University of Bradford eThesis

This thesis is hosted in Bradford Scholars – The University of Bradford Open Access repository. Visit the repository for full metadata or to contact the repository team.

© University of Bradford. This work is licenced for reuse under a Creative Commons Licence.
MACHINE LEARNING-BASED FEATURE SELECTION AND OPTIMISATION FOR CLINICAL DECISION SUPPORT SYSTEMS

Luca PARISI

Submitted for the Degree of

PhD by Published Work

Faculty of Engineering and Informatics
University of Bradford

2019
Abstract

Luca Parisi

Machine Learning-based Feature Selection and Optimisation for Clinical Decision Support Systems

Optimal Data-driven Feature Selection Methods for Binary and Multi-class Classification Problems: Towards a Minimum Viable Solution for Predicting Early Diagnosis and Prognosis

Keywords: Machine Learning; Genetic Algorithms; Feature Selection; Optimisation; Clinical Decision Support.

This critical synopsis of prior work by Luca Parisi is submitted in support of a PhD by Published Work. The work focuses on deriving accurate, reliable and explainable clinical decision support systems as minimum clinically viable solutions leveraging Machine Learning (ML) and evolutionary algorithms, for the first time, to facilitate early diagnostic predictions of Parkinson's Disease and hypothermia in hospitals, as well as prognostic predictions of optimal postoperative recovery area and of chronic hepatitis. Despite the various pathological aetiologies, the underlying capability of ML-based algorithms to serve as a minimum clinically viable solution for predicting early diagnosis and prognosis has been thoroughly demonstrated. Feature selection (FS) is a proven method for increasing the performance of ML-based classifiers for several applications. Although advances in ML, such as Deep Learning (DL), have denied the usefulness of any extrinsic FS by incorporating it in their architectures, e.g., convolutional filters in convolutional neural networks, DL
algorithms often lack the required explainability to be understood and interpreted by clinicians within the context of the diagnostic and prognostic tasks of interest. Their relatively complicated architectures, the hardware required for running them and the limited explainability or interpretability of their architectures, the decision-making process – although as assistive tools driven by the algorithms’ training and predictive outcomes have hindered their application in a clinical setting. Luca Parisi’s work fills this translational research gap by harnessing the explainability of using traditional ML- and evolutionary algorithms-based FS methods for improving the performance of ML-based algorithms and devise minimum viable solutions for diagnostic and prognostic purposes. The work submitted here involves independent research work, including collaborative studies with Marianne Lyne Manaog (MedIntellego®) and Narrendar RaviChandran (University of Auckland). In particular, conciliating his work as a Senior Artificial Intelligence Engineer and volunteering commitment as the President and Research Committee Leader of a student-led association named the “University of Auckland Rehabilitative Technologies Association”, Luca Parisi decided to embark on most research works included in this synopsis to add value to society via accurate, reliable and explainable, hence clinically viable applications of AI. The key findings of these studies are: (i) ML-based FS algorithms are sufficient for devising accurate, reliable and explainable ML-based classifiers for aiding prediction of early diagnosis for Parkinson’s Disease and chronic hepatitis; (ii) evolutionary algorithms-based optimisation is a preferred method for improving the accuracy and reliability of decision support systems aimed at aiding early diagnosis of hypothermia; (iii) evolutionary algorithms-based optimisation
methods enable to devise optimised ML-based classifiers for improving postoperative discharge; (iv) whilst ML-based algorithms coupled with ML-based FS methods are the minimum clinically viable solution for binary classification problems, ML-based classifiers leveraging evolutionary algorithms for FS yield more accurate and reliable predictions, as reducing the search space and overlapping regions for tackling multi-class classification problems more effectively, which involve a higher number of degrees of freedom. Collectively, these findings suggest that, despite advances in ML, state-of-the-art ML algorithms, coupled with ML-based or evolutionary algorithms for FS, are enough to devise accurate, reliable and explainable decision support systems for performing both an early diagnosis and a prediction of prognosis of various pathologies.
First, I would like to thank my supervisors Professor Daniel Neagu and Professor Felician Campean for their time, invaluable inputs and well-rounded advice, continuous support, and inspiration in guiding me throughout this fantastic journey. Without their vision and belief, distilling the essence of my contributions to the field of Applied Machine Learning (ML) for clinical decision support systems and synthesising it in this thesis cohesively would have not been possible.

An important mention goes to my two academic mentors, Dr Mansour Youseffi from the University of Bradford (UK) and Dr Rhys Pullin from Cardiff University (UK), who have been a constant light in front of me of guidance and support throughout my academic and professional career in industry as well, helping me grow as a person and as an engineer.

A warm thank you goes to my parents for their love, their guidance by being my epitomes of honest hard work and sacrifice with a vision, continuous support, for always believing in me and trusting in my ability to identify and leverage opportunities in life to maximise my contribution to science and society.

Most importantly, a unique and wholehearted thank you to my unique half, my wife, for her unconditional love and not only for supporting me throughout the journey of life and this chapter, but also for being my constant memento of the incommensurable and beautiful value life has when shared with the right person for us.
I would like to thank everyone who has supported me in my journey in AI so far, both in academia and industry, especially the University of Bradford (UK) and the International Institute for Advanced Scientific Studies (IIASS) from Italy, which offered me the first opportunity to learn about AI, and Accenture, which has given me the opportunity to maximise the impact of my applications as a Specialist first and now as an Associate Manager in AI via Applied ML at scale.
Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Luca Parisi
Publications and Presentations

Papers in Peer-Reviewed Journals and Conference Proceedings


Conference Abstracts and Presentations


Support Tool to Aid Prediction of Prognosis in Patients with Hepatitis. The 16th International Conference on Biomedical Engineering, National University of Singapore (NUS), Singapore.
# Table of Contents

Abstract.................................................................................................................................................. i

Acknowledgements................................................................................................................................ iv

Declaration.............................................................................................................................................. vi

Publications and Presentations........................................................................................................... vii

Table of Contents .................................................................................................................................. ix

List of Figures .......................................................................................................................................... xi

List of Tables .......................................................................................................................................... xii

Glossary .................................................................................................................................................. xiii

1 Introduction ............................................................................................................................................. 1

1.1 Justification for the work................................................................................................................. 1

1.2 Context and scope of the work ...................................................................................................... 8

1.3 End-user feedback to inform system requirements................................................................. 10

1.4 Research aims and research questions ....................................................................................... 13

1.5 Methodology ................................................................................................................................ 15

1.6 Background discussion of own publications ............................................................................ 20

1.7 Thesis outline ................................................................................................................................. 22

2 State of the art ..................................................................................................................................... 24

2.1 Research background ................................................................................................................... 24

2.1.1 Machine Learning to aid early diagnosis of Parkinson’s Disease ........................................... 24
2.1.2 Machine Learning to aid prediction of prognosis in patients with chronic hepatitis ............................................................. 29

2.1.3 Machine Learning to improve postoperative discharge decision making 33

2.2 Conclusions ........................................................................................................................................................................ 37

3 Contributions to Machine Learning and Evolutionary Algorithms for clinical decision-making support ........................................... 39

3.1 Narrative framework ......................................................................................................................................................... 39

3.2 Preliminary work ............................................................................................................................................................. 40

3.3 Machine Learning and Evolutionary Algorithms for decision support .................................................................................................. 43

3.4 Machine Learning for intrinsic feature selection and Genetic Algorithm for optimisation ......................................................................................................................... 47

3.5 Clinically viable solution to predict diagnosis and prognosis .......... 52

3.6 Discussion ........................................................................................................................................................................ 55

3.6.1 Machine Learning to augment clinical decision making ........ 55

3.6.2 Machine Learning for feature selection and Evolutionary Algorithms for optimisation .......................................................... 57

3.6.3 Clinician-aided evaluation to achieve a clinically viable diagnostic and prognostic tool ......................................................... 61

3.6.4 Considerations for applications at scale ........................................ 65

3.7 Conclusions ........................................................................................................................................................................ 69

4 Research findings meeting aims and objectives............................... 71
4.1 Salient points addressing research questions................................. 71

5 Conclusions and future research ..................................................... 74

5.1 Conclusions..................................................................................... 74

5.2 Limitations and value from published work ..................................... 76

5.3 Future work..................................................................................... 79

References........................................................................................... 82

Appendix A – Statement of independence .............................................. 93

A.1 Specific contributions by publication............................................... 94

A.2 Co-authors confirmation.................................................................. 96

A.3 Approvals from publishers to share published work....................... 103

A.4 Relevant milestones for published work to date............................. 108

Appendix B – Published works for consideration.................................. 112

Appendix C – Supporting documents.................................................... 147

List of Figures

Figure 1. Selected Parkinson’s Disease-related features.......................... 17

Figure 2. Decision support system to aid early diagnosis of Parkinson’s Disease ................................................................. 17

Figure 3. Selected chronic hepatitis-related features .............................. 18

Figure 4. Decision support system to aid early prediction of prognosis of chronic hepatitis................................................................. 19
Figure 5. Decision support system to improve postoperative discharge decision making ................................................................. 21
Figure 6. Validation of performance of the decision support system to improve postoperative discharge decision making ........................................................................... 44
Figure 7. Validation of performance of the decision support system to aid early diagnosis of Parkinson's Disease ........................................................................... 49
Figure 8. Validation of performance of the decision support system to aid early prediction of prognosis of chronic hepatitis ......................................................... 64

List of Tables

Table 1. Published journal papers submitted for inclusion in the PhD ........... 5
Table 2. Supporting materials ........................................................................... 7
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the receiver operating characteristic Curve</td>
</tr>
<tr>
<td>DL</td>
<td>Deep Learning</td>
</tr>
<tr>
<td>DRL</td>
<td>Deep Reinforcement Learning</td>
</tr>
<tr>
<td>EA</td>
<td>Evolutionary Algorithm</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>LSVM</td>
<td>Lagrangian Support Vector Machine</td>
</tr>
<tr>
<td>M-LSVM</td>
<td>Lagrangian Support Vector Machine for multi-class classification</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-Layer Perceptron</td>
</tr>
<tr>
<td>M-MLP</td>
<td>Multi-Layer Perceptron for multi-class classification</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s Disease</td>
</tr>
<tr>
<td>RL</td>
<td>Reinforcement Learning</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>M-SVM</td>
<td>Support Vector Machine for multi-class classification</td>
</tr>
<tr>
<td>GA-M-LSVM</td>
<td>GA-LSVM for multi-class classification</td>
</tr>
<tr>
<td>AvA</td>
<td>All-vs-All multi-class classification approach</td>
</tr>
<tr>
<td>OvA</td>
<td>One-vs-All multi-class classification approach</td>
</tr>
<tr>
<td>th</td>
<td>Threshold for calculating the area under the receiver operating characteristic curve (ROC-AUC)</td>
</tr>
<tr>
<td>UCI</td>
<td>University-California Irvine</td>
</tr>
<tr>
<td>FD</td>
<td>Fault Diagnosis</td>
</tr>
<tr>
<td>SOM</td>
<td>Self-Organising Map</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean-Squared Error</td>
</tr>
<tr>
<td>ACC</td>
<td>Classification Accuracy (on the testing set)</td>
</tr>
<tr>
<td>SN</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>SP</td>
<td>Specificity</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>HBV</td>
<td>Hepatitis B</td>
</tr>
</tbody>
</table>
HCV  
Hepatitis C

HDV  
Hepatitis D

kNN  
k-Nearest Neighbour

CV  
Cross-Validation

SGOT  
Serum Glutamic-Oxaloacetic Transaminase

PCA  
Principal Component Analysis

OSH  
Optimal Separating Hyperplane

MQAS  
MedIntellego® Quality Assessment Scale (for Studies on Applied Artificial Intelligence in Healthcare)

RCT  
Randomised Clinical Trial

2D  
Two-dimensional

3D  
Three-dimensional

PSO  
Particle Swarm Optimisation

LPC  
Linear Predictive Analysis Coefficient

MFCC  
Mel-Frequency Cepstral Coefficient

TPR  
True Positive Rate

TNR  
True Negative Rate

FS  
Feature Selection

RBF  
Radial Basis Function
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPDRS</td>
<td>Unified Parkinson’s Disease Rating Scale</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>L-CORE</td>
<td>Patient’s internal temperature</td>
</tr>
<tr>
<td>L-SURF</td>
<td>Patient’s surface temperature</td>
</tr>
<tr>
<td>SURF-STBL</td>
<td>Patient’s surface temperature</td>
</tr>
<tr>
<td>CORE-STBL</td>
<td>Patient’s core temperature</td>
</tr>
<tr>
<td>BP-STBL</td>
<td>Patient’s blood pressure</td>
</tr>
</tbody>
</table>
This document is a critical synopsis of the published work of Luca Parisi (hereafter referred to as the author) relating to the enablement and facilitation of the predictions of an early diagnosis of Parkinson’s Disease and hypothermia in hospitals, as well as a prognosis of chronic hepatitis, via tailored applications leveraging Machine Learning (ML) for feature selection (FS) and Evolutionary Algorithms (EAs) for optimisation in presence of binary and multi-class classification problems respectively, which is herewith submitted for a PhD by Published Work.

1.1 Justification for the Work

The author’s contributions, specified in Appendices A.1 and A2, aim to advance the understanding and feasibility of ML and Evolutionary Algorithms for FS and optimisation to derive viable clinical decision support systems for diagnostic and prognostic purposes within the context of binary and multi-class classification problems correspondingly.

In recent years ML techniques have been increasingly used to support diagnosis of chronic hepatitis and that of hypothermia in hospitals, as well as prognostic assessment of patients with Parkinson’s Disease (PD). However,
as reported in Parisi et al. (2018a, 2018b, 2019), there are no accurate, reliable and explainable methods, collectively referred to as ‘clinically viable methods’ in this thesis, towards binary classification tasks for aiding early diagnostics, e.g., an early detection of PD, and early prediction of prognosis for patients with chronic hepatitis, as well as prognostic tasks involving multi-class classification problems, such as predicting the optimal postoperative recovery area. Devising and deploying such clinically viable methods would help to maximise patient benefit in a clinical setting. These three examples of diagnostic and prognostic problems, although not exhaustive, have been addressed by researchers via ML but they have not been able to derive easy-to-understand and clinically viable ML-based recommendations to clinicians.

Thus, the above-mentioned studies of Parisi et al. (2018a, 2018b, 2019), whose full-length papers, further to the approvals from the respective publishers in Appendix A.3, have been made available in Appendix B, seek to fill the following research gaps:

1. It is not clear whether intrinsic ML-based FS may be the preferred approach to help with solving binary classification problems (whether diagnostic or prognostic).

2. There is lack of evidence whereby EAs for optimisation may be the preferred approach to help with solving multi-class classification problems (whether diagnostic or prognostic).

3. There are currently no clinically viable methods leveraging the above-mentioned techniques, which are accurate, reliable, and explainable enough for them to be applied in a clinical setting.
et al. ML for some medical applications can be treated as a tool for solving binary classification problems, characterised by two states, that is ‘healthy’ or ‘pathological’. For example, early diagnosis of PD, i.e., whether an individual is likely to suffer from PD or not based on speech patterns (Parisi et al., 2018a), or early prognostics in patients with chronic hepatitis, i.e., whether a patient is likely to survive or die due to chronic hepatitis based on haematological patterns and medical history (Parisi & Manaog, 2017a; Parisi et al. 2019). To aid in solving such binary classification problems, the significance of Luca Parisi’s work demonstrates that intrinsic ML FS is effective for devising optimal ML-based classifiers (Parisi et al., 2018a; Parisi et al., 2019), with accurate, reliable and explainable results for clinicians, as well as retaining the required explainability in their architecture and decision-making process for clinicians to understand them. These findings help in filling the first above-mentioned research gap.

For multi-class classification problems, intrinsic ML FS was not enough; however, EAs, such as Genetic Algorithm (GA), were found to maximise the classification performance and reduce the computational cost of an LSVM for multi-class classification (M-LSVM) (Parisi et al., 2018b). The advantage of such methods was demonstrated for the prediction of the optimal postoperative recovery area based on benchmark clinical data, derived from the UCI ML database (Parisi et al., 2018b), which helps address the second high-level research gap listed above.

As Hansen & McDonald (2001) had pointed out, EAs, such as GA, provide an objective method for unconstrained optimisation of hyperparameters of ML-
based classifiers. Thus, GAs are not regarded as a sub-branch of ML but as a separate sub-field of computer science that can be leveraged in ML. Differently from heuristic and statistical hyperparameter optimisation methods (Hansen & McDonald, 2001), GA has been demonstrated to lead to consistently higher accuracy and reliability of ML-based multi-class classification methods for aiding prediction of postoperative discharge (Parisi et al., 2018b).

This hybrid multi-class classifier, optimised via GA, was proven accurate, reliable, and explainable to assist clinicians and nurses in improving postoperative discharge decision making (Parisi et al., 2018b).

Collectively, findings on these novel hybrid classifiers suggest they are clinically viable methods, thus filling the third high-level research gap reported above too. They provide innovative insights into the usefulness of intrinsic ML-based FS or optimisation via EAs to improve the performance of traditional ML classifiers for binary and multi-class classification problems, respectively.

In summary, the work demonstrates the clinical viability of using the simplest, the most explainable but the best performing (the most accurate and reliable) ML approaches for both diagnostic and prognostic tasks. Such solutions (Parisi et al., 2018a, 2018b, 2019) are not only more accurate and reliable but also more understandable than algorithms from the literature, which supports their clinical application, as described, and discussed in chapter 2.

The body of published work submitted for this PhD is itemised in Table 1. Such research outputs include three journal papers, one of which published by Springer-Nature and the other two published by Elsevier in two of the top 30 peer-reviewed journals in AI worldwide as per the most up to date SCImago
journal ranking in Artificial Intelligence. These papers have been selected because they summarise the novel methodology, whose Luca Parisi is the principal author, developed and validated to fill the three above-mentioned high-level research gaps were filled, as discussed in this synopsis. In fact, his research demonstrates that intrinsic ML-based FS and EAs-based optimisation, coupled with ML classifiers, enables to develop clinically viable decision support systems to aid prediction of early diagnosis and prognosis of various pathologies.

Table 1. Published journal papers submitted for inclusion in the PhD.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal Paper</th>
<th>Contribution</th>
</tr>
</thead>
</table>

PR**

P = available in the public domain; R = refereed journal paper; ** = principal author.

Additional supporting material is provided, as summarised in Table 2, which includes three conference abstracts on seminal work leading to the main findings and novelties hereby presented to fill the three main research gaps of interest, along with a further conference abstract in support of the mathematical usefulness of EAs-based optimisation for multi-class ML-based classifiers.

Moreover, in further support of the above-mentioned considerations, relevant milestones and accolades are included in Appendices A.4 and C.
Table 2. Supporting materials.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Conference Abstract</th>
<th>Contribution</th>
</tr>
</thead>
</table>
P = available in the public domain; R = refereed conference abstract; ** = principal author.

The conference abstracts in Table 2 (Parisi & Manaog, 2016; Parisi & Manaog, 2017a; Parisi & RaviChandran, 2018) and one presentation on the study of Parisi & Manaog (2017a) have been included in Appendices B and C. As further evidence of the impact of the author’s research, Appendix C also mentions the keynote talk given at the MATLAB Conference 2017 – the author was the only one presenting not working for MathWorks, given his work leveraged MATLAB was deemed worthy of being discussed at this conference - and the Young Investigator nomination from the National University of Singapore (NUS) in 2016.

1.2 Context and scope of the work

The author is a medical engineer who has worked in the field of AI research since 2014 and who collaborated with two organisations: a company specialised in providing personalised AI-based solutions in healthcare named MedIntellego®, New Zealand (Marianne Lyne Manaog’s team); University of Auckland Rehabilitative Technologies Association (UARTA), New Zealand (with Narrendar RaviChandran).

His role in these collaborations was a translational one: developing and implementing an AI engineering approach to mine clinical data and perform data-driven modelling on diagnostic and prognostic data. The aim of the work was to evaluate whether traditional ML algorithms, enhanced with FS methods based on ML or optimised via EAs, can represent clinically viable decision
support systems to aid in binary and multi-class classification problems correspondingly, e.g., to aid prediction of early diagnosis and prognosis of various pathologies, including PD and chronic hepatitis, as well as predict the optimal postoperative recovery area to improve clinical outcomes.

The author has led the conceptual and practical work on the development, optimisation, testing, and validation of the hybrid ML algorithms used in these research works.

The work has involved: (i) consulting experts in a clinical setting in understanding their requirements with respect to accuracy, reliability and explainability, designing minimum viable ML-based solutions to meet them with regards to computational architecture and underlying decision-making process to aid prediction of early diagnosis and prognosis of various pathologies, as well as selecting methods and metrics to evaluate their results whilst retaining explainability; (ii) strategic presentations to educate and inform the teams at MedIntellego® and UARTA, as well as both technical and the general audience, such as during the MATLAB Conference held in Auckland, New Zealand, in 2017, regarding the advantages of leveraging intrinsic ML-based FS methods and EAs for optimisation, to improve the accuracy and reliability of ML-based classifiers for diagnostic and prognostic purposes, whilst preserving their explainability; (iii) critical appraisal of the theoretical and practical research work of others, to inform the direction of the collaborative research; (iv) developing and implementing advanced ML techniques and EAs for both FS and optimisation to aid prediction of early diagnosis and prognosis based on clinical data.
The author’s knowledge of medical engineering, along with his expertise in ML and data science, have enabled him to make key contributions to the body of work submitted here. As such, he has pioneered the application of intrinsic ML FS for ML-based classification and demonstrated that these methods or EAs for optimisation, coupled with traditional ML classifiers, are enough for devising clinically viable decision support systems, for aiding predictions of early diagnosis of PD and prognosis of chronic hepatitis and patients with postoperative hypothermia respectively.

The work presented here represents Luca Parisi’s contributions to knowledge, which demonstrate that traditional ML algorithms, when improved by adequate, intrinsic ML-based FS or EAs-based optimisation methods, can be accurate, reliable, and explainable decision support systems.

1.3 End-user feedback to inform system requirements

As per the studies of Parisi et al. (2018a, 2018b, 2019) and the MedIntellego® Quality Assessment Scale (MQAS) for Studies on Applied Artificial Intelligence in Healthcare (Appendix, Parisi et al., 2019), the following primary end-user requirements for clinical viability were considered and well received by the end-users:

1. classification **accuracy** on the testing set equal to or higher than 80%, quantifying the predictive power of discriminating between the classes of interest, the only one that is clinically important, as it is based on the testing set of unknown data to the classifier;
2. **reliability** - always measured on the testing set - equal to or higher than 85%, defined as the predictive capability of attributing an appropriate degree of certainty on the classification outcome, quantifiable via sensitivity and specificity, and, more comprehensively, via the area under the receiver operating characteristic curve (ROC-AUC) based on a pre-set threshold. Generally, for binary classification problems, such as aiding early diagnosis of PD (Parisi *et al.*, 2018a) or early prediction of survival in patients with chronic hepatitis (Parisi *et al.*, 2019), the outcome for each patient is either ‘true’ or ‘false’, thus resulting in four possible scenarios (Grunkemeier & Jin, 2001). However, for a multi-class classification problem, such as aiding postoperative discharge decision making involved in Parisi *et al.* (2018b), the overall outcome is derived from each binary classification.

3. **explainability**, which is referred to the interpretability of the decision-making process leading to the binary or multi-class classification sought after, as well as of the results from such an algorithmic decision-making process. Whilst the first type of interpretability guides the design of the classification architecture to ensure it is the simplest possible without compromising the above-mentioned performance-related requirements, the second type of interpretability ensures the top features or factors leading to the classification sought after are can be clearly visualised and understood in terms of their relative importance towards the required classification. Interpretability of the decision-making process can be ensured by being able to qualitatively and quantitatively evaluate which features in the data are the most
influential ones and how important they are towards the required classification, e.g., ranking and visualising the most important features to aid early diagnosis of PD (Parisi et al., 2018a). Instead, the explainability of the results lies in visualising metrics computed using methods that are well understood by and regarded as clinical gold standard methods, such as accuracy, sensitivity, specificity, and ROC-AUC.

As the ROC-AUC is a more discriminating performance metric than accuracy, the former was the main criterion for deciding which classifier was the best performing one (Ling et al., 2003).

As a secondary requirement for an algorithm to be clinically viable, without compromising the above-mentioned three primary end-user requirements, a relatively low computational cost was also deemed important. The term “computational cost” was used in the studies of Parisi et al. (2018a, 2018b, 2019) to include both the hardware-dependent metric of computational time to execute an algorithm and the software-dependent number of training iterations (epochs) prior to reaching convergence. As the latter measure is not related to the hardware whereby the algorithm was evaluated, but it is directly associated with the algorithm itself, the number of epochs was more important than the computational time when assessing the performance of ML-based algorithms. In fact, a low enough number of epochs would ensure the algorithm could run on normal CPU hardware typical in public health for widespread applications of this technology.
In summary, high accuracy, reliability and explainability, ideally whilst maintaining a low computational cost, were the main criteria, as specified above, applied to select the most clinically viable algorithm for aiding the diagnostic and prognostic binary and multi-class classification problems of interest (Parisi et al., 2018a, 2018b, 2019).

1.4 Research aims and research questions

The overall aim of the work has been to devise clinically viable solutions as decision support systems for various diagnostic and prognostic purposes, such as predicting early diagnosis of PD, survival in patients with chronic hepatitis and the optimal postoperative recovery area to improve postoperative discharge decision making and prevent hypothermia in hospitals.

The published works aim to gain new insights into the appropriateness of either ML for FS or EAs for optimisation to maximise the predictive performance of patients’ diagnosis and prognosis considering various pathologies.

The research questions the author’s research and this thesis are addressing are the following:

R1. Can supervised ML be used to devise a clinically viable decision support system for aiding early medical diagnoses and prediction of prognosis in patients with different pathologies?

R2. Can ML-based FS improve classification performance (accuracy and reliability) when considering diagnostic and prognostic clinical decision
support systems, whilst retaining explainability of the underlying decision-making process?

R3. Can EAs-based optimisation improve classification performance (accuracy and reliability) when considering diagnostic and prognostic clinical decision support systems, whilst retaining explainability of the underlying decision-making process?

R4. How can ML and EAs improve classification performance (accuracy and reliability) when used for FS and optimisation, whilst retaining explainability?

The underlying aim of such a demonstration is to focus on accelerating their translational applications in a clinical setting and evaluating their performance for clinical viability. In fact, the ML-based classifiers of Parisi et al. (2018a, 2018b, 2019) are clinically viable diagnostic and prognostic solutions. The classification performance (accuracy and reliability) of such methods has been validated against the best performing algorithms from previous studies to date and commercially available tools, such as Microsoft Azure ML Studio and Alyuda NeuroIntelligence, on benchmark clinical datasets (Parisi et al., 2018a, 2018b, 2019).

Thus, Luca Parisi’s work is purposefully not focused on progressively incrementing the accuracy and reliability of existing methods by 1% or less than that with theoretically useful studies involving more advanced ML algorithms, such as DL and RL, as these delay the translational applications of research findings and, therefore, their clinical benefit.
1.5 Methodology

The specific methodology adopted as an approach to answering the research questions in 1.3, included the following steps:

O1. Study the performance of ML methods for supervised classification, develop and validate a novel methodology for aiding in early diagnosis of PD against the best performing algorithms from previous studies, using benchmark clinical datasets – related to R1, R2 and R4.

O2. Study the performance of ML methods for supervised classification, develop and validate a novel methodology for aiding prediction of prognosis in patients with chronic hepatitis against the best performing algorithms from previous studies, using benchmark clinical datasets – related to R1, R2 and R4.

O3. Study the performance of ML methods for supervised classification, develop and validate a novel methodology for aiding prediction of optimal postoperative recovery area to improve postoperative discharge decision making against the best performing algorithms from previous studies, using benchmark clinical datasets – related to R1, R2 and R4;

O4. Study the influence on predictive performance of the novel decision support systems developed in O1, O2 or O3 due to ML-based FS or GA-based optimisation – related to R1-R4.

O5. Devise and validate novel methodologies for ML-based FS or GA-based optimisation to improve the predictive performance (accuracy and reliability)
of the decision support systems developed in O1, O2 or O3 – related to R1-R4.

O6. Synthesise the work to gain deeper insights into mechanisms associated with improvements in learning performance due to either ML-based FS or GA-based optimisation – as O5, related to R1-R4.

The three examples of diagnostic and prognostic problems mentioned in O1-O3, although not exhaustive, can be solved via ML and, as a result, an additional ML-based recommendation could be provided to clinicians to assist them in their multi-factorial decision making.

Luca Parisi’s work demonstrates that intrinsic ML FS is suitable for devising optimal ML-based classifiers for diagnostic and prognostic applications involving binary classification problems (Parisi et al., 2018a).

In fact, 20 diagnostic features derived from ranking the weights computed via a Multi-Layer Perceptron (MLP) (Fig. 1) were used as inputs to a Lagrangian Support Vector Machine (LSVM) (Fig. 2) for aiding early diagnosis of PD with 100% classification accuracy and ROC-AUC, whilst retaining the lowest computational cost amongst all classifiers considered in the study and from the literature on the same benchmark dataset from the University of California-Irvine (UCI) ML database (Parisi et al., 2018a). The computational cost, although it was deemed of lower clinical importance with respect to the performance metrics to improve (accuracy and reliability) and explainability, included both hardware- and software-dependent metrics, i.e., computing time and number of epochs prior to convergence (Parisi et al., 2018a).
Figure 1. Pie chart illustrating the initial input features (N = 27) and the discarded features in red (N = 7) as per the feature ranking method based on the weights from a Multi-Layer Perceptron (MLP) algorithm. MLP-based weights were scaled as 51.3 units:100% for ease of visualisation of the results (Parisi et al., 2018a).

Figure 2. The hybrid algorithm leveraging a Multi-Layer Perceptron (MLP) for feature selection and a Lagrangian Support Vector Machine (LSVM) for classification to aid early diagnosis of Parkinson’s Disease (PD) mainly based on speech features. TPR: true positive rate; TNR: true negative rate; H: healthy; P: pathological (Parisi et al., 2018a).
16 prognostic features (Fig. 3) were, instead, selected via the weights derived by applying the geometric margin maximisation criterion of a LSVM as inputs to an MLP (Fig. 4), which was also able to achieve 100% accuracy and ROC-AUC in predicting the prognosis of patients with chronic hepatitis on two independent, benchmark datasets from the UCI ML database, the first one with the same follow-up time across all patients, the second with variable follow-up times as well (Parisi et al., 2019).

