School of Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Internet: office@ecs.vuw.ac.nz

Rapid detection and identification of contamination within fish products using rapid ionisation evaporative mass spectrometry

Jesse Wood

Supervisors: Bach Hoai Nguyen, Bing Xue, Mengjie Zhang

Submitted in partial fulfilment of the requirements for Doctorate of Philosophy - Artificial Intelligence.

Abstract

This document gives some ideas about how to write a project proposal, and provides a template for a proposal. You should discuss your proposal with your supervisor.

1. Introduction

- Scope - place the problem in the world.
- Specifics to New Zealand, sustainability.
- Fish processing - automation, quality control, contaimination.
- Current state-of-the-art
  - GC-MS, manual, time consuming, expensive, destructive, instrumental drift.
- Aquaculture
- REIMS
- Between 1950 and 2018, aquaculture has increased tenfold [1]
- Foundational work and inspiration for this project is [2]

2. Literature Review

- Mass spectrometry [**?**]
- REIMS
- Classification
- Feature Selection
- Interpretable ML
- Genetic Programming
- Transfer Learning

3. Prelimary Work

- Automated Fish Classification on GC-MS data.
- CNN for Fish classification on GC-MS data.
- Genetic Programming (GP) for GC-MS data
  - Single-Tree GP
  - Multi-tree GP
- REIMS exploratory data analysis

4. Contributions

- Each research question applies to the REIMS and Data Infusion .
- For each dataset, hoki and mackeral.

These are the research questions from Plant and Food Research.

- Can REIMS data be used to classify different species tissues? What variables are responsible?
  - Classification
  - Feature Importance - Interpretable
- Can REIMS data detect mixed-species contaminiation in fish tissues? At what concentration? What varaibles are responsible?
  - Classification
  - Regression
  - Feature importance - Interpretable
- Can REIMS data detect mineral oil contamination in fish? At what concentration? What variables are responsible?
  - Classification
  - Regression
  - Feature importance - Interpretable
- Can REIMS data be used to distinguish between different fish individuals? What variables are responsible?
  - Identification
  - Feature imporance - Interpretable

5. Milestones

- Literature Review
- EDA
- Preprocessing
- Classification
- Cross-species Contaminiation
- Mineral-oil Contaminiation
- Individual Identification
- Auto ML
- Thesis

6. Thesis Outline

1. Introduction
2. Background
    - Mass Spectrometry
    - REIMS
    - Classifcation / Regression
    - Interpretable ML
3. Preparations
    - Exploratory Data Analysis
    - Preprocessing
4. Applications
    - Classification
    - Contaminiation Detection
    - Individual Identification
    - Auto ML
5. Discussion
6. Conclusion

7. Resources

In this section you will detail any resource requirements such as hardware, software or access to subjects.

- Hardware
    - ECS Grid Compute
    - Rapoi
    - Niwa HPC - via Auckland University
- Software
    - Repository - Github
    - Project Management - Github Projects
    - Programming language - Python
    - Documentation - Read the Docs
- Experience
    - Field-trip to Callaghan Innovation to see REIMS
    - Field-trip to NZ Plant and Food Research (if necessary for future datasets).

Glossary

aquaculture Farming things in water, especially for good, e.g. Snapper are often raised in captivity.. 1

REIMS Rapid Evaporative Ionisation Mass Spectrometry. 1–3

Bibliography

[1] FAO, *The State of World Fisheries and Aquaculture, 2020*. FAO, 2020.

[2] 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.