



Australian Government
**Department of Agriculture
and Water Resources**



A Software as a Solution to enhance the Physi-Trace program

Final Report APL Project 2020/0062

May 2022

Graeme Crook & Hannes Bernhardt
InterpreData
20a McKenzie Street,
Albany, WA, 6330

Acknowledgements

This project is supported by funding from Australian Pork Limited and the Department of Agriculture and Water Resources.

Executive Summary

This concept project is to demonstrate the enhancement of Physi-Trace with TraceBase Artificial Intelligence (AI) to provide a cost-effective and highly efficient method of determining the provenance of pork samples.

Fifteen samples were collected from three different farms in Western Australia to create a trace elemental fingerprint for each site using the Physi-Trace method of analysis. Manual interpretation of these results is a very time-consuming process and requires a great deal of skill and knowledge to perform accurately. TraceBase uses three different models and the method is calculated and fully auditable. TraceBase was able to clearly identify the significant analytes and separate the samples from the three different producers within 10 minutes of uploading the dataset, in 1-4% of the time of the manual process which is a very significant cost-saving.

The unknown samples were also very quickly and easily associated with one of the farms as part of the classification process.

An additional benefit of TraceBase is the ability to track changes over time. This would allow the interpretation of as few as 2-3 samples per producer annually as opposed to the 10 samples normally required to maintain the farm fingerprint. If the deviation in fingerprint annually is minimal, we can further delay the testing period and avoid unnecessary cost until significant feed or production changes occur, which would require re-sampling to establish the new baseline for the farmgate fingerprint. A full history of the farm would be available for trend analysis.

TraceBase will also increase the efficiency and accuracy of interpretation by removing any subjective interpretation of samples. This interpretation can now be undertaken at the Abattoir or APL with minimal training and will greatly reduce the overall sampling cost to the industry.

The TraceBase AI interpretive method is self-validated by comparing the results from three different machine learning techniques against each other, providing irrefutable provenance.

This POC provided the compelling evidence required to propose connection to TraceBase with the historical Physi-Trace data. These data could now be successfully imported into TraceBase to generate a fingerprint for each producer, leveraging the considerable investment that the Pork industry through APL and government have made in the collection of the existing dataset.

Having an accurate, validated, fast and easy to use system to identify samples will greatly enhance the current Physi-Trace system and allow the industry to benefit from scientifically verifiable branding claims.

Table of Contents

Acknowledgements	2
Executive Summary	3
Background to Research	7
Objectives of the Research Project	8
Introductory Technical Information	9
Research Methodology	10
Results	14
Discussion	23
Implications & Recommendations	25
Intellectual Property	26
Technical Summary	26
Literature cited	26
Publications Arising	26

List of Tables

Table 4.5.1 Chemical fingerprinting concentrations I I	
Table 4.5.2 Chemical fingerprinting concentrations I I	
Table 5 Results for individual machine learning models	14
Table 5.1.1.1 First unknown sample prediction using Analyte Compliance Range Prediction	15
Table 5.1.1.2 Significant analytes for farm 1 2591 I	15
Table 5.1.2.1 Second unknown sample prediction using Analyte Compliance Range Prediction	16
Table 5.1.2.2 Significant analytes for farm 2 8614	16
Table 5.1.3.1 Third unknown sample prediction using Analyte Compliance Range Prediction	17
Table 5.1.3.2 Significant analytes for farm 3 Corrigin meatworks	17
Table 5.2.1 First unknown sample prediction using LDA	19
Table 5.2.2 Second unknown sample prediction using LDA	20
Table 5.2.3 Third unknown sample prediction using LDA	20
Table 5.3.1 Neural Network Hyperparameters	21
Table 5.3.2 Neural Network Statistics and Results	22

List of Figures

Figure 4.5.3.1 Visual representation of the classification algorithms available in TraceBase	12
Figure 4.5.3.2 Visual representation of the selected classification	12
Figure 4.5.4.1 Deep Belief Neural Network using Bernoulli Function	13
Figure 4.5.4.2 Visual representation of the selected classification for the network	13
Figure 5.1.1 First unclassified sample plotted against predicted farm fingerprint	15
Figure 5.1.2 Second unclassified sample plotted against predicted farm fingerprint	16
Figure 5.1.3 Third unclassified sample plotted against predicted farm fingerprint	17
Figure 5.1.4 Third unclassified sample plotted against two farm fingerprints	18
Figure 5.1.5 Third unclassified sample plotted against farm I 2591 I	18
Figure 5.2.1 First unclassified sample plotted as a linear discriminant analysis scatter chart	19
Figure 5.2.2 Second unclassified sample plotted as a linear discriminant analysis scatter chart	20
Figure 5.2.3 Third unclassified sample plotted as a linear discriminant analysis scatter chart	21
Figure 5.2.4 Third unclassified sample plotted as a linear discriminant analysis scatter chart in 3D	21

Background to Research

The Australian Pork Industry has been at the forefront of traceability with the development of Physi-Trace over the last ten years. The development of the Black Box technology through Pork CRC enhanced that offering. However, technology has greatly advanced in recent years especially in the cloud deployment, database development and AI capabilities. The existing Physi-Trace system is no longer available nor usable on modern computers.

There are issues within the Physi-Trace system which include:

1. having no database to store changes
2. no tracking of the interpretation by the analyst did to produce the final dataset – BlackBox technology
3. The software can only analyse a fixed set of isotopes and a single CRM
4. The web interface is not HTML5 compliant so it cannot work in modern web browsers.
5. no mechanism for classification so it must be done manually with 3rd party statistical software that has no AI capabilities
6. the analyst does not have a mechanism to establish the quality and separability of the analysed batch while the samples are still in the auto sampler
7. The source code remains in India and the mechanism of duplicating and registering the software has been retired years ago. There are additional problems in that the Indian company that made the software no longer exists, so any backup or support is no longer available.

To allow the industry to return to the forefront of traceability and to leverage the considerable investment made in this area, it was proposed to conduct a proof of concept for TraceBase to fill the current gap in functionality

Objectives of the Research Project

The objective of this project was:

- To evaluate the capacity of InterpreData's unique TraceBase platform to provide irrefutable farm gate provenance of Pork samples from three separate farms using cross validation techniques that include Chemistry methodology, Linear Discriminate Analysis and its Deep learning Neural network.
- To provide confidence to APL and DAVE that TraceBase can build on this Proof of Concept phase and make use of the data previously collected during Physi-Trace development in a subsequent project.
- To demonstrate the cost-saving available through automation by using TraceBase for interpretation of Physi-Trace data, making testing affordable for the Pork Industry

Benefits

With an extensive database of sites/farms, hosted and stored by InterpreData, each producer will benefit by having 24x7 secure access to their data to provide the irrefutable provenance that customers are seeking as well as ensuring that their sustainability claims can be scientifically verified should there be any challenges.