**Figure 3.** Pie chart illustrating the initial input features (N = 19) and the features excluded (N = 3) following feature selection via a Lagrangian Support Vector Machine (LSVM) algorithm. LSVM-based weights were scaled as 2.1 units:100% for ease of visualisation of the results. SGOT: serum glutamic oxaloacetic transaminase (Parisi et al., 2019).
Figure 4. The hybrid algorithm leveraging a Lagrangian Support Vector Machine (LSVM) for feature selection and a Multi-Layer Perceptron (MLP) for classification to aid prediction of survival in patients with chronic hepatitis mainly based on haematological data (Parisi et al., 2019).

Conversely, for multi-class classification, EAs, such as GA, were found to improve classification performance and reduce computational cost to predict the optimal area in the hospital in which patients should recover postoperatively based on benchmark clinical data (Parisi et al., 2018b).

To summarise, whilst intrinsic ML-based FS was found to aid in solving binary classification tasks, such as predicting early diagnosis of PD (Parisi et al., 2018a) and prognosis in patients with chronic hepatitis (Parisi et al., 2019), EAs for optimisation were found to improve multi-class classification performance, such as that required to assist with postoperative discharge decision making (Parisi et al., 2018b).
1.6 Background discussion of own publications

Whilst researchers are still striving to devise less explainable algorithms (e.g., DL and RL) to improve on previous results that are already satisfactory, the clinical translation of such ML-based applications has been delayed by such a theoretical effort.

Thus, the author's research aims to bring own translational applications to life, shedding light onto clinically viable solutions leveraging traditional ML algorithms that are not only accurate and reliable but also explainable for them to be applied in a clinical setting, also reducing the computational cost and the hardware required with respect to the DL and deep reinforcement learning (DRL) algorithms that have been recently suggested.

Existing decision-making support systems for aiding postoperative discharge heavily rely on statistical-based methods that lack objectivity in predicting optimal recovery area on a subject-specific basis. Instead, GA enabled to achieve an objective and ‘controlled’ All-vs-All multi-class classification to predict optimal postoperative recovery area by optimising the kernel and the hyperparameters (v and α) of the M-LSVM, which maximised its geometric margin and improved the delineation of an optimal separating hyperplane (OSH) to discriminate the input features into the target classes more linearly (Parisi et al., 2018b) (Fig. 5). In this instance, the novel hybrid algorithm (GA-M-LSVM) was found to yield the highest classification accuracy, improving the highest accuracy from the literature by almost 12% on the same benchmark clinical data (Parisi et al., 2018b).
Accurate and reliable postoperative discharge decision-making, as an example of multi-class classifier achievable via this hybrid, GA-optimised but explainable decision-making support system, can not only improve postoperative patient outcomes and survival, but also the management of hospital resources, both financial and human ones.

**Figure 5.** The hybrid algorithm leveraging a Genetic Algorithm (GA)-optimised All-vs-All (AvA)-based multi-class classification (M-)LSVM to improve postoperative discharge, as compared to traditional, ‘uncontrolled’ or non-parameterised One-vs-All (OvA) SVM. (Parisi et al., 2018b).
1.7 Thesis outline

This PhD thesis by Published Work is organised as follows:

• Chapter 2 includes a review of the relevant literature, to assist the readers in understanding the background and context of the research hereby presented, focusing on identifying and discussing knowledge gaps in the field, limitations of previous studies and which of these are tackled in this thesis;

• Chapter 3 outlines the original contributions to the fields of ML and EAs for aiding and augmenting clinical decision making, describing the novel FS methods and decision support systems used to devise clinically viable clinical decision making support tools to aid prediction of early diagnosis of PD, prognosis in patients with chronic hepatitis, and optimal postoperative recovery area to improve postoperative discharge decision making;

• Chapter 4 describes and discusses how research findings answer the research questions posed in chapters no. 1 (section 1.3) and no. 2.

• Chapter 5 provides both a conclusion summarising the added value from the published work to the body of knowledge in ML and EAs for devising clinically viable clinical decision support tools, along with providing suggestions for future opportunities from the research hereby presented.

• Appendix A includes a summary of the specific contributions by publication, confirmation on percentage of contribution towards the
published works included in this thesis from all co-authors, approvals from publishers to share them and relevant milestones achieved.

- Appendix B presents the published works included in this thesis for the reader’s reference and consideration.

- Appendix C provides supporting documents towards the work included in this thesis.
This chapter presents an overview of the state of the art underpinning the research work of the author and his collaborators. The review provided seeks to assist the readers in understanding the background and context of the research hereby presented, whilst identifying and discussing knowledge gaps in the field, limitations of prior work and which of these are tackled in this thesis.

2.1 Research background

2.1.1 Machine Learning to aid early diagnosis of Parkinson’s Disease

7-10 million people have PD worldwide (Hermanns, 2011). PD is a neurodegenerative disease that leads to sensorimotor symptoms, including tremors, mobility, and speech impairments. Conventional treatments for PD involve invasive surgery and deep brain stimulation (Engel et al., 2005). Non-invasive treatments for PD include medical therapies to enhance sensorimotor synchronisation but cannot change the course of the disease significantly (Satake et al., 2009). Moreover, in the elderly, who are most affected by PD, it is challenging to differentiate early signs of PD from senility; in fact, 20% of the elderly is undiagnosed (Schrag et al., 2002; Das, 2010) and more than
90% of patients with PD has speech disorders that are also undiagnosed (Das, 2010).

Amongst speech impairments due to PD, dysphonia and dysarthria are the most common ones, which are associated with a reduction in loudness and stability of the voice (Baker et al., 1998). Nevertheless, hypertonia can also affect PD patients, leading to unexpected interruptions in vocal sounds and, therefore, impairments in vocal quality (Stelzig et al., 1999).

The extent of vocal impairments due to PD can be quantified via dysphonic metrics, which are, therefore, biomarkers for the disease. Speech processing is typically deployed to recognise pathophysiological voice patterns and it is clinically used as one of the most preferred non-invasive diagnostic methods. Pathological vocal features are conventionally analysed via signal processing techniques, which lack the ability to recognise their underlying patterns; however, they can be identified in an automated fashion by applying ML on relevant features of sustained vowel sounds, such as amplitude and frequency (Sakar et al., 2013).

ML offers a higher reliability in dealing with noise and outliers in voice-related signals, thus leading to a greater generalisation (Sakar et al., 2017). Although previous studies (Alhussein, 2017; Sakar et al., 2017) did not lead to a clinically viable classification performance due to the application of extrinsic data-driven approaches alone, such advances in isolated and continuous speech processing via ML have enabled to ponder on their potential for clinical applications to aid early diagnosis of PD to personalise treatments and improve clinical outcomes. An intrinsic ML-based FS method may improve
classification performance (accuracy and reliability) and explainability with respect to the classifiers from the above-mentioned studies.

ML provides non-invasive and highly specific automated methods for detecting anomalies in voice signals as indicators of PD (Alhussein, 2017). Nevertheless, the architecture and results from such previous studies are not explainable, as data pre-processing and extrinsic feature engineering methods, e.g., principal component analysis (PCA) (Parisi, 2014a), modify the underlying patterns in voice signals, thus affecting the performance of speech processing methods (Parisi & Manaog, 2017a) and impairing their explainability for them to be applied in a clinical setting.

Conversely, automated pattern recognition via ML, improved via intrinsic ML-based FS, can aid in detecting important patient-specific biomarkers in speech data to differentiate between healthy and pathological patterns due to PD accurately, reliably and in an explainable manner (Parisi & Manaog, 2017a; Parisi et al., 2018a).

Although advances in ML have facilitated the prognostic assessment of patients with PD, they either lack a clinically viable classification performance (Alhussein, 2017; Sakar et al., 2017) or are limited in their explainability and difficult to interpret (Gill & Johnson, 2009; Cantürk & Karabiber, 2016). Moreover, the reliability and explainability of the leave-one-subject-out (LOSO)-SVM of Benba et al. (2017), the best performing classifier on the benchmark speech dataset on patients with PD of Sakar et al. (2013), are impaired by the relatively low number of input features (N = 15) selected via an extrinsic method used for FS, i.e., Human Factor Cepstral Coefficients.
Instead, using a more representative number of features (N = 20) that were retrieved as the most important features via an intrinsic ML-based FS method (the MLP-related weights) of Parisi et al. (2018a) led to an AUC of 0.6 in the LOSO-SVM (Benba et al., 2017) and of 1.0 in the MLP-LSVM (Parisi et al., 2018a), and a considerably higher computational cost (130 epochs in the LOSO-SVM, as opposed to 3 in the MLP-LSVM), although of secondary importance with respect to the three metrics defining clinical viability, which are accuracy, reliability and explainability. This comparison of performance demonstrates the advantage of using intrinsic ML-based FS (via the MLP) to improve the performance of ML binary classifiers (LSVM in Parisi et al., 2018a), instead of using extrinsic FS algorithms (e.g., HFCC in Benba et al., 2017).

Moreover, the cross-validation (CV) algorithm used in the LOSO-SVM of Benba et al. (2017), i.e., the LOSO, is not as exhaustive as the k-fold CV deployed by Parisi et al. (2018a), whose k folds cover more permutations of the input data, thus enabling the MLP-LSVM to avoid overfitting, and have a reduced variance in the classification performance.

These methodological limitations in previous studies that hinder the predictive performance and explainability of previous algorithms (e.g., Benba et al., 2017) are due to the application of data-driven approaches ‘as is’ or extrinsic feature-driven methods (e.g., HFCC in Benba et al., 2017); instead, intrinsic feature-driven approaches, leveraging classifiers’ own feature engineering methods (e.g., via the MLP, in Parisi et al., 2018a) can improve classification performance and ensure higher explainability (Parisi et al., 2018a).
In fact, the author developed an intrinsic FS method to one classifier (MLP), i.e., using its own weights (Parisi et al., 2018a), instead of conventional extrinsic FS methods. Intrinsic FS led to improve linear separability of the input voice signals, thus improving the classification performance of a second classifier (LSVM) and ensuring explainability in the hybrid classification system generated to aid early diagnosis of PD based on speech data accurately and reliably, whilst retaining explainability (MLP-LSVM, Parisi et al., 2018a).

The author and his collaborators applied ANN (Parisi, 2014b) for intrinsic FS before classifying the selected voice-related features by the Lagrangian SVM or LSVM (Mangasarian & Musicant, 2001; Parisi et al., 2015; Parisi & Manaog, 2016; Parisi et al., 2018a). Thus, the author’s main novel high-level contribution lies in demonstrating that accuracy, reliability and explainability in diagnostic tasks can be achieved via an assistive, decision support tool that leverages intrinsic ML-based FS for ML-based binary classification. In PD-related research, the author’s low-level contribution lies in the development and validation of a hybrid algorithm that demonstrates the above-mentioned clinical viability by blending intrinsic FS from an ANN (MLP) and the best performing ML classifier on benchmark speech data (LSVM) to aid early detection of PD (Parisi et al., 2018a). This contribution seeks to fill the knowledge gap in the literature whereby intrinsic ML-based FS can be the preferred approach to improve the performance (accuracy and reliability) of ML-based binary classifiers, whilst retaining explainability for clinically viable applications.
Differently from algorithms from previous studies tested on the same dataset, the hybrid classifier of Parisi et al. (2018a) was also validated not only against the diagnostic features from speech data expected to be clinically important based on evidence-based research but also against out-of-the-box solutions in commercially available software. Relevant in-depth methodological novelties in the author’s research will be outlined in sub-chapter 3.3-3.5.

2.1.2 Machine Learning to aid early prediction of prognosis in patients with chronic hepatitis

Even when attempting to aid prediction of prognosis in patients with viral hepatitis, as grouped into hepatitis B (HBV), C (HCV) and D (HDV) based on the pathophysiology, existing ML-based solutions that leverage extrinsic FS methods either lack the required predictive performance (accuracy and reliability) as clinical decision support tools (Ulutasdemir & Dagli, 2010; Hayashi & Fukunaga, 2016) or their explainability is not enough for them to be clinically viable (Tan et al., 2009; Bascil & Oztekin, 2012). These limitations helped the author identify the second knowledge gap being tackled, which is the capability of intrinsic ML-based FS to aid in a further binary classification task, i.e., early prognostics – not only in early diagnostics as per the MLP-LSVM for aiding early diagnosis of PD (Parisi et al., 2018a).

Although 400 million people worldwide are affected by viral hepatitis, most patients are not aware of having this pathology or the viral infection is often undiagnosed, as it remains asymptomatic for decades (World Health Organization, 2016). As the chronicity and the severity of the inflammation
increase, hepatitis can be diagnosed via assessing patients’ symptoms (e.g., fatigue, malaise, and hepatic hypertrophy) and by analysing haematological data (Desmet et al., 1994).

Nevertheless, assessment of prognosis of patients with chronic hepatitis heavily relies on a posteriori evaluation of patient morbidity by biopsies (Castera, 2012) or non-invasive techniques aimed at measuring indirect biomarkers of the disease, such as levels of serum and the extent of hepatic fibrosis (Salkic et al., 2014) via computerised tomography or non-ionising medical imaging techniques, e.g., magnetic resonance elastography (Castera, 2012).

Since such methods, regardless of their invasiveness, only provide biochemical or tissue-related data individually (Castera, 2012; Salkic et al., 2014), the multi-factorial clinical decision making is still unable to predict such a prognosis before patient symptoms become manifest.

Conversely, Artificial Intelligence (AI), including ANN, such as the MLP (Rumelhart et al., 1986; Parisi, 2014a; Parisi, 2014b; Parisi et al., 2015), and ML, e.g., the LSVM (Parisi & Manaog, 2016), can detect the underlying pathological patterns in haematological data to understand the course of the disease based on salient features influencing it, thus, aiding to perform a prognostic prediction (Bascil & Oztekin, 2012; Nahato et al., 2015; Parisi & Manaog, 2016).

Although studies that applied ML to haematological data for prognostic predictions reported accuracies higher than 95% (Calisir & Dogantekin, 2011; Kaya & Uyar, 2013; Nahato et al., 2015), they either presented algorithms that
are not interpretable enough to be applicable in a clinical setting (Kaya & Uyar, 2013) or lack appropriate data pre-processing and feature engineering considering clinical decision making, thus limiting their clinical viability. The lack of interpretability is mainly due to the extrinsic methods used for feature selection, which helped the author identify the second opportunity to contribute to fill this knowledge gap by demonstrating the advantage of leveraging intrinsic ML-based in binary classification tasks involved in prognostics too.

Moreover, such previous studies did not mention the use of any adequate data pre-processing techniques or computational cost involved in the algorithm aiding prediction of prognosis of hepatitis, which, although of secondary importance with regards to accuracy, reliability and explainability, further hinder the clinical viability of such ML-based classifiers.

Conversely, the author’s research, beside innovating in the ML-based decision support systems devised, is focused on ensuring they are interpretable, clearly presenting standardisable data pre-processing, intrinsic ML-based FS to improve accuracy and reliability of ML classification systems, thus overcoming the above-mentioned limitations from previous studies.

Validating a hybrid classifier using haematological and demographic information to predict survival of patients with chronic hepatitis early (Gong, 1988) by leveraging appropriate and interpretable data pre-processing and intrinsic ML-based feature engineering (via a LSVM, Parisi et al., 2019), the author aims at overcoming the above-mentioned limitations from previous studies. By enhancing the multi-factorial clinical decision making via automated pattern recognition, ML-based FS can further aid physicians in
predicting prognosis in patients with chronic hepatitis more accurately, reliably and in a more explainable manner (Parisi et al., 2019).

The proposed hybrid decision support system is not only accurate and reliable but it is also explainable, as testified by its validation using two publicly available benchmark datasets to help in predicting the survival in patients with hepatitis regardless of whether the follow-up times are variable or consistent amongst patients, differently from ML-based classifiers from published studies leveraging extrinsic ML-based feature selection and out-of-the-box solutions available in commercial software, e.g., Microsoft Azure ML Studio (Parisi et al., 2019).

Moreover, in an attempt of developing standards to devise ML-based classifiers to be applied for medical applications, either diagnostic or prognostic, the author and his collaborators developed a quality assessment scale (MQAS), which can be also used to assess the quality, scientific rigour and clinical viability of published studies deploying supervised ML in healthcare (Parisi et al., 2019).

Relevant in-depth methodological novelties in the author’s research will be outlined in sub-chapter 3.3-3.5.

In summary, the author demonstrated the clinical viability of decision support systems leveraging intrinsic ML-based FS to improve binary classification tasks involved in both diagnostics (e.g., early diagnosis of PD in 2.1.1, Parisi et al., 2018a) and prognostics (e.g., early prediction of prognosis in patients with chronic hepatitis in 2.1.2, Parisi et al., 2019).
2.1.3 Machine Learning to improve postoperative discharge decision making

Current multi-factorial clinical decision making, hindered by information overload as well, impairs the reliability of postoperative discharge, as physicians alone cannot always predict in which postoperative recovery area patients should be sent to next further to undergoing a surgical operation. This is an example of a complex multi-class classification problem that occurs in a clinical setting, which can be aided via ML-based classifiers leveraging EAs-based optimisation (Parisi et al., 2018b). In doing so, the author seeks to fill the third high-level knowledge gap in the literature being tackled in this thesis.

Postoperative discharge decision making is complex, as several variables influence it with varying degrees of significance for each patient, e.g., demographics, physical condition, body temperature and blood pressure, and patient discomfort upon discharge (Summers et al., 1993).

Inaccuracies in making this postoperative discharge-related decision lead to postoperative complications and, in some cases, they may affect patient survival and hospital resources significantly. Moreover, beds are generally limited in the general ward and, especially, in the intensive care unit (ICU); therefore, it is important to manage all resources effectively to prioritise patients who require a higher medical attention.

As during surgery, anaesthetics and extrinsic factors may lower patient’s body temperature, hypothermia can usually occur as a postoperative complication, wherein patient’s body temperature can be lower than the expected temperature physiologically (36 °C) (Vaughan et al., 1981). Hypothermia is
often undiagnosed (Davidson & Grant, 1981) and may lead to a higher consumption of oxygen and impair survival (Woolery et al., 1991; Karalapillai et al., 2013; Saad & Aladawy, 2013).

To improve recovery, the optimal postoperative recovery area where patients should be sent to next after surgery should be predicted to help in minimising the risk of hypothermia (Poveda et al., 2009; Saad & Aladawy, 2013).

As most methods currently at the physicians’ disposal in a clinical setting to aid postoperative discharge decision-making are statistical and extrinsic to ML, e.g., PCA, they do not have the capability of recognising the underlying patterns of such highly nonlinear, dimensional, multivariate data, thus limiting their generalisation to unseen patient data.

Conversely, ML has been proven to detect the underlying patterns of such medical data in the multi-class classification problem it entails (Parisi & Manaog, 2017b). ML has been successfully applied to facilitate recovery post-surgery by aiding temperature control (Woolery et al., 1991), minimising the rate of postoperative complications and improving the utilisation of hospital resources (Dang et al., 2002; Abuaqel et al., 2017).

The multi-class classification problem tackled in the author’s research involves predicting the optimal postoperative recovery area where patients should be sent to next further to undergoing surgery, amongst the patient’s home, the general ward and the ICU, to avoid hypothermia (Hsieh et al., 2015) and improve clinical outcomes.
However, to predict such an optimal postoperative recovery area, researchers have applied either subjective or heuristic methods to optimise hyperparameters of ML-based classifiers (Hansen & McDonald, 2001), which are thus not objective and loosely coupled to them, thus being inaccurate, not reliable and not explainable enough.

Instead, EAs, such as GA, provide the ability to solve unconstrained optimisation problems, such as a tightly coupled optimisation of hyperparameters for ML-based classifiers to predict the optimal postoperative recovery area, to improve classification performance and explainability of the optimisation involved in the decision support process. Moreover, per se, optimising postoperative discharge decision making improves clinical outcomes and survival; therefore, although this additional hypothesis is at very early stages, the author hypothesised this problem as having ‘evolutionary’ nature, thus further suggesting GA as being a suitable method to optimise the hyperparameters of an ML classifier to predict optimal postoperative recovery area.

Hsieh et al. (2015) used an EAs, namely ‘particle swarm optimisation’ (PSO), to enable a fuzzy classifier to retrieve crisp rules from the same benchmark postoperative data (Summers et al., 1993) to achieve an ACC of 84% but lacking interpretability of the architecture for it to be applied in a clinical setting. With a similar shortcoming, Luukka (2010) developed a fuzzy logic-based classifier based on features derived from PCA to predict the optimal recovery area on the benchmark dataset developed by Summers et al. (1993) with 66.22% of test classification accuracy (ACC), which is not clinically viable.
Forghani & Yazdi (2015) classified the same benchmark dataset (Summers et al., 1993) by a more complex, less interpretable algorithm named ‘fuzzy min–max neural network with symmetric margin’ (FMNWSM) and applied it to classify the same data set (Summers et al., 1993) with an ACC of 72.99%, with respect to 74.75% via SVM, which are also not clinically viable, but with a far lower computational cost than SVM (2.1 ms for FMNWSM as compared to 4583 ms for SVM).

Although the FMNWSM achieved a faster convergence with respect to the conventional fuzzy min-max neural networks due to the absence of peculiar nodes in superimposed areas (Forghani & Yazdi, 2015), its architecture lack explainability to be a decision support tool for physicians; moreover, its ACC (72.99%), the only accuracy clinically important as it based on the test set or ‘unseen’ data, is not clinically viable.

Via 10-fold cross validation (CV), Abuaqel et al. (2017) applied SVM and ANN on the same postoperative data of interest (Summers et al., 1993), classifying it with 88.54% and 82.81% of ACC, respectively. Before the author’s research outcomes (Parisi et al., 2018b), the SVM used by Abuaqel et al. (2017) for multi-class classification was the most accurate classifier on this benchmark dataset (Summers et al., 1993), although used as a black-box classifier in Weka. In this case, whilst a clinically viable multi-class classification performance was partly achieved (88.54% ACC via SVM, Abuaqel et al., 2017), the black-box implementation and use lack explainability.

Thus, the ML-based decision support systems from the literature used to predict the optimal postoperative recovery area on a subject-specific basis are
either too inaccurate and unreliable (Luukka, 2010; Forghani & Yazdi, 2015) or not explainable (Hsieh et al., 2015; Abuaqel et al., 2017) to be clinically viable.

Given the absence of clinical translation of such classifiers (Luukka, 2010; Forghani & Yazdi, 2015; Hsieh et al., 2015; Abuaqel et al., 2017), the author’s research aims to develop and validate a more accurate, reliable and explainable novel hybrid ML-based classifier that is optimised via EAs to predict postoperative discharge, although its performance might be limited by the presence of unbalanced classes (e.g., 2 patients whose optimal postoperative recovery area was the ICU).

Relevant in-depth methodological novelties in the author’s research will be outlined in sub-chapter 3.3-3.5.

2.2 Conclusions

Although previous research involved the use of advanced extrinsic techniques for FS (either statistical or ML-based) to aid detection of PD, the author’s research sheds light onto intrinsic ML-based FS approaches that improve binary classification performance for facilitating early diagnosis of PD in a clinical setting and are explainable to clinicians (Parisi et al., 2018a).

The author’s novel contribution also supplies a similar intrinsic ML-based FS methodology to aid early prediction of the prognosis in patients with chronic hepatitis, thus clearly filling a noticeable knowledge gap in the literature (Parisi et al., 2019). Moreover, a hybrid, explainable feature engineering based on
intrinsic ML-based FS and clinical gold-standard prognostic biomarkers from meta-analyses is deployed to improve the prediction of prognosis further (Parisi et al., 2019). Differently from previous studies, the author also validated such novel methods against out-of-the-box algorithms available in commercial software, such as Microsoft Azure ML Studio (Parisi et al., 2019).

In multi-class classification problems instead, such as predicting the optimal postoperative recovery area on a patient-specific basis, the author's research supports the use of EAs for optimising multi-class ML-based classifiers for such decision support systems to be clinically viable (Parisi et al., 2018b). Such an approach not only has the highest classification performance in the literature on the same benchmark data but is also more interpretable than algorithms from previous studies (Parisi et al., 2018b).
This chapter describes the original contributions of the author’s research to the fields of ML and EAs for aiding clinical decision-making by devising accurate, reliable, and explainable decision support systems, with focus on medical applications to predict early diagnosis and prognosis of various pathologies. The relevance and added values of the author’s work are also discussed in the context of the findings of other researchers.

3.1 Narrative framework
To contextualise the author’s research work and to derive the salient points from his findings, a narrative framework has been outlined, which describes the work presented in a cohesive manner. For ease of reference and
understanding, this narrative categorises the author’s papers into separate sub-sections, with each of them dealing with a separate ‘contribution’, which is mapped against the relevant methodological approaches outlined in section 1.4. To aid the reader, at the beginning of each narrative section, a short paragraph is included summarising and contextualising the work undertaken.

3.2 Preliminary work

This section presents and discusses preliminary work performed by the author, focused on devising clinically viable decision support systems to aid early diagnoses and prediction of prognosis in presence of different pathologies via either intrinsic ML-based FS or EAs-based optimisation to improve the performance and explainability of binary and multi-class classifiers respectively. To demonstrate the clinical viability of such classifiers, experiments were performed to aid early diagnosis of PD, early prediction of prognosis in patients with chronic hepatitis and in patients with postoperative hypothermia.

To provide an objective method to aid early diagnosis of PD, the author and his main collaborators had devised a clinically viable solution leveraging ML for both isolated and continuous speech processing (Parisi & Manaog, 2017a). In such an instance, the author had evaluated the clinical usefulness of a four-dimensional self-organising map (SOM), an LSVM and an MLP with ten sigmoid hidden and Softmax output neurons (Parisi & Manaog, 2017a). Dysphonia metrics and the “Unified Parkinson’s Disease Rating Scale” (UPDRS) that were retrieved from the UCI database, which are biomarkers for
PD used in a clinical setting, were analysed and the most salient features retained based on a comparison of three different FS algorithms: two extrinsic FS methods, i.e., the ReliefF and the Partial Least Square (PLS) methods, and one intrinsic feature engineering algorithm derived from the MLP’s own method for ranking features based on its weights (Parisi & Manaog, 2017a). Thus, in agreement with published studies, the UPDRS, the mean pitch and the standard deviation of the period were ranked as the three most important features or factors towards the diagnosis of PD (Parisi & Manaog, 2017a).

Whilst SOM and LSVM led to a classification accuracy on the testing set of 84.52% and 98.27% respectively, the MLP achieved 100% accuracy and an AUC of 1 in aiding early diagnosis of PD (Parisi & Manaog, 2017a).

Moreover, Parisi & Manaog (2016) preliminarily validated the use of the LSVM learning classifier (Mangasarian & Musicant, 2001) as a clinical decision-making support tool to assist with predicting the prognosis in patients with chronic hepatitis.

Haematological data of forty-six (46) male patients (age: 44 ±7 years, twenty-three survivors and twenty-three deceased) derived from the "Hepatitis Data Set" of the UCI Repository of ML (Gong, 1988) were used. In such a study (Parisi & Manaog, 2016), two artificial neural networks (ANN), i.e., MLP with learning and momentum terms set to 0.6 (Barton and Lees, 1995) and a Radial Basis Function (RBF) (Mackay, 1997), a four-dimensional unsupervised self-organising map (SOM) (Kohonen, 1982, 1997) and an LSVM (Mangasarian & Musicant, 2001) were devised, implemented and validated.
Haematological features with the highest relative importance towards defining patient prognosis were selected via ReliefF (Robnik-Šikonja and Kononenko, 2003), PLS- and PCA (Maitra and Yan, 2008; Parisi & Manaog, 2016).

The performance of the above-mentioned ML-based classifiers was assessed and compared based on the classification accuracy on the testing set, with the LSVM being the most accurate algorithm amongst those that were tested (LSVM: 84.78%, SOM: 78.26%, RBF: 63.21 %, MLP: 55.57%) (Parisi & Manaog, 2016).

Based on a comparison between the results obtained via the three above-mentioned FS methods (ReliefF, PLS and PCA), an enlarged liver, variceal bleeding, spider's blood, antiviral agents and a lower-than-normal level of blood albumin were consistently ranked amongst the nine most relevant features towards predicting the prognosis in patients with chronic hepatitis (Parisi & Manaog, 2016). These findings are also in agreement with published studies (Parisi & Manaog, 2016; Parisi et al., 2019). Thus, the LSVM was preliminarily found to be a clinically viable ML-based decision support system for assisting in predicting the prognosis in patients with chronic hepatitis (Parisi & Manaog, 2016).

As EAs can aid in the minimisation of the loss function in ML-based classifiers, Parisi & RaviChandran (2018) devised a mathematical approach that leverages GA, i.e., a type of EAs, to select optimal kernel and hyperparameters of SVM.

The author’s framework leveraging EAs and SVM can lead to a significant improvement in maximising the geometric margin of the OSH, thus increasing
the number of true positives and true negatives, whilst minimising that of false positives and false negatives correspondingly, thus improving the classification accuracy on the testing set and reliability substantially.

3.3 Machine Learning and Evolutionary Algorithm for decision support

This section presents and discusses extensive published work performed by the author, focused on leveraging intrinsic ML-based FS or EAs-based optimisation to improve the performance and explainability of supervised ML algorithms. These are requirements to devise clinically viable decision support systems, which are accurate, reliable, and explainable for them to be applied in a clinical setting.

As hypothermia is a common concern for patients who underwent surgical interventions, which further complicates postoperative discharge decision making, the author and his colleagues (Parisi et al., 2018b) used benchmark data on body temperature measurements and indicators of patient (dis)comfort upon discharge from the UCI ML repository on ninety (N = 90) patients (Summers et al., 1993). A hybrid multi-class classification algorithm was thus devised to improve postoperative discharge decision making by aiding prediction of optimal postoperative recovery area amongst home (N = 24), the general hospital ward (N = 64) or the ICU (N = 2): the GA-M-LSVM (Parisi et al., 2018b). GA, a type of EA, thus enabled the M-LSVM to be a clinically viable decision support system for aiding postoperative discharge decision making.
To help prevent the SVM from being increasingly trapped at local *minima* non-linearly with the number of epochs in an 'uncontrolled' manner, thus improving classification performance (accuracy and reliability), necessary for decision-making support systems, the augmented Lagrangian formulation of SVM (LSVM) was used (Mangasarian & Musicant, 2001; Parisi *et al.*, 2018b). The M-LSVM achieved linear separability of the input features by finding an OSH to maximise the geometric margin further (Parisi *et al.*, 2018b).

