance so you want to be sure you what you are eating.

Many steps from ocean-to-plate in producing fish products, prone to human error and ripe for criminal activity.

In a 2016 meta-analysis [2] found an average mislabelling rate of 30% in seafood.



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.

[1] Aussies. No surprises there!

Background

Catfishing: when you order Australian Dory, you were served Vietnamese catfish[1]

For quality assurance so you want to be sure you what you are eating.

Many steps from ocean-to-plate in producing fish products, prone to human error and ripe for criminal activity.

In a 2016 meta-analysis [2] found an average mislabelling rate of 30% in seafood.

Background

Fish oil has omega-3 fatty acids, nutritionally important fats often missing from Western diets [3].

Causing high consumer demand for omega-3 supplements, produced from a wide range of marine biomass [4].

We can find which fish species (or body parts) are suitable for use in omega-3 supplements by analyzing their chemical composition with Gas Chromatography [5].

However, preparing/analyzing fatty acid data requires domain experts and is a very expensive and time-consuming process.

Previous works [6, 7] employ high-accuracy black-box CNNs. We need to understand models to build trust in their predictions and to verify/troubleshoot them.

Data

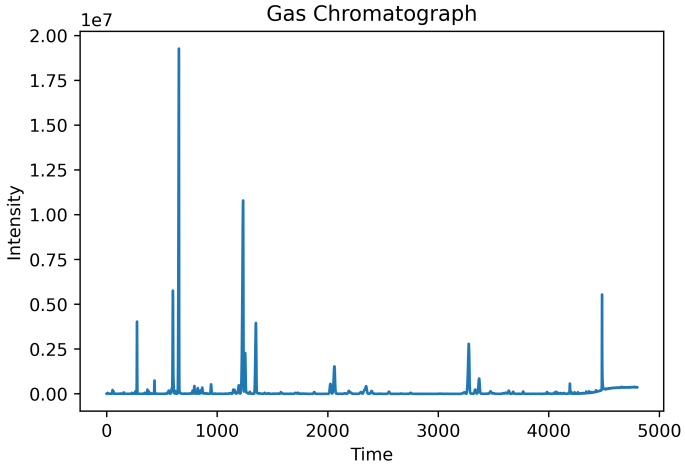
Gas Chromatography (GC) is an analytical chemistry method that produces high-dimensional low-sample data.

It is an expensive and time-consuming task to prepare/analyze Gas Chromatography data [8].

Steps [5]:

1. Apply heat to liquid.
2. Evaporate into gas.
3. Travel through long tube.
4. Detector measures intensity.

Gist: Molecules have distinct and different boiling points, these correspond to known timestamps.



GC for Snapper fish species

Classification

Task: to predict Fish Species and Body Parts, two datasets that share the same features.

Method:

- average balanced classification accuracy
- over 10-fold cross-validation.
- 30 independent runs for each