Brands will be protected as will the industry in general as any food substitution will be detectable throughout the supply chain. Queries can be made for provenance by accessing the database and classifying an unknown sample against all the farms in the dataset.

In the event of a food safety issue or a biosecurity breach from an unknown source, tracking back will provide a further level of assurance for the industry.

The more cost-effective and efficient analytical methods, the machine learning and the time savings will enable more of the producers to undertake the testing and to benefit from the technology. Something that Physi-Trace was unable to achieve in its past format.

Question

Can the TraceBase AI correctly detect the trace elements used to predict the farm of origin for a pork sample using the Physi-Trace analytical method?

The three main success criteria were:

1. **Accuracy** - Will all the machine learning models predict the farm associated with an unknown sample?
2. **Speed** - Can this be done within minutes rather than hours or days?
3. **Ease of use** - Can the system be operated by someone with limited analytical experience?

Introductory Technical Information

A plant or animal has a distinct elemental signature relative to its location. Plants that are grown in a location will absorb tiny trace elements that are very specific to its location. Animals will consume plants with a specific elemental signature that will also be reflected in an animal sample. By analysing a sample, looking for these unique characteristics, we can accurately verify the origin of that sample.

TraceBase provides a unique elemental fingerprint for a farm gate, horticultural, mining, pharmaceutical, wildlife product and any other substance requiring traceability.

What is it, and what does it do?

TraceBase is a highly sophisticated analytical database capable of organising and interpreting datasets producing quality data for any product, organic or inorganic. TraceBase uses 108 specifically created algorithms, a Linear Discriminate Analysis capability and a neural network that provides artificial learning to provide classification data.

The software capability provides clean data from uploaded csv files after samples are processed on ICP-MS and ICP-AES spectrometers. The software automatically identifies and removes interference and background noise in order to provide a clean dataset for our classification models with the purpose of establishing provenance.

The software can provide the results from 200 samples in approximately 10 minutes, which can take an analytical chemist anywhere between six hours to a couple of days, (industry average for an analytical chemist is \$230 per hour) depending on the skill level of the chemist. A laboratory assistant can arrive at the same result as an analytical chemist due to the laboratory standard operating procedures being embedded in the system.

It chemically analyses organic and inorganic substances, storing the results in a data warehouse with a connected GeoDatabase for GIS analysis. Samples are used to sequence and categorise the distinct patterns of trace elements. With every dataset that is entered into the system, we exponentially increase our knowledge base of elemental fingerprints for a specific location. There is no other program in existence that performs this function.

TraceBase identifies trace elements in parts per billion and the analysis provides a unique, elemental fingerprint that provides irrefutable evidence of place of origin. It is impossible to cheat on the fingerprint, adding a unique, unforgeable foundation layer in the quest for traceability of primary products.

It helps protect producers from food substitution and provide confidence in the supply chain and for which clients have paid a premium.

In improving the Physi-Trace offering, the standards and analytical methods have been updated and refined over the last couple of years, so the accuracy of TraceBase is much higher. Errors in the normalization section of the analytical methods have been corrected and that is not reflected in the old system

Research Methodology

1.1 Purpose

The purpose of the test is to establish whether the unknown sample was from one of the farms already sampled or if comes from a different location or source.

1.2 Sample Preparation

Microwave digestion using nitric acid (7-10mL) and hydrogen peroxide (1.5-3mL) and then made up to volume with ultrapure MQ water.

1.3 Summary

15 samples were collected from three different farms in Western Australia to create a trace elemental fingerprint for each farm. The samples were supplied by Linley Valley and Corrigin Meatworks.

The unique fingerprint for each farm can then be used to trace back unknown meat samples to the producer. The process consists of two phases, analysis and interpretation. During the analysis phase samples are collected and analysed in a lab using ICP-MS and ICP-AES instruments to identify all the trace elements contained in each sample. Concentrations of trace elements down to 1 part per billion are detected.

Pork samples are digested in an acid solution with a PerkinElmer Titan MPS. Exothermic reactions during the digestion, caused by a high amount of polyunsaturated fatty acids, can be avoided by a pre-reaction at room temperature for several hours.

Notes: This application serves only as a guideline and may need to be optimized for your sample. *This application is designed for the digestion of 8 samples. Decrease the power at the first step by 5% for each sample less than 8. Minimum power is 40% regardless of the number of samples digested.

1.4 Sample Analysis

The sample analysis was performed by ChemCentre as per the Physi-Trace methodology.

1.5 Sample Interpretation

The first sample of each batch was used as a blind sample. The TraceBase AI was trained with the remaining samples for each class. Each class or farm in this test requires a separate set of significant analytes to predict the class correctly and consistently. The significant analytes and its concentration make up the elemental fingerprint.

1.5.3 Linear Discriminant Analysis

For this test the algorithms below were tested, and we selected Linear as our classifier for predictive analysis as can be seen in the bottom figure.

Figure 4.5.3.1 Visual representation of the classification algorithms available in TraceBase