Further to optimising hyperparameters of the M-LSVM via GA (Fig. 5), its classification performance was higher than that of algorithms from published studies tested on the same benchmark dataset and that of state-of-the-art classifiers, e.g., M-MLP, M-SVM and M-LSVM across all four measures of clinical viability used for comparison (100% ACC, SN and SP, and 1 as AUC) (Fig. 6, Parisi *et al.*, 2018b).
Figure 6. 4D plot comparing the classification performance between the hybrid algorithm with GA-optimised hyperparameters (GA-M-LSVM) and that of multi-class classifiers with default hyperparameters (M-MLP, M-SVM and M-LSVM) with and without PCA for feature selection (Parisi et al., 2018b).

Besides multi-class classification problems, such as the postoperative discharge-related one tackled as described above (Parisi et al., 2018b) via EAs-based optimisation coupled with a multi-class ML classifier, other binary classification problems can be considered clinically, such as early diagnosis of PD and early prognostics in patients with chronic hepatitis. Such binary classification problems were aided via intrinsic ML-based FS instead.

As PD is still difficult to diagnose early, relevant vocal samples selected by neurologists, including sustained vowels (‘a’ and ‘o’), numbers, words and short sentences, from 40 subjects (20 patients with PD, age: 64.86 ±8.97 years, gender: 6 females, 14 males; 20 healthy controls, age: 62.55 ±10.79 years, gender: 10 females, 10 males) were used (Sakar et al., 2013) to develop a classifier to aid early diagnosis of PD (Parisi et al., 2018a). Clinical scores of disease severity, such as the gold standard metric named ‘Unified Parkinson’s Disease Rating Scale’ (UPDRS: 13 ±15.89), were also used as input data. Speech features of 28 additional patients with PD (62.67 ±10.96) were extracted for further testing (Sakar et al., 2013; Parisi et al., 2018a).

Further to performing normalisation, standardisation and removal of outliers, the ML-based classifier LSVM was applied with an intrinsic MLP-based FS method for aiding detection of PD in the above-mentioned patients (Parisi et al., 2018a) (Fig. 2).
Besides the above-mentioned diagnostic task, intrinsic ML-based FS was hypothesised to lead to clinically viable binary classifiers also when aiding prediction of prognosis, thus contributing to augment clinical decision making in a well-rounded manner.

For instance, the author demonstrated how intrinsic ML-based FS (via an LSVM) can aid the prediction of survival of patients with chronic hepatitis via an MLP, thus devising and validating a clinically viable hybrid novel binary classifier (LSVM-MLP, Parisi et al., 2019). In such a study, haematological, demographic (age and gender) and outcome-related (e.g., fatigue, malaise, anorexia) patient data on one hundred fifty-five (N = 155) patients with chronic hepatitis (140 men, 15 women; age 41.2 ± 12.5 years) (Gong, 1988) were used, one hundred thirty-two (N = 132) of which were survivors and twenty-three (N = 23) were deceased (Parisi et al., 2019).

After performing intrinsic FS via an LSVM by leveraging its weights to infer the relative importance of the individual features towards the required classification, the reduced data (N = 16 features) were then input to an MLP to discriminate between patients with hepatitis who are likely to survive and those who are not (Parisi et al., 2019) (Fig. 2).
3.4 Machine Learning for intrinsic feature selection and Genetic Algorithm for optimisation

This section presents and discusses published work undertaken by the author, focused on deploying ML algorithms for leveraging intrinsic ML-based FS mechanisms, such as those based on weights from ANN (e.g., MLP in Parisi et al., 2018a) and those based on selecting relevant features to improve linear separability of the input data into the target classes (e.g., LSVM in Parisi et al., 2019). Moreover, it also discussed the author’s work focused on deploying GA for optimisation aimed at maximising the performance of multi-class classifiers for improving postoperative discharge decision making.

To improve the accuracy and reliability of the LSVM classifier for aiding automated early detection of PD from speech data, an MLP was deployed to retain only those features with the highest relative importance towards the classification of interest out of the 27 features available (Parisi & Manaog, 2017a,b; Parisi et al., 2018a). Such salient features were retrieved by leveraging the gradient descent training algorithm, whose gradient is determined via the backpropagation method whilst minimising the training error, such as the mean squared error (MSE), until a specific target error is reached (e.g., MSE = 0.05) (Parisi et al., 2018a).

The hybrid model of Parisi et al. (2018a) used an MLP with only one hidden layer and the hyperbolic tangent sigmoid transfer function (Parisi & Manaog, 2017a,b) for intrinsic FS, as it can account for any Boolean-bounded function mathematically and a higher range of inputs (Cybenko, 1989).
The *moduli* of the weights from the MLP were deployed to rank the input features towards their relative importance in discriminating them into the target classes, i.e., healthy, or PD-related voice patterns (Parisi *et al.*, 2018a).

The most salient input features ($N = 20$, **Fig. 1**) were then used as inputs to maximise the margin of an LSVM deployed for classifying (Mangasarian & Musicant, 2001; Parisi & Manaog, 2016) between healthy and PD-related data (voice signals and UPDRS scores) (Parisi *et al.*, 2018a). The whole modelling workflow deployed in this study is illustrated in Fig. 1 of Parisi *et al.* (2018a) (**Fig. 2** of this thesis).

The *moduli* of the weights in the MLP were not affected by input resampling, as those deployed for intrinsic FS were extracted further to the last epoch prior to convergence and achieving the highest classification performance without overfitting ($\text{ACC} = 100\%; \text{AUC} = 1$) (Parisi *et al.*, 2018a) (**Fig. 7**).
Thus, the novel hybrid algorithm (MLP-LSVM) could outperform not only state-of-the-art classifiers when used alone (e.g., MLP and LSVM) with the least computational cost (Parisi et al., 2018a) but also those from published studies when tested on the same benchmark dataset of healthy and PD-related speech signals. Moreover, by virtue of its intrinsic ML-based FS, its explainability is also higher.
Using only the reduced features ($N = 20$, **Fig. 1**) led to a decrease in computational cost, as quantified via the software-dependent metric of the number of epochs required prior to convergence (from 7 to 3, Table 1 of Parisi *et al.*, 2018a) and a hardware-related metric, i.e., its computational time, which was reduced by five times (from 0.05 to 0.01, Table 1 of Parisi *et al.*, 2018a). The computational cost was considered of secondary importance with respect to the main criteria of interest defining clinical viability, which are accuracy, reliability and explainability. Nevertheless, it was a positive by-product of devising explainable decision support systems.

As opposed to loosely coupled, extrinsic statistical-based methods for FS, such as PCA (Calisir & Dogantekin, 2011), which may impair classification performance (Parisi, 2014a), the study of Parisi *et al.* (2019) further demonstrates the considerable improvements in predictive performance when deploying an intrinsic, tightly coupled ML-based method for FS to another ML classifier. Whilst Parisi *et al.* (2018a) supports the use of intrinsic ML-based FS for aiding diagnostics involving binary classification problems, e.g., aiding early prediction of Parkinson’s Disease, Parisi *et al.* (2019) also suggests its clinical viability for prognostic applications, always in the presence of a binary classification task, e.g., predicting survival in patients with chronic hepatitis.

Whilst adjusting MLP-related weights and biases is influenced by resampling input data, especially when k-fold cross-validation (CV) is deployed, using an LSVM for FS for an MLP, the resulting hybrid algorithm (LSVM-MLP, **Fig. 4**) was able to obviate this shortcoming and improve classification performance (Parisi *et al.*, 2019).
For multi-class classification problems, e.g., predicting the optimal postoperative recovery area amongst home, ward and ICU, intrinsic ML-based FS was not enough, thus suggesting exploring EAs-based optimisation to improve classification performance (Parisi et al., 2018b).

A computationally efficient GA with the data of interest, i.e., the unified non-dominated sorting genetic algorithm (U-NSGA-III) (Seada and Deb, 2015), was applied to optimise the hyperparameters of an M-LSVM to build a hybrid multi-class classification algorithm (GA-M-LSVM) to aid prediction of recovery area further to undergoing a surgical intervention aimed at improving clinical outcomes (Parisi et al., 2018b). In such a study, GA was applied to optimise the hyperparameters and the kernel of the M-LSVM to determine an OSH that further improves linear separability of the input features into the three target classes (Parisi et al., 2018b). The mathematical implementation of the hybrid algorithm (GA-M-LSVM) consists of equations (1-2) in the study of Parisi et al., 2018b, which seek to optimise the hyperparameters of the LSVM as well, described in equations (3-4) in Parisi et al. (2018b).

The GA-based optimisation led to improve the predictive performance of the LSVM, as quantified by the AUC, whilst minimising the computational cost (Parisi et al., 2018b). The novel hybrid algorithm (GA-M-LSVM, Fig. 5), via a 'controlled' All-vs-All (AvA) multi-class classification, attained 100% of both accuracy (ACC) and reliability (SN, SP and AUC) (Fig. 6) with the least number of epochs required (N = 3) and computational time (0.2 s on an Intel® Core™ i5-4460T 2.70Ghz Processor) (Parisi et al., 2018b).
3.5 Clinically viable solution to predict diagnosis and prognosis

This section presents and discusses published work performed by the author, focused on devising a clinically viable decision support system to aid early diagnosis of PD and prediction of prognosis in patients with chronic hepatitis, as well as predicting the optimal recovery area to improve postoperative discharge decision making.

In the study of Parisi et al. (2018b), a computationally efficient GA-based optimisation method was deployed to find the optimal kernel function and hyperparameters to maximise the classification performance (ACC and AUC) and minimise computational cost in a multi-class LSVM, which resulted in a novel hybrid algorithm, i.e., the GA-M-LSVM. This optimisation also ensured an improved linear separability with highly unbalanced classes in the input features used (Summers et al., 1993), reducing areas of ambiguity or overlapping regions with respect to uncontrolled AvA and OvA multi-class classification methods (van den Burg & Groenen, 2016).

The classification performance of the proposed hybrid model (GA-M-LSVM), quantified via accuracy on the testing set (ACC) and reliability (SN, SP and AUC), was higher than that of state-of-the-art classifiers and algorithms from published studies on the same benchmark dataset (Parisi et al., 2018b).

For the hybrid algorithm to be clinically viable, although of secondary importance with respect to the main criteria of interest (accuracy, reliability and explainability), the computational cost, including both the number of epochs prior to convergence without overfitting and the computational time, had to be
the lowest one amongst the models considered (Parisi et al., 2018b). This consideration is also important as the computational hardware available in clinics, especially in the public healthcare sector, is usually limited.

To enable early diagnosis of PD via ML-based speech processing, given the ability to deal with long-range and highly dimensional and nonlinear data, a novel hybrid algorithm was devised, wherein an MLP (Parisi & Manaog, 2017a,b) was used for intrinsic FS based on its weights and an LSVM (Mangasarian & Musicant, 2001; Parisi & Manaog, 2016) for the binary classification involved in the diagnostic task of interest, i.e., to discriminate between healthy and PD-related speech patterns (Parisi et al., 2018a).

With selected features (N = 20, Fig. 1) via the MLP, the hybrid algorithm (MLP-LSVM) attained 100% of classification accuracy with the least computational cost too (three epochs and 0.01s of computational time) (Parisi et al., 2018a). It is worth noting the number of epochs and the computational time in the LSVM decreased by more than double and by five times respectively (Parisi et al., 2018a).

Via a sensitivity analysis wherein the number of selected features was made vary (N = 15 and 16 as per the extrinsic FS method of HFCC in Benba et al. (2017), 20 – the selected features via the hybrid model and 27 – all initial features), the hybrid model, leveraging intrinsic ML-based FS instead (Parisi et al., 2018a) was proven to be more accurate, reliable and explainable, i.e., more clinically viable, than the best performing algorithm from published studies, i.e., the LOSO-SVM of Benba et al. (2017).
The need for custom modelling implementation was justified by its higher classification performance (accuracy and reliability) and explainability with respect to models in commercial software, such as Alyuda NeuroIntelligence (2.2 (577), Cupertino, California, USA), adopting the same data pre-processing workflow (Parisi et al., 2018a).

In its novelty, whilst retaining explainability, the hybrid algorithm (MLP-LSVM, Fig. 2) was devised to aid clinical decision-making processes to enable early diagnosis of PD mainly based on speech patterns (Parisi et al., 2018a).

A reverse architecture (LSVM-MLP, Fig. 4) was devised and deployed for aiding early prediction of survival in patients with chronic hepatitis (Parisi et al., 2019). Based on the number of reduced input features (Parisi & Manaog, 2017b), 16-fold CV was applied iteratively to train the proposed hybrid algorithm (LSVM-MLP) and its performance was compared when 5-fold CV was used to verify whether the computational cost could be reduced whilst retaining the same or a similar classification performance (accuracy and reliability) (Parisi et al., 2019).

The novel hybrid classifier was validated against when leveraging the 16 most important features from a clinical standpoint, as per the meta-analysis of Friedrich-Rust et al. (2008) on chronic hepatitis, whereby age, fatigue and histology were not statistically significant biomarkers for patient prognosis (Parisi et al., 2019).
3.6 Discussion

In this section, the work of the author is discussed and evaluated considering findings from other researchers. In particular, the work of the author seeks to highlight how ML alone can aid and augment clinical decision making via a tailored intrinsic ML-based FS process or via GA-based optimisation. The resulting hybrid systems, devised and evaluated by the author against clinical requirements and compared against the best performing decision support systems from the literature on the same benchmark datasets, represent clinically viable tools for aiding both early diagnosis and prediction of prognosis of various pathologies. Moreover, salient points to be considered when applying such systems at scale are also listed and discussed.

3.6.1 Machine Learning to augment clinical decision making

To augment postoperative discharge decision making, as an example of a complex multi-class classification problem tackled clinically, a GA-optimised and ‘controlled’ AvA multi-class classifier was leveraged. Initially, the performance of an MLP was evaluated in predicting the optimal postoperative recovery area based on all benchmark clinical data used (N = 8 features) (Parisi et al., 2018b). The ACC of 92.31% and the AUC of 0.65 proved its capability in classifying these highly dimensional and nonlinear data. A linear M-SVM was deployed but only reached an ACC of 71.11% and an AUC of 0.42 (95%CI: 0.36–0.48) when PCA was applied for dimensionality reduction (Parisi et al., 2018b). As the LSVM improves margin maximisation via an augmented Lagrangian formulation with respect to traditional SVM
(Mangasarian & Musicant, 2001; Parisi et al., 2015; Parisi & Manaog, 2016; Parisi et al., 2018b), it was also applied for this multi-class classification problem as a baseline multi-class classifier (M-LSVM). In fact, the M-LSVM resulted in a higher overall classification performance (AUC M-LSVM: 1, 95%CI 1–1) than the conventional SVM, whilst retaining the same optimal linear kernel to ensure a fair comparison of classification performance (Parisi et al., 2018b).

GA, applied to optimise the kernel and hyperparameters of the M-LSVM, enabled to achieve a ‘controlled’ AvA M-LSVM via the determination of a more accurate OSH to maximise the separation of the input features into the target classes (GA-M-LSVM) as compared to conventional uncontrolled AvA or One-vs-All (OvA) multi-class decision support systems (Rifkin & Klautau, 2004; Parisi et al., 2018b) (Fig. 5).

It is worth noting the testing classification accuracy of the hybrid algorithm (GA-M-LSVM) was much higher (100%, Parisi et al., 2018b) than that of the most accurate algorithm from published studies (88.54%, Abuaqel et al., 2017).

To truly augment clinical decision making, although ML is not deterministic when trained and tested on new data, such a ‘controlled’ multi-class classification enables consistently high classification performance with new postoperative discharge-related data.
3.6.2 Machine Learning for feature selection and Evolutionary Algorithms for optimisation

To discriminate between PD and healthy voice patterns, which is the first binary classification task aided by intrinsic ML-based FS (Parisi et al., 2018a), pitch-related parameters (Little et al., 2009), or frequency-, pulse-, amplitude- and harmonicity-related parameters have been used (Sakar et al., 2013).

Given the importance of helping physicians distil the salient features from those collected for diagnostic purposes (Cantürk & Karabiber, 2016; Sakar et al., 2017), FS before classification by ML is still deemed essential.

For instance, Alhussein (2017) deployed an SVM with an RBF kernel to improve the classification performance of an extreme learning machine (ELM), achieving a 97.2% classification accuracy on the testing set (ACC) and an AUC of 0.971 in discriminating between PD and healthy speech patterns. Although Benba et al. (2017) attained 100% classification accuracy in classifying the same benchmark dataset, they used the first 15 and 16 extrinsic HFCCs features based on only three sustained vowels, not only lacking interpretability from a clinical standpoint but also without using the entire dataset, as inputs to an SVM with polynomial kernel via the LOSO cross-validation algorithm.

Thus, such hybrid classifiers from published studies have relied on non-interpretable, extrinsic FS algorithms, which make their architecture not explainable, thus not being clinically viable (Parisi et al., 2018a).

Instead, in the study of Parisi et al. (2018a), intrinsic ML-based FS was used to derive an interpretable classification architecture as a decision support system to aid early diagnosis of PD based on speech data in a clinical setting.
In fact, the weights following the last training iteration of an MLP that attained 100% average classification accuracy (AUC = 1, 95% CI: 1–1) were used to derive the 20 most important features and then input to an LSVM for classification.

Such a classification system is explainable, because, differently from those from previous studies wherein an extrinsic FS algorithm is used, it deploys another classifier with baseline hyperparameters (e.g., MLP) to derive the selected input features for classification via a further ML-based classifier (e.g., LSVM) (Parisi et al., 2018a).

Moreover, the clinical viability of such an intrinsic ML-based FS method for the binary classification of interest is confirmed by comparing the 7 variables discarded (Fig. 1) against those from published studies, e.g., Alhussein (2017). As matching the expected low clinical importance of such biomarkers, the discarded input features account for 8.3% of the total importance of all input features towards discriminating PD and healthy voice patterns from speech data (Sakar et al., 2013; Parisi et al., 2018a) (Fig. 1).

The resulting novel hybrid algorithm (MLP-LSVM) not only attained 100% of average classification accuracy on the benchmark dataset of interest (Sakar et al., 2013) but also with the lowest computational cost (3 epochs and 0.01 s as computational time) (Parisi et al., 2018a).

The increase in classification performance, the explainability and the decreased computational cost of the hybrid algorithm (MLP-LSVM) confirmed the clinical viability of the intrinsic MLP-based FS method to increase the linear separability of the input data into the two target classes (PD and healthy), thus
facilitating the delineation of the OSH in discriminating between such features via the LSVM classifier (Parisi et al., 2018a).

The MLP-LSVM (Parisi et al., 2018a) has a higher reliability than the LOSO-SVM classifier of Benba et al. (2017), since its classification performance was consistently higher both when using selected features and when considering the entire dataset, as well as it leveraged intrinsic and not extrinsic FS, which makes the MLP-LSVM more explainable too.

The lack of reliability in the LOSO-SVM built by Benba et al. (2017) is tangible when deploying the 20 most important features based on the MLP of Parisi et al. (2018a), which led to a low AUC (0.6) and a high number of epochs (N = 130), whilst the MLP-LSVM of Parisi et al. (2018a) attained a perfect classification performance (AUC = 1) in only 3 epochs.

Furthermore, the LOSO CV algorithm deployed in Benba et al. (2017) does not exhaustively cover all permutations of input data as the k-fold CV used in Parisi et al. (2018a), wherein all data points are considered in randomised order when creating the number (k) of folds, instead of discarding all observations for a single subject at a time, as in Benba et al. (2017).

Moreover, the LOSO CV in the classifier of Benba et al. (2017) led to a higher variance in the classification performance as compared to the MLP-LSVM of Parisi et al. (2018a).

These results further demonstrate the clinical viability of deploying intrinsic ML-based FS for aiding binary classification tasks in a clinical setting, such as the MLP used to improve the classification performance of an LSVM in aiding early diagnosis of PD (Parisi et al., 2018a).
Considering the promising results when using intrinsic ML-based FS (Parisi et al., 2018a), the author applied it for prognostic tasks too (Parisi et al., 2019). In fact, a reverse architecture (LSVM-MLP) was built to predict the prognosis of patients with chronic hepatitis, specifically their survival (Parisi et al., 2019). The coefficients of the OSH derived via the LSVM were regarded as the weights or relative importance of each feature for selecting them (Parisi et al., 2019).

After discarding three clinical features (antivirals, histology, and prothrombin time or “protime”) from the initial dataset via the weights derived by the LSVM (Fig. 3), 16 selected features were fed as inputs to an MLP (Parisi et al., 2019). The above-mentioned features that were excluded from the initial dataset only explained 2% of the saliency of the input data towards the required classification of prognosis in the patients with chronic hepatitis considered (Gong, 1988; Parisi et al., 2019) (Fig. 3).

In agreement with part of the medical literature whereby there is no significant histological discrepancy with regards to the prothrombin time when patients have similar staging of chronic hepatitis (Shahangian et al., 2006), this finding from Parisi et al. (2019) suggests that these features may not be significant when predicting the prognosis of such patients (Parisi et al., 2019).

The hybrid algorithm, leveraging intrinsic LSVM-based FS (LSVM-MLP, Fig. 4), attained 100% of average classification accuracy when using 5-fold CV and 10 hidden neurons in the only hidden layer of the MLP, with the lowest computational cost (9 epochs, 0.09s of computational time) (Parisi et al., 2019). This prognostic task is the second binary classification problem whose resolution has been aided by a binary ML classifier leveraging intrinsic ML-
based FS, which further demonstrates how the author has clearly filled the related knowledge gap in the literature.

3.6.3 Clinician-aided evaluation to achieve a clinically viable diagnostic and prognostic tool

In this thesis and in the studies of the author and his collaborators, deriving a clinically viable diagnostic or prognostic tool involved the development, validation and clinician-aided evaluation of an ML-based decision support system that, whilst retaining the highest classification performance (accuracy and reliability) and explainability – the primary end-users’ requirements as per section 1.3, also has the lowest computational cost, i.e., the secondary end-users’ requirement always described in section 1.3.

An example of such a clinically viable prognostic tool is the GA-M-LSVM, as a ‘controlled’ AvA multi-class classification system developed by Parisi et al. (2018b) (Fig. 5), which is able to predict the optimal recovery area for patients who underwent surgery with 100% of average classification accuracy (Fig. 6) and the lowest computational cost (3 epochs, 0.2s of computational time). A particular type of EA (GA) was used for optimising a multi-class classifier (M-LSVM), thus supporting the hypothesis whereby EA for optimisation is useful when tackling multi-class classification problems, such as aiding postoperative discharge decision making (Parisi et al., 2018b). This is a further knowledge gap filled by the author’s research work.

The GA-based optimisation of kernel and hyperparameters of the M-LSVM led to a higher linear separability of the input data into the target classes, thus aiding the M-LSVM implicitly in deriving an OSH that further improves margin
maximisation and predicts optimal postoperative recovery area in a more accurate and reliable manner (Parisi et al., 2018b).

With regards to its interpretability, the architecture of the GA-M-LSVM is more explainable (Fig. 5) than common fuzzy-based classifiers from the literature (Luukka, 2010; Hsieh et al., 2015; Saastamoinen, 2016), which are not only complex architecturally but did not have clinically viable classification performance (ACC < < 80%, as specified in section 1.3) (Parisi et al., 2018b). Thus, in binary classification problems, intrinsic ML-based FS was enough to devise a clinically viable ML-based decision support system for diagnostic and prognostic purposes, respectively facilitating early diagnosis of PD (Parisi et al., 2018a) and prediction of early prognosis in patients with chronic hepatitis (Parisi et al., 2019), whilst retaining explainability in their architectures (Figs. 2 and 4) and results (Figs. 7 and 8) to meet end-users’ requirements as per section 1.3.

To aid early diagnosis of PD based on voice signals, the MLP-LSVM classifier was developed and validated on the same PD benchmark dataset (Sakar et al., 2013) against ML-based classifiers from published studies (Parisi et al., 2018a). Moreover, unlike the classifier of Benba et al. (2017), the reliability and explainability of the MLP-LSVM was not only higher based on selected features from three sustained vowels only, but on those from the whole dataset, thus including numbers, words and short sentences, which further supports its clinical viability.

Moreover, to further demonstrate the MLP-LSVM as a clinically viable ML-based solution to aid early diagnosis of PD, both accuracy (ACC) and reliability (SN, SP and AUC) were reported and with confidence intervals (Parisi et al.,
i.e., not only partially and as point estimates only, such as in all previous studies deploying the same benchmark dataset (Cantürk & Karabiber 2016; Alhussein, 2017; Benba et al., 2017; Sakar et al., 2017). Moreover, the most important features were also visualised in an explainable manner from a clinical standpoint too (Fig. 1, Parisi et al., 2018a).

Thus, there is strong evidence supporting the application of the MLP-LSVM for tele-diagnostic and tele-monitoring systems for achieving a semi-automated early diagnosis of PD (Little et al., 2009), as well as devising patient-specific, more effective treatments (Parisi et al., 2018a).

With a reverse architecture but retaining the same degree of explainability, the LSVM-MLP model (Fig. 4) was designed and validated (100% of ACC and AUC, 9 epochs, 0.09s of computational time) to help physicians achieve an early prediction of survival in patients with hepatitis, thus assisting them in personalising treatments to maximise clinical outcomes (Parisi et al., 2019).

The LSVM-based feature reduction implicitly optimised the gradient descent algorithm for the MLP classifier to converge to an optimal solution and with the lowest computational cost (Parisi et al., 2019). The LSVM-MLP, whose performance was compared in an explainable manner against algorithms from previous studies and commercially available solutions based on clinical gold standard metrics of accuracy (ACC) and reliability (SN, SP and ROC-AUC) (Fig. 4), was confirmed as a clinically viable solution for aiding early prediction of survival in patients with chronic hepatitis. Furthermore, the most relevant features were also evaluated in an explainable manner to be clearly understood by physicians (Fig. 3, Parisi et al., 2019).
In fact, the LSVM-MLP is the most accurate, reliable and the most explainable ML-based classification architecture as compared to the only classifier from the literature of comparable performance (Kaya & Uyar, 2013) and the most similar algorithm to it that could be built via commercially available software, i.e., an SVM-MLP implemented via Microsoft Azure ML Studio (Parisi et al., 2019) (Fig. 8).

**Figure 8.** Four-dimensional (4D) plot comparing the classification performance of the customised algorithms (MLP and LSVM-MLP), the SVM-MLP implemented in Microsoft Azure, and the AI-based classifiers used in published studies, as assessed via AUC, ACC, SN and SP. GA: genetic algorithm; WK: wavelet kernel; ELM: extreme learning machine; RS: rough set (Parisi et al., 2019).

The lack of the augmented Lagrangian formulation of the SVM (LSVM) in Microsoft Azure is a slight limitation of the above-mentioned evaluation,
although considering the findings from multiple evaluations, the LSVM-MLP can be safely regarded as a clinically viable ML-based solution for assisting with early prediction of the prognosis in patients with chronic hepatitis (Parisi et al., 2019).

Besides meeting the end-users’ requirements as specified in 1.3 and discussed in this section, to enable the prospective design and validation of novel ML-based algorithms deployed for either diagnostic or prognostic applications in the medical field, the author and his collaborators developed a quality assessment scale (MQAS), which can be also used to evaluate the quality of published studies deploying AI in healthcare (Parisi et al., 2019).

### 3.6.4 Considerations for applications at scale

Optimising kernel and hyperparameters via GA led to the maximum classification performance (ACC/SN/SP: 100%; AUC: 1) in the hybrid algorithm GA-M-LSVM in being able to predict optimal postoperative recovery area regardless of unbalanced classes (2 patients to be sent to the ICU, 24 to be discharged home and the remaining 64 to be sent to a general hospital ward; Summers et al., 1993) (Parisi et al., 2018b). This finding, although limited at the same time by the sample size that is not entirely representative of the patient population (considering only 2 patients that were sent to the ICU), demonstrates the reliability of the hybrid decision support system in handling such highly heterogeneous underlying patterns in the initial postoperative data in either cohorts with regards to the degree of severity in patient status.
GA enabled each of the binary classifiers in the ‘controlled’ AvA multi-class classifications system to improve their classification performance, with seamless integration with the loss function to ensure a higher reliability, even in presence of unbalanced classes, and explainability of such a knowledge-based system (Parisi et al., 2018b).

Besides the testing classification accuracy (ACC), the SN, SP and AUC further demonstrate its reliability regardless of the peculiarity of the clinical cases postoperatively.

Differently from previous studies that deployed the same dataset (Luukka, 2010; Forghani & Yazdi, 2015; Hsieh et al., 2015; Saastamoinen, 2016; Abuaqel et al., 2017), Parisi et al. (2018b) reported all classification outcomes with their corresponding confidence intervals, instead of their point estimates alone, thus enabling to assess the reliability of the methods developed more transparently.

The accuracy and reliability of the proposed ML-based classifiers, along with their explainability, demonstrate their clinical viability to allow for automated prediction of optimal postoperative recovery area in a patient-specific manner (Parisi et al., 2018b).

Unlike the best performing algorithm from previous studies to aid early diagnosis of PD based on the same benchmark speech dataset, i.e., the LOSO-SVM (Benba et al., 2017), the classification performance of the hybrid algorithm (MLP-LSVM, Parisi et al., 2018a) was consistently higher regardless of the number of input features selected. For instance, when deploying the 20 features associated with the highest weight towards the classification of
interest, the LOSO-SVM (Benba et al., 2017) had an AUC of 0.6 and it took 130 training iterations or epochs to converge, whilst the MLP-LSVM (Parisi et al., 2018a) converged in only 3 epochs and reached an AUC of 1 without overfitting.

To further evaluate the reliability of the proposed method (MLP-LSVM) for aiding early diagnosis of PD using voice signals, its classification accuracy and AUC were consistently higher - by 27.83% and 33.33% respectively - than those of an MLP-SVM implemented in a commercially available software, Alyuda NeuroIntelligence (2.2 (577), Cupertino, California, USA) (Parisi et al., 2018a).