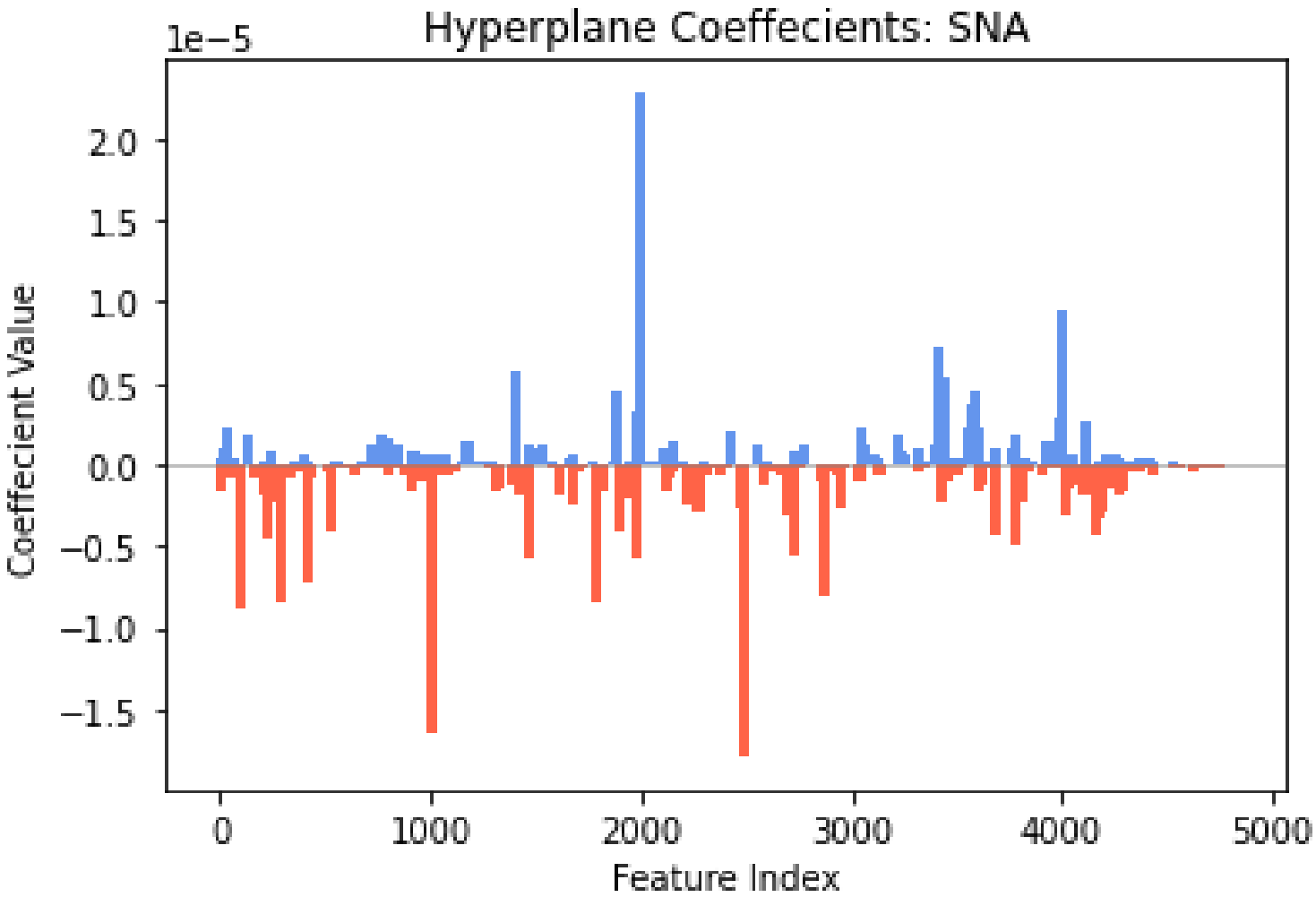
Results: Linear SVM performed best, with **near-perfect** accuracy for fish species, and best accuracy for body parts.

Why it matters? An accurate and interpretable model for determining species and body parts, can identify high-value fish oil [4] and avoid catfishing/mislabelling [1, 2].

Classification: Results

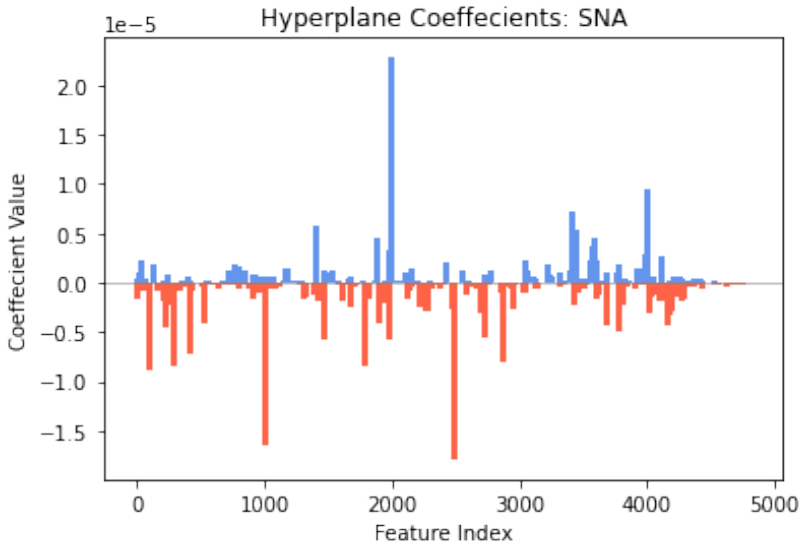
Dataset	Method	Train	Test
Species	KNN	83.57	74.88
	RF	100.0	85.65
	DT	100.0	76.98
	NB	79.54	75.27
	SVM	**100.0**	**98.33**
Parts	KNN	68.95	43.61
	RF	100.00	72.60
	DT	100.00	60.14
	NB	65.54	48.61
	SVM	**100.00**	**79.86**