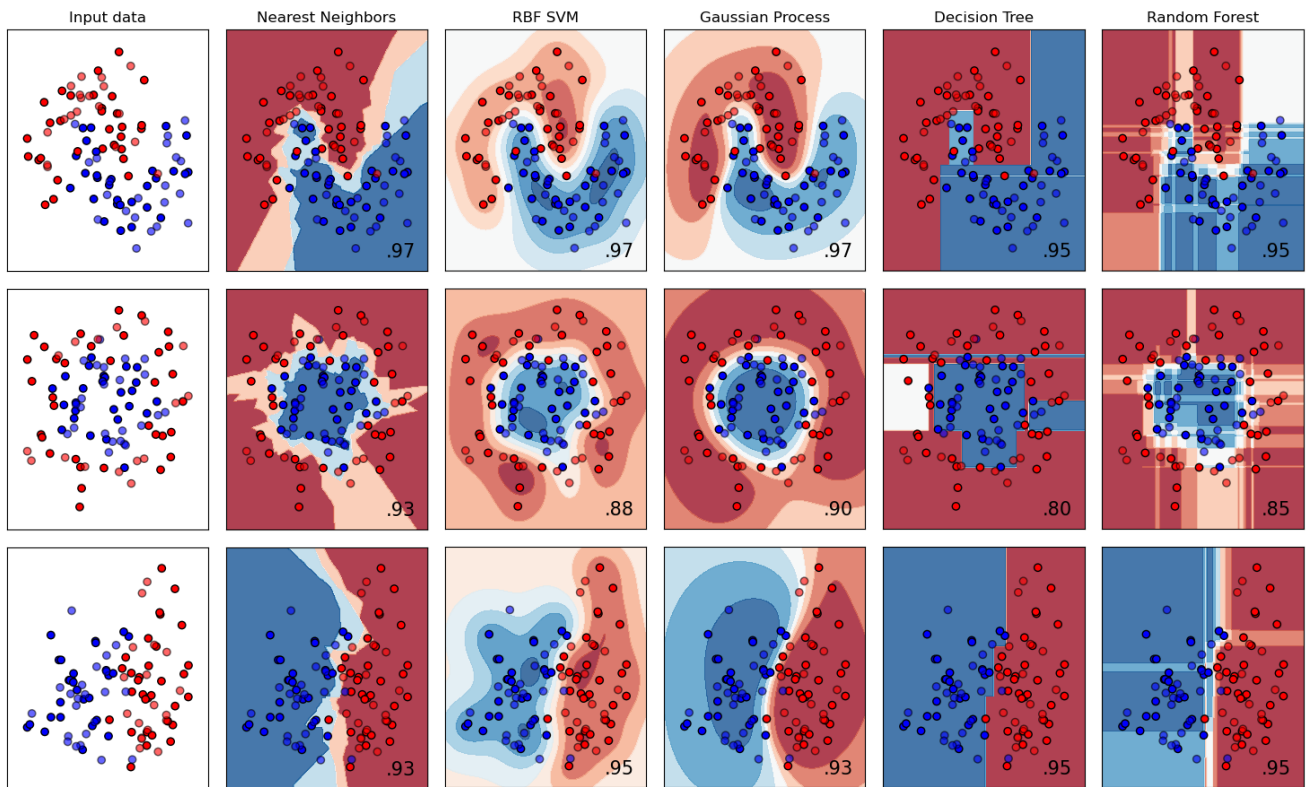
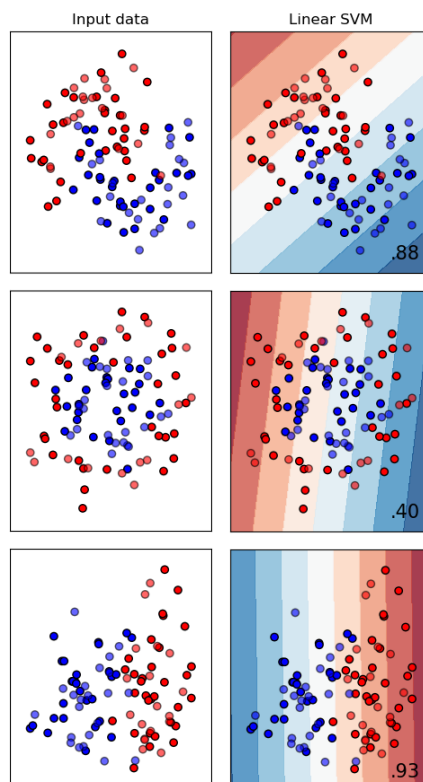


Figure 4.5.3.2 Visual representation of the selected classification



1.5.4 Neural Network Prediction

The deep learning neural network selected and trained for this prediction is the developed. The neural network architecture selected for this test is a fully connected deep belief network using the Bernoulli function.

For each sample prediction a new model is trained and calibrated specifically for the selected input dataset.

Figure 4.5.4.1 Deep Belief Neural Network using Bernoulli Function

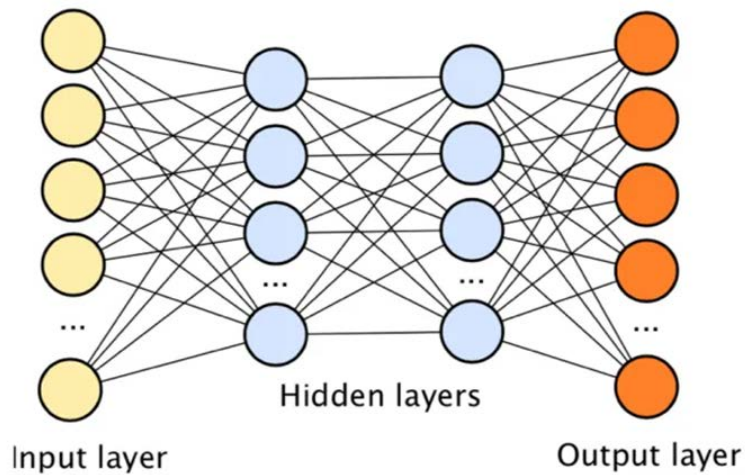
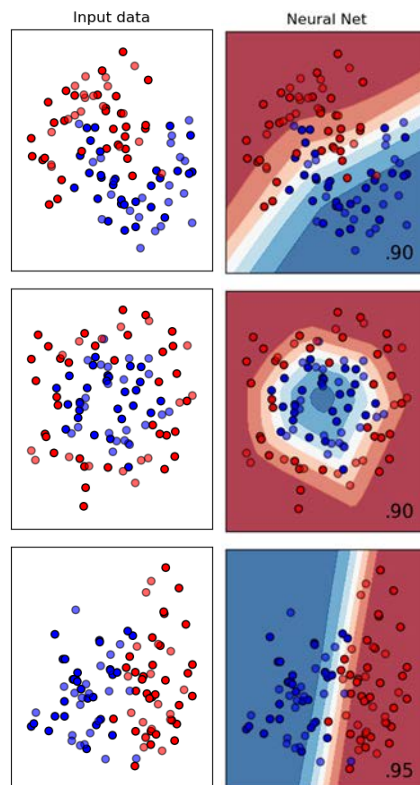


Figure 4.5.4.2 Visual representation of the selected classification for the network



Results

The key to achieving high predictability scores in class prediction using chemical analytes is that the data must be relevant to the predictive model you are building. Trace elemental concentration can answer several questions for example individual animal or grouped by feed, farm, region or country. In addition to comparing subsets of groupings or classes requires that only the relevant analytes or trace elements are selected for each prediction. Predictive model performance and accuracy greatly improve by selecting only the relevant analytes with the associated class.

Table 5 Results for individual machine learning models

Blind Sample Number	Predicted Class	Analyte Compliance Range Prediction	LDA Prediction Confidence	Neural Network Prediction Confidence
Batch #: 21S1883 Sample #: PT163009	25911	100	100	100
Batch #: 21S1883 - Sample #: PT162924	8614	100	99.9996	100
Batch #: 21S1883 - Sample #: 23321	Corrigin meatworks	100	99.9983	100

The table above illustrates the confidence each of the three models trained and used to predict the correct class in this case the farm of origin. The predicted class in all cases are a True Positive meaning all of them are predicted accurately.

The results are further explained in the sections for each of the three unknown samples using three very different machine learning models:

- Analyte Compliance Range Prediction
- Linear Discriminant Analysis (LDA) Prediction
- Neural Network Prediction

These machine learning models form the TraceBase AI and all of them must reach consensus on the class before a sample is considered correctly classified.