The limitation in the commercial software consisting in not offering the out-of-the-box capability to build an MLP-LSVM slightly impairs the generalisation of the findings but it does not impair the superior accuracy, reliability and explainability of the proposed algorithm (Parisi et al., 2018a).

To further consolidate its clinical viability, the number of training iterations and the computational time were decreased by 95.52% and 99.6% (Parisi et al., 2018a).

Furthermore, the reduced confidence interval shows a remarkably increased confidence in the solution when deploying the MLP-LSVM (Parisi et al., 2018a).

Such improvements in classification performance are due to the improved margin maximisation and higher linear separability of the input data into the target classes considered (patients with PD and healthy subjects), as well as
the hyperparameter optimisation of the number of hidden neurons in the optimal MLP (N = 10) as compared to the best performing MLP in Alyuda (N = 3) (Parisi et al., 2018a).

Therefore, considering its classification performance and explainability, the hybrid algorithm (MLP-LSVM) is deemed a clinically viable decision support system for aiding early diagnosis of PD based on speech data, but its performance will have to be further demonstrated for other healthcare-related applications by considering all performance measures used in Parisi et al. (2018a), i.e., classification accuracy on the testing set (ACC), reliability (SN, SP and AUC) and computational cost (number of epochs and computational time).

Reverting the selection on which algorithm was used for intrinsic FS and which one for classification has also enabled to devise a highly accurate, reliable and explainable decision support system (LSVM-MLP), but for a different binary classification task, i.e., to predict the survival in patients with chronic hepatitis considering demographic information (age and gender) and haematological data (Parisi et al., 2019).

The reliability of the algorithm was demonstrated by its consistently high classification performance on a further benchmark dataset on patients with chronic hepatitis from the UCI ML repository (Santos et al., 2015), wherein the follow-up time was one (N = 1) year for all patients, to verify whether the classification performance of the proposed algorithm was consistently higher than other algorithms used for comparison regardless of the patient dataset
considered and the duration of follow-up time at which survival was assessed (Parisi et al., 2019).

To further evaluate its reliability, the classification performance of the LSVM-MLP was also higher than that of an SVM-MLP built via Microsoft Azure ML Studio (Parisi et al., 2019), applying the same data pre-processing. In Microsoft Azure ML Studio, the augmented Lagrangian formulation of the SVM is not available, thus a linear SVM was used (Parisi et al., 2019), which partly limits the reliability of such a comparison, although the algorithmic architecture is as close as possible to that of the customised algorithm.

3.7 Conclusions
The author's preliminary research (Parisi & Manaog, 2016) demonstrated the clinical viability of the LSVM as an ML-based decision support tool to aid early prediction of prognosis in patients with chronic hepatitis. By using selected features based on the average ranking derived from three extrinsic FS algorithms, the author's subsequent research also proved another ML algorithm - the MLP - to be a clinically viable decision support system for the above-mentioned purpose (Parisi & Manaog, 2017a).

Nevertheless, to devise a clinically viable solution to be applied in a clinical setting, the author pursued this research further and derived a hybrid model that leverages intrinsic ML-based FS (LSVM-MLP) to improve the prediction of prognosis in patients with chronic hepatitis, which involves a binary classification task aimed at predicting which patients survive or die based on demographics and hematological data (Parisi et al., 2019). Via a reverse
approach (MLP-LSVM), the author demonstrated the clinical viability of intrinsic ML-based FS for a further binary classification problem, i.e., facilitating early diagnosis of PD in a clinical setting via a more accurate, reliable and explainable algorithm (Parisi et al., 2018a).

Therefore, the author’s research demonstrates how intrinsic ML-based FS can improve binary classification performance with regards to both diagnostics (Parisi et al., 2018a) and prognostics (Parisi et al., 2019).

Conversely, for multi-class classification problems, the author found GA, a type of EAs, to be a clinically viable tool for optimising ML-based classifiers to improve postoperative discharge decision making (Parisi et al., 2018b). In such an instance, a computationally efficient GA was used for optimising the kernel and the hyperparameters of an M-LSVM, thus deriving a novel hybrid architecture (GA-M-LSVM) (Parisi et al., 2018b).

Considering their superior classification performance and higher explainability with respect to algorithms from previous studies (e.g., the LOSO-SVM from Benba et al., 2017) and from commercial software (e.g., Microsoft Azure ML Studio and Alyuda NeuroIntelligence) on the same benchmark data, the author’s algorithms (Parisi et al., 2018a, 2018b, 2019) are not only novel and thoroughly validated but clinically more viable.

The suitability of such methods to different applications and pathophysiology (Parisi et al., 2018a, 2018b, 2019), along with their high performance and low computational cost, makes them clinically viable.
4
Research findings meeting aims and objectives

4.1 Salient points addressing research questions

In this section, salient findings derived from the work of the author are assessed to understand and appreciate how they answered the research questions in 1.3 via the methodological approaches of his research, presented in section 1.4.

The salient points addressing the research questions in 1.3. (R1-R4) and corresponding methodological approaches in 1.4 (O1-O6) are the following:

(i) Applying an intrinsic FS from an ML-based classifier (MLP) to select the most important feature vectors, which improve linear separation of the input data into target classes, as inputs to another ML algorithm (LSVM) led to a novel hybrid model (MLP-LSVM) that significantly improved classification performance in early detection of PD from speech data (Parisi et al., 2018a). This is the first noteworthy binary classification problem wherein intrinsic ML-based FS was pivotal in improving predictive performance of an ML-based
classifier, whilst retaining explainability, which fulfils O1 and addresses R1, R2 and R4 (Parisi et al., 2018a).

(ii) Deploying an intrinsic feature ranking algorithm based on LSVM to select the reduced feature set, which improves the delineation of the OSH to maximise the geometric margin, as inputs to another ML algorithm (MLP) led to the creation of a hybrid algorithm (LSVM-MLP) that considerably increased classification performance in predicting survival in patients with chronic hepatitis (Parisi et al., 2019). This is the second noteworthy binary classification problem wherein intrinsic ML-based FS was pivotal in improving predictive performance of an ML-based classifier, whilst ensuring interpretability, which meets O2 and tackles R1, R2 and R4 (Parisi et al., 2019).

(iii) Leveraging a tightly coupled EA technique (GA) as an optimisation method for a multi-class ML-based classifier (M-LSVM) led to an improvement in performance of a hybrid algorithm (GA-M-LSVM) in predicting early diagnosis of hypothermia (Parisi et al., 2018b). This is the first noteworthy multi-class classification problem wherein tightly coupled GA-based optimisation was pivotal in improving predictive performance of an ML classifier, whilst preserving explainability, which covers O3 and pertains to R1, R2 and R4 (Parisi et al., 2018b).

(iv) Using a tightly coupled GA-based optimisation method led to improve the prediction of the optimal postoperative recovery area to improve patient outcome via an ML-based classifier accurately and reliably (Parisi et al., 2018b). This is the second noteworthy classification problem involving multiple
target groups wherein tightly coupled GA-based optimisation was fundamental in improving predictive performance of an ML classifier, whilst retaining explainability, which satisfies O4 and relates to R1-R4 (Parisi et al., 2018b).

(vi) Research showed that applying either intrinsic ML-based FS or tightly coupled GA-based optimisation generally leads to improvements in the predictive performance in binary (i, ii) (Parisi et al., 2018a, 2019) and multi-class classification (iii, iv) (Parisi et al., 2018b) problems respectively, whilst preserving interpretability, as per the hybrid decision support systems developed, which contributes to attain O6 and is referred to R1-R4 (Parisi et al., 2018a, 2018b, 2019):

(vii) Intrinsic ML-based FS was found to lead to improvements in performance and explainability in supervised learning (i-ii), which provides further evidence to meet O6 and address R1-R4 (Parisi et al., 2018a, 2019).

(viii) Tightly coupled GA-based optimisation was found to lead to improvements in performance in supervised learning (iii-iv), whilst ensuring explainability, which further supports the above-mentioned findings to fulfil O6, covering R1-R4 (Parisi et al., 2018b).

(ix) Findings indicate intrinsic ML-based FS may lead to improvements in binary classifiers (i-ii) (Parisi et al., 2018a, 2019), whilst GA-based optimisation leads to improvements in multi-class classifiers (iii-iv) (Parisi et al., 2018b), whilst retaining interpretability, which leads to satisfy O6 and covers R1-R4 (Parisi et al., 2018a, 2018b, 2019).
Conclusions and future work

In this section, relevant and important conclusions are derived from the salient findings reported in the work of the author, highlighting their added values both from a theoretical and a practical standpoints, respectively contributing to enrich and expand the body of knowledge in the field of AI and Applied ML in healthcare. In particular, this section focuses on the importance of the author’s research in demonstrating how intrinsic ML-based FS and tightly coupled GA-based optimisation were pivotal respectively in aiding binary and multi-class classification problems, regardless of whether they were diagnostic or prognostic.

5.1 Conclusions

Generic conclusions are provided below to wrap up the author’s research in the hope of aiding the reader to appreciate the answer to the research questions in Chapter 1. This summative section further highlights how the knowledge gaps in the literature presented in Chapter 2 were filled. Moreover, specific contributions to knowledge and salient findings as per Chapters 3 and 4 respectively were hereby summarised.
Previous research studies applied extrinsic FS to aid diagnostics, e.g., of PD, and prognostics, e.g., of chronic hepatitis and to optimise post-operative discharge decision making, without achieving clinically viable classification performance (Luukka, 2010; Ulutasdemir & Dagli, 2010; Hayashi & Fukunaga, 2016; Alhussein, 2017; Sakar et al., 2017) and explainability (Gill & Johnson, 2009; Tan et al., 2009; Calisir & Dogantekin, 2011; Bascil & Oztekin, 2012; Kaya & Uyar, 2013; Forghani & Yazdi, 2015; Hsieh et al., 2015; Nahato et al., 2015; Cantürk & Karabiber, 2016; Abuaqel et al., 2017; Benba et al., 2017).

Moreover, many previous studies did not report the use of any adequate data pre-processing techniques or the computational cost involved in the required ML-based classification (Calisir & Dogantekin, 2011; Kaya & Uyar, 2013; Nahato et al., 2015). Conversely, the author demonstrated that intrinsic ML-based FS can improve such classification for both diagnostics (for PD – Parisi et al., 2018a) and prognostics (for chronic hepatitis – Parisi et al., 2019), as well as its interpretability too.

This is an important contribution to fill such a research gap with regards to binary classification. Moreover, hybrid feature engineering syncretising intrinsic ML-based FS and clinical biomarkers was proven to further improve the prediction of patients with chronic hepatitis, even against out-of-the-box commercially available ML solutions (Parisi et al., 2019).

With respect to multi-class classification, e.g., the prediction of post-operative recovery area to maximise patient outcome, the author validated the use of EA to achieve a clinically viable ML-based decision support system, thus filling a further important knowledge gap in the literature (Parisi et al., 2018b).
particular, optimisation leveraging a sub-group of EA algorithms, i.e., GA, was found to improve such a multi-class classification performance (Parisi et al., 2018b).

Moreover, the author’s research presented standardised data processing methods that further contribute to the explainability of the proposed binary (Parisi et al., 2018a; Parisi et al., 2019) and multi-class classifiers (Parisi et al., 2018b). The author also produced and validated a quality assessment scale (MQAS) that can be applied not only to retrospectively assess the quality of published studies using supervised ML in healthcare but also as a set of guidelines to drive the prospective development of ML algorithms for aiding both health diagnostics and prognostics (Parisi et al., 2019).

To summarise, clinical viability, involving accuracy, reliability and interpretability, was achieved in binary and multi-class ML-aided diagnostics and prognostics by leveraging intrinsic ML-based FS and EA respectively. This is additional, important knowledge to the current body of the literature in both theoretical and applied AI, which can enable their clinical translation.

5.2 Limitations and value from published work

The classification performance and explainability of the novel hybrid models involving intrinsic ML-based FS to aid in binary classification tasks, e.g., early detection of PD from voice signals (MLP-LSVM, Parisi et al., 2018a) and to predict prognosis in patients with chronic hepatitis (LSVM-MLP, Parisi et al., 2019), demonstrates their clinical viability. Particularly, as a measure of its impact in the scientific research community, the work in Parisi et al. (2018a)
has been able to attract citations from two journal articles published in ‘Expert Systems with Applications’ (Ali et al., 2019; Wang and Xing, 2019), one in Nature’s ‘Scientific Reports’ (Shahtalebi et al., 2020) and another one in ‘Pattern Recognition Letters’ (Almeida et al., 2019). Thus, intrinsic ML-based FS, when coupled with binary ML-based classifiers, can assist in both diagnostic and prognostic tasks in a clinical setting.

Instead, for multi-class classification problems, a type of EA (GA) was found to optimise multi-class classifiers, evaluated by the author in aiding the prediction of the optimal postoperative recovery area, thus improving prevention of hypothermia and the overall postoperative discharge decision making and patient outcome (GA-M-LSVM, Parisi et al., 2018b). This work in Parisi et al. (2018b) has been able to attract a citation from a journal article published in ‘Soft Computing’ (Wang et al., 2020). Thus, tightly coupled GA-based optimisation, along with multi-class ML-based classifiers, was found clinically viable, hence being able to assist in both diagnostic and prognostic tasks in a clinical setting.

The classification performance (accuracy and reliability) and explainability of the author’s hybrid decision support systems was thoroughly evaluated against state-of-the-art ML-based classifiers, both with customised implementations and readily available in commercial software from leading cloud vendors (e.g., Microsoft Azure), as well as algorithms from selected published studies. The proposed hybrid algorithms proved to be the most accurate, reliable (100% of ACC and 1 of AUC), explainable and
computationally the fastest ML-based classifiers for both binary and multi-class diagnostic and prognostic tasks (Parisi et al., 2018a, 2018b, 2019).

Such novel hybrid algorithms (Parisi et al., 2018a, 2018b, 2019) can be deemed clinically viable as assistive tools for decision support in a clinical setting for both diagnostic and prognostic applications. Moreover, considering the demonstrated advantage in reliability of the intrinsic ML-based FS (Parisi et al., 2018a; Parisi et al., 2018b) and of GA-based optimisation (Parisi et al., 2019) respectively for binary and multi-class classifiers, such methods have the advantage of being applicable to further diagnostic and prognostic classification problems, with healthcare data and beyond too.

Although with a limited sample size for one of the three classes considered in one study (2 patients whose optimal postoperative recovery area was the ICU, Parisi et al., 2018b) and considering the potential advantage of using decision tree-related algorithms with regards to explainability for the classification problems tackled, a clinically viable predictive performance was achieved whilst retaining explainability. In fact, such a limitation in sample size available from the publicly available benchmark data sources in Parisi et al., (2018b) represented a challenge initially in understanding which type of GA-based optimisation had to be leveraged, as well as to define which multi-class classification strategy to apply (e.g., between AvA and OvA). Thus, if possible, it is recommended to perform the required prospective data collection or further retrospective data gather that is required to meet the requirements of the experimental design of the ML-based decision support system being devised.
Moreover, despite the limited performance metrics assessed in the author’s published studies (Parisi et al., 2018a, Parisi et al., 2018b, Parisi et al., 2019), such as accuracy (ACC) and reliability (SN, SP, ROC-AUC), further metrics could be used to evaluate classification performance more reliably in presence of imbalanced dataset, such as precision, F1-beta score and the precision-recall area under the curve.

Interpretability was not only attained for the assistive decision-making process but also for the most salient features leveraged in it from the benchmark datasets of interest, which further contribute to the clinical viability of the proposed algorithms.

5.3 Future work

This section highlights the author’s future vision, plan to tackle limitations of his published works to date and further knowledge gaps, as well as potential avenues for those who seek to progress the work undertaken by the author and his collaborators.

The classification performance and explainability of the proposed algorithms make them clinically viable decision-making support systems for both diagnostic and prognostic tasks (Parisi et al., 2018a, 2018b, 2019). In fact, findings from recent published works that cited the author’s research demonstrated the application of similar decision support systems for online applications (Wang and Xing, 2019) and for improving deep brain stimulation in PD patients (Camara et al., 2019). Nevertheless, the author is planning to investigate further performance metrics that may be more suitable to assess
the performance of the classifiers developed more reliably, as mentioned in 5.2. Furthermore, the author will be working on extending the validation of the data processing pipeline and quality assessment scale devised (MQAS) in further binary and multi-class classification problems in healthcare. An assessment of relatively novel methods, such as data augmentation, against traditional undersampling and oversampling methods to handle imbalanced datasets, typical in healthcare, will be required. Such an analysis will then contribute to expand the data processing pipeline devised and corroborate its reliability.

There is also future potential to assist with early diagnostic and prognostic prediction of other pathologies, e.g., mental health-related disorders, such as depression and anxiety, and haematological conditions, e.g., anaemia, bleeding disorders and blood cancers. The author, as an Editorial Board Member and Lead Guest Editor for the Springer-Nature journal ‘Neural Computing and Applications’, is planning to publish a special issue, now named ‘Topical Collection’, which will attract high-quality full-length journal articles featuring the above-mentioned applications, as well as extended to other use cases in healthcare and other industries, such as financial services.

Overall, the future vision is to extend the application of intrinsic ML-based FS methods and EA-based optimisation to a wider range of diagnostic and prognostic binary and multi-class classification problems respectively, involving mainly mental health-related disorders and neurodegenerative diseases. In particular, the focus of the author will be on applying the proposed
binary classifiers to help diagnose and treat depression and anxiety, as well as improve early detection and prognosis in patients with PD.

Future work can involve real-time implementations of the hybrid algorithms developed, tested, and validated in these studies (Parisi et al., 2018a, 2018b, 2019) for demonstrating their accuracy and reliability further in the above-mentioned applications and other research domains too.
References


41. Parisi L, Manaog ML (2016). Preliminary validation of the Lagrangian support vector machine learning classifier as clinical decision-making support tool to aid prediction of prognosis in patients with hepatitis. In:
The 16th international conference on biomedical engineering, National University of Singapore (NUS).


Appendix A – Statement of independence

To undertake the work hereby submitted, a multi-disciplinary collaboration with other experts in medical applications of Artificial Intelligence from both academia and industry was thus required. Therefore, it was essential for the author to cooperate as a part of a cohesive team on every project wherein he contributed.

In his collaborations, the author led the design and development of novel decision support systems to be used in a clinical setting presented in the submitted work. This work involved: (i) applying design thinking strategies and cutting-edge machine learning-based techniques to devise novel decision support systems; (ii) liaising with experts from industry to ensure the decision support systems devised were minimum viable solutions for diagnostic or prognostic purposes based on the application of interest; (iii) delivering technical and strategic presentations to educate and inform the teams at MedInteligible and the Biomechatronics Engineering Group at the University of Auckland regarding the potential of applying machine learning and evolutionary algorithms for medical applications; (iv) critically appraising the work of others, to inform the direction of the collaborative work. The author’s knowledge of machine learning and evolutionary algorithms, along with his expertise in data science, medical engineering, applied mathematics, advanced statistics and computational modelling enabled him to provide key outputs as part of the body of the work hereby submitted.
A.1 Specific contributions by publication


**Description**: This work presents a novel machine learning-based decision support system to aid early diagnosis of Parkinson’s Disease, which was conceived and devised by the author. The author led the work with ML Manaog and N RaviChandran, which was mainly executed by the author.

**Independent contribution(s)**: analysis, interpretation and preprocessing of the data, feature engineering and selection, machine learning modelling design and implementation, main contribution in writing of the paper.


**Description**: This work presents a novel hybrid machine learning-based decision support system to aid early prediction of prognosis in patients with hepatitis, which was conceived and devised by the author. The author led the work with ML Manaog and N RaviChandran, which was mainly executed by the author.
Independent contribution(s): analysis, interpretation and pre-processing of the data, feature engineering and selection, machine learning modelling design and implementation, conceiving and mainly devising a novel quality assessment scale (MQAS) for evaluating the quality and rigour of peer-reviewed journal articles that present machine learning models applied on healthcare data for either diagnostic or prognostic purposes, main contribution in writing of the paper.


**Description**: This work presents a novel multi-class classification decision support system to aid early prediction of hypothermia and improve postoperative discharge decision making, which was conceived and devised by the author. The author led the work with ML Manaog and N RaviChandran, which was mainly executed by the author.

**Independent contribution(s)**: analysis, interpretation and pre-processing of the data, feature engineering and selection, machine learning modelling design and implementation, main contribution in writing of the paper.
A.2 Co-authors confirmation

In support of the above-mentioned statements, supporting letters from the author’s main collaborators, Marianne Lyne Manaog, Narrendar RaviChandran, are provided below.
Department of Mechanical Engineering
Mechatronics Engineering Research Group
University of Auckland
20 Symonds Street
Auckland CBD
Auckland, 1010
New Zealand

To whom it may concern:

This letter is as an acknowledgement that I have read and agree with “Appendix A – Statement of independence” in Luca Parisi’s PhD thesis entitled “Machine Learning-Based Feature Selection and Optimisation for Clinical Decision Support Systems”.

As a co-author, I confirm that the overall descriptions and specific contributions are true and correct to the best of my knowledge. In particular, I can confirm the veracity of the statements for the specific works I co-authored with Luca Parisi listed in Tables 1 and 2.

Table 1. Published journal papers and a conference paper submitted for inclusion in the PhD.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal Paper</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Conference Abstract</td>
<td>Contribution</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>

P = available in the public domain; R = refereed journal paper; ** = principal author.

Table 2. Supporting materials.
Thank you very much for your consideration in this matter. If you have any further queries, please do not hesitate to contact me.

Very truly yours,

R.C. Narrendar

Narrendar RaviChandran, BEng (Hons), MSc, PhD Candidate (final year)

PhD Research Scholar,
Department of Mechanical Engineering
Mechatronics Engineering Research Group
University of Auckland
20 Symonds Street
Auckland CBD
Auckland, 1010
New Zealand
16 October, 2019
Re: Mr Luca Parisi

Division of Applied Artificial Intelligence in Healthcare,
MedIntellego®
4 Colonel Barton Glade,
St Johns,
Auckland 1072,
New Zealand

To whom it may concern:

I acknowledge that I have read and agree with “Appendix A – Statement of independence” in Luca Parisi’s PhD thesis entitled “Machine Learning-Based Feature Selection and Optimisation for Clinical Decision Support Systems”.

As a co-author, I confirm that the descriptions and contributions are true and correct to the best of my knowledge. I can also confirm the veracity of the statements for the works I co-authored with Luca Parisi referenced in Tables 1 and 2.

Table 1. Published journal papers and a conference paper submitted for inclusion in the PhD.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal Paper</th>
<th>Contribution</th>
</tr>
</thead>
</table>


P = available in the public domain; R = refereed journal paper; ** = principal author.

Table 2. Supporting materials.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Conference Abstract</th>
<th>Contribution</th>
</tr>
</thead>
</table>

101

P = available in the public domain; R = refereed conference abstract; ** = principal author.

Thank you for your request on this matter.

If you have any further enquiries, please do not hesitate to contact me.

With best wishes,

Marianne Lyne Manaog, BSc (Hons), PGDip
CEO,
Division of Applied Artificial Intelligence in Healthcare, MedIntelegeo®
4 Colonel Barton Glade,
St Johns,
Auckland 1072,
New Zealand
A.3 Approvals from publishers to share published work

This sub-section of Appendix A includes the written approval obtained from the publishers, Elsevier for Parisi et al. (2018a) and Parisi et al. (2018b), Springer for Parisi et al. (2019), to share the full papers in this thesis, as a part of the submission and publication of this PhD thesis itself.
**Order Completed**

Thank you for your order.

This Agreement between University of Bradford -- Luca Parisi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

**printable details**

<table>
<thead>
<tr>
<th>License Number</th>
<th>462662342848</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Jul 12, 2019</td>
</tr>
<tr>
<td>Licensed Content</td>
<td>Springer Nature</td>
</tr>
<tr>
<td>Publication</td>
<td>Neural Computing &amp; Applications</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Luca Parisi, Narrendar Ravichandran, Marianne Lyne Manag</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jan 1, 2019</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requester type</td>
<td>academic/university or research institute</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>full article/chapter</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>no</td>
</tr>
<tr>
<td>Circulation/distribution</td>
<td>&gt;50,000</td>
</tr>
<tr>
<td>Author of this Springer Nature content</td>
<td>yes</td>
</tr>
<tr>
<td>Title</td>
<td>PhD by Published Work</td>
</tr>
<tr>
<td>Institution name</td>
<td>University of Bradford</td>
</tr>
<tr>
<td>Expected presentation date</td>
<td>Jan 2023</td>
</tr>
<tr>
<td>Order reference number</td>
<td>UoB</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>University of Bradford</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 GBP</td>
</tr>
</tbody>
</table>

**ORDER MORE** | **CLOSE WINDOW**
A.4 Relevant milestones for published work to date

This sub-section of Appendix A includes the main milestones achieved to date for the peer-reviewed journal papers listed in Table 1, including the number of citations for the most cited paper (Parisi et al., 2018a), Elsevier's article metrics for Parisi et al. (2018a) and Parisi et al. (2018b) and the number of downloads to date for the most recent paper of Parisi et al. (2019).
Feature-driven machine learning to improve early diagnosis of Parkinson's disease

Luo Peng, 
M. Z. Al-Ismail, 
A. M. Khan, 
M. A. Al-Bloushi, 
N. A. Al-Kharusi, 
M. A. Al-Maskari 

https://doi.org/10.1016/j.eswa.2018.06.043

Highlights

- We developed a novel hybrid algorithm for aiding early diagnosis of Parkinson's disease.
- We compared the performance of the hybrid algorithm against commercially available software.
- We also compared the performance of the hybrid algorithm against models from published studies.
- Customized algorithmic implementation can improve the overall classification performance.
- Feature reduction via artificial neural networks increased the performance of the hybrid model.
Decision support system to improve postoperative discharge: A novel multi-class classification approach

Highlights

- We developed a novel hybrid algorithm for improving postoperative discharge decision making.
- We compared the performance of the hybrid algorithm against state-of-the-art classifiers.
- We also compared the performance of the hybrid algorithm against models from published studies.
- Optimization of kernel and hyperparameters can improve the overall classification performance.
- Genetic Algorithm led to a higher control over the decision boundary in the hybrid model.
A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis

Luca Parigi, Narrendar Revichandran & Marianne Lyne Manaog

Neural Computing and Applications (2019) | Cite this article

Abstract

This study investigated the application of a novel hybrid artificial intelligence (AI)-based classifier for aiding prediction of the prognosis in patients with chronic hepatitis. Nineteen biomarkers on 155 patients with hepatitis from the University California Irvine Machine Learning repository were used as input data. Weights derived by applying the geometric margin maximisation criterion of a Lagrangian support vector machine (LSVM) were used for selecting the features associated with the highest relative importance towards the required classification, i.e. to predict whether a patient with hepatitis would have survived or died. Thus, the 19 initial features were reduced to the 16 most important prognostic factors and were fed into various AI-based classifiers. Results indicated an overall classification accuracy and area under the receiver operating characteristic curve of 100% for the proposed hybrid algorithm, the LSVM multilayer perceptron (MLP), thus demonstrating its potential for aiding prediction of prognosis in patients with hepatitis in a clinical setting.
Appendix B – Published works for consideration

This appendix includes the author’s main peer-reviewed journal papers as listed in Table 1. Each document is reproduced as published.
Feature-driven machine learning to improve early diagnosis of Parkinson’s disease

Luca Parisi\textsuperscript{a,b,c}, Narreddar Ravichandran\textsuperscript{a,cd}, Marianne Lyne Manaog\textsuperscript{d}

\textsuperscript{a}Auckland School of Engineering (ASSE), University of Auckland, 70 Symonds Street, Grafton, Auckland 1010, New Zealand
\textsuperscript{b}Department of Electrical Engineering, University of Auckland, 70 Symonds Street, Grafton, Auckland 1010, New Zealand
\textsuperscript{c}Division of Oral Artificial Intelligence in HealthCare, Ministadent, 4 Cranwell Farms Road, St Johns, Auckland 0622, New Zealand
\textsuperscript{d}University of Auckland Bioengineering Technology Associates (AUBTA), University of Auckland, 70 Symonds Street, Grafton, Auckland, 1010, New Zealand

\textbf{ARTICLE INFO}

Article history:
Received 28 January 2018
Revised 31 May 2018
Accepted 1 June 2018
Available online 5 June 2018

Keywords:
Machine learning
Deep learning
Parkinson’s disease
Diagnosis
Feature selection
Multi-layer perception

\textbf{ABSTRACT}

Although advances in speech processing have facilitated the prognostic assessment of patients with Parkinson’s Disease (PD), there is no objective method towards an early detection in a clinical setting. This study investigated the application of a novel hybrid Artificial Intelligence-based classifier for aiding early diagnosis of PD. Data on dysphonic measures and clinical scores on 68 subjects were obtained from the University of California, Irvine Machine Learning Database. Weights derived from a Multi-Layer Perceptron (MLP) were applied for feature selection and their models were used to rank the input features based on their relative importance towards discriminating between physiological and pathological data patterns. Thus, the initial 27 features were reduced to 20 selected diagnostic factors. This reduced feature set was then input to a Logistic Support Vector Machine (LSVM) for classification. The overall performance of this hybrid feature-driven algorithm (MLP-LSVM) was then compared against commercially available software and classifiers from similar studies. Results indicate an overall classification accuracy and the area under the receiver operating characteristic curve of 100% for the proposed feature-driven algorithm (MLP-LSVM), with relatively faster convergence, thus demonstrating its potential for aiding early diagnosis of PD in a clinical setting.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Pattern recognition to aid diagnosis of Parkinson’s disease

Parkinson’s disease (PD) is a neurodegenerative disorder that affects 7–10 million people worldwide (Herrmann, 2011). As PD directly affects the nervous system, it often leads to tremors and other symptoms including impairments in locomotion, posture and speech. Conventional invasive methods for treating PD involve surgical interventions, such as pallidotomy, and deep brain stimulation (Englot, Moll, Fried, & Ojemann, 2005). Conversely, medical treatments that aim to suppress pathophysiological nerve impulses and improve sensorimotor synchronisation are the most common noninvasive treatment for PD; however, medications have not been found to help reduce the progression of PD (Satpute et al., 2009). It is worth noting that 20% of the people with PD remain undiagnosed (Schrag, Ben-Shlomo, & Quinn, 2002). In fact, it is clinically challenging to discern common signs of physiological ageing from subtle symptoms of PD in older adults (Das, 2010).

