



An Information Technology-based platform for microbiological quality



Karadeniz, O.¹ Wong, A.¹, Heffer, S.¹, Fengou, L.², Schultz, N.³, Carstensen, M.³, Nychas, G.² and **Mohareb, F.^{1*}**
1 Bioinformatics Group, Cranfield University, College Road, Cranfield, Bedfordshire MK43 0AL, UK
2 Laboratory of Microbiology and Biotechnology of Foods, Agricultural University of Athens, Iera Odos 75, Athens 11855, Greece

Background

Rapid and efficient solutions to assess food quality throughout the food chain are essential to be fulfil the consumer’s expectations while ensuring quality and safety. Several rapid, non-destructive analytical methods such as multi- and hyper-spectral imaging have been successfully applied for monitoring food quality, often by coupling their output to a machine learning or pattern recognition predictive model.

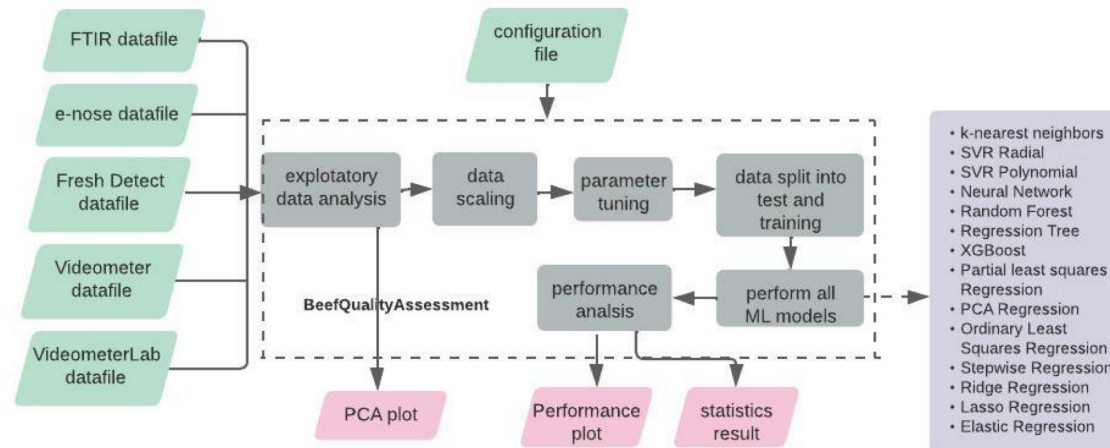
SorfML (www.sorfml.com) is a data science platform for food science, to automate the process of identifying, optimising and validating the most suitable machine learning algorithms for a given analytical platform and food product.

The “BeefQualityAssessment” R Package

SorfML’s functionalities were expanded via the implementation of a novel R package BeefQualityAssessment, which provides a comprehensive list of ML algorithms for the user to choose from. The package can be deployed as part of the SorfML platform, or as a standalone library in R. The user can choose between different ML algorithms to build the quality prediction model, or apply and rank them according to their predictive power and performance.

The package was tested using sensor data provided by AUA, Greece for monitoring beef quality. This included datasets from 5 analytical platforms viz. are FTIR, e-nose, MSI, Videometer and fresh detect.

The R package automate the process of data scaling, parameter tuning and cross-validation.



	RMSE_Statistics				
	Videometer	ENOSE	FTIR	Fresh Detect	VideometerLab
Neural Network	0.18	0.29	0.24	0.23	0.21
k-nearest neighbors	0.22	0.28	0.24	0.21	0.21
SVM-Radial	0.06	0.03	0.07	0.11	0.05
SVM-Polynomial	0.12	0.41	0.24	1.10	0.19
Regression Tree	1.38	1.73	1.51	1.46	1.51
XGBoost	0.05	0.10	0.44	0.01	0.08
Random Forest	1.38	1.87	1.21	1.15	1.06
Partial least squares Regression	0.21	0.26	0.22	0.23	0.18
PCA Regression	0.22	0.26	0.23	0.23	0.25
Ordinary Least Squares Regression	1.22	1.88	1.47	1.84	1.03
Stepwise Regression	1.23	1.78	1.44	1.7	1.01
Ridge Regression	1.04	1.65	1.32	1.11	1.05
Lasso Regression	1.09	1.64	1.29	1.13	1.01
Elastic Regression	2.15	1.86	2.19	2.14	2.16

- XGBoost and SVM-Radial were the best performers accors the board, while, while Elastic Regression was the worst, followed by Ordinary Least Squares Regression.
- Among the tree-based algorithms, XGBoost outperformed the others
- Regression Tree failed to provide good prediction, which is likely due to the overfitting problem.
- Neural Network and k-nearest neighbours performed similarly

Availability: <https://github.com/ozlemkaradeniz/BeefQualityAssessment>

Email: f.mohareb@cranfield.ac.uk

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 861915.