Analyte Compliance Range Prediction

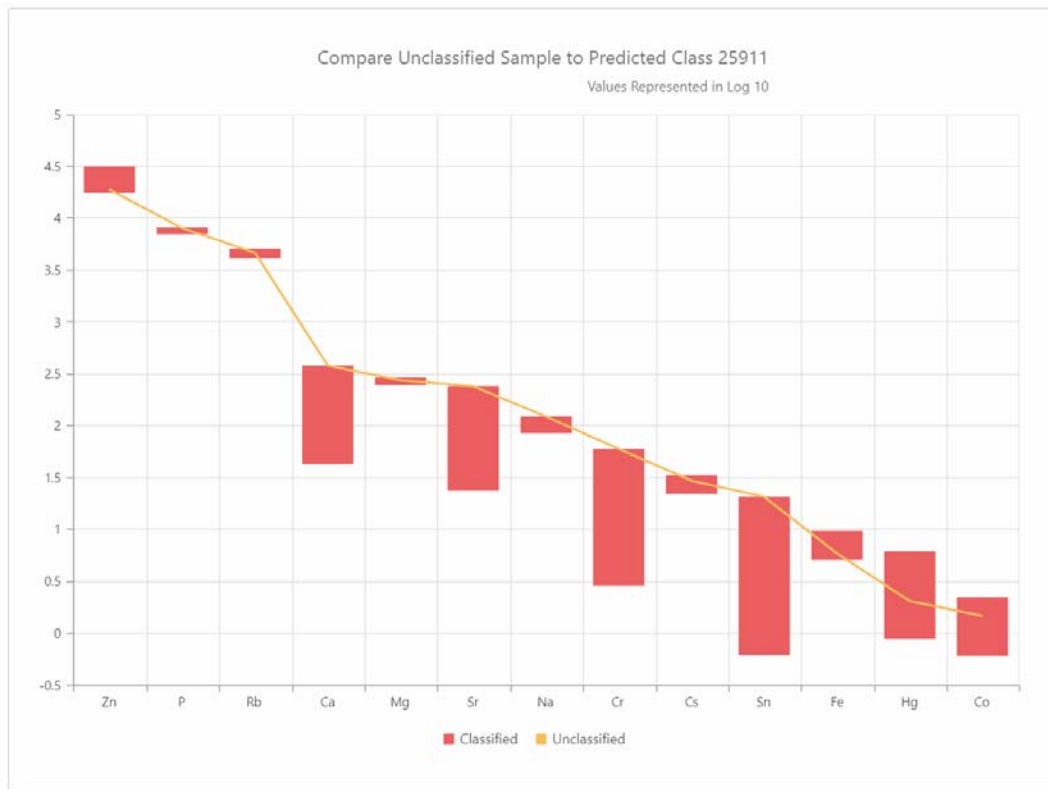
1.5.5 Selected unknown Sample 1

The table below is the results for the Analyte Compliance predicted class and the confidence as a percentage. The prediction result was correct as can be seen in the figure below the yellow line for the first sample taken as the 1st blind sample fit completely within the range of farm 1 (25911) for all the significant analytes.

Table 5.1.1.1 First unknown sample prediction using Analyte Compliance Range Prediction

Unclassified Sample Number	Predicted Class	Compliance
Batch #: 21S1883 - Sample #: PT163009	25911	100

Figure 5.1.1 First unclassified sample plotted against predicted farm fingerprint



The y axis is the concentration values in Log10, and the x axis is the significant trace elements.

Table 5.1.1.2 Significant analytes for farm 1 25911

Class	Significant Analytes	Compliance (%)	Be	Cr	Co	Zn	Rb	Sr
25911	13	100	0	1	1	1	1	1
8614	4	59.33	0	0.2525	0.8095	1	0.3572	0.627
Corrigin meatworks	5	67.46	0	1	1	0.6912	0.4715	0.371

1.5.6 Selected unknown Sample 2

The table below is the results for the Analyte Compliance predicted class and the confidence as a percentage. The prediction result was correct as can be seen in the figure below the yellow line for the first sample taken as the 2nd blind sample fit completely within the range of farm 2 (8614) for all the significant analytes.

Table 5.1.2.1 Second unknown sample prediction using Analyte Compliance Range Prediction

Unclassified Sample Number	Predicted Class	Predicted Class
Batch #: 21S1883 - Sample #: PT162924	8614	100

Figure 5.1.2 Second unclassified sample plotted against predicted farm fingerprint