The two main vocal impairments due to PD are dysphonia, which is the inability to generate normal vocal sounds, and dysarthria, i.e., being hardly able to pronounce words. Both dysphonia and dysarthria are associated with decreased vocal loudness, instability and frequency abnormality (Sakur, Renig, Luschei, & Smith, 1998); conversely, hypophonia, which also occurs in PD patients, can lead to voice breaks and impaired vocal quality (Stelzig, Hochhaus, Gall, & Henneberg, 1999). Dysphonic metrics can aid the assessment of vocal impairments due to PD, thus being quite effective biomarkers to aid its diagnosis. Speech processing is used to detect anomalies in physiological speaking and it is often preferred in a clinical setting due to its non-invasiveness. Most of the conventional methods for aiding automated detection of PD-related vocal features are based on sustained vowel phonation, where subjects are asked to generate a steady vowel sound with respect to both amplitude and frequency (Sakur et al., 2005).

Over 90% of the patients with PD experience vocal disorders that are often undiagnosed (Das, 2010). Conventional ap...
proaches applied to speech-related data for aiding automated di-
agnosis of PD dealt with such highly non-linear data only from a
signal processing perspective, thus failing to recognise their un-
derlying patterns. Current speech processing methods heavily rely
on semi-automated extraction of temporal and frequency-related
features from the speech, such as the linear predictive analysis (DPC)
and the mel-frequency cepstral coefficients (MFCC) (Gass & Hanaz,
2011). These features indicate changes in loudness, breathi-
bspness, tone, amplitude, and stability. Considering the highly nonlin-
ear inherent noise in speech-related signals, adequate and highly
signal-specific data processing methods are required. Conversely,
Machine Learning (ML)-based algorithms provide a higher gener-
alisability, thus coping better with noise and outliers in vocal sig-
nals (Sakar, Seren, & Sakar, 2017). However, only recently, ad-
vances in automated speech processing using ML-based classifiers
have shown considerable potential to aid clinicians in perform-
ing an early diagnosis of Parkinson’s Disease to timely medical
 treatments to individual patients, improving clinical outcome. Such
methods for automated diagnosis of PD from anomalies in voice
signals are not only noninvasive but also highly specific in de-
tecting vocal impairments due to PD (Albashaw, 2017). Approp-
riate data pre-processing and feature selection highly influence the
performance of such speech processing algorithms (Pariit & Man-
ang, 2017a). Furthermore, similarly to principal component analy-
sis, as demonstrated in a recent study (Pariit, 2016a), such tech-
niques may alter the underlying patterns of speech data, thus
impairing the reliability of speech processing algorithms further.
As a complementary method for processing speech data, auto-
mated pattern recognition can help capture the underlying vocal
characteristics to discriminate between physiological and impaired
speech patterns (Pariit & Manang, 2017a). Therefore, pattern recog-
nition can be also used to derive more subject-specific indicators of
impaired speech that may help detect PD more accurately and reli-
ably.

1.2. Rationale and aim

Previous studies (Albashaw, 2017; Sakar et al., 2017) lack a clin-
ically significant classification performance. It was hypothesised
that this shortcoming is mainly due to the adoption of data-driven
approaches alone. Such approaches have still left some challenges
open, as mentioned in 1.1. An additional feature-driven approach
may lead to a higher classification performance, thus ensuring a
higher generalisability. In such a feature-driven classification ap-
proach, only selected features are given as inputs to the classifier.
In this instance, a feature selection (FS) framework intrinsic to the
classifier itself (using its own weights) was developed, rather than
using external FS methods, which are, instead, widely applied. Such
an internal FS would optimise linear separability of the input vocal
signals and improve the training of the proposed model. Since pre-
vious studies have preliminarily validated the combined use of Ar-
tificial Neural Networks (ANN) and Support Vector Machines (SVM)
for automatic detection of PD (ACC=.905, Gil & Johnson, 2009),
and have also demonstrated the potential of using SVM for auto-
mated detection of PD from vocal parameters (Sakar et al., 2017),
in this study, ANN (Pariit, 2014b; Pariit & Kavichandran, 2017)
was used for FS prior to classifying speech-related features via the
augmented Lagrangian formulation of SVM (SVM, Mangasian & Ma-
sicant, 2001; Pariit, Riggs, Whitting, & Holt, 2015; Pariit & Man-
ang, 2016), which optimises margin maximisation further with re-
spect to conventional SVM (Mangasian & Multari, 2001). Thus,
the main novelty of this study lies in using intrinsic FS to im-
prove early diagnosis of Parkinson’s Disease from speech-related
data.

2. Methods

2.1. Data and pre-processing

Data on vocal parameters (sustained vowels, numbers, words
and short sentences) and clinical scores from the unified Parkin-
son’s disease rating scale (UPDRS: 13 ± 15.89) of 40 subjects, i.e.,
20 patients with PD (age: 64.86 ± 8.97 years; gender: 6 females, 14
males) and 20 healthy controls (age: 62.55 ± 10.79 years; gender:
10 females, 10 males), were obtained from the University Califor-
nia Irvine Machine Learning repository (Sakar et al., 2015). An in-
dependent t-test was performed to verify whether there was a sta-

tistically significant gender-based discrepancy in the vocal param-
ters of the subjects considered in this study. The UPDRS is in the gold
standard assessment tool for characterising motor impairments in
people with PD (Song et al., 2009). Specific vocal samples were se-
lected by neurologists, as known clinical biomarkers for PD, from
trials of exercises that patients were asked to perform (Sakar et al.,
2013). Speech features from 168 recordings sampled at 96kHz of
28 additional patients with PD (62 ± 7.93), who were asked to
pronounce only the sustained vowel ‘a’ and ‘e’ three times each,
were also used (Sakar et al., 2015). All the recordings were col-
clected via a Trust MC-1500 microphone with frequencies ranging
from 98Hz to 331kHz (Sakar et al., 2013). The development and
testing of these algorithms were performed via MATLAB R2017b.
The MathWorks, Inc. on an Intel® Core™ i5-4460 2.70GHz Pro-
cessor. All input features and targets were normalised by mapping
their minimum and maximum values in the range between −1 and
+1 for the MEP as per the range of its hyperbolic tangent sigmoid
transfer function, for the SVM used in this study and described in
2.2, values were normalised between 0 and 1. Since the ranges
of the vocal parameters were far different from each other, varying
widely from 0 to 1.690 in value, they were then standardised as z-
scores to remove any feature scaling and improve convergence of
the classifiers thereafter (Ioffe & Szegedy, 2015), i.e., their values
were transformed to have a mean of zero and a standard deviation
of 1. Feature datasets for every subject that were two or more
standard deviations away from the mean were considered potential
outliers and, therefore, were removed accordingly. Finally, the ML-
based classifier was deployed along with a ML-based FS method
for automated detection of PD in the above-mentioned patients.

2.2. A hybrid machine learning-based algorithm for feature selection
and classification

To improve automated detection of PD, we propose to use the
Multi-Layer Perceptron (MLP) to increase the number of initial input features (N=27) to the most relevant ones for the classification of interest. Such a reduced feature set is obtained by applying the gradient descent training algorithm, whose gradient is computed with the backpropagation method.

Via the gradient descent training algorithm, the MLP aims to
minimise the error term ϵ as described via the following cost func-
tion (1):

\[ \epsilon(w_m, \theta) = \frac{1}{2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \]  

(1)

where,

\[ z_i = g(w_{m,i} \cdot x_i + b_i) \]  

(2)

where \( z_i \) is the hyperbolic tangent sigmoid transfer function, \( N \) is the number of training iterations or epochs and \( y_i \) \( \hat{y}_i \) is the output from the hidden layer for weights \( w_i \) and inputs \( x_i \). Training

involves the computation of weights \( w_m \) for which the error

\( \epsilon \) is the difference between the actual and the target outputs. The
Fig. 1. The proposed novel hybrid algorithm using the Multi-Layer Perceptron (MLP) for feature selection and the Lagrangian Support Vector Machine (LSVM) for classification in an early diagnosis of Parkinson's Disease (PD) from vocal parameters and clinical scores. TP: true positive rate; TN: true negative rate; H: healthy; P: pathological (patients with PD).

iterative training algorithm is stopped when the target MSE (0.05) is reached.

In the back-propagation algorithm, as mathematically represented in the FS stage of Fig. 1, the initially randomised weights inputs are propagated forward, whilst the errors are iteratively propagated backwards until the adjustment of weights generates the least mean squared error (MSE) (Hummel, Hinton, & Williams, 1986), thus maximising classification performance. As any Boolean bounded function can be mathematically described by a MLP with one hidden layer (Cybenko, 1989), only one hidden layer was implemented in the proposed hybrid model. To account for a larger range of inputs, the hyperbolic tangent sigmoid function was used as a transfer function in the hidden layer:

\[ a_i = \tanh(w_{ij}x_j + b_i) \]

\[ w_{ij} = w_{ij} - \eta \frac{\partial E}{\partial w_{ij}} \]

\[ \eta \in \mathbb{R} \quad [0 < \eta < 1] \]

The weights of the MLP-based weights can be interpreted as the relative importance of the input features towards the discrimination of the input data into the target output classes of interest, i.e., healthy controls and PD patients.

The MLP-based weights were used to rank the input features and find the optimal reduced feature set minimising the cost function (5) and yielding the maximum classification performance, as assessed via four gold standard performance measures, i.e., testing classification accuracy (ACC), sensitivity (SN), specificity (SP) and area under the receiver operating characteristic (ROC) curve (AUC). Since AUC is a more discriminating performance measure than accuracy, ACC was the ultimate criterion for deciding which algorithm performed best (Ling, Huang, & Zhang, 2003). Thus, it was hypothesised that this implementation not only would have maximised the overall classification performance in the hybrid algorithm, as quantified mainly by ACC and AUC, but also minimised its computational cost, as assessed via the number of training iterations (epochs) prior to convergence along with the computational time.

To obtain an optimal reduced feature set, the objective function (5) was devised to rank the relative importance of individual inputs in the initial feature set (F) towards the required classification:

\[ F = \text{minimise} \sum_{j=1}^{n} \left( \frac{TP}{TP+FN} \right) \left( \frac{TN}{TN+FP} \right) \frac{1}{\theta_b} \]

\[ F \text{ is the feature set } (|F| \leq |F_{\text{sort(weights)}}|) \text{ that includes all the } n \text{ input features to the MLP (Fig. 1). These features were ranked in decreasing order of importance towards the classification based on their weights (Fig. 1). In (5), TP stands for true positives, TN for true negatives, FP for false positives and FN for false negatives (Grankiewicz & Grankiewicz, 2001). Above in (5), \theta_b \text{ is the threshold whereby the receiver operating characteristic (ROC) curve is built.}

The reduced input features (A = 20) were then input to an ML-based classifier for discriminating which patients had PD based on vocal parameters and the UPDRS score. Considering their capability to cope with long-range, highly dimensional and highly nonlinear data, the proposed hybrid algorithms involve the MLP (Pardos & Manang, 2017a,b) in its feature reduction stage and the Lagrangian Support Vector Machine (LSVM) (Mangasarian & Musicam, 2001; Pardos & Manang, 2016) for classification. This hybrid model (MLP-LSVM) has been developed, in its simplicity but uniqueness, to assist physicians in clinical decision-making processes for aiding early diagnosis of PD and, therefore, helping them tailor treatments to individual patients with PD to improve patient outcome from the first-line treatment.

The LSVM algorithm is part of a category of supervised ML-based learning classifiers named “Support Vector Machines” (SVMs) that aims to generate an optimal separating hyperplane (OSH) (Fig. 1) in the decision-making surface boundary for a classification, by maximising the margin between the classes of the input dataset and minimising their training error (Cortes & Vapnik, 1995). SVMs are computationally efficient and generalise well as they can generate the OSH based on only a few inputs that lie on the decision boundary, named “support vectors” (Cortes & Vapnik, 1995). Nonlinearly separable data are typically projected over a higher dimensional space via a kernel function, thus facilitating hyperplane-based separation further. With respect to conventional SOA, the iterative Lagrangian augmentation formulation of SVM (LSVM) optimises the geometric margin maximisation for classification further and, therefore, it was selected for this study (Mangasarian & Musicam, 2001).

Applying the MLP for FS, the extent of discrimination of the input data into the two target classes (healthy controls and PD patients) was enhanced. By increasing the discrimination, the MLP-based reduced feature set guided the LSVM-related augmented Lagrangian algorithm to converge faster. Since the margin maximisation inherent to the LSVM algorithm was implicitly optimised via the MLP-based reduced input feature set, the error was also minimised. Furthermore, the adjustment of the weights in the MLP was not affected by the resampling of the input data, since the weights used for feature selection were taken at the last training iteration or epochs not only prior to convergence but further to achieving the maximum classification performance, as defined by 100% ACC and an AUC of 1. Therefore, by applying the optimal MLP in the feature selection stage of the LSVM, the hybrid algorithm (MLP-LSVM) could obviate this shortcoming and improve the classification performance of the initial LSVM. To the best of the authors’ knowledge, the MLP was deployed for the first time for FS.
of the LSVM in a study to aid early diagnosis of PD from vocal parameters and clinical scores on the motor impairment due to this disease. The learning parameter and the momentum term of the MLP were set to 0.0 and 0.8 respectively (Paris, 2014a). Considering the number of selected input features based on the weights derived from the MLP (N=20) (Paris & Mangun, 2017b), 20-fold cross-validation algorithm was used at each iteration to train the AI-based classifiers and the classification performance of the proposed hybrid algorithm (MLP-LSVM) was evaluated.

The entire modelling workflow adopted in this study is shown in Fig. 1.

2.3. Comparison of classification performance between the proposed hybrid model, commercially available software and published studies

To assess the need for customised algorithmic implementation, the proposed model was compared against existing models (MLP-SVM) in the black-box, commercially available software in Aplyosa Neurointelligence (2.2 (577), Cupertino, California, USA), and classifiers from published studies. A sensitivity analysis was also carried out to compare the performance of the proposed algorithm (MLP-LSVM) against the best performing classifier from published studies (leave-one-subject-out (LOSO)-SVM; Renna, Jilhão, & Ham mutual, 2017); the number of features was selected from the best performing classifier from the literature (N=15, 16; Renna et al., 2017) that is selected by the proposed classifier (N=20), and when all initial features (N=27) were utilised.

The same data pre-processing methodology in 2.1 was applied when assessing the performance of readily available binary learning-based classifiers in the commercial software.

Furthermore, the classification performance, as assessed via the four gold standard performance measures mentioned in 2.2.

High accuracy and reliability, but a low computational cost, were sought in determining which algorithm would have been suitable for clinical applications aimed at aiding early diagnosis of PD via automated speech processing. The term "computational cost" connotes the study to include both the computational time to execute an algorithm and the number of training iterations (epochs) prior to reaching convergence. Since the former metric is hardware-dependent, it was discarded from the radar plot in Fig. 3, whilst the latter measure was included as it does not depend on the hardware via which the algorithm has been tested, but only on the algorithm itself. Therefore, the number of epochs is a more important than the computational time when assessing the performance of AI-based algorithms.

3. Results

3.2. Feature selection to improve classification performance

Further to verifying that there was no statistically significant association between the gender of the subjects considered in this study and their vocal parameters (Liang et al., 2016), the reduced features via the MLP-based weights (N=20) were used for predicting PD between the patients (N=48) and the healthy controls (N=20)—considered in this study by considering their vocal parameters (Sakar, 2013). Considering the number of reduced input features (N=20), 20-fold CV was selected (Paris & Mangun, 2017b) and 10 hidden neurons were used in the hidden layer of the MLP (Table 1). Initially, the MLP and the LSVM were applied for classifying all input data features (N=27). Further to performing FS via the MLP-based weights, the classification performance of the hybrid algorithm MLP-LSVM was compared against that of the MLP and the LSVM algorithms (Table 1). Table 1 shows a quantitative comparison of classification performance between the MLP and the LSVM when considered individually, and the proposed hybrid classifier (MLP-LSVM), with the testing accuracy referred to as "ACC", the sensitivity as "SN", the specificity as "SP", and the area under the receiver operating characteristic curve as "AUC", the number of training iterations (epochs) and the computational time.

When the MLP achieved the highest average classification accuracy (100%) when all input features were used, the LSVM achieved 99.57% accuracy. Further to performing FS based on the weights of the MLP as described in 2.2, the novel hybrid algorithm (MLP-LSVM) also reached 100% of average classification accuracy but in a reduced number of training iterations (epochs) and decreased computational time (Table 1). In fact, by using only selected features via the MLP (N=20), the LSVM reduced both the number of epochs required prior to convergence (from 7 to 3, Table 1) and its computational time by five times (from 0.05 to 0.0), Table 1).

Specifically when using the novel hybrid algorithm (MLP-LSVM) as compared to the MLP (Table 1), the following percentage changes in classification performance measures were noted:

- An increase of 0.43% in training accuracy.
- An increase of 1.72% in ACC.
- An increase of 0.72% in average classification accuracy.
- An increase of 1.83% in SN.
- An increase of 1.58% in SP.
- An increase of 2% in AUC.
- A decrease of 51.94% in the number of epochs.
- A decrease of 90.1% in computational time.

3.2. A comparison of classification performance between the hybrid algorithm and commercially available software

Furthermore, the classification performance of the hybrid algorithm (MLP-SVM-MLP) was compared against a similar classifier but implemented using the commercially available software mentioned in 2.3, i.e., the hybrid MLP-SVM algorithm (Table 1). One limitation of such a software is the lack of algorithmic implementation of the augmented SVM; therefore, a conventional linear SVM was used. Consequentially, the customised MLP-LSVM had an increased overall classification performance as quantified by the following percentage changes in the performance measures used (Table 1):

- An increase of 25.08% in training accuracy.
- An increase of 27.18% in CV accuracy.
- An increase of 31.46% in ACC.
- An increase of 27.81% in average classification accuracy.
- An increase of 38.47% in SN.
- An increase of 25.99% in SP.
- An increase of 33.33% in AUC.
- A decrease of 95.32% in epochs.
- A decrease of 99.6% in computational time.

A qualitative comparison of the algorithms tested (MLP, LSVM and the MLP-SVM) and the MLP-SVM in Aplyosa Neurointelligence was also performed via the following:

- A 2D plot comparing the testing classification accuracy (ACC) and the number of training iterations prior to convergence (epochs) (Fig. 2), with the best performing classifier being in the top-left corner (MLP-SVM) and the least performing algorithm at the bottom-right corner (MLP-SVM Aplyosa Neurointelligence).
- A radar plot comparing the testing classification accuracy (ACC), reliability (SN, SP, SN, AUC) and the computational efficiency (the number of epochs) (Fig. 3), with the perfect classifier resulting in the maximum area for the triangles no. 1–3 (in red).
and the best performing classifier tested (MLP-LSVM) describing the area delimited by the green lines.

### 3.3. A comparison of classification performance between the hybrid algorithm and classifiers from published studies

In Table 2, ACC, SN, SP and AUC of all algorithms tested (the initial MLP and SVM, the proposed MLP-LSVM, the MLP-SVM in Ayala NeuroIntelligence) were compared quantitatively against published studies using the same PD dataset. Table 2 also reports the number of epochs and the computational time for each algorithm.

### Table 1

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>Hidden</th>
<th>Training</th>
<th>CV</th>
<th>Accuracy</th>
<th>Average</th>
<th>SN</th>
<th>SP</th>
<th>AUC (0-1)</th>
<th>Epoths</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>neurons</td>
<td>accuracy</td>
<td></td>
<td></td>
<td>accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>20</td>
<td>10</td>
<td>100</td>
<td>100%</td>
<td>100%</td>
<td>100</td>
<td>100</td>
<td>(99-100)</td>
<td>1 (1-1)</td>
<td>5</td>
</tr>
<tr>
<td>LSVM</td>
<td>20</td>
<td>–</td>
<td>100</td>
<td>98.28%</td>
<td>98.28%</td>
<td>98.28</td>
<td>98.28</td>
<td>(98-99)</td>
<td>8 (8-8)</td>
<td>0.01</td>
</tr>
<tr>
<td>MLP-LSVM</td>
<td>20</td>
<td>10</td>
<td>100</td>
<td>100%</td>
<td>100%</td>
<td>100</td>
<td>100</td>
<td>(100-100)</td>
<td>1 (1-1)</td>
<td>5</td>
</tr>
<tr>
<td>MLP-SVM</td>
<td>20</td>
<td>3</td>
<td>75.23</td>
<td>78.68%</td>
<td>78.68%</td>
<td>78.68</td>
<td>78.68</td>
<td>(78-80)</td>
<td>57 (57-57)</td>
<td>2.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Feature selection</th>
<th>N</th>
<th>ACC (95% CI)</th>
<th>SN (95% CI)</th>
<th>SP (95% CI)</th>
<th>AUC (0-1)</th>
<th>Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balzarini (2017)</td>
<td>MLP</td>
<td>50</td>
<td>98</td>
<td>96</td>
<td>0.74</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Berdia et al. (2017)</td>
<td>LSVM</td>
<td>15</td>
<td>89</td>
<td>100</td>
<td>100</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Benzaidi and Stassen (2017)</td>
<td>SVM (RBF)</td>
<td>35</td>
<td>84</td>
<td>90</td>
<td>0.81</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Cantrak and Cândolle (2015)</td>
<td>MLP</td>
<td>12</td>
<td>65.94</td>
<td>74.15</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Saleh et al. (2017)</td>
<td>MLP</td>
<td>50</td>
<td>96</td>
<td>90</td>
<td>0.74</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>This study</td>
<td>MLP</td>
<td>20</td>
<td>98.94</td>
<td>90.47</td>
<td>0.73</td>
<td>57</td>
<td>2.33</td>
</tr>
</tbody>
</table>

Fig. 2. A comparison of testing classification accuracy (ACC) and the number of training iterations prior to convergence (epochs) between the algorithms tested (MLP-LSVM and the MLP-SVM) and the MLP-SVM implemented in Ayala NeuroIntelligence. The ACC-related confidence intervals are represented by error bars.

Fig. 3. Scatter plot comparing the classification performance between the algorithms tested (MLP-LSVM and the MLP-SVM) and the MLP-SVM implemented in Ayala NeuroIntelligence. The scatter plot compares the ACC, SN, and SP for each algorithm.

The results showed that the MLP-LSVM algorithm achieved the highest ACC (98.94% with a confidence interval of 96.94-99.94%), SP (90.47% with a confidence interval of 88.47-92.47%), and AUC (0.73 with a confidence interval of 0.70-0.76) compared to the initial MLP and SVM algorithms. The sensitivity analysis performed in 2.3 also revealed that the MLP-LSVM algorithm was consistently better than the other algorithms tested.
of the best performing algorithm from the literature (LOSO-SVM; Benha et al., 2017). Significant differences in performance were noted when using the 20 most important features as inputs, in which instance the LOSO-SVM (Benha et al., 2017) led to an AUC of 0.6 in 100 epochs, whilst the proposed algorithm (MLP-LSVM) resulted in an AUC of 1 and in 3 epochs only.

4. Discussion

4.1. The advantage of adequate feature selection

Feature selection (FS) enables to quantify the relative importance of each input in the data towards characterising their underlying patterns. Particularly, pitch-related parameters can aid discrimination between PD patients and healthy controls, e.g., the pitch period entropy was used for automatic detection of PD via the ML-based support vector machine (SVM) with 95.45% accuracy (Little, McGhany, Hunter, Spornosen, & Rurang, 2006). Furthermore, frequency-, pulse-, amplitude- and harmonicity-related parameters have been found to be valuable biomarkers for detecting voice disorders that may indicate early PD (Sakar et al., 2013). Considering the relative importance of individual vocal parameters to detect PD, and the difficulty in distilling the essential features from all these collected for diagnostic purposes, PS prior to performing classification via ML is deemed vital.

Most of the previous studies have used FS, highlighting its significance towards improving the overall classification performance. Since hybrid algorithms must be highly problem- and data-specific, adequate PS is required to improve the accuracy of classifiers (Canterik & Karabiber, 2016; Sakar et al., 2017).
A sensitivity analysis was performed to select the optimal set of reduced input features via the MLP, which yielded the highest classification performance in discriminating between physiological speech-related features and indicators of PD. As a result, twenty (N = 20) features were selected for classification and seven features (N = 7), i.e., "number of pauses", "pitch (local)", "voicing of locally unvoiced frames", "shutter (bp/s)", "jitter (percent)", "jitter (ppq5)", "degree of voice breaks" and "standard deviation of pitch", were discarded from the input features (Fig. 4). The pie chart in Fig. 4 shows an iconic representation of the input patterns in the initial PD dataset, thus showing that, according to the MLP-based weights, the discarded features only accounted for 8.6% of the importance of the input features towards characterising the underlying patterns of the PD dataset of interest (Sarker et al., 2013). This finding suggests that such variables may not be extremely important for diagnosing PD based on speech data, confirming results from recent published studies (Albaret et al., 2017). 20-fold CV was selected as 20 were the reduced features further to performing feature selection via the MLP-based weights. To enable a fair comparison on the classification performance, such a CV was used in all classifiers tested.

Since the average classification accuracy of the LSVm was close to 100% on the initial data set (99.28%) and considering that the LSVm had a reduced number of epochs (N = 7) prior to convergence with respect to the MLP (N = 13) (Table 1), it was hypothesised that, by performing FS via the MLP, which had achieved the highest performance in discriminating between PD-related vocal features and physiological vocal, such a reduced feature set had increased the linear separability of the input classes into the two target classes. Therefore, to test this hypothesis, the reduced features were input to the LSVm, resulting in a novel hybrid algorithm (MLP-LSVm), which not only achieved 100% of average classification accuracy but also with the lowest number of epochs (N = 3) and computational time (0.005 s) required prior to convergence (Table 1). The increase in classification performance and the decreased number of training iterations (epochs) required to classify the input data (Table 1) verified the hypothesis whereby the MLP weights-based FS based on gradient descent increased their linear separability, thus assisting the LSVm indirectly in finding a more optimal separating hyperplane, improving margin maximisation and classification of the input data between healthy controls and PD patients. The decrease in computational time (Table 1) can be explained by the reduced feature set used to achieve the same maximum classification performance.

Having also validated the proposed model (MLP-LSVm) not only on the sustained vowels, but also on the numbers, words and short sentences, which are of interest, enhanced its generalisation over the model of Benga et al. (2017), which, although achieved almost perfect accuracy when 15 or 16 features were considered (Table 1), did so at the cost of reducing the generalisability of the classification outcomes obtained. The lack of generalisation in the LSOV-SVM algorithm developed by Benga et al. (2017) is evident when using the 20 most important features, which led to a low AUC (0.6) and a high number of epochs (N = 130), whilst the proposed algorithm (MLP-LSVm) resulted in a perfect classification performance (AUC = 1) in this instance too, converging in 3 epochs only. Moreover, the LSOV cross-validation (CV) used in Benga et al. (2017) is not as exhaustive as the k-fold cross-validation algorithm used in this study, wherein all data points were considered when creating the number (k) of folds, rather than discarding all observations for a single subject at a time, such as in Benga et al. (2017). The lack of exhaustiveness of the LSOV CV algorithm is clear given the higher variance in the solutions obtained when varying the number of selected features in the LSOV-SVM classifier (Benga et al., 2017), as compared to the proposed algorithm (MLP-LSVm), whose performance was consistently high (Table 3).

Qualitatively, the improved classification performance of the hybrid model (MLP-LSVm) against the MLP and the LSVm, when tested individually, can be assessed by Figs. 2-5. This finding demonstrates the usefulness of applying MLP for FS, thus increasing the linear separability of the input data to be fed into the LSVm for classification. This increase in linear separability of the input features via the MLP-based FS indirectly optimises the augmented Lagrangian algorithm of SVM to converge to a more optimal solution and faster; thus, using the MLP to assist the LSVm in avoiding tripping at local minima and improving the overall classification performance.

4.2. The need for customised, test-specific algorithmic implementation

To further test whether the use of the proposed hybrid algorithms (MLP-LSVm) could be recommended to detect PD-related speech features and to assess the need for customised algorithmic implementation, the classification performance of the MLP-LSVm was compared against a similar algorithm (MLP-SVM) that was implemented via a commercially available software, Alyuda NeuroIntelligence (2.2.577), Cupertino, California, USA. As mentioned in 2.2, the limitation in the software implementation of the findings derived from the comparison reported in this study. Nevertheless, the proposed customised MLP-LSVm algorithm resulted in a significantly higher average classification accuracy and AUC with respect to those in the commercial software, respectively by 27.88% and 33.33% (Table 1), thus demonstrating the need for customised algorithmic implementation. In further support of this finding, the number of epochs and the computational time were reduced by 95.52% and 99.6%.

Qualitatively, by achieving a higher accuracy and a faster convergence, the classification performance of the hybrid model (MLP-LSVm) was far higher than the MLP-SVM in Alyuda (Fig. 2). Furthermore, the decreased confidence interval shows a considerably increased confidence in the solution when using the proposed model (Fig. 2). Such improvements in classification performance (Table 1) are due to the optimised margin maximisation and subsequent increase in linear separability of the input data into the two target classes of interest (PD patients and healthy controls) as well as the parameterisation of the number of hidden neurons used in the customised MLP (N = 10) (Table 1) when compared to the best performing MLP architecture (N = 3) (Table 1) as found automatically via Alyuda. Therefore, for healthcare-related applications targeted to improve early diagnosis of PD, findings suggest the need for customised algorithmic implementations tailored for such a specific application, e.g., the MLP-LSVm, and thorough validation via all performance measures adopted in this study (ACC, SN, SP, AUC, number of epochs and computational time).

