Innovative feature-driven machine learning and deep learning for finance, education, and healthcare

Luca Parisi[1] · Marianne Lyne Manaog[2]

1 Introduction

There is a considerable body of literature on feature extraction and selection algorithms that could enable feature-driven ML for specific applications; however, it presents either theoretical mathematical algorithms that do not lead to tangible applications or approaches that lack feasibility in the field of interest, not being reusable in other domains either.

Thus, this special issue on the "'Babel Fish' for Feature-driven Machine Learning" (ML) sought for publications highlighting novel practical or potential applications of feature-driven ML to use cases including but not limited to financial services, wherein it has been a proven technology in industry, to more cutting-edge applications in healthcare, public sector, engineering, and education. The authors, including the Lead Guest Editor, Luca Parisi (PhD, MBA), would like to thank the Co-Guest Editors for their input in setting up the special issue and their support: Alin Ungureanu (MBA), Lee Andrew Kissane (MD, PhD), and Richard Tranter (MD, PhD). Furthermore, the Guest Editors would like to thank Professor John MacIntyre, Editor in Chief of the journal "Neural Computing and Applications," for his support and the opportunity to create this special issue.

The Guest Editors welcomed submissions of original research articles involving methods outlined in the Sect.2 encompassing both real-life applications and those experiments or simulations showing potential and feasibility for real-life applications. Although the topics could be those mentioned in the Sect. 3 or even broader, preference was given to those applications involving mental health and psychiatry for healthcare-related articles, as well as capital investment, trading, and cybersecurity for the financial services-related studies.

2 Objectives

The feature-driven applications presented should have leveraged one or more of the following to help in solving either classification or regression problems: feature selection, feature extraction, feature engineering. The problems defined should have been generalised enough for the solutions developed to be so too, such that they have been or can be applied in other domains as well. The feature-driven approaches must have led to the following requirements being met in the resulting ML-based decision support systems: accuracy, completeness, reliability and explainability, i.e., ease of interpretability from a user standpoint, e.g., clinicians for healthcare-related applications, business professionals for financial services-related ones.

3 Topics of interest

The Guest Editors invited the submission of high-quality papers related to one or more of the following topics:

- Machine Learning for feature extraction
- Machine Learning for feature engineering
- Machine Learning for feature selection
- Hybrid feature engineering
- Feature-driven Machine Learning for Financial Services - in particular, to improve capital investment, trading and for cybersecurity.

✉ Luca Parisi
luca.parisi@ieee.org

Marianne Lyne Manaog
marianne.manaog@ieee.org

1 Faculty of Business and Law (Artificial Intelligence Specialism), Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom

2 School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

- Feature-driven Machine Learning for Healthcare - in particular, to support personalised treatments and predict treatment response for patients with mental health disorders
- Feature-driven Machine Learning for Public Sector
- Feature-driven Machine Learning for Engineering
- Feature-driven Machine Learning for Education
- Feature-driven Machine Learning for Enhanced Pattern Recognition
- Feature-driven Machine Learning for Image and Video Classification
- Feature-driven Machine Learning for Isolated and Continuous Speech Processing
- Feature-driven Machine Learning for Text Classification
- Feature-driven Machine Learning for Anomaly Detection
- Ensembles of Feature-driven Machine Learning approaches

The Guest Editors were delighted to receive eleven high-quality submissions; nevertheless, considering novelty, scientific rigour, and practical usefulness as key criteria for acceptance, along with further ones as per the journal "Neural Computing and Applications", they had to be selective and could only accept seven articles (63.64% of those received) for publication.

4 Accepted articles and high-level assessment of quality

The seven accepted and published papers in this special issue were holistically evaluated based on their scientific rigour and potential for practical applications via the MedIntellego Quality Assessment Scale (MQAS), as per Table 1. As the maximum start rating (18/31) of the seminal article on the MQAS is only 3.23% lower than the minimum (17/31) of the articles in this special issue (Table 1), these results demonstrate that the quality of the seven accepted and published papers in this special issue is objectively very good (median +/− median absolute deviation = 19 +/− 2 (out of 31), i.e., 61.29% +/− 6.45%). Table 1 shows that the articles of Cousins et al. and Selvam and Rajendran were the top two papers overall.