Interpretable

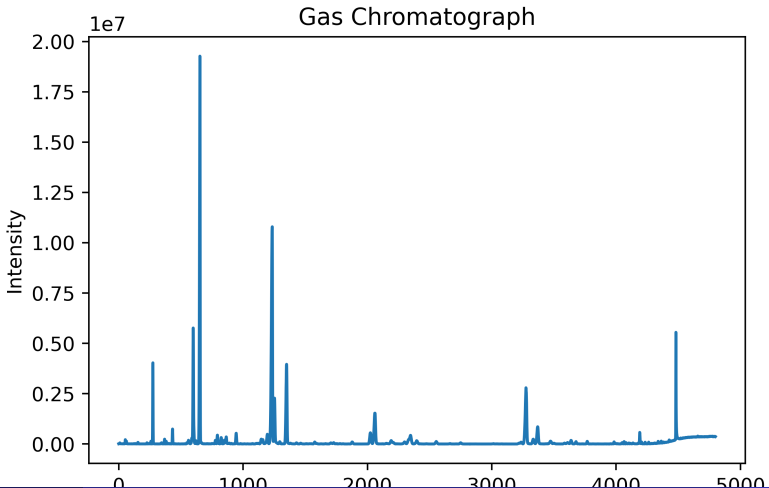


Linear SVM hyperplane coefficients for the Snapper

Interpretable



post hoc analysis to build trust in the prediction 



Feature Selection

Task: To reduce the dimensionality, improve computation efficiency, and (even) improve classification performance.

Method:

- Each **feature selection** method for number of features in { 50, 100, ..., 4800 }.
- PSO was evaluated on 30 independent runs.
- Give best run for each method on next slide.

Results: PSO has **improves** accuracy for fish species, using 25% of the features making it **4 times faster**.

Why it matters? Only using important features leads to simpler models, we can build trust in models we understand, giving a smaller problem space for troubleshooting/diagnosis [9].

Feature Selection: Results

Dataset	Method	# Features	Train	Test
Species	ReliefF	359	100.0	98.33
	mRMR	**1500**	**100.0**	**99.17**
	χ^2	3250	100.0	98.33
	PSO	**1192**	**100.0**	**99.17**
	Full	4800	100.0	98.33
Parts	ReliefF	1650	100.0	84.44
	mRMR	**1500**	**100.0**	**86.94**
	χ^2	1550	100.0	82.50
	PSO	1223	100.0	84.31
	Full	4800	100.0	79.86

TLDR;

Linear SVM can accurately predict fish species, **PSO** makes that process 4 times faster, producing an **accurate**, **interpretable** and **efficient** model for **Gas Chromatography**.

[1] 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-inst html

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

[3] 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.

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

[5] K. Eder, "Gas chromatographic analysis of fatty acid methyl esters," *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 671, no. 1-2, pp. 113–131, 1995.

[6] 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.

[7] 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.

[8] 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.

[9] Z. Zhao, R. Anand, and M. Wang, "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform," in *2019 IEEE international conference on data science and advanced analytics (DSAA)*. IEEE, 2019, pp. 442–452.