4.3. An effective algorithm for clinical translation

To further validate the use of the proposed hybrid algorithm for aiding early diagnosis of patients with PD, the classification performance of the MLP-LSVm was quantitatively compared against ML-based models tested on the same PD dataset (Sarker et al., 2013), from published studies (Table 2). As quantitatively and qualitatively assessed in Table 2 respectively, findings strongly support the translational application of the proposed hybrid classifier (MLP-LSVm) in a clinical setting, considering its highest classification performance and extremely low computational cost (epochs and time) in aiding early diagnosis of patients with PD. Furthermore, the MLP-LSVm lends itself to further development into a clinically valuable software due to its simple architecture and higher generalisability (consistently higher AUC and lower number of epochs). Table 3) as compared to the best performing classifier from the
literature (Brook et al., 2017) of comparable performance, where further reduced feature set on only three sustained vowels (N = 15, N = 16) rather than the whole dataset might have led to 100% classification accuracy at the cost of discarding further data (numbers, words and short sentences) to perform a reliable diagnosis, as mentioned in 4.1. To enable a more thorough comparison of findings for validating the use of novel algorithms for medical applications, and ultimately to enable clinical translation, it is recommended to report findings more comprehensively, i.e., reporting both the testing classification accuracy (ACC) and reliability (SN, SP and AUC) of the algorithms used in all Tables 1-2, as well as reporting results in confidence intervals rather than only as point estimates (Table 2), such as in all previous works using the same data set compared in this study (Albouhaini, 2017; Borkhate & Shahrkani, 2022; Borkhane et al., 2017; Cantekin & Karabiber 2016; Sakar et al., 2017).

Such ML-based algorithms could be applied to devise accurate and reliable automated tele-diagnostic and tele-monitoring systems to improve outcome in PD patients (Little et al., 2009), as well as patient-safety and effectiveness of the treatments delivered by physicians.

5. Conclusion

Following a comprehensive analysis of the classification performance of the proposed hybrid model (MLP-LSVM) against the MLP, the LSVM, the SVM-MLP implemented in commercially available software and the ML-based classifiers from selected published studies, the MLP-LSVM proved to be the most accurate, reliable and computationally the fastest ML-based classifier for detecting PD from speech-related data. Therefore, the MLP-LSVM can aid early diagnosis of patients with PD in a clinical setting, thus improving the accuracy and reliability of current automated speech processing systems.

Noteworthy, the novel hybrid algorithm (MLP-LSVM) not only achieved 100% of ACC (and 1 of AUC) in discriminating between healthy controls and PD patients using speech-related parameters, but it did so with the lowest computational cost, as quantified by the number of training iterations (epochs) prior to reaching convergence and the computational time. These findings further confirm the usefulness and viability of the MLP-LSVM classifier for aiding early diagnosis of PD based on readily available data.

The MLP-LSVM’s intuitiveness of the classification procedure and outcomes lends itself to further development into a valuable clinical tool and/or a-making support tool. There is also future potential to assist with assessment of prognosis in patients with PD, thus aiming at optimizing patient outcome more broadly.

Declaration of Interest

The authors declare no conflicts of interest.

Contributors

All authors directly participated in the planning, execution and analysis in the study. All authors also approved the final version of the manuscript, and this submission for possible publication in Expert Systems with Applications.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors LP, and N.R. would like to thank the University of Auckland for giving them the opportunity to carry out their Ph.D. research projects. The authors LP, and N.R. would like to thank the University of Auckland Rehabilitation Technologies Association (UATRA), and MediTechLab® for giving them the chance of developing this collaborative research work.

References


A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis

Luca Parisi

Received: 19 February 2018 / Accepted: 23 January 2019
© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract

This study investigated the application of a novel hybrid artificial intelligence (AI)-based classifier for aiding prediction of the prognosis in patients with chronic hepatitis. Nineteen biomarkers on 155 patients with hepatitis from the University California Irvine Machine Learning repository were used as input data. Weights derived by applying the geometric margin maximisation criterion of a Lagrangian support vector machine (LSVM) were used for selecting the features associated with the highest relative importance towards the required classification, i.e., to predict whether a patient with hepatitis would have survived or died. Thus, the 19 initial features were reduced to the 16 most important prognostic factors and were fed into various AI-based classifiers. Results indicated an overall classification accuracy and area under the receiver operating characteristic curve of 100% for the proposed hybrid algorithm, the LSVM multilayer perceptron (MLP), thus demonstrating its potential for aiding prediction of prognosis in patients with hepatitis in a clinical setting.

Keywords Multilayer perceptron - Decision support - Prognosis - Hepatitis

1 Introduction

1.1 Hepatitis and current clinical challenges

Based on the chronicity and the type of virus leading to inflammation of the liver, hepatitis is categorised into 8

 Luca Parisi
 luca.parisi@here.org

 Narendar Ravi Chandram
 narendar@see.org

 Marianne Lyne Manaog
 marianne@medicintellego.com

 1 Auckland Bioengineering Institute (ABEI), University of Auckland, 70 Symonds Street, Grafton, Auckland 1010, New Zealand

 2 Department of Mechanical Engineering, University of Auckland, 20 Symonds Street, Grafton, Auckland 1010, New Zealand

 3 Division of Applied Artificial Intelligence in Healthcare, Medicintellego, 4 Colonel Burton Glade, St Johns, Auckland 1072, New Zealand

 4 University of Auckland Rehbilitative Technologies Association (UARTA), University of Auckland, 11 Symonds Street, Auckland 1010, New Zealand

 HBV, C (HCV) and D (HDV). Viral hepatitis affects 400 million people worldwide, with 240 and 130–150 million people being infected with HBV and HCV, respectively [1]. In United States (USA) alone, 1.2 million people are chronically infected with HBV [2] and over 3.2 million people with HCV; however, most of the patients affected by viral hepatitis are currently unaware of having the disease [1]. In patients developing HBV or HCV, the infection is often undiagnosed. This lack of diagnosis occurs because the infection remains asymptomatic upon its onset and it takes decades for the symptoms to manifest. When the infection becomes chronic, it increases the severity of the inflammation [1]. Typically, hepatitis is diagnosed by patients’ symptoms (e.g., physical fatigue, malaise, anorexia, palpable spleen, a firm and hypertrophic liver) and blood tests [3]. However, it is still clinically challenging to differentiate early between the main symptoms of chronic hepatitis and those of other viral forms of hepatitis. Moreover, the prognosis of patients with chronic HBV and HCV still relies on the assessment of patient morbidity, such as the extent of liver fibrosis, oesophageal varices, hepatocellular carcinoma and the risk for cirrhosis [4], as assessed via biopsy. In the last decades, non-invasive techniques [5] have replaced the need for liver biopsy
to perform a prognostic assessment of the actual outcome of such patients with hepatitis, i.e. mortality and duration of follow-up [4]. These relatively new methods quantify serum levels and liver stiffness as potential biomarkers. The degree of fibrosis or liver stiffness is conventionally assessed via computerised tomography or non-contrast medical imaging methods, such as ultrasound or magnetic resonance elastography [4]. Although physicians can perform a diagnosis of hepatitis based on such indirect biomarkers [4, 5], their capability of objectively predicting the prognosis in such patients is still limited. As these methods can provide only the biochemical or tissue-related information separately [4, 5], they are unable to capture the underlying patterns that may be associated with other aetologies. Recognising such pathological patterns is essential for performing an accurate and reliable prediction of prognosis.

1.2 Artificial Intelligence for aiding prognostic assessment

Clinical decision making heavily relies on several criteria that make the estimation of the patients’ prognosis challenging. Differently from conventional prognostic systems, artificial intelligence (AI)-based decision-making support tools, which include machine learning (ML)- and artificial neural networks (ANN)-based algorithms, can perform pattern recognition on several input features that may influence the course of the disease [6]. Even the human reasoning of the most expert clinician could not handle such highly dimensional criteria and their underlying associations. Instead, ML can capture such subtle pathological patterns and cope with highly nonlinear, highly dimensional and ill-behaved solution spaces, thus showing potential in aiding such a prognostic assessment in a clinical setting. Whilst unsupervised ANNs-based learning classifiers, such as self-organising maps (SOMs) [7] and k-nearest neighbours (kNN), can classify input patient data without preliminarily knowing their classes, supervised AI-based learning classifiers, such as the ANNs-based multilayer perceptron (MLP) [8–11] and the ML-based Lagrangian support vector machine (LSVM) [6] assume that the patient input data’s true classes are preliminarily known. Whilst learning, an ANN is being trained iteratively to recognise the underlying patterns of the input data to improve its classification performance. Transfer functions facilitate learning in ANN, where, in the training algorithm, weights are iteratively adjusted based on the target outputs [11]. This iterative learning process leads the ANN to minimise its training error and maximise its classification performance. Ultimately, the moduli of the weights determine the degree of relative importance towards the required classification. Along with the second

stage of cross-validation (CV), the ANN is trained to apply the learnt relationship in the third and last step, i.e. testing, to classify unknown data. In the field of hepatology, the potential of applying ANN to learn to predict the prognosis of patients with hepatitis has been shown [12, 13].

1.3 Related studies and study rationale

We hypothesised that: AI-based learning classifiers could assist in performing pattern recognition-related tasks for predicting the prognosis of patients with hepatitis, as they have high predictive capability when dealing with several types of healthcare data [8–10].

Nevertheless, such studies lack adequate data pre-processing, which impairs the reliability of their algorithms proposed to predict the prognosis of patients with hepatitis. Sophisticated algorithms have been developed in the literature to obtain 100% testing classification accuracy [14], but they are, therefore, not applicable in a clinical setting.

Moreover, there was no mention of the computational time involved in running such classifiers.

By presenting clear and standardisable data pre-processing, feature selection and classification approaches, this study aims to overcome such shortcomings from published studies.

Lack of awareness on the prevalence of chronic viral hepatitis and the limitations with the adoption of preventive and diagnostic methods contributes to perpetuate the disease. Lack of early screening also complicates patient management and impacts overall healthcare outcomes worldwide [15]. As compared to invasive and intensive testing protocols, non-invasive and recent techniques have shown to provide viable biomarkers for aiding early prognosis of hepatitis [5]. Considering the prerequisite of a clinician’s expert opinion to infer the underlying pathophysiology of hepatitis by using these methods, the adoption of AI-based algorithms can help physicians predict the course of the disease more accurately. This is new information that is not currently at their disposal. Feature reduction via AI can also improve the overall classification accuracy and reliability of such algorithms, thus making them more suitable for applications in a clinical setting. To predict the survival of patients with chronic hepatitis based on haematological parameters and demographics [16], this study aims to develop and validate a hybrid model, whose classification performance is higher than other ML-based
classifiers from previous works [6], by optimising the
selection of relevant input features for classification. This
study seeks to provide clinicians with an accurate and
reliable decision-making support tool for predicting the
prognosis of patients with chronic hepatitis (survival rate),
thus improving the objectivity of current clinical practice.

The target of this study is to demonstrate how to opti-
mise the use of publicly available data sets to build deci-
dion support systems to help predict the mortality in
patients with hepatitis.

2 Methods

2.1 Data collection and pre-processing

Haematological parameters, patient demographics (age and
gender) and outcomes on physical status (e.g. physical
fatigue, malaise, anorexia) on one hundred fifty-five
(N = 155) patients with chronic hepatitis (140 men, 15
women; age 41.2 ± 12.5 years) from the University of
California Irvine (UCI) ML repository were used [16]. 132
of which were survivors and 23 were deceased. This data
set was used as it is one of the most comprehensive data
sets with prognostic data on patients with hepatitis publicly
available and it is derived from a trusted source for
designing machine models on, i.e. the UCI repository. Such
features were collected as known prognostic biomarkers of
chronic hepatitis by Carnegie-Mellon University [16].
Several AI-based classifiers were deployed along with
feature reduction techniques for predicting survival of the
above-mentioned patients. The development and testing of
these algorithms were performed via MATLAB (R2017b).
The MathWorks, Inc.) on an Intel® Core™ i5-4460T
2.70GHz Processor, Mean imputation (“Appendix”) was
applied to replace all missing values from the original
hepatitis dataset [10]. Data in the feature vector named
“SGOT” (serum glutamic-oxaloacetic transaminase) were
set to remain as real numbers. Accordingly, the corre-
sponding means were rounded towards the nearest integer.
Subsequently, all input features and targets were nor-
malised by mapping their minimum and maximum values
in the range between −1 and +1 for all classifiers (as per
the range of their transfer functions). For the SVM, values
were normalised between 0 and +1. Input values were then
standardised as z-scores, i.e. their values were transformed
to have a mean of zero and a standard deviation of 1.

2.2 Feature selection

Typically, statistical-based methods, such as principal
component analysis (PCA), are used for feature selection
[15], thus retaining the most important features towards the
prediction in the input data prior to performing classifica-
tion. However, a recent study [18] demonstrated that using
such statistical-based approaches for feature selection may
impair classification performance. This study was pivotal in
demonstrating the need for an intrinsic, tightly coupled
ML-based method to the ML classifier, as opposed to an
extrinsic, loosely coupled statistical method, such as PCA.
This impairment in classification performance is due to the
lack of control over the feature reduction using PCA.
Therefore, to optimise feature reduction, this study pro-
poses the development and validation of a novel hybrid
algorithm that uses ML-based methods for feature selection
prior to performing classification via ANN, as further
explained in 2.3 (Fig. 1). The MLP in Fig. 1 was trained by
back-propagation with k-fold cross-validation algorithm,
where k was made vary between 5 and 16, i.e. the number
of selected inputs, as further explained in 2.3.

2.3 A hybrid AI-based algorithm for feature
selection and classification

To overcome the limitations of statistical-based methods
for feature selection, as described in 2.2, we propose to use
the LSVM [17] to reduce the initial number of input feat-
ures (N = 19) to the most relevant ones for the classifi-
cation of interest. Such a reduced feature set is obtained by
applying the geometric margin maximisation criterion,
which is optimised further by the LSVM with respect to
conventional SVM [17]. The LSVM iterative algorithm
applies a SVM with an augmented Lagrangian formulation
[17]. Eventually, the transpose of the input data (dataT),
the target classes (classes) and the optimal solution resulting
from the iterative LSVM algorithm to optimise margin
maximisation (x*opt) are used to compute the vector of
coefficients (wLSVM) for the optimal separating hyperplane
(OSH):

\[ w_{LSVM} = data^T \cdot classes - x_{opt} \]  

(1)

The moduli of the weights wLSVM in (1) can be inter-
preted as the relative importance of the input features
towards the linear separability of the input data into the
target output classes of interest, i.e. towards the
classification.

A sensitivity analysis was performed considering the
optimal reduced feature set by the LSVM yielding the
highest overall classification performance in the MLP,
as assessed via four performance measures, i.e. testing clas-
sification accuracy (ACC), sensitivity (SN), specificity (SP)
and area under the receiver operating characteristic (ROC)
curve (AUC). Thus, an iterative algorithm was designed to
minimise the cost function in (2), maximising the overall
classification performance in the hybrid LSVM-MLP
algorithm.
To obtain an optimal reduced feature set, the objective function (2) was devised to rank the relative importance of individual inputs in the initial feature set \( F \) towards the classification performance:

\[
F = \min \sum_{i=1}^{N} \left( \frac{TP}{TP + FP} \ln \left( \frac{TP}{TP + FP + TN} \right) \right) + \lambda \sum_{i=1}^{N} \ln \left( \frac{TP}{TP + FP + TN} \right)
\]

\( \lambda \) is the threshold whereby the ROC is built.

The category of ML-based learning classifiers named SVM aims to generate an OSH as the decision-making surface boundary for a classification by maximising the margin between the classes of the input dataset and minimising their training error \( \epsilon \). The reduced input features \( N = 16 \) were then used as inputs to an ANN-based classifier for discerning which patients with hepatitis would have survive or died. The proposed hybrid algorithm involves the LSVM in its feature reduction stage and an ANN for classification.

This model has been designed, in its simplicity but uniqueness, to assist physicians in clinical decision-making processes for aiding early prediction of prognosis and, therefore, helping them tailor treatments to individual patients with hepatitis to improve patient outcome. Considering their capability to cope with large, high-dimensional and highly nonlinear data, the following AL-based learning classifiers were used to predict the prognosis of patients with hepatitis, and their classification performance was compared with that of the proposed hybrid algorithm via four gold standard performance measures [18]:

- The supervised ANN MLP [6, 8–10, 19, 20];
- The proposed hybrid algorithm (LSVM-MLP), applying the supervised ML-based algorithm LSVM [6, 17] for feature selection and the MLP for classification.

The MLP neural network was trained via the backpropagation algorithm, according to which the initial randomly weighted inputs are propagated forward, whilst the errors are iteratively propagated backwards until the weights’ adjustment generated the least mean squared error (MSE) [11]. As any Boolean bounded function can be mathematically described by a MLP with one hidden layer [21], only one hidden layer was implemented in the proposed hybrid model (LSVM-MLP). To account for a larger range of inputs, the hyperbolic tangent function was used as a transfer function in the hidden layers:

\[
eti = \tanh(\netai + bi), \quad \{ ai \in \mathbb{R} \} - 1 \leq \etai \leq +1 \}
\]

\[
\etai = \sum_{j=1}^{N} \omega_{ij} \cdot x_{i}
\]
where the subscript \( h \) indicates the number of hidden layers in the network, the subscript \( i \) represents the number of inputs, \( k \) is the transfer function used in linear neurons, \( N \) is the number of training iterations or epochs, \( y_h(x_i, z_j) \) is the output from the hidden layer for weights \( w_h \) and inputs \( z_j \). Training involves the computation of weights \( w_{h,j} \) for which the error \( e \) is the difference between the actual and the target outputs and the iterative training algorithm is stopped when the target MSE (0.05) is reached. By applying the LSVM for feature reduction, the linear separability of the input data was enhanced. By maximising the margin, the LSVM-based reduced feature set guided the MLP-related gradient descent algorithm to converge faster. Since the gradient descent’s cost function was implicitly minimised via the reduced input feature set, the error term \( e \) was also minimised. Furthermore, whilst the adjustment of the weights and biases was highly affected by the resampling of the input data in the MLP, particularly when \( k \)-fold CV is applied, when the LSVM was applied in the feature selection stage of the MLP, the hybrid algorithm (LSVM-MLP) was able to obviate this shortcoming and improve the classification performance of the initial MLP. To the best of the authors’ knowledge, the LSVM was deployed for the first time in a haematological study in the feature selection stage of the MLP to aid prediction of prognosis of patients with hepatitis on the sole basis of haematological parameters. The learning parameter and the momentum term of the MLP were set to 0.6 and 0.8, respectively [8]. Considering the number of selected input features [20], 5-fold cross-validation algorithm was used at each iteration to train the AI-based classifiers and the classification performance of the proposed hybrid algorithm (LSVM-MLP) was also compared when \( k \)-fold CV was applied, to verify the hypothesis whereby, retaining the same or a similar accuracy and reliability, the computational cost could be reduced to make the hybrid algorithm (LSVM-MLP) more suitable for clinical applications.

The performance of the selected algorithm was assessed when using the same number of features (\( N = 16 \)) but retained from an expert-based perspective, i.e. as extracted from the meta-analysis of Friedrich-Rust et al. [34] on chronic hepatitis, whereby age, fatigue and histology were not statistically significant prognostic biomarkers.

The selected algorithm was also tested on an additional dataset (dataset 2) [36], wherein the follow-up time was the same, i.e. 1 year, to verify whether the performance of the proposed algorithm was consistently higher than other algorithms used for comparison.

To further assess the need for customised algorithmic implementation, the proposed model was compared against existing models (SVM-MLP) in the black-box, commercially available software in Microsoft Azure. The same data pre-processing methodology in 2.1 was applied when assessing the performance of readily available binary learning-based classifiers in Microsoft Azure ML Studio.

2.4 MedIntelligence™ Quality Assessment Scale for Studies on Applied Artificial Intelligence in Healthcare

The “MedIntelligence™ Quality Assessment Scale for Studies on Applied Artificial Intelligence in Healthcare” (MQAS) in “Appendix” is applicable to any research articles, regardless of their type (e.g. full-length scientific articles or conference papers), dealing with AI applied to clinical and/or healthcare-related data. The authors L.P. and M.L.M. developed the MQAS collaboratively, based on the recent pioneering study by Luo et al. [22], which attempted to describe how ML algorithms should be devised and the findings should be reported when such ML algorithms are applied in biomedical research. The MQAS not only provides such qualitative guidelines but also a quantitative approach, via clearly defined star-rating criteria, to assess the quality of published studies (retrospectively) that used AI, not only ML, with healthcare data. The MQAS has been also designed to guide AI engineers in undertaking any research (prospectively) that involves the development and validation of AI-based algorithms for applications in healthcare. The authors L.P. and M.L.M. also validated the use of the MQAS by applying it to select relevant high-quality scientific articles dealing with the same hepatitis dataset [16] used in this study for comparing the classification performance of the proposed hybrid classifier (LSVM-MLP) against the best performing AI-based algorithms from selected studies from the literature. This comparison has been carried out in previous studies in the absence of a similar quality assessment scale, thus impairing the reliability of the conclusions drawn from previous studies using the same hepatitis dataset [16]. The MQAS was designed to be the equivalent in AI research in healthcare of the “Newcastle-Ottawa Quality Assessment Scale for Cohort Studies” [23], which is used to assess the quality of articles reporting findings from randomised clinical trials (RCTs).

3 Results

Initially, all input features (\( N = 19 \)) were used for predicting the survival of patients with hepatitis considered in this study [16] via a MLP, where the \( k \)-fold CV and the number of hidden neurons were made to vary (Table 1). Therefore, further to objectively discard three features (antivirals, irrelevant as patients were likely to be under...
Table 1 Classification performance of the MLP when using the initial 19 input features

<table>
<thead>
<tr>
<th>k-folds for CV</th>
<th>Hidden neurons</th>
<th>Training accuracy %</th>
<th>CV accuracy %</th>
<th>AUC</th>
<th>AUC (95% CI)</th>
<th>Average accuracy %</th>
<th>Average accuracy (95% CI)</th>
<th>SN</th>
<th>SP</th>
<th>AUC</th>
<th>Epochs</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>89.21</td>
<td>100</td>
<td>100 (100–100)</td>
<td>96.40</td>
<td>100 (100–100)</td>
<td>1.1 (1–1)</td>
<td>14</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>87.77</td>
<td>87.50</td>
<td>87.59</td>
<td>87.62</td>
<td>87.56</td>
<td>0.87</td>
<td>19</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>85.68</td>
<td>90.23</td>
<td>87.62</td>
<td>39</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>96.49</td>
<td>88.43</td>
<td>95.63</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>32</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>83.74</td>
<td>93.75</td>
<td>92.50</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>11</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>88.62</td>
<td>94.12</td>
<td>87.62</td>
<td>27</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>100</td>
<td>100</td>
<td>95.39</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>18</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

interferon therapy at that time, histology and prostate) from the initial dataset using the LSVM-based weights, the reduced feature set (N = 16) was input to a MLP by varying the same above-mentioned parameters (Table 2).

Tables 1 and 2 show a quantitative comparison of classification performance between the MLP and the proposed hybrid classifier (LSVM-MLP), with the number of training iterations or epochs and the computational time.

Whilst the highest average classification accuracy for the MLP was 96.40% when fivefold CV and 32 hidden neurons in the MLP were used (Table 1), the novel hybrid algorithm (LSVM-MLP) reached 99.80% of average classification accuracy when fivefold CV and 10 hidden neurons in the MLP were used (Table 2). By using only selected features via the LSVM, the MLP reduced both the number of epochs required prior to convergence (from 14 to 9, Table 1 as compared to Table 2), along with its computational time (from 0.19 to 0.09, Table 1 as compared to Table 2).

Specifically, when using the novel hybrid algorithm (LSVM-MLP) as compared to the MLP (Table 1), the following percentage changes in average classification accuracy were noted:

- An increase of 2.49% when fivefold CV and 32 hidden neurons in the MLP were used;
- An increase of 13.89% when fivefold CV and 10 hidden neurons in the MLP were used;
- An increase of 10.33% when fivefold CV and 10 hidden neurons in the MLP were used;
- An increase of 5.42% when fivefold CV and 5 hidden neurons in the MLP were used;
- An increase of 4.3% when 19- and 16-fold CV were used, respectively (due to the different number of input

Table 2 Classification performance of the proposed hybrid algorithm, the LSVM-MLP, when using the 16 reduced features via the LSVM as inputs to the MLP

<table>
<thead>
<tr>
<th>k-folds for CV</th>
<th>Hidden neurons</th>
<th>Training accuracy %</th>
<th>CV accuracy %</th>
<th>AUC</th>
<th>AUC (95% CI)</th>
<th>Average accuracy %</th>
<th>Average accuracy (95% CI)</th>
<th>SN</th>
<th>SP</th>
<th>AUC</th>
<th>Epochs</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>96.40</td>
<td>100</td>
<td>100 (100–100)</td>
<td>98.80</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>15</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>96.40</td>
<td>100</td>
<td>100 (100–100)</td>
<td>98.80</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>19</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>90.28</td>
<td>100</td>
<td>100 (100–100)</td>
<td>99.76</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>9</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>99.26</td>
<td>99.76</td>
<td>100 (100–100)</td>
<td>99.76</td>
<td>100 (100–100)</td>
<td>1 (1–1)</td>
<td>15</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>89.43</td>
<td>93.75</td>
<td>97.02</td>
<td>11</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>19</td>
<td>91.06</td>
<td>94.93</td>
<td>96.02</td>
<td>12</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
features, 19 in the MLP, 16 in the LSVM-MLP), but always with 32 hidden neurons in the MLP:
• An increase of 4.04% when 19- and 16-fold CV were used, respectively (due to the different number of input features, 19 in the MLP, 16 in the LSVM-MLP), but always with 19 hidden neurons in the MLP:
• A slight decrease of 0.94% when 19- and 16-fold CV were used, respectively (due to the different number of input features, 19 in the MLP, 16 in the LSVM-MLP), but always with 10 hidden neurons in the MLP:
• A slight increase of 0.66% when 19- and 16-fold CV were used, respectively (due to the different number of input features, 19 in the MLP, 16 in the LSVM-MLP), but always with 5 hidden neurons in the MLP.

Furthermore, the classification performance of the hybrid algorithm (LSVM-MLP) was compared against a similar classifier but implemented via a commercially available software, Microsoft Azure ML Studio, i.e. the hybrid SVM-MLP algorithm (Table 3). In Microsoft Azure, the augmented Lagrangian formulation of the SVM is not available, thus a conventional linear SVM was used. Comparatively, the LSVM-MLP had an increased overall classification performance as quantified by the following percentage changes in the performance measures used (Table 3):
• An increase of 29.28% in average classification accuracy;
• An increase of 44.01% in SN;
• An increase of 39.28% in SP;
• An increase of 28.21% in AUC;
• A decrease of 91.09% in epochs;
• A decrease of 99.15% in computational time.

A qualitative comparison of the algorithms tested (MLP and the LSVM-MLP) and the SVM-MLP in Microsoft Azure was also performed via the following:
• a conventional two-dimensional (2D) convergence plot comparing classification error (100-ACC) % and the number of epochs that the classifier underwent prior to reaching convergence (Fig. 2), with the best performing classifier having zero error and the lowest number of epochs (LSVM-MLP, as represented by a blue circle);
• a three-dimensional (3D) convergence plot comparing AUC, epochs and computational time (Fig. 3), with the hybrid classifier (LSVM-MLP, as represented by a blue circle) having the highest classification performance that lies at the top of the 3D plot, with the least number of epochs and computational time, whilst the least performing one lies at the bottom of it.

In Table 4, ACC, SN, SP and AUC of the proposed hybrid algorithm were compared quantitatively against

<table>
<thead>
<tr>
<th>Table 3: Comparison between the proposed hybrid algorithm, the LSVM-MLP and the SVM-MLP models implemented using Microsoft Azure.</th>
<th>SN</th>
<th>SP</th>
<th>AUC</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>LSVM-MLP</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
</tr>
<tr>
<td>SVM-MLP</td>
<td>89.0</td>
<td>89.0</td>
<td>89.0</td>
<td>89.0</td>
</tr>
<tr>
<td>LSVM-MLP</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
</tr>
<tr>
<td>SVM-MLP</td>
<td>89.0</td>
<td>89.0</td>
<td>89.0</td>
<td>89.0</td>
</tr>
</tbody>
</table>

---

128
published studies using the same hepatitis dataset [16], as selected via the "MedIntelligence® Quality Assessment Scale for Studies on Applied Artificial Intelligence in Healthcare" in "Appendix". The feature selection algorithm and the number of selected features are also reported in Table 4. A qualitative comparison of the classification performance of the proposed classifier against models from previous studies was also performed via a four-dimensional (4D) plot comparing the classification performance of the AI-based classifiers tested, showing AUC, ACC, SN, and SP (where available) (Fig. 4). As the hybrid algorithm of Kaya and Uyar [14], the novel hybrid classifier (LSVM-MLP) using selected features (N = 16) (Fig. 5) achieved the maximum classification performance across all four measures used for comparison (100% ACC, SN, and SP, as well as 1 as AUC).

4 Discussion

Initially, the MLP was used to predict survival of patients with chronic hepatitis based on all input features (N = 19) from the initial hepatitis dataset [16]. However, the classification performance of the MLP was highly affected by the choice of the number (k) of folds in the CV algorithm used, as well as the number of hidden neurons (Table 1).

Given that SVMs can find an OSH to linearly separate input data into target classes and considering that the augmented Lagrangian formulation of SVM (LSVM) improves margin maximisation for classification further, the coefficients of the OSH found via LSVM were considered as weights for feature selection, i.e. the higher the modulus of the LSVM-based weight, the higher the relative importance of the feature towards the required classification, i.e. to discriminate between survivors and deceased patients with chronic hepatitis. The novelty of the proposed model lies in the utilisation of the LSVM for feature selection and the supervised MLP for the subsequent classification, which is an advanced tool in decision making.

As a result, sixteen (N = 16) features were selected for classification, and three features, i.e. "antivirals", "histology" and prothrombin time ("protime"), were discarded from the input features (Fig. 5). The pie chart in Fig. 5 shows an isometric representation of a map of input patterns in the initial dataset, thus showing that, according to the LSVM-based weights, the discarded features only accounted for 2% of the importance of the input features towards explaining the underlying patterns of the hepatitis dataset of interest [16]. In agreement with part of the medical literature, whereby there is no significant difference histologically and with regards to the prothrombin time when patients have similar staging of hepatitis, with the latter variable being biased by subjective response [35], this finding suggests that, although such variables may be important for diagnosing hepatitis [3], they may be less significant for predicting its prognosis.