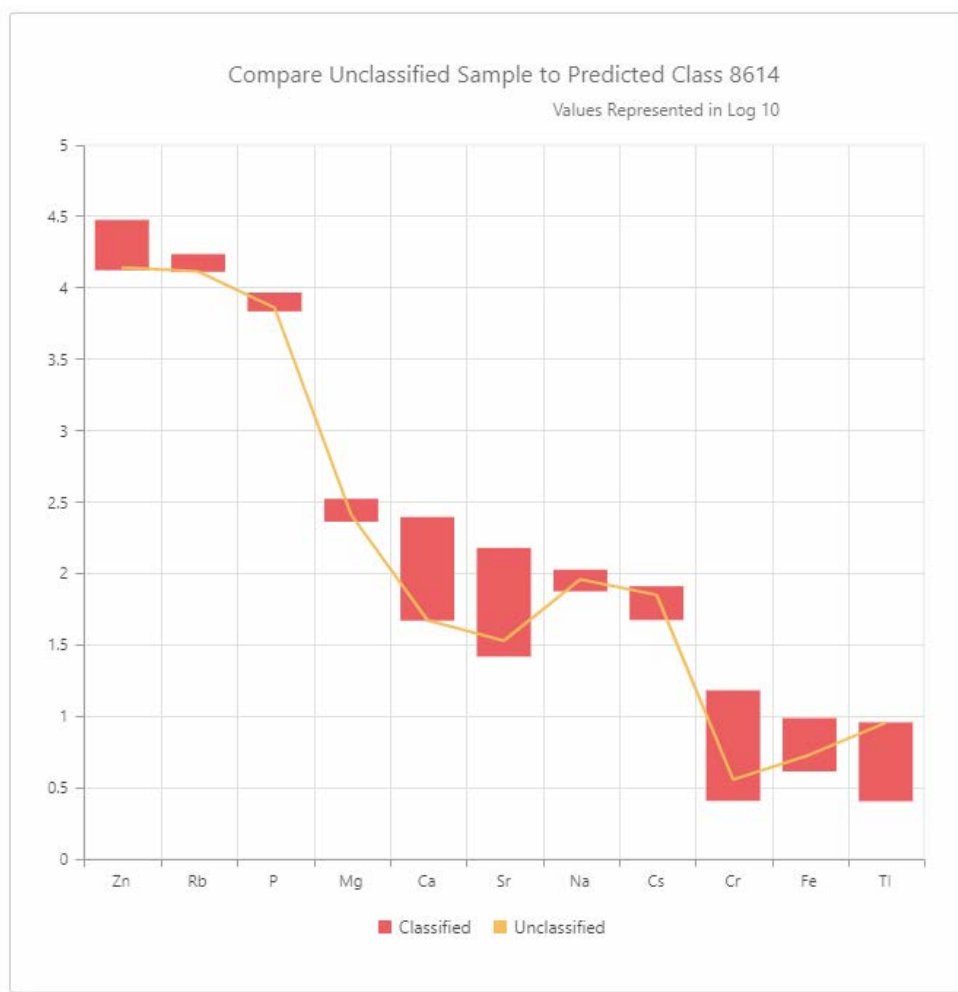


Table 5.1.2.2 Significant analytes for farm 2 8614

Class	Significant Analytes	Compliance (%)	Be	Cr	Co	Zn	Rb	Sr	Nb	Ag	Li
25911	7	66.17	0	1	0	0.7976	0.391	1	0	0	
8614	11	100	0	1	0	1	1	1	0	0	
Corrigin meatworks	6	79	0	0.663	0	0.51	1	1	0	0	

1.5.7 Selected unknown Sample 3

The table below is the results for the Analyte Compliance predicted class and the confidence as a percentage. The prediction result was correct as can be seen in the figure below the yellow line for the first sample taken as the 3rd blind sample fit completely within the range of farm 3 (Corrigin meatworks) for all the significant analytes.

Table 5.1.3.1 Third unknown sample prediction using Analyte Compliance Range Prediction

Unclassified Sample Number	Predicted Class	Compliance
Batch #: 21SI883 - Sample #: 23321	Corrigin meatworks	100

Figure 5.1.3 Third unclassified sample plotted against predicted farm fingerprint

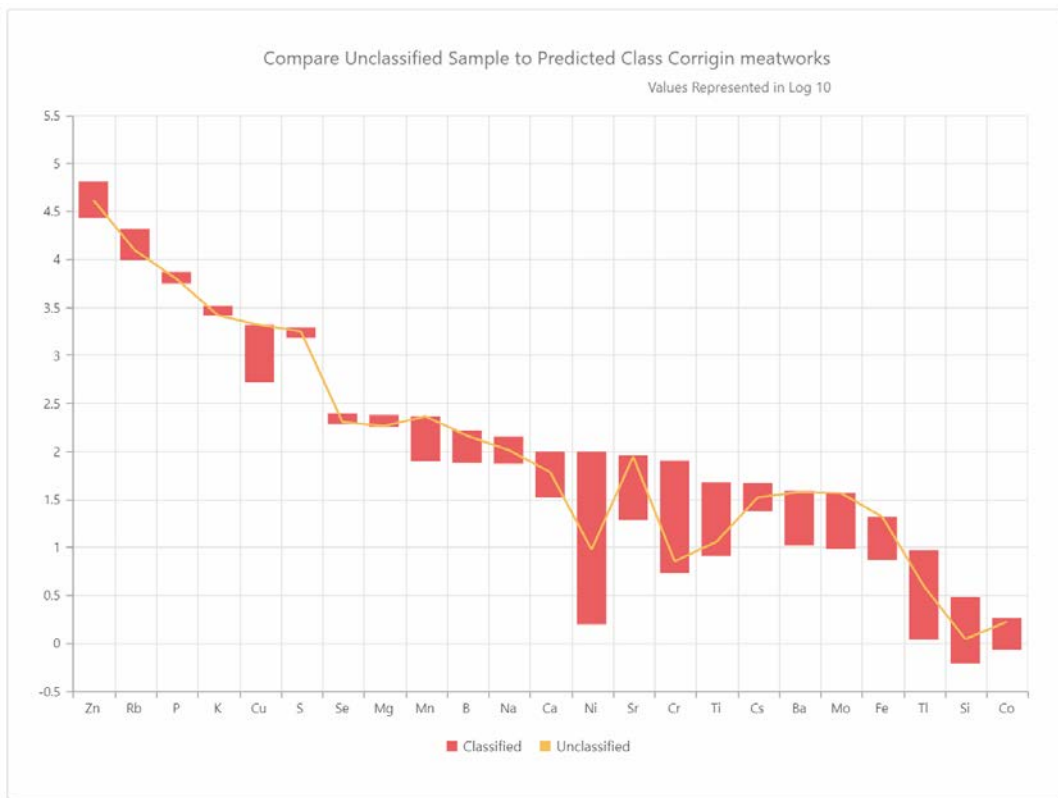


Table 5.1.3.2 Significant analytes for farm 3 Corrigin meatworks

Class	Significant Analytes	Compliance (%)	<input checked="" type="checkbox"/> B	<input checked="" type="checkbox"/> Sc	<input checked="" type="checkbox"/> Ti	<input checked="" type="checkbox"/> V	<input checked="" type="checkbox"/> Cr	<input checked="" type="checkbox"/> Mn	<input checked="" type="checkbox"/> Co	<input checked="" type="checkbox"/> Ni	<input checked="" type="checkbox"/> Cu	<input checked="" type="checkbox"/> Zn	<input checked="" type="checkbox"/> Ga	<input checked="" type="checkbox"/> Ge	<input checked="" type="checkbox"/> As	<input checked="" type="checkbox"/> Se	<input checked="" type="checkbox"/> Rb	<input checked="" type="checkbox"/> Sr	<input checked="" type="checkbox"/> Nb	<input checked="" type="checkbox"/> Mo	<input checked="" type="checkbox"/> Sn	<input checked="" type="checkbox"/> Pb
25911	11	72.27	0.774	0	1	0	1	0.8081	1	1	0.5318	0.7553	0	0.7037	0	1	0.408	1	0.6923	0.336	0	0
8614	11	80.29	0.8032	0	1	0	1	0.5856	0.7083	1	0.4924	0.7112	0	0.6667	0.5556	1	1	1	0	0.378	0	0
Corrigin meatworks	23	100	1	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	0	1	0	0

The “Corrigin meatworks” class has a more complex trace elemental fingerprint. It uses 23 analytes to accurately predict the classification.

Note: Every farm has a unique set of concentrations as well as elements that form part of the fingerprint

1.5.8 Selected unknown Sample 3 plotted against Farm 1 25911

The figures below are used in the classification process to verify that there is a significant difference between the predicted class and another class. By selecting another class to compare the sample to we can clearly see that only 11 of the 23 analytes are within the range therefore class 25911 is clearly not a match for sample 3.