A further sub-set analysis was performed to assess the:

- Theoretical potential based on the stars (out of 17) obtained in the sections "pre-processing", "methods", "validation", and "performance" of the MQAS, as per Table 2.
- Practical usefulness based on the stars (out of 17) obtained in the sections "rationale", "objective", "classification", "advantages", "potential for

translational impact", and "practical implications" of the MQAS, as per Table 3.

Table 2 shows that the articles of Selvam & Rajendran and Nouri-Moghaddam et al. were the top two papers overall when considering their theoretical potential.

Table 3 shows that the articles of Cousins et al. and Huang et al. were the top two papers overall when considering their practical usefulness alone.

5 Detailed review of the accepted articles

The first paper of Abdullah et al. created the dataset "E-learningDJUST" that contains information on the progress with four e-learning courses undertaken by students enrolled at the Jordan University of Science and Technology (JUST) during the COVID-19 pandemic (from 2020 to 2021). They identified a statistically significant and high correlation between e-learning-related events and the students' final grade and were able to predict students' performance via both decision tree-based (Random Forest and XGBoost) and a feed forward neural network-based models with RMSE values ranging from 7 to 17. Leveraging feature selection led to the Random Forest yielding the highest predictive performance quantified by an RMSE ranging from 0 to 0.20. Nevertheless, the percentage of students achieving high grades during the pandemic was like that prior to it.

The second paper by Cousins et al. used a psychiatry-related dataset including baseline and self-completed reviews from the web-based tool "Psynary" involving 458 subjects across Australia, New Zealand, and Japan. They coupled a fully connected neural network-based feature extractor with a Long-Short Term Memory (LSTM) deep neural network to predict remission in patients with any subtypes of depression with 76.5% accuracy by solely leveraging two reviews. Noteworthily, they were able to predict changes in medications with 97.4% precision and 71.4% recall, which helps to reduce the time taken to optimise treatments for patients with depression, thus enabling earlier remission and enhancing their quality of life.

The third paper by Huang et al. used datasets from two Portuguese secondary schools to predict students' performance via a feature weighted-support vector machine (SVM) and an artificial neural network (ANN). The weighted features of the SVM were determined by leveraging information gain ratios and then the SVM was used to perform a coarse-grained binary classification to predict pass or fail (RMSE of 0.2364 when using the main features only. Thereafter, grades were split from D to A+, and the ANN was leveraged to perform a fine-grained, multi-class

Table 1 MQAS-based stars attributed to published studies in the special issue "'Babel Fish' for Feature-driven machine learning: from financial services to healthcare"

Article	Rationale	Objective	Classification			Pre-processing				Methods								Validation		Performance			Advantages				Potential for translational impact		Practical implications			Total
	1	1	1	2	3	1	2	3	4	1	2	3	4	5	6	7	8	1	2	1	2	3	1	2	3	4	1	2	1	2	3	
Abdullah et al.	☆	☆	☆	☆	☆	–	–	–	–	☆	–	☆	☆	☆	☆	–	☆	☆	–	–	–	☆	☆	☆	–	☆	–	–	☆	–	☆	17
Cousins et al.	☆	☆	☆	☆	☆	☆	–	–	–	☆	☆	☆	–	☆	☆	–	☆	☆	–	–	–	☆	☆	–	☆	☆	–	–	☆	☆	☆	22
Huang et al.	☆	☆	☆	☆	☆	☆	–	☆	☆	☆	–	–	☆	☆	☆	☆	☆	☆	☆	–	–	☆	☆	–	☆	☆	–	☆	☆	–	☆	21
Nouri–Moghaddam et al.	☆	☆	☆	☆	☆	☆	☆	–	–	☆	–	☆	☆	☆	☆	☆	☆	☆	☆	–	–	☆	☆	–	☆	☆	–	–	–	–	–	19
Selvam and Rajendran	☆	☆	☆	–	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	–	☆	–	☆	☆	–	–	☆	☆	–	☆	–	–	22
Wang et al.	☆	☆	☆	☆	☆	–	☆	–	☆	☆	–	☆	–	☆	☆	☆	☆	☆	☆	–	–	☆	☆	–	–	☆	–	–	☆	–	☆	18
Zeng et al.	☆	☆	☆	☆	☆	☆	☆	☆	–	☆	–	☆	☆	☆	☆	☆	☆	–	–	–	–	–	☆	–	☆	☆	–	–	☆	–	–	18