Using only the selected input features (N = 16), as obtained by ranking the initial input variables according to the LSVM-based weights, the overall classification performance of the hybrid algorithm (LSVM-MLP) was not only increased to reach 100% average classification accuracy (Table 2), differently from the MLP in Table 1 (ACC = 93.34 ± 2.86%), but was consistently very high (ACC = 97.96 ± 1.82%) across all cases of k-fold CV and number of hidden neurons considered (Table 2). The performance of the proposed algorithm was consistently higher than other algorithms used for comparison when tested on an additional dataset on prognostic data of patients with hepatic disease (dataset 2) [36], wherein,
### Table 4: Comparison of the feature selection algorithms, number of input features, classifier and performance measures between the proposed hybrid classifier, the LSVMM-MLP, and the AI-based classifiers used in published studies

<table>
<thead>
<tr>
<th>References</th>
<th>Feature selection</th>
<th>Input features N</th>
<th>Classifier</th>
<th>ACC (% 95% CI)</th>
<th>SN (% 95% CI)</th>
<th>SP (% 95% CI)</th>
<th>ADC</th>
</tr>
</thead>
<tbody>
<tr>
<td>[24]</td>
<td></td>
<td>19</td>
<td>Re-RX</td>
<td>83.24</td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>[25]</td>
<td></td>
<td>19</td>
<td>PART</td>
<td>84.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[26]</td>
<td>One-R/Relief</td>
<td>19</td>
<td>NB</td>
<td>85.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[27]</td>
<td>GA</td>
<td>15</td>
<td>SVM</td>
<td>80.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[14]</td>
<td>RS</td>
<td>4</td>
<td>ELM</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>[12]</td>
<td></td>
<td>19</td>
<td>PNN</td>
<td>91.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[28]</td>
<td></td>
<td>19</td>
<td>GRNN</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[29]</td>
<td>LDA</td>
<td>8</td>
<td>ANFIS</td>
<td>94.16</td>
<td>96.66</td>
<td>91.66</td>
<td>N/A</td>
</tr>
<tr>
<td>[30]</td>
<td>CBR</td>
<td>19</td>
<td>PSO</td>
<td>94.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[15]</td>
<td>PCA</td>
<td>10</td>
<td>LSSVM</td>
<td>90.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[13]</td>
<td>RS</td>
<td>8</td>
<td>MLP</td>
<td>97.30</td>
<td>98.32</td>
<td>97.28</td>
<td>0.95</td>
</tr>
<tr>
<td>[31]</td>
<td></td>
<td>19</td>
<td>GA-WK-ELM</td>
<td>95.56</td>
<td>96.89</td>
<td>98.23</td>
<td>0.96</td>
</tr>
<tr>
<td>[32]</td>
<td>k-means</td>
<td>19</td>
<td>MLP</td>
<td>97.93</td>
<td>99.18</td>
<td>90.90</td>
<td></td>
</tr>
<tr>
<td>[33]</td>
<td>Expert-based</td>
<td>19</td>
<td>SS-SVM</td>
<td>98.73</td>
<td>95.57</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>16</td>
<td>MLP</td>
<td>96.27</td>
<td>95.24</td>
<td>88.89</td>
<td>0.90</td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>16</td>
<td>MLP</td>
<td>94.02-97.68</td>
<td>76.18-99.86</td>
<td>70.84-97.68</td>
<td>0.87-0.92</td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>16</td>
<td>MLP</td>
<td>100 (100-100)</td>
<td>100 (100-100)</td>
<td>100 (100-100)</td>
<td>1 (1-1)</td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>16</td>
<td>MLP</td>
<td>97.00 (91-99.72)</td>
<td>97.56</td>
<td>97.83</td>
<td>0.97</td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>16</td>
<td>MLP</td>
<td>91 (97-99.94)</td>
<td>87.14-99.94</td>
<td>88.67-99.94</td>
<td>0.90-0.99</td>
</tr>
</tbody>
</table>


**Fig. 4**  Four-dimensional (4D) plot when comparing the classification performance of the customised algorithms (MLP and LSVMM-MLP), the SVM-MLP implemented in Microsoft Azure, and the AI-based classifiers used in published studies, as assessed via AUC, ACC, SN and SP. GA: genetic algorithm; WK: wavelet kernel; ELM: extreme learning machine; RS: rough set.

differently from the first dataset [16], in which prognostic data might have been collected at different time points, the follow-up time was the same, i.e. 1 year.

Qualitatively, the improved classification performance of the hybrid model (LSVMM-MLP) against the MLP can be assessed by Figs. 2, 3 and 4. This finding demonstrates the usefulness of applying LSVMM for feature selection, thus increasing the linear separability of the input data to be fed into the MLP for classification. This increase in linear separability of the input features via the LSVMM-based feature reduction indirectly optimises the gradient descent algorithm to converge to a more optimal solution and faster, thus using the LSVMM to assist the MLP in avoiding trapping at local minima and improving the overall classification performance.

The need for performing feature selection via LSVMM and that for customised algorithmic implementation were assessed by comparing the classification performance of the proposed hybrid model (LSVMM-MLP) against a similar

---

130
algorithm (SVM-MLP) that was implemented via Microsoft Azure ML Studio. The augmented Lagrange formulation of the SVM (LSVM) used in the customised algorithm (LSVM-MLP) is not available in Microsoft Azure; thus, it may be regarded as a limitation of the conclusions that can be drawn from the comparison reported in this study.

Nevertheless, the proposed customised LSVM-MLP algorithm resulted in a significantly higher average classification accuracy and AUC with respect to those in the SVM-MLP, respectively, by 29.28% and 28.21% (Table 3), thus demonstrating the need for customised algorithmic implementation. Noteworthy, in further support of the clinical translation and application of the proposed hybrid model, the number of epochs and the computational time were reduced by 91.09% and 99.15%, respectively, when using the customised algorithm (LSVM-MLP) as compared to the SVM-MLP implemented in Microsoft Azure. Qualitatively, the improved classification performance of the hybrid model (LSVM-MLP) against the SVM-MLP in Microsoft Azure can be assessed by Figs. 2, 3 and 4. Such improvements in classification performance (Table 3) in the proposed hybrid model (LSVM-MLP), with respect to a similar hybrid model (SVM-MLP) in Microsoft Azure, are due to the optimised feature selection carried by the LSVM as compared to the SVM in Microsoft Azure, as well as the parameterisation of k folds in the CV algorithm and the number of hidden neurons used in the customised MLP when compared to the MLP in Azure. Therefore, for healthcare-related applications, findings suggest the need for customised algorithmic implementations tailored for specific applications and thorough validation via all performance measures adopted in this study (ACC, SN, SP, AUC, number of epochs and computational time).

To further validate the use of the proposed hybrid algorithm for predicting prognosis of patients with chronic hepatitis, the classification performance of the LSVM-MLP was quantitatively compared against AI-based models tested on the same hepatitis dataset [16] from published studies (Table 4), as selected via the MQAS described in 2.4 and Table 5, reported in “Appendix”. A qualitative comparison of the classification performance of the LSVM-MLP was also performed against the three best performing classifiers (according to the comprehensive classification performance measure AUC) from published studies [13, 14, 31] in Fig. 4, always as selected via the MQAS. As quantitatively and qualitatively assessed in Table 4 and Fig. 4 respectively, findings strongly support the translational application of the proposed hybrid classifier (LSVM-MLP) in a clinical setting, considering its highest classification performance in predicting survival of patients with chronic hepatitis, along with its simple architecture and ease of interpretability as compared to the best performing classifier from the literature [14] of comparable performance. To enable a more thorough comparison of findings for validating the use of novel algorithms for medical applications, the adoption of the MQAS is strongly recommended, not only to assess the quality of published studies applying AI in healthcare but also to guide prospective AI-based translational research using healthcare data for either diagnostic or prognostic purposes.

5 Conclusion

Following a comprehensive analysis of the classification performance of the proposed hybrid model (LSVM-MLP) against the MLP, the SVM-MLP implemented in commercially available software and the AI-based classifiers from selected published studies, the LSVM-MLP proved to be the most accurate, reliable and computationally the fastest AI-based classifier for predicting survival of patients with chronic hepatitis. Furthermore, its simple architecture
and ease of interpretability facilitate clinical translation. Therefore, the L SVM-MLP is deemed clinically useful for aiding prediction of prognosis of patients with chronic hepatitis in a clinical setting.

Noteworthily, the novel hybrid algorithm (LSVM-MLP) achieved 100% of ACC in discriminating between survivors and deceased patients with chronic hepatitis using haematological data and demographics. These findings further confirm the clinical usefulness and viability of the LSVM-MLP classifier for aiding prediction of survival in patients with chronic hepatitis based on readily available data routinely collected in a clinical setting.

The LSVM-MLP’s intuitiveness of the classification procedure and outcomes lends itself to further development into a valuable clinical decision-making support tool. There is also future potential to assist with assessment of other haematological pathologies, such as anaemia, bleeding disorders and blood cancers, thus aiming at optimising patient outcomes more broadly.

Acknowledgements The authors Luca Patini and Narinder Ravi Chandran would like to thank the University of Auckland for giving them the opportunity to carry out their Ph.D. research projects. The authors Luca Patini and Narinder Ravi Chandran would like to thank the University of Auckland High-Performance Technologies Association (UARTA) and Medintello® for giving them the chance of developing this collaborative research work.

Authors’ contributions All authors directly participated in the planning, execution and analysis of the study. All authors also approved the final version of the manuscript, and this submission for possible publication in Neural Computing and Applications.

Funding This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

Appendix

Mean imputation to replace missing values from the initial hepatitis dataset

The “mean imputation” method was applied to replace all missing values that were represented by a question mark in the initial hepatitis dataset (“?”), as per the dataset description [18] of all of the following features by the mean of all available instances:

- fourth (4th) feature, “steroid”, with one (1) missing value;
- sixth (6th) feature, “fatigue”, with one (1) missing value;
- seventh (7th) feature, “malaise”, with one (1) missing value;
- eighth (8th) feature, “anorexia”, with one (1) missing value;
- ninth (9th) feature, “liver big”, with ten (10) missing values;
- tenth (10th) feature, “liver firm”, with eleven (11) missing values;
- eleventh (11th) feature, “spiders palpable”, with five (5) missing values;
- twelfth (12th) feature, “spiders”, with five (5) missing values;
- thirteenth (13th) feature, “ascites”, with five (5) missing values;
- fourteenth (14th) feature, “varices”, with five (5) missing values;
- fifteenth (15th) feature, “bilirubin”, with six (6) missing values;
- sixteenth (16th) feature, “alkaline phosphate”, with twenty-nine (29) missing values;
- seventeenth (17th) feature, “SGOT”, with four (4) missing values;
- eighteenth (18th) feature, “albumin”, with sixteen (16) missing values;
- nineteenth (19th) feature, “protime”, with sixty-seven (67) missing values.

Medintello® Quality Assessment Scale (MQAS) for Studies on Applied Artificial Intelligence in Healthcare

Note: A maximum of two stars can be attributed for Pre-processing, Performance and Clinical Implications. A maximum of three stars can be attributed for Advantages.

1. Rationale

(1) Does the study have a clearly identified clinical goal?
   (a) Yes ∗
   (b) No

2. Objective

(1) Is it clear how the predictive task may help fulfill the clinical goal?
   (a) Yes ∗
   (b) No

3. Classification

(1) Is the measurement for the predictive task clearly defined?
   (a) Yes ∗
4. Pre-processing

(1) Is data cleaning performed?
   (a) Yes ✓
   (b) No

(2) Is data transformation (e.g., normalisation, standardisation) performed?
   (a) Yes ✓
   (b) No

(3) Are outliers identified and removed?
   (a) Yes ✓
   (b) No

(4) Are the methods used for handling missing values clearly described?
   (a) Yes ✓
   (b) No

5. Methods

(1) Clearly defined data sources
   (a) Which data are used ✓
   (b) Presence or lack of missing values ✓
   (c) Ethical approvals obtained or not required as data sources are correctly cited ✓

(2) Clearly defined feature selection algorithm
   (a) Which feature selection algorithm is used ✓

(3) Clearly defined artificial intelligence-based models
   (a) Which learning-based classifier is used ✓
   (b) Which parameters are used ✓
   (c) Which learning algorithm is used ✓
   (d) Which transfer function is used ✓

6. Validation

(1) Are the validation metrics (e.g., mean squared error, sensitivity, specificity and area under the receiver operating characteristic curve) clearly defined?
   (a) Yes ✓
   (b) No

(2) Is the cross-validation set created from the initial data?
   (a) Yes ✓
   (b) No

7. Performance

(1) Are the results reported in confidence intervals?
   (a) Yes ✓
   (b) No

(2) Are results compared with the literature based on confidence intervals?
   (a) Yes ✓
   (b) No

(3) Are results compared with the literature based on accuracy and at least one performance measure (e.g., mean squared error, sensitivity, specificity, area under the receiver operating characteristic curve)?
   (a) Yes ✓
   (b) No

8. Advantages

(1) Are any assumed input and output data format explicitly mentioned?
   (a) Yes ✓
   (b) No

(2) Are there any potential pitfalls in interpreting the model?
   (a) Yes ✓
   (b) No

(3) Are there any potential bias of the data used in the model?
   (a) Yes ✓
   (b) No

(4) Are the findings reporting clinically relevant test classification accuracy (> 80%)?
   (a) Yes ✓
   (b) No
Table 5 MQAS-based stars attributed to published studies on applications of classification algorithms on the same UCI hepatitis dataset [18]

<table>
<thead>
<tr>
<th>Articles</th>
<th>Rationale</th>
<th>Objective</th>
<th>Classification</th>
<th>Pre-processing</th>
<th>Methods</th>
<th>Validation</th>
<th>Performance</th>
<th>Advantages</th>
<th>Clinical translation</th>
<th>Clinical implications</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Are the results confirmatory of what was expected? (2) Is the amount of money, which could be saved, via a better prediction, evaluated? (3) Is the amount of money, which could be saved, via a better prediction, evaluated? (4) Are any clinical implications derived from the study? (5) Are any clinical implications derived from the study? (6) Are any clinical implications derived from the study? (7) Are any clinical implications derived from the study? (8) Are any clinical implications derived from the study? (9) Are any clinical implications derived from the study? (10) Are any clinical implications derived from the study? (11) Are any clinical implications derived from the study? (12) Are any clinical implications derived from the study? (13) Are any clinical implications derived from the study? (14) Are any clinical implications derived from the study? (15) Are any clinical implications derived from the study? (16) Are any clinical implications derived from the study? (17) Are any clinical implications derived from the study? (18) Are any clinical implications derived from the study? (19) Are any clinical implications derived from the study? (20) Are any clinical implications derived from the study? (21) Are any clinical implications derived from the study? (22) Are any clinical implications derived from the study? (23) Are any clinical implications derived from the study? (24) Are any clinical implications derived from the study? (25) Are any clinical implications derived from the study? (26) Are any clinical implications derived from the study? (27) Are any clinical implications derived from the study? (28) Are any clinical implications derived from the study? (29) Are any clinical implications derived from the study? (30) Are any clinical implications derived from the study?
References

6. Patni L, Manoj ML (2016) Preliminary validation of the Lagnagar support vector machine learning classifier as clinical decision-making support tool in aid prediction of prognosis in patients with hepatitis B. J: The 10th international conference on biomedical engineering, National University of Singapore (NUS)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Decision support system to improve postoperative discharge: A novel multi-class classification approach

Luca Parisi, Narreddar Ravichandran, Marianne Lyne Manaog

Abstract

Postoperative discharge decision-making is a critical process that determines not only patient outcomes and, in some cases, their survival, but also the management of the hospital resources, both financial and human ones. Existing decision-making support systems for aiding postoperative discharge mainly rely on statistical-based methods that lack objectivity in predicting optimal recovery area on a subject-specific basis. Machine Learning (ML)-based methods can enable these predictions, but current modeling implementations are inaccurate to be applied clinically or too sophisticated for the relatively low gain in classification performance. As an accurate and reliable method to predict where patients in a postoperative recovery area should be sent to next, the clinical potential of a novel hybrid multi-classification algorithm was assessed. Data on 90 patients regarding their body temperature, oxygen saturation, blood pressure and perceived comfort upon discharge were obtained from the University of California-Irvine (UCI) Machine Learning Repository. A multi-class classification was performed on such data using a "truncated" All-vs-All approach by optimizing kernel and hyperparameters via Genetic Algorithm. The novel hybrid algorithm was found to yield the highest classification accuracy, improving the highest accuracy from the literature by almost 12%. Achieving maximum accuracy and reliability, whilst retaining the lowest computational cost amongst the classifiers tested, the hybrid model is deemed an accurate, reliable and clinically visible solution to assist clinicians and nurses in improving postoperative discharge decision making.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

II. The need for objective, automated postoperative decision making

Currently, physicians are left alone to assess the clinical status of the patients postoperatively based on several factors to decide on where they should be sent to next after undergoing a surgical intervention. Such a postoperative decision making may affect the incidence of postoperative complications, impairing the survival in some patients and the management of the hospital resources (financial resources, e.g., equipment, hospital beds, etc.; human resources, e.g., clinicians and nurses, etc.). Resources in the intensive care unit (ICU) are limited and, therefore, it is vital to prioritize them for patients who need critical medical attention. In a general ward, hospital resources are very limited too, e.g., hospital beds, thus an appropriate management of such resources is required. The high volume of patients under observation and the induced stress by working for long shifts often result in work-related fatigue in clinicians, which may impair the objectivity in their decision making, further complicating patient management and, in some cases, increasing mortality [1].

Hypothermia is a condition that occurs whenever the patient's core body temperature is lower than the physiological temperature (36°C). Postoperative hypothermia is common, as, during surgery, administration of anaesthetics and environmental factors can lower a patient's body temperature [26]. Hypothermia may lead to increased oxygen consumption, increased patient discomfort and physiological instability [27]. Persistent hypothermia is associated with increased mortality in both patients undergoing cardiac [22] and non-cardiac surgery [9]. Thus, early detection and prevention of hypothermia may improve postoperative recovery. Maintaining a physiological body temperature is not only essential for survival [22] but can also help reduce the length of hospital stay, the incidence of surgical site infections, postoperative blood

* Corresponding author.
  Email addresses: luca.parisi@andrew.cmu.edu (L. Parisi), narreddar@ieee.org (N. Ravichandran), marianne@andrew.cmu.edu (M.L. Manaog).

https://doi.org/10.1016/j.knosys.2018.01.003
0098-7295/© 2018 Elsevier B.V. All rights reserved.
transfusions, pressure ulcers, as well as patient discomfort and, ultimately, mortality [20]. Nevertheless, hypothermia is often undiagnosed and left untreated [1].

Postoperative discharge decision making, as any clinical decision-making process, is complex, since it involves several variables that may be associated with each other and with different levels of significance, such as demographics, physical status, body temperature, blood pressure, level of discomfort upon discharge [24], etc. Most of the current systems available in hospitals to aid postoperative discharge decision-making are based on statistical tools, which, however, lack the capability of capturing the underlying patterns of such highly nonlinear and dimensional data to make accurate and reliable predictions.

1.2. A multi-class approach to postoperative discharge decision making

Machine Learning (ML) has been successfully used to improve prediction of postoperative discharge, thus showing potential in improving management of hospital resources and reducing the incidence of postoperative complications [1], e.g., pulmonary-related issues, such as bronchitis, pneumonia, pleural effusions and prolonged mechanical ventilation leading to acute respiratory failure [22]. As compared to statistical-based methods, such as principal component analysis (PCA), ML has been proven to capture the underlying patterns of clinical data further [13].

Postoperative discharge decision making inherently involves a multi-class classification problem, whereby there may be potentially multiple clinical outcomes associated with the patient status, baseline and postoperative characteristics, thus resulting in more than two potential output classes, e.g., further to undergoing surgery, patients may be discharged home or, alternatively, be sent to a general ward if further observation and assistance are required during their recovery, or, should the conditions of the patients deteriorate critically following surgery, sent to the ICU. ML applies the learnt relationship from the input data used for training, once validated on a separate dataset (the cross-validation data), to predict on additional data samples, namely the testing set.

As Hansen and McDonald [7] had pointed out, there is still no objective method for optimising hyperparameters of Machine Learning-based classifiers. Currently, parameterization of such learning parameters heavily relies on the subjective experience of Machine Learning engineers and heuristic methods, which are strictly task- and data-specific and, therefore, not generalizable. Conversely, Genetic Algorithm (GA)-based unconstrained optimization enables to select optimal hyperparameters [15] that yield consistently higher classification performance than heuristic and statistical methods [7], thus showing potential in improving the accuracy and reliability of ML-based multi-class classification methods for aiding prediction of postoperative discharge.

1.3. Related studies

In an attempt of dealing with such uncertainties on postoperative data, linguistic data were represented as fuzzy numbers and principal component analysis (PCA) was used for feature reduction in an attempt of improving classification performance in previous studies [11].

Luso [11] implemented a fuzzy- and PCA-based similarity classifier and managed to predict the most appropriate recovery area (home, general hospital ward or ICU) in the same dataset of 90 patients developed by Summers et al. [24] with 80.22% testing classification accuracy. Forghani and Yaafid [5] proposed a more complex algorithm, the fuzzy min-max neural network with symmetric margin (FMNWSM) and applied it to classify the same data set [24], achieving 72.99% of ACC, as compared to 74.75% with Support Vector Machine (SVM), but with a considerably lower computational cost than SVM (2.1 ms for FMNWSM algorithm against 4.083 ms for the SVM). Although FMNWSM converges faster than conventional fuzzy min-max neural networks due to the lack of special nodes in overlapped regions [5], its architecture is too complex for it to be used in a clinical setting, considering that its accuracy (72.99%) is not clinically valuable.

ML has been successfully applied to avoid hypothermia by predicting the most appropriate recovery area for each patient (home, general ward or ICU) [8]. Heile et al. [8] used a particle swarm optimization (PSO)-based fuzzy classifier to retrieve the crisp rules from the same postoperative data set [24], where such rules were used to aid prediction of clinical decision making, which was 84% accurate.

Using 50-fold cross validation (CV), Abouel et al. [1] applied SVM and Artificial Neural Networks (ANN) in Weka to predict postoperative discharge on the same data set of interest [24], obtaining 88.54% and 82.81% of classification accuracy respectively.

To the best of authors' knowledge, the M-SVM (SVM for multi-class classification) used by Abouel et al. [1] is the best performing classifier on this data set [24] to date.

1.4. Rationale and aim

Temperature control post-surgery can aid early recovery, thus increasing patient comfort, decreasing shivering and associated oxygen saturation, facilitating physiological stability and leading to better postoperative outcomes [27]. We hypothesise that ML can assist physicians and nurses in making an evidence-based informed decision on where patients should be sent to next after surgery, thus optimising clinical outcome and management of hospital resources, also improving survival in some cases. Improved physiological stability can reduce the length of hospitalisation in an expensive recovery room substantially, thus also saving additional costs to the hospital, insurance companies and, in some cases, to the patient themselves [27].

Therefore, such expert systems using ML can help prevent hypothermia, thus also reducing information overload, typical in healthcare as described in Sections 1.1 and 1.2, and, therefore, improving the quality of patient care and clinical outcomes [27]. Nevertheless, the underlying engineering process for devising a ML-expert system with predictive capability on selecting the most appropriate recovery area post-surgery for each patient is still either too complex to be applied in a clinical setting [5,31] or inadequately described [8].

Considering that the lack of clinical translation of previous research findings [39,31], has left the clinical challenges discussed in 1.1 still open, this study aims to develop and validate a novel hybrid algorithm that can accurately and reliably predict postoperative discharge in case of unbalanced classes, typical in healthcare data of this type.

2. Methods

2.1. Data pre-processing and encoding methods

Postoperative discharge-related data from the University California-Irvine (UCI) Machine Learning repository on ninety (N = 90) patients were used [24]. Since hypothermia is a concern post-surgery, the input features are related to body temperature measurements.

The eight input features for classification are the following:

1. Patient's internal temperature (\(-0.5^\circ\mathrm{C}\) to +3°C), high (+37), and low (+36 and +35); low (36.5 and +35); low (<+35);
3. Oxygen saturation (O2-Sat in %), excellent (98-100), good (95-98) and (92-95), fair (90-95), poor (<90).
4. Last measurement of blood pressure (BPMeasure), high (>130/90), and (120/90), low (<90).  
7. Stability of patient's blood pressure (BP-STRIB), stable, moderately stable, unstable. 
8. Patient's perceived comfort at discharge (COMFORT), measured as an integer between 0 and 20.

The three missing values in the above-mentioned feature no. 8 were replaced by mean imputation.

The next step is to improve clinical outcomes using the above-mentioned input data, the classification task consisted of predicting whether patients in a recovery area post-surgery should be sent to: home (N = 243), to a general hospital ward (N = 64) or to the ICU (N = 2).

To this purpose, multi-class classification was performed on the above-mentioned dataset via the following algorithms in MATLAB (R2017b), The MathWorks, Natick, United States, on an Intel® CoreTM i5-4460T, 2.9GHz Processor:

- The supervised ANN Multi-Layer Perceptron for multi-class classification (M-MLP), with learning rate 0.5 and momentum 0.8 [13,14].
- The supervised ML SVM for multi-class classification (M-SVM) and its Lagrangian augmentation (M-SVM), which improves margin maximisation further [12].

Prior to classification, the input data were randomised. Thereafter, input data were cleaned (ie., outliers were removed), standardised (ie. z-scores to have mean 0 and standard deviation 1) and split into train and test prior to classifying them. As the number of k subsets in the k-fold cross-validation (CV) algorithm deployed should be equal to the number of the input features [N = 8] [4,17], k-fold CV was applied to all ML-based classifiers tested.

2.2. A hybrid multi-class classification algorithm to improve postoperative discharge decision making: GA-M-SVM

Typically, extending binary SVM to solve multi-class classification problems involves either the use of heuristic methods or large optimisation algorithms. In this study, a computationally efficient GA-based optimisation method was used to find the optimal kernel function. At input parameters that would yield the highest classification performance in the multi-class classification SVM (M-SVM) tested in this study. The All-vs-All (AVA) approach reduces a multi-class classification to a series of binary classifications, where each classifier learns to discriminate between pairs of classes, until all possible combinations of classification are attempted. All M-SVM (with and without the GA-based optimisation of hyperparameters) were modelled with the same optimal kernel to enable a fair comparison of their classification performance. Subsequently, the hypothesis whereby a linear kernel in an AVA multi-class classification would have yielded optimal hyperparameters in all binary classifications involved in this study was verified by applying a GA-based optimisation and then, testing on which kernel function would have led to the highest classification performance, while minimising the computational cost. Concurrently, this optimisation would have ensured a higher linear separability in case of highly unbalanced classes in the input data, such as in this data set developed by Summers et al. [24]. Thus, it was hypothesised that such an optimisation would have naturally led to an improvement of the AUC and AUC of the resulting M-SVM in such an AVA configuration.

As opposed to the conventional One-vs-All (OVA) approach, which leads to a higher computational cost and larger regions of ambiguity [25], AVA reduces a multi-class classification to a series of binary classifications minimising both the computational cost and reducing the areas of ambiguity. In an AVA-based multi-class classification, each classifier learns to discriminate between pairs of classes, until all possible combinations of classification are attempted (three combinations in this study, considering these classes, ie., binary classifications between the first class and the second one, between the first class and the third one, between the second class and the third one).

Traditional, uncontrolled or non-parametrised SVM may not converge due to the high percentage of nonlinearly separable cases in the input data leading to high misclassification rates. The SVM would thus be trapped at local minima, theoretically for an infinite number of training iterations or epochs. To help prevent the predictive error (between the actual and the target outputs) from increasing nonlinearly with the number of epochs, the augmented Lagrangian formulation of SVM (LSVM) was used [12], with

\[ v - k_n \leq 0, \text{where } k_n \text{ is a real number (by default, } k_n = 1) \text{ such that the global linear convergence criterion for the LSVM (0 \leq v < 2\gamma) is satisfied, is } \gamma \text{ the convergence parameter, which is equal to } 12\gamma \text{ } [12]. \]

Such an augmented formulation enables the M-SVM to ensure linear separability of the input data via an optimal separating hyperplane (OSH) that maximises the margin further, as compared to traditional M-SVM.

Furthermore, Genetic Algorithms (GAs) were used to optimise not only the hyperparameters of the M-SVM but also its kernel, which ultimately determines the OSH as a decision boundary. This optimisation was performed for each of the three multi-class classification (LSVM-M-SVM) used in this study to improve their margin maximisation and the linear separability of the input data into the three target classes of interest (patients to be discharged home, those to be sent to a general hospital ward and those to be admitted to the ICU). GAs were also used to help avoid convergence issues and trapping of the solution at local minima instead of converging to the desired global minimum, thus improving the overall classification performance of the M-SVM.

In the proposed implementation for a GA-optimised M-SVM, the fitness functions (1) and (2) were defined by input (the input data to be classified), v (the augmented Lagrangian parameter) in (3), nopt (the maximum number of training iterations or epochs) and u (the Lagrangian convergence parameter) in (4). The optimised set of solutions (v, u) was sought to maximise the overall classification performance. The training was stopped when the maximum number of epochs prior to convergence (nopt) was reached, i.e., when the target error (v) was lower than 10⁻⁵: m × 1 in (3) is the column vector of the input matrix inputs. The optimal set of kernels and hyperparameters (v and u) for each of the three binary SVM classifiers used in an AVA configuration to discriminate between each pair of classes were selected to maximise the overall classification performance, as quantified by the AUC if multiple sets of kernels and hyperparameters leading to the maximum predictive performance were found via the GA-based optimisation, the one achieving it while retaining the lowest computational cost was eventually selected as the optimal one.

\[ f(v, u, \text{input kernel}) = v \cdot (\text{input, kernel}) \]

\[ f(\text{err, opt v, nopt, kernel}) = -\text{err} \cdot (\text{opt v, kernel}) \]

\[ v \leq -\frac{k_n}{\text{epochs}} \left[ \text{epochs} \leq 10^{-5} \right] \]

\[ u = \frac{\text{epochs}}{\text{opt v}} \]

\[ a = \frac{k_n}{\text{epochs}} \]
The All-vs-All (Ava) multi-class classification approach using the proposed GA-optimisation for M-LSVM is shown in Fig. 1, as opposed to the conventional Ova approach.