Figure 5.1.4 Third unclassified sample plotted against two farm fingerprints

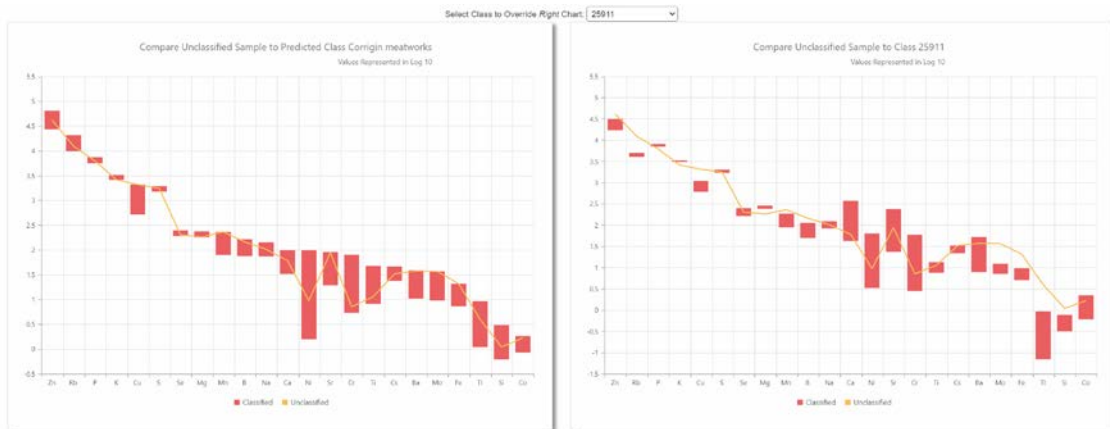
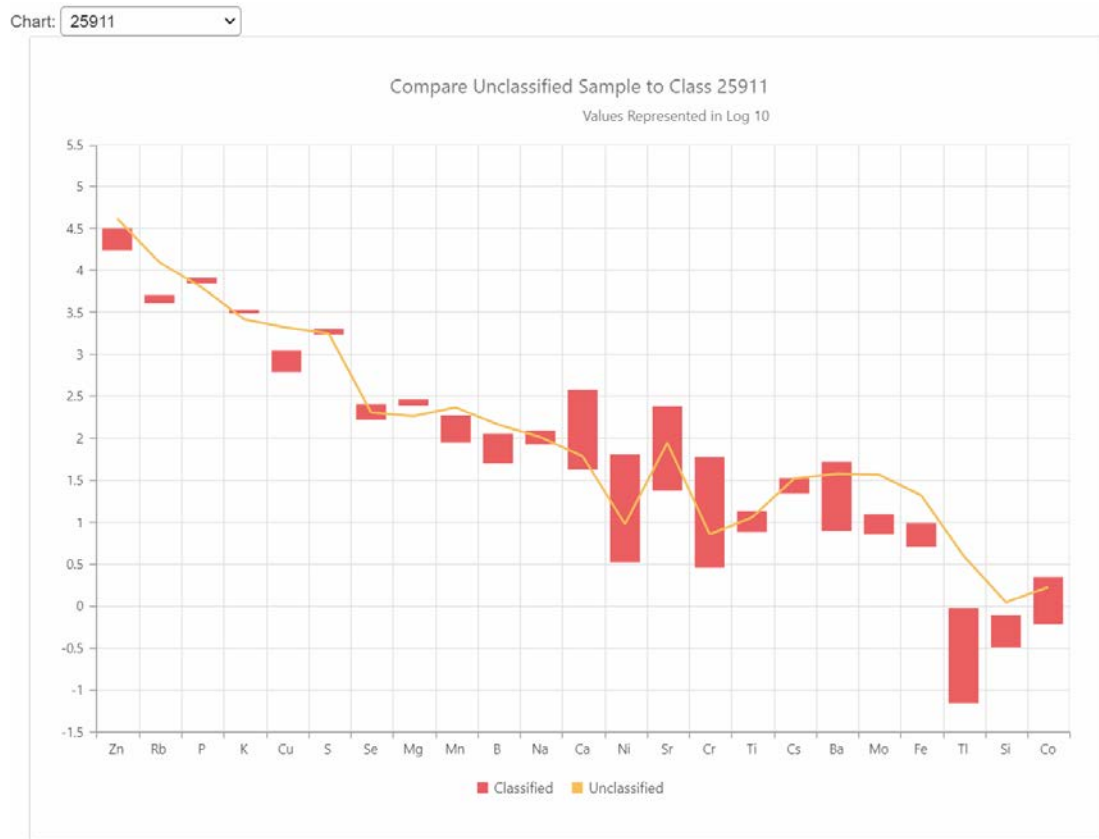


Figure 5.1.5 Third unclassified sample plotted against farm 1 25911



The figure above illustrates what an unmatched fingerprint looks like. The yellow line is not within all the red concentrations of the elements and therefore not a match.

Linear Discriminant Analysis (LDA) Prediction Confidence

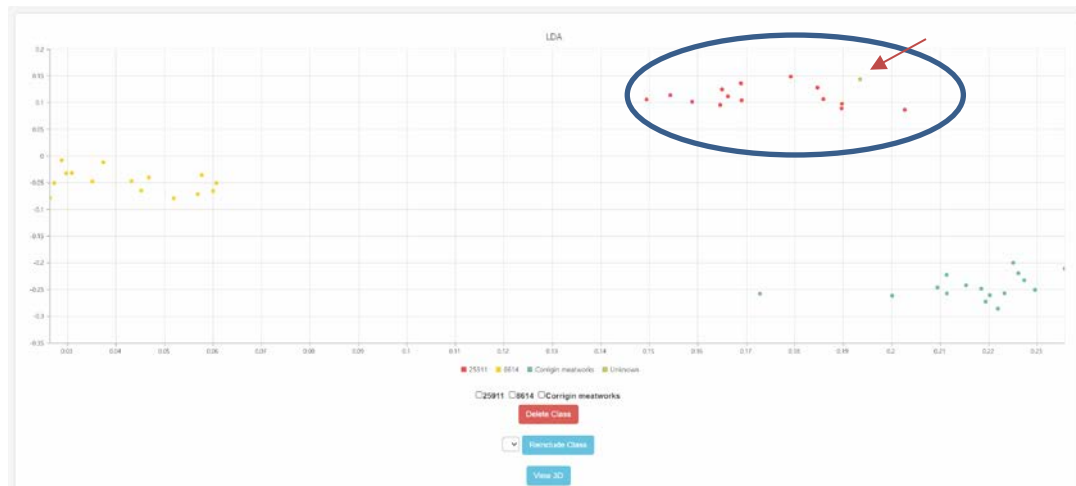
1.5.9 Selected unknown Sample 1

The table below is the results for the LDA predicted class 25911 and the confidence score as a percentage. The prediction result was correct as can be seen in the figure below the red arrow for the first sample taken as the 1st blind sample fit completely within the blue oval of farm 1 (25911) for all the significant analytes were the analytes as reduced to an x, y value. By only using the significant analytes we were able to clearly separate the classes and therefor achieve very high accuracy.