Table 2 The seven accepted and published papers ranked based on their theoretical potential alone (out of 17 stars)

Article	Sub-total out of 17 (%)
Abdullah et al.	8 (47.06%)
Cousins et al.	11 (64.71%)
Huang et al.	11 (64.71%)
Nouri-Moghaddam et al.	12 (70.59%)
Selvam and Rajendran	14 (82.35%)
Wang et al.	9 (52.94%)
Zeng et al.	10 (58.82%)

Table 3 The seven accepted and published papers ranked based on their practical usefulness alone (out of 14 stars)

Article	Sub-total out of 14 (%)
Abdullah et al.	9 (64.29%)
Cousins et al.	11 (78.57%)
Huang et al.	10 (71.43%)
Nouri-Moghaddam et al.	7 (50%)
Selvam and Rajendran	8 (57.14%)
Wang et al.	9 (64.29%)
Zeng et al.	8 (57.14%)

classification (RMSE of 0.5472 when using the main features only) of the pass and fail classes obtained previously via the SVM.

The fourth paper by Nouri-Moghaddam et al. used microarray data to select appropriate genes via a hybrid solution leveraging a multi-filter and adaptive chaotic multi-objective forest optimisation algorithm (AC-MOFOA) along with an ensemble classifier. They used the multi-filter ensemble model to reduce the dimensionality of the data, along with other five filter-based methods to detect and remove redundant and irrelevant genes. Leveraging the reduced features from AC-MOFOA, they used the Kernel Extreme Learning Machine (KELM) to classify nine publicly available microarray datasets, achieving a higher performance than state-of-the-art ensemble methods on seven of these nine datasets.

The fifth paper by Selvam and Rajendran developed and validated an automatic feature engineering framework for modelling the trend-cycle (toffee-tree) to aid time-series forecasting. They were able to demonstrate that 286 domain-agnostic features could be engineered to model the trend-cycle. Thereafter, they leveraged a Light Gradient Boosted Machine (LightGBM) and SHapley Additive exPlanations (SHAP) to select relevant features. They then

compared the performance of a Multi-Layer Perceptron (MLP) and an LSTM, and they found that the LSTM achieved 20% better predictive performance, as measured via the Symmetric Mean Absolute Percentage Error (SMAPE).

The sixth paper by Wang et al. developed and validated an unpaired sample augmentation and multi-scale network to enhance the segmentation of the ribs based on chest X-ray images. Pneumonia-related texture changes were learned and augmented samples were generated. Thereafter, a multi-scale network was leveraged to learn hierarchical features, which were leveraged to improve the reliability of multi-organ segmentation networks. Noteworthily, the Dice Similarity Coefficient (DSC) of their proposed algorithm (88.03%) was 3% higher than that of state-of-the-art methods, achieving a precision of 88.25% and a recall of 88.36%, thus being more robust in presence of overlapping and fuzzy regions of multiple organs.

In the seventh and last paper of this special issue, Zeng et al. developed a novel deep belief network (DBN)-based multi-task learning model to classify Alzheimer's disease (AD) and mild cognitive impairment (MCI), including further differentiating between progressive MCI (pMCI) and stable MCI (sMCI). Via dropout and zero masking, they were able to avoid overfitting and enhance the generalisation and reliability of the DBN model. Further to pre-processing Magnetic Resonance Imaging (MRI) scans, they leveraged principal component analysis (PCA) to reduce the dimensionality of the data and multi-task feature selection to retain relevant features. Using the ADNI dataset, their DBN model achieved accuracies ranging from 87.78% to 99.62% across all the above-mentioned classification tasks, thus being a promising tool to aid the diagnosis of AD.