Table 5.2.1 First unknown sample prediction using LDA

Unknown Sample Number	LDA Predicted Class	Score
Batch #: 21S1883 - Sample #: PT163009	25911	100

Figure 5.2.1 First unclassified sample plotted as a linear discriminant analysis scatter chart



Dimensionality reduction is a technique where you project the features of multiple analytes in our case between 11 and 23 depending on the class and projecting it in two dimensions as seen in the above figure as the x and y axis for the LDA scatter charts.

When we cannot separate two classes in a multi class model we can delete the class with the best separation and run the model again. In this case it was not necessary as the classes are separate and the LDA model can very easily predict the class associated with the unknown sample.

1.5.10 Selected unknown Sample 2

The table below is the results for the LDA predicted class 8614 and the confidence score as a percentage. The prediction result was correct as can be seen in the figure below the red arrow for the first sample taken as the 2nd blind sample fit completely within the blue oval of farm 2 (8614) for all the significant analytes were the analytes as reduced to an x, y value. By only using the significant analytes we were able to clearly separate the classes and therefor achieve very high accuracy.

Table 5.2.2 Second unknown sample prediction using LDA

Unknown Sample Number	LDA Predicted Class	Score
Batch #: 21S1883 - Sample #: PT162924	8614	99.9996

Figure 5.2.2 Second unclassified sample plotted as a linear discriminant analysis scatter chart



PCA > LDA

The selected sample is excluded from the LDA training set and then plotted after the fact as not to influence the supervised machine learning model. Technically we are feeding unsupervised machine learning principal component analysis (PCA) model into a supervised machine learning linear discriminant analysis (LDA) model without the unknown sample.

This method allows us to see the unknown sample in relation to the known samples without it influencing the prediction.

1.5.11 Selected unknown Sample 3

The table below is the results for the LDA predicted class Corrigin meatworks and the confidence score as a percentage. The prediction result was correct as can be seen in the figure below the red arrow for the first sample taken as the 3rd blind sample fit completely within the blue oval of farm 3 (Corrigin meatworks) for all the significant analytes were the analytes as reduced to an x, y value. By only using the significant analytes we were able to clearly separate the classes and therefore achieve very high accuracy.

Table 5.2.3 Third unknown sample prediction using LDA

Unknown Sample Number	LDA Predicted Class	Score
Batch #: 21S1883 - Sample #: 23321	Corrigin meatworks	99.9983

Figure 5.2.3 Third unclassified sample plotted as a linear discriminant analysis scatter chart

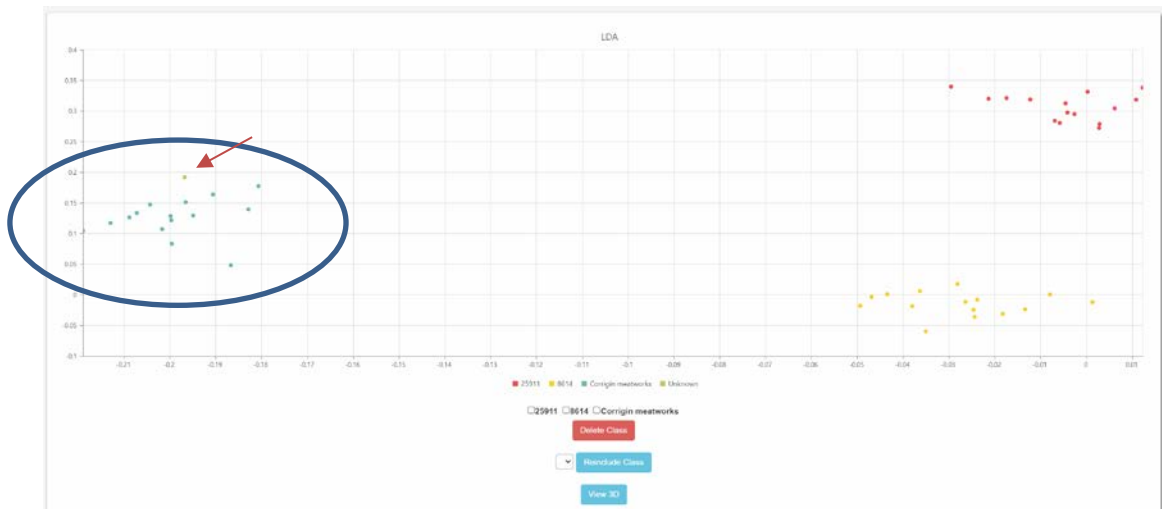
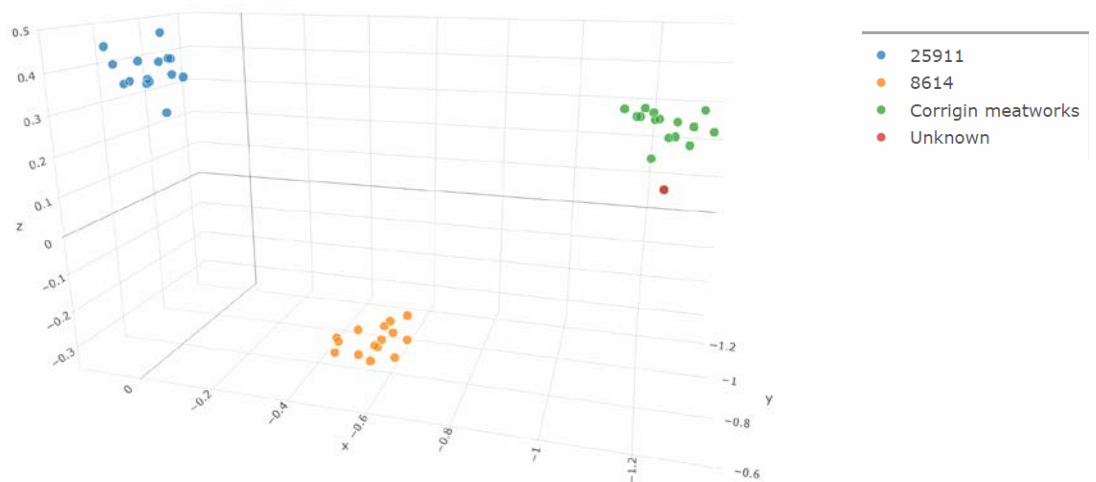


Figure 5.2.4 Third unclassified sample plotted as a linear discriminant analysis scatter chart in 3D



TraceBase has the ability to extent the projection to three dimensions x, y and z in an interactive report that allows the analyst to zoom and rotate in 3D space.

Neural Network Prediction Confidence

To achieve a high level of confidence in the predictions a neural network's hyperparameter or shape and design must directly relate to the type and scale of the data. Every time we run a prediction of an unknown sample the system will automatically retain a model this the optimal size and shape related to the input data.

Table 5.3.1 Neural Network Hyperparameters

Description	NN Confidence
Input Layers	24
Hidden Nodes	3312
Hidden Layers	83
Output Nodes	3
Iterations	5000

The size of your neural network is crucial. Too small and it will have insufficient synapsis connections to accuracy predict an output. If you make it too big the network will diverge bringing your overall confidence down.

Table 5.3.2 Neural Network Statistics and Results

Blind Sample Number	Classification	Time to Execute (Seconds)	NN Confidence
Batch #: 21S1883 Sample#: PT163009	25911	1.5	100%
Batch #: 21S1883 Sample#: PT162924	8614	1.33	100%
Batch #: 21S1883 Sample #: 23321	Corrigin Meatworks	1.43	100%

As can be seen in the table above a neural network is a very fast and accurate model to predict pork data.

This model uses a 80 – 20 rule to train the network. The data is split into two groups for each class: a train and a test set. The test data consists of 20% of the reference set and the train data is the remaining 80% of the set. The test data is randomly selected from each of the output classes. The network is then trained with the 80% of the data and then the test data is then used to verify the accuracy of the network. Once the best possible predictive confidence level is through an iterative process the model is then considered trained. The next step is to pass the unknown sample into the network and the network produce a prediction to one of the output nodes and a confidence calculation for the selected output node or in our case a farm of origin.

Discussion

One of the biggest challenges with the interpretation of the results is the labour-intensive process of interpreting them and another is finding skilled Analytical Chemists to do that interpretation. The process requires you to compare every element against every other element for each sample and then each of the farms to determine the list of elements and their respective concentrations that will allow you to group the farms mathematically. Once you have identified the significant elements you are able to match an unknown sample using those pre-determined elements unique to that farm. This is a very time-consuming process and requires a great deal of skill and knowledge to perform accurately.

The manual process of interpretation utilises a single machine learning method and the selected analytes used for this test are subjective to the Chemical Analyst and could cause some inconsistencies. The manual process can take between 4 hours and 2 days given the complexity of the samples, and the experience of the analyst.

TraceBase AI uses three different models and the method is calculated and auditable. The TraceBase AI was able to clearly identify the significant analytes and separate the samples from the three different producers mathematically within 10 minutes of uploading the dataset. It removes the subjective nature of the manual method, providing irrefutable provenance of the samples if the farm is included in the dataset.

TraceBase AI consists of a scientific database specifically designed to store each element or analyte individually so that you can interrogate any analyte to any other analyte for each or all the samples. TraceBase is connected to a powerful AI that automatically trains three different machine learning models based on the input data. In data science one of the key criteria for success in predictive models is relevant data, and TraceBase AI identifies the significant analytes automatically.

Research conducted by Meat & Livestock Australia (MLA) (Project Code: V.MFS.0447 published August 2020) proposes a need to validate the fraud issue for Australian red meat in global markets and supply chains, and the preparation of an industry response, stating the nature and extent of fraudulent misrepresentation of Australian red meat is not accurately known. Similar can be said about the Australian pork industry.

MLA also published a report on (Project Code: P.PSH.1170) - Compositional Traceability – Origin Fingerprints for Australian Beef and Lamb using elemental fingerprinting.

The quality of the Physi-Trace data received from ChemCentre is sufficient to mathematically separate the pork samples into farms (not only country of origin). With an unknown sample TraceBase can accurately predict the farm of origin using TraceBase AI.

Another major benefit of TraceBase is the ability to track changes over time, which could require only 2 – 3 samples per producer per year as opposed to the 10 samples normally required to maintain the farm fingerprint. If the deviation in fingerprint annually is minimal, we can further delay the testing period and avoid unnecessary costs until significant feed or production changes occur, which would require re-sampling to establish the new baseline for the farmgate fingerprint.

Due to the low variance in chemical concentrations between pork samples from the same state using existing methodologies would be too time consuming to interpret commercially. Using a single machine learning technique does not provide you with the verification or confidence that the specific algorithm used, was able to correctly classify your sample as it is only a prediction.

Implications & Recommendations

Governments and industry associations around the world are grappling with traceability, counterfeiting and substitution issues. Consumers are increasingly demanding to know where products come from.

Whether it is gold, diamonds, beef, prawns, fish or pork, consumers and producers are demanding provenance. TraceBase provides a pathway to provenance that currently does not exist and can be applied in any industry or product.

TraceBase can be easily integrated into Block Chain products, QR codes, labelling programs and any other developments aimed at preventing the loss of revenue from fraud and substitution.

Next Steps

This POC provides compelling evidence that TraceBase can produce meaningful and effective fingerprints for the farms already covered in the existing Physi-Trace dataset. The benefit of doing this is identifying gaps in the current data preventing the creation of the unique fingerprint for producers.

TraceBase will classify samples taken from the same farm over different periods to see the variation or drift between the intervals. This will allow TraceBase to identify the frequency and volume of sampling needed to maintain the reference dataset for each producer. If the deviation in fingerprint is minimal, we can extend the testing period until significant feed or process changes occur to require further sampling and re-establish the baseline for the farmgate fingerprint.

Benefits

Cost reduction

Current test/analysis costs are \$100/sample with 10 samples currently required for each producer, hence \$1000 annual cost/producer. By using TraceBase to optimise the number of samples for the effective producer database maintenance, the cost will be reduced to \$300 or maybe even \$0 depending on the frequency of testing required.

Current Manual Interpretation rate using an analytical chemist is \$200/hr for a period of 4 to 16 hours depending on the complexity of the data and the experience of the chemist. The recent POC project would have utilised 4 hrs.

In comparison TraceBase AI rate is \$50/hr and the time taken depend on volume and complexity) eg POC - (10 minutes)

Time saving

The interpretation of the results is virtually instantaneous after the analysis data is made available in the platform from the testing centre so the Total time to produce a traceback is greatly reduced. This could be even further reduced with the use of an App on a mobile device.

Future proofing

Future proofing by ensuring a premium quality product which is sustainably and ethically produced, traceable and trackable through the complete supply chain, and verifiable through AI

Meat grading and Quality assurance

InterpreData is currently working with a partner on a proposal with MSA to automate the Meat grading process for Beef. With the Marbling score link with eating quality, Interpredata will provide the industry with an automated grading process guaranteeing Premium for producers and top quality for consumers.

Additional services

Industry protection through traceback for biosecurity, residue detection and breed definition.

Intellectual Property

All background IP is under patent to the relevant participating organisation.

No project IP was generated

Technical Summary

We used existing software and data science methods to prove trace elemental analysis of pork for provenance and traceability.

Literature cited

Machine Learning

1. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

Fishers Linear Discriminant Analysis

2. Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems" (PDF). *Annals of Eugenics*. 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227.

Neural Network

3. Hinton, G. E., Osindero, S. and Teh, Y. A fast learning algorithm for deep belief nets. *Neural Computation* 18, pp 1527-1554. <https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf>

Publications Arising

No publications have resulted from this proof of concept