2.3. Comparison of classification performance between the proposed model and classifiers from published studies

To assess the need for a GA-based optimisation of the kernel function and hyperparameters for the M-LSVM, the classification performance of the proposed hybrid model (GA-M-LSVM) was compared against that of classifiers from published studies. The same data pre-processing methodology in 2.1 was applied to enable a fair comparison of their classification performance, which was assessed via the following two gold standard performance criteria of accuracy and robustness of ML-based classifiers [6]:

1. Accuracy, quantifying the predictive power of discriminating between the classes of interest, measured by the testing classification accuracy (ACC) (5), the only one that is clinically important, as it is based on the testing set of unknown data to the classifier.

2. Reliability, the predictive capability of attributing an appropriate degree of certainty on the classification outcome, quantifiable via sensitivity (Sn) (6) and specificity (Sp) (7), and, more comprehensively, via the area under the curve (AUC) based on a pre-set threshold (t) (8).

Generally, for a binary classification, the outcome for each patient is either ‘true’ or ‘false’, thus resulting in four possible scenarios (5). However, for a multi-class classification problem, such as the one involved in this study, the overall outcome is derived from each binary classification, as shown in Table 1.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (7)
\]

\[
\text{Area under the curve} = \int_0^1 \frac{TP}{TP + FN} (t) \left( \frac{1}{\sum_{i=1}^K p_i} \right) \left( \frac{1}{h} \right) \, dt \quad (8)
\]

Where TP, TN, FP and FN are the means of the cumulative true positives (TP_i, TP_s, TP_t, TP_f), true negatives (TN_i, TN_s, TN_t, TN_f), false positives (FP_i, FP_s, FP_t, FP_f), false negatives (FN_i, FN_s, FN_t, FN_f) in Table 1.

High accuracy and reliability, but a low computational cost, were sought to select the most suitable algorithm to aid prediction of postoperative discharge via ML in a clinical setting. As mentioned in 2.2, the term “computational cost” has been used in this study to include both the computational time to execute an algorithm and the number of training iterations (epochs) prior to reaching convergence. Since the former metric is hardware-dependent, it was discarded from the radar plot in Fig. 6, whilst the latter measure was included, as it does not depend on the hardware whereby the algorithm has been tested, but only on the algorithm itself. Therefore, the number of epochs is deemed more important than the computational time when assessing the performance of ML-based algorithms.
### Table 1

Confusion matrix for a three-class classification problem, whereas TP, TN, FP and FN cases are referred to the patients involved in this study for each of the three binary classifications (mA, 1, 2) derived from the initial multi-class classification problem shown in Fig. 1.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual (mA)</th>
<th>Adj. TP</th>
<th>Adj. TN</th>
<th>Adj. FP</th>
<th>Adj. FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (P)</td>
<td>True (TP)</td>
<td>True (TN)</td>
<td>False (FP)</td>
<td>False (FN)</td>
<td></td>
</tr>
<tr>
<td>Negative (N)</td>
<td>True (TP)</td>
<td>True (TN)</td>
<td>False (FP)</td>
<td>False (FN)</td>
<td></td>
</tr>
<tr>
<td>Positive (P)</td>
<td>True (TP)</td>
<td>True (TN)</td>
<td>False (FP)</td>
<td>False (FN)</td>
<td></td>
</tr>
<tr>
<td>Negative (N)</td>
<td>True (TP)</td>
<td>True (TN)</td>
<td>False (FP)</td>
<td>False (FN)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

Comparison of classification performance of all AI-based classifiers tested in this study.

<table>
<thead>
<tr>
<th>AI Classifiers</th>
<th>ACC (95% CI)</th>
<th>SN (95% CI)</th>
<th>SP (95% CI)</th>
<th>AUC (95% CI)</th>
<th>Epochs</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-MLP No PCA</td>
<td>94.31 (93.87-94.90)</td>
<td>80 (80-80)</td>
<td>95.98 (94.36-96.91)</td>
<td>0.48 (0.44-0.52)</td>
<td>24 (19-29)</td>
<td>59 (45-72)</td>
</tr>
<tr>
<td>PCA</td>
<td>94.47 (93.98-94.95)</td>
<td>79 (76-81)</td>
<td>96.09 (94.57-96.85)</td>
<td>0.50 (0.46-0.54)</td>
<td>32 (23-41)</td>
<td>68 (56-78)</td>
</tr>
<tr>
<td>M-SVM No PCA</td>
<td>94.79 (94.34-95.33)</td>
<td>75 (73-77)</td>
<td>96.98 (95.46-97.96)</td>
<td>0.59 (0.55-0.63)</td>
<td>27 (22-31)</td>
<td>52 (40-63)</td>
</tr>
<tr>
<td>PCA</td>
<td>94.91 (94.46-95.46)</td>
<td>76 (74-78)</td>
<td>96.95 (95.42-97.97)</td>
<td>0.56 (0.52-0.60)</td>
<td>32 (25-37)</td>
<td>57 (43-68)</td>
</tr>
<tr>
<td>M-SVM No PCA</td>
<td>94.79 (94.34-95.33)</td>
<td>75 (73-77)</td>
<td>96.98 (95.46-97.96)</td>
<td>0.59 (0.55-0.63)</td>
<td>27 (22-31)</td>
<td>52 (40-63)</td>
</tr>
<tr>
<td>PCA</td>
<td>94.91 (94.46-95.46)</td>
<td>76 (74-78)</td>
<td>96.95 (95.42-97.97)</td>
<td>0.56 (0.52-0.60)</td>
<td>32 (25-37)</td>
<td>57 (43-68)</td>
</tr>
<tr>
<td>GA-M-SVM No PCA</td>
<td>94.79 (94.34-95.33)</td>
<td>75 (73-77)</td>
<td>96.98 (95.46-97.96)</td>
<td>0.59 (0.55-0.63)</td>
<td>27 (22-31)</td>
<td>52 (40-63)</td>
</tr>
<tr>
<td>GA-M-SVM PCA</td>
<td>94.91 (94.34-95.33)</td>
<td>76 (74-78)</td>
<td>96.95 (95.42-97.97)</td>
<td>0.56 (0.52-0.60)</td>
<td>32 (25-37)</td>
<td>57 (43-68)</td>
</tr>
</tbody>
</table>

3. Results

3.1. GA-based optimisation of machine learning hyperparameters to improve classification performance

GA-based optimisation was applied to select the optimal hyperparameters α and β in the M-LSMVM used for classification in this study. Considering the number of input features (N = 8), 8-fold CV was selected [10] and 8 hidden neurons were used in the hidden layer of the M-MLP. Initially, the M-LSMVM was applied for classifying all input data features (N = 8) with default hyperparameters, i.e., α = 10, β = 1, and γ = 1 [12]. Further, to performing a GA-based optimisation to optimise α and β in all three M-LSMVM algorithms individually, as used in an All-vs-All (AvA) configuration for multi-class classification, the classification performance of the hybrid algorithm GA-M-LSMVM was compared against that of the M-MLP [13], a traditional M-SVM and the M-LSMVM algorithm ([Table 2]).

Table 2 shows a comparative analysis of classification performance between the M-MLP, M-SVM and the M-LSMVM when considered individually, and the proposed hybrid classifier (GA-M-LSMVM), with the testing accuracy referred to as "ACC", the sensitivity in "SN", the specificity as "SP", the area under the receiver operating characteristic curve as "AUC", the number of training iterations as "Epochs" and the computational time as "Time".

While the M-MLP achieved the highest testing classification accuracy (ACC: 94.79, 95%CI: 94.34-95.33) when PCA was not applied, the M-LSMVM achieved the highest overall classification performance (AUC: 0.59, 95%CI: 0.55-0.63). The Lagrangian augmentation of SVM extended to multi-class classification (M-LSMVM) led to a higher ACC (0.75, 95%CI: 0.72-0.78). The GA-M-LSMVM achieved an AUC (0.62, 95%CI: 0.55-0.69) and MCP (0.75, 95%CI: 0.67-0.83) with the highest accuracy and sensitivity, respectively, compared to the traditional SVM (ACC: 0.55, 95%CI: 0.43-0.68, AUC: 0.42, 95%CI: 0.36-0.48).

Further to performing a GA-based optimisation of the hyperparameters of the M-LSMVM, as in 2.2, the novel hybrid algorithm (GA-M-LSMVM) reached 100% of both accuracy (ACC) and reliability (SN, SP and AUC), in a considerably reduced number of training iterations (3 epochs) and decreased computational time (0.2 s) (Table 2). In fact, by optimising hyperparameters via GA, the M-LSMVM reduced both the number of epochs required prior to convergence (from 5 to 3, Table 2) and its computational time by almost half (from 0.38 s to 0.2 s, Table 2).

Specifically, when using the novel hybrid algorithm (GA-M-LSMVM) as compared to the M-LSMVM (Table 2), the following percentage changes in classification performance measures were noted when PCA was not applied:

- An increase of 14.29% in ACC;
- An increase of 13.33% in SN;
- An increase of 15.38% in SP;
- An increase of 29.23% in AUC;
- A decrease of 40% in the number of epochs;
- A decrease of 47.37% in computational time.

A qualitative comparison of the classification performance of the proposed hybrid algorithm with GA-based optimised hyperparameters (GA-M-LSMVM) against that of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM and M-LSMVM) was also performed via the following:

- A conventional two-dimensional (2D) convergence plot comparing testing classification error (0.18-0.05) and the number of epochs that the classifiers underwent prior to reaching convergence (Fig. 2), with the best performing classifier having zero error and the lowest number of epochs (GA-M-LSMVM without PCA, as represented by the pink circle at the bottom-left corner of Fig. 2);
- A 2D plot comparing the testing classification accuracy (ACC) and the number of training iterations prior to convergence.
[310x65]141

Fig. 2. 2D convergence plot when comparing training classification error (100-ACC %) and the number of epochs between the proposed hybrid algorithm with GA-based optimized hyperparameters (GA-M-LSVM) and those of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM and M-LSVM) with and without PCA for feature selection.

3.2. Comparison of classification performance between the hybrid algorithm and classifiers from published studies.

In Table 3, ACC, SN, SP and AUC of all algorithms tested (the initial M-MLP, M-SVM and M-LSVM, as well as the proposed GA-M-LSVM) were compared quantitatively against published studies using the same dataset on postoperative discharge [24]. Table 3 also reports the type of multi-class classification applied and the kernel used for classification. The numbers of epochs and the computational time are also reported in Table 3. Differently from published studies, the novel hybrid classifier (GA-M-LSVM) (Fig. 1) achieved the maximum classification performance across all four measures used for comparison (100% ACC, SN and SP and 1 as AUC).

4. Discussion

4.1. The advantage of 'unsupervised' All-vs-All for multi-class classification of data on postoperative discharge.

The M-MLP, M-SVM and M-LSVM were first used individually for predicting the optimal recovery based on all input postoperative discharge related data (N = 8); the M-MLP could reach 92.31% of testing classification accuracy (AUC = 0.65, 95% CI: 0.61-0.68) (Table 2), as also demonstrated in a previous study [18], which satisfies the hypothesis whereby the MLP is a suitable algorithm for classifying highly dimensional and nonlinear data. However, the decrease in specificity (60%, 95%CI 14.66-94.75) does not make the M-MLP a reliable classifier for applications in a clinical setting. Thus, it was assumed that a linear SVM could have improved the linear separability of the unbalanced input data into the three target classes (discharge home, to a general hospital ward or admission to the ICU) by making use of the margin maximization. However, a slightly reduced ACC was noted in the M-SVM when PCA was not applied for feature selection [70, 95%CI: 50.43%–73.21%]; noteworthy, feature selection via PCA was only beneficial when using the M-SVM classifier, resulting in an increase in ACC by 11.5%, although the overall classification performance, as more comprehensively measured by the AUC (Table 2), decreased (AUC for M-SVM with PCA: 0.42, 95%CI: 0.36–0.48; AUC for M-SVM without PCA: 0.53, 95%CI: 0.27–0.80). The classification performance of the M-LSVM was also tested on the same data set [24], as the L SVM improves margin maximization via an augmented Lagrangian formulation with respect to traditional SVM [12,15,16].

The M-LSVM resulted in a higher overall classification performance (AUC: 0.63, 95%CI 1–1; AUC: 0.79, 95%CI 0.75–0.80) than the conventional SVM, and the same optimal linear kernel was used for enabling a fair comparison between the respect-
Fig. 3. A comparison of testing classification accuracy (ACC) and the number of training iterations prior to convergence (epochs) between the proposed hybrid algorithm with GA-based optimized hyperparameters (GA-M-LSVF) and those of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM, and M-LSVM) with and without PCA for feature selection. The ACC-related confidence intervals are represented by error bars.

Fig. 4. 3D convergence plot when considering classification performance (AUC), against time and epochs between the proposed hybrid algorithm with GA-based optimized hyperparameters (GA-M-LSVF) and those of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM, and M-LSVM) with and without PCA for feature selection.
Fig. 5. 3D plot when comparing the classification performance between the proposed hybrid algorithms with GA-based optimised hyperparameters (GA-M-LSVM) and those of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM, M-LSVM) with and without PCA for feature selection.

Fig. 6. Radar plot comparing the classification performance between the proposed hybrid algorithms with GA-based optimised hyperparameters (GA-M-LSVM) and those of multi-class classification algorithms with default hyperparameters (M-MLP, M-SVM, M-LSVM) with and without PCA for feature selection.
Table 3: Comparison of characteristics and classification performance between the proposed hybrid classifier (GA-M-L SVM) and the Machine learning-based classifiers used in published studies.

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Type of multi-class classification</th>
<th>Kernel</th>
<th>Classifier</th>
<th>M (ACC-S</th>
<th>N (%)</th>
<th>M (SN-S</th>
<th>N (%)</th>
<th>M (SP-ES</th>
<th>N (%)</th>
<th>M (AUC-SSC</th>
<th>N (%)</th>
<th>Epochs</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>OVA</td>
<td>RBF</td>
<td>M-SVM</td>
<td>80.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>[2]</td>
<td>GA-based</td>
<td>RBF</td>
<td>M-SVM</td>
<td>74.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>[4]</td>
<td>FS- and fuzzy-based</td>
<td>Similarity</td>
<td>FHMNN</td>
<td>76.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>[5]</td>
<td>FS- and Hybrid</td>
<td>Similarity</td>
<td>New Similirity-based</td>
<td>66.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>This study</td>
<td>0.0%</td>
<td>Linear</td>
<td>GA-M-L SVM</td>
<td>95.00</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.82</td>
<td>100</td>
<td>0.20</td>
<td>100</td>
<td>5</td>
<td>600</td>
</tr>
</tbody>
</table>

* - not reported; OVA: One-Vs-All; RBF: Radial Basis Function; M-SVM: Multi-class Support Vector Machine; GA: Genetic Algorithm; M-LP: Multi-Layer Perceptron; FS: Filter-based feature selection; HFMNN: Hybrid Fuzzy Neural Network; PCA: Principle Component Analysis; Avg. 16-Vs-16.

time performance measures (ACC, SN, SP, AUC, epochs and computational time). Optimal kernel and hyperparameters were selected via GA for each of the classifiers used in the GA-M-L SVM in an All-Vs-All (Ava) configuration, thus achieving a finer control over the decision boundary for classification with respect to the conventional Ava or One-Vs-All (OVA) multi-class classifiers [21].

Considering partial classification outcomes achieved by the first of the three classifiers used in the Ava configuration for all classifiers tested with default parameters (M-LP, M-SVM and M-L SVM), the partial binary discrimination between patients to be discharged home and those to be discharged to a general hospital ward was the lowest one amongst all three partial binary classifications that the initial multi-class classification problem was decomposed in to an Ava configuration, thus resulting in an average partial classification accuracy of about 70% across the three classifiers tested.

This finding suggests that the heterogeneity in either cohorts with regards to the degree of severity in patients conditions compounds postoperative discharge decision making more than the presence of unbalanced classes (2 patients to be sent to the ICU, 30 to be discharged home and the remaining 64 to be sent to a general hospital ward) [24]. However, optimising kernel and hyperparameters via GA enabled the hybrid algorithm GA-M-L SVM to capture such highly heterogeneous underlying patterns in the initial data and achieve the maximum classification performance (ACC95.0%; SN95.0%; AUC; 1), also in presence of such unbalanced classes of patients. Such a maximum classification performance can be explained by accounting for different weights of the errors in misclassification that were minimised by a GA-specific optimisation for each of the binary classifiers in the Ava configuration of the GA-M-L SVM and integrated in the loss function, enabling a higher generalisability of the resulting multi-class classification knowledge-based system. To enable a fair comparison on the classification performance, the Ava configuration was adopted in all classifiers tested. In the literature on multi-class classification, the study conducted by Rifkin and Klautau [21] is often reported as evidence that the Ava configuration leads to a higher classification performance in SVM [36]; nevertheless, findings suggest that Ava is a promising method for multi-class classification when using SVM in a ‘controlled’ modelling scenario, such as that obtained by a GA-based optimisation of the SVM-related kernel and hyperparameters in this research study.

Since the SP of the M-L SVM was the highest one across all classifiers with default hyperparameters tested (SP: 85.68, 95.03, 59.54–89.34), and considering that the M-L SVM without PCA had a reduced number of epochs (N = 5) to converge with respect to the M-LP without PCA (N = 24) and the M-SVM when PCA was applied (N = 7) (Table 2), it was hypothesised that, by optimising hyperparameters via GA, which had achieved the highest performance in predicting optimal recovery area (AUC: 0.77, 95%CI 0.75–0.80), such an optimised margin maximisation would have increased the linear separability of the input classes to the three target classes further. Therefore, to test this hypothesis, the GA-optimised hyperparameters were input to the three M-L SVM used in an Ava configuration respectively, thus optimising discrimination between each pair of classes when considered at a time, resulting in a novel hybrid algorithm (GA-M-L SVM), which not only achieved 100% of average classification accuracy, but outperforming all other classifiers tested, but also with the lowest number of epochs (N = 5) and computational time (0.2 sec) required prior to convergence (Table 2). The increase in classification performance and the decreased number of training iterations (epochs) required to classify the input data (Table 2) verified the hypothesis whereby the GA-optimised hyperparameters v and w in the M-L SVM increased the linear separability of the input data, thus assisting the M-SVM indirectly in finding a more optimal separating hyperplane, improving margin maximisation and classification of the input data to predict optimal recovery area more accurately and reliably. The decrease in computational time (Table 2) can be explained by the optimised augmented Lagrangian margin maximisation used to achieve the same maximum classification performance.

Qualitatively, the improved classification performance of the hybrid model (GA-M-L SVM) against the M-LP, M-SVM and the M-L SVM, when tested individually, can be assessed by Figs. 2-6. This finding demonstrates the usefulness of applying GA to optimise hyperparameters for the M-L SVM classifier, thus increasing the linear separability of the input data to be fed into the SVM for classification. This increase in linear separability of the input features via GA indirectly optimises the augmented Lagrangian algorithm of SVM to converge to a more optimal solution and faster, thus using GA to assist the SVM in avoiding trapping at local minima and improving the overall classification performance.

4.2. An effective algorithm for clinical translation

To further validate the use of the proposed hybrid algorithm for automated prediction of optimal recovery area based on postoperative discharge-related data, the classification performance of the GA-M-L SVM was quantitatively compared with ML-based models tested on the same dataset [24] from published studies (Table 3). The testing classification accuracy of the GA-M-L SVM was far higher (95%) than that of the best performing classifier from published studies (88-84%) [1]. As quantitatively and qualitatively assessed in Tables 2-3 and Figs. 2-6 respectively, findings strongly support the translational applicability of the proposed hybrid classifier (GA-M-L SVM).
Furthermore, the architecture of the GA-M-LSVM is more intuitive than common fuzzy-based classifiers from the literature [8][12][23], whose degree of complexity did not lead to clinically valuable classification performance (ACC < 85%). To enable a more thorough comparison of findings for validating the use of novel algorithms for medical applications, it is recommended to re-run port findings comprehensively, i.e. reporting both the testing classification accuracy (ACC) and reliability (SN, SP and AUC) of the algorithm using the data sets in Tables 2-3, as well as reporting results in confidence intervals rather than only as point estimates (Table 3), such as in all previous works using the same data sets compared in this study [12][20][11][23].

Such ML-based algorithms could be applied to devise accurate and reliable software-based systems to enable automated prediction of optimal recovery area for patients following surgery.

5. Conclusion

Following a comprehensive analysis of the classification performance of the proposed hybrid model (GA-M-LSVM), against the M-MFL, the M-SVM and the M-LSVM, as well as the best performing ML-based classifiers from published studies, the GA-M-LSVM proved to be the most accurate, reliable and computationally the fastest ML-based classifier for predicting the optimal recovery area on a patient-specific basis, using postoperative discharge-related data. Therefore, the GA-M-LSVM can aid prediction of optimal recovery in a clinical setting, thus improving the accuracy and reliability of postoperative discharge decision making.

Noteworthily, the novel hybrid algorithms (GA-M-LSVM) not only achieved 100% of ACC (and 1 of AUC) in discriminating between three different classes of patients, i.e., those discharged home, those to be discharged to a general hospital ward and those requiring treatment in the ICU, using postoperative discharge-related data, but it did so with the lowest computational cost, as quantified through the number of training iterations (epochs) prior to reaching convergence and the computational time. These findings further confirm the usefulness and viability of the GA-M-LSVM classifier for automated prediction of optimal recovery area based on readily available data.

The GA-M-LSVM’s intuitiveness of the classification procedure and outcomes lends itself to further development into a valuable clinical decision-making support tool.

Acknowledgment

The authors L.P. and N.R. would like to thank the University of Auckland for giving them the opportunity to carry out their PhD research projects. The authors L.P. and N.R. would like to thank the University of Auckland Rehabilitation Technologies Association (UARTA) and Medinefis®, for giving them the chance of developing this collaborative research work.

Declaration of interest

The authors declare no conflicts of interest.

Contributors

All authors directly participated in the planning, execution and analysis in the study. All authors also approved the final version of the manuscript, and this submission for possible publication in Knowledge-Based Systems.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[16] L. Pari, M.L. Maiy, Preliminary validation of the laparoscopic support vector learning support machine learning as a clinical decision-making tool for postoperative complications in patients with laparoscopy. The 8th International Conference on Biomedical Engineering, National University of Singapore (Singapore), 2016.
Evolutionary Algorithms for Margin Maximisation of Support Vector Machine

Luca Parisi$^{1,2}$, Narrendar RaviChandran$^{2,3}$

$^1$Auckland Bioengineering Institute (ABI), University of Auckland
luca.parisi@ieee.org

$^2$University of Auckland Rehabilitative Technologies Association (UARTA), University of Auckland

$^3$Department of Mechanical Engineering, University of Auckland

The capability of support vector machines to classify highly non-linear, multi-dimensional data heavily relies on the choice of kernel cost parameters. Appropriate kernel function parameters influence the determination of the decision boundary.

Support vector machine uses hyperplanes as a decision boundary for data classification. Particularly, the selection of an optimal hyperplane determines the accuracy, sensitivity and specificity of such learning-based classification algorithms.

Given that evolutionary algorithms can help minimise the cost function, we have formulated a mathematical approach that uses evolutionary algorithms to optimise kernel function parameter selection of support vector machines.

This algorithmic implementation can lead to a considerable improvement in maximising the margin of the optimal hyperplane, accounting for higher true positive and true negative rates, thus maximising the classification accuracy.
Appendix C – Supporting documents

The author delivered an oral presentation entitled "A Machine Learning–Based Speech Processing Solution for Facilitating Early Diagnosis of Parkinson’s Disease" (Parisi & Manaog, 2017) at the MATLAB Conference 2017 held in Auckland, New Zealand. It is worth noting he was the only selected author who was not working for MathWorks®.

The above-mentioned outcomes were derived via a follow-on study from that of Parisi & Manaog (2016), which granted the author one of the six nominations towards the “Young Investigator Award” at the 16th International Conference on Biomedical Engineering (ICBME) organised by the National University of Singapore (NUS).

These two studies set the foundations for the pioneering study of Parisi et al. (2018a), wherein a novel AI-based system was devised and validated as a clinical decision support tool for aiding early diagnosis of Parkinson’s Disease based on speech data and clinical scores of disease severity.
<table>
<thead>
<tr>
<th>Time</th>
<th>Location</th>
<th>Programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00</td>
<td>LT02</td>
<td>Opening Ceremony and Welcome</td>
</tr>
<tr>
<td>08:30</td>
<td>LT02</td>
<td>- The Keynote Talk: A New Paradigm in Cancer Treatment - Dr. Jane Doe, University of Cambridge</td>
</tr>
<tr>
<td>09:15</td>
<td>LT02</td>
<td>- Innovative Technologies in Cancer Research - Dr. John Smith, Cancer Research Institute</td>
</tr>
<tr>
<td>10:00</td>
<td>LT02</td>
<td>- Break - Coffee Break</td>
</tr>
<tr>
<td>10:30</td>
<td>LT02</td>
<td>- Recent Advances in Cancer Immunotherapy - Dr. Emma Brown, University of Oxford</td>
</tr>
<tr>
<td>11:15</td>
<td>LT02</td>
<td>- The Role of Genomics in Cancer Diagnosis and Treatment - Dr. Michael Johnson, National Cancer Institute</td>
</tr>
<tr>
<td>12:00</td>
<td>LT02</td>
<td>- Lunch - Buffet Lunch</td>
</tr>
<tr>
<td>13:30</td>
<td>LT02</td>
<td>- The Impact of Artificial Intelligence in Cancer Research - Dr. Sarah Wilson, IBM Research</td>
</tr>
<tr>
<td>14:15</td>
<td>LT02</td>
<td>- The Future of Cancer Research: Challenges and Opportunities - Dr. David Taylor, Cancer Research UK</td>
</tr>
<tr>
<td>15:00</td>
<td>LT02</td>
<td>- Break - Tea Break</td>
</tr>
<tr>
<td>15:30</td>
<td>LT02</td>
<td>- The Importance of Public Health in Cancer Prevention - Dr. Emily Green, American Cancer Society</td>
</tr>
<tr>
<td>16:15</td>
<td>LT02</td>
<td>- The Economic Impact of Cancer - Dr. Robert Lee, World Health Organization</td>
</tr>
<tr>
<td>17:00</td>
<td>LT02</td>
<td>- Closing Remarks - Dr. Jane Doe, University of Cambridge</td>
</tr>
</tbody>
</table>

**Day 2 - Wednesday, 8 December 2016**

<table>
<thead>
<tr>
<th>Time</th>
<th>Location</th>
<th>Programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00</td>
<td>LT02</td>
<td>Opening Ceremony and Welcome</td>
</tr>
<tr>
<td>08:30</td>
<td>LT02</td>
<td>- The Keynote Talk: A New Paradigm in Cancer Treatment - Dr. Jane Doe, University of Cambridge</td>
</tr>
<tr>
<td>09:15</td>
<td>LT02</td>
<td>- Innovative Technologies in Cancer Research - Dr. John Smith, Cancer Research Institute</td>
</tr>
<tr>
<td>10:00</td>
<td>LT02</td>
<td>- Break - Coffee Break</td>
</tr>
<tr>
<td>10:30</td>
<td>LT02</td>
<td>- Recent Advances in Cancer Immunotherapy - Dr. Emma Brown, University of Oxford</td>
</tr>
<tr>
<td>11:15</td>
<td>LT02</td>
<td>- The Role of Genomics in Cancer Diagnosis and Treatment - Dr. Michael Johnson, National Cancer Institute</td>
</tr>
<tr>
<td>12:00</td>
<td>LT02</td>
<td>- Lunch - Buffet Lunch</td>
</tr>
<tr>
<td>13:30</td>
<td>LT02</td>
<td>- The Impact of Artificial Intelligence in Cancer Research - Dr. Sarah Wilson, IBM Research</td>
</tr>
<tr>
<td>14:15</td>
<td>LT02</td>
<td>- The Future of Cancer Research: Challenges and Opportunities - Dr. David Taylor, Cancer Research UK</td>
</tr>
<tr>
<td>15:00</td>
<td>LT02</td>
<td>- Break - Tea Break</td>
</tr>
<tr>
<td>15:30</td>
<td>LT02</td>
<td>- The Importance of Public Health in Cancer Prevention - Dr. Emily Green, American Cancer Society</td>
</tr>
<tr>
<td>16:15</td>
<td>LT02</td>
<td>- The Economic Impact of Cancer - Dr. Robert Lee, World Health Organization</td>
</tr>
<tr>
<td>17:00</td>
<td>LT02</td>
<td>- Closing Remarks - Dr. Jane Doe, University of Cambridge</td>
</tr>
</tbody>
</table>
Preliminary Validation of the Lagrangian Support Vector Machine (LSVM) Learning Classifier as Clinical Decision-Making Support Tool to Aid Prediction of Prognosis in Patients with Hepatitis

Luca Parisi1, Marianne Lyne Manaog2
1Auckland Bioengineering Institute (ABI), University of Auckland
lucapa@ieee.org
2MedIntelligence, Auckland, New Zealand

Purpose

It is hoped to provide a preliminary validation of the Lagrangian Support Vector Machine (LSVM) learning classifier (Mangasarian and Musicant, 2000) for aiding clinical decision-making processes to predict prognosis of patients with hepatitis.

Methods

Data derived from blood tests of forty-six (46) male patients (age: 44 ±7 years) were selected (twenty-three survivors and twenty-three deceased) (“Hepatitis Data Set” of the University of California Irvine Repository of Machine Learning Databases, 1988; Gong, 1988). The following classifiers were deployed to predict the prognosis of patients with hepatitis:

- the supervised Multilayer Perceptron (MLP) and Radial Basis Functions (RBF) (The University of Edinburgh, School of Informatics, 1999),
- four-dimensional unsupervised self-organising maps (SOM) (Kohonen, 1982, 2001), and

The learning parameter and the momentum term were both set to 0.6 (Barton and Lee, 1995). Feature selection algorithms, i.e., ReliefF (Robnik-Šikonja and Kononenko, 2003), partial least square (PLS) and principal component analysis (PCA)-based (Malina and Yan, 2008), were used to quantify the relative importance of each of the data variables toward the patient outcome.

Results

The Artificial Intelligence (AI) classifiers’ performance were compared via the following test/out-of-sample classification accuracies expressed in percentage:

- MLP: 55.57 %,
- RBF: 63.21 %,
- SOM: 78.26 %, and
- LSVM: 84.78 %.
The presence of an enlarged liver, variceal bleeding, spider's blood, antiviral agents and a lower-than-normal level of blood albumin resulted to be amongst the nine most important factors toward the prediction of prognosis based on the feature selection algorithms adopted, i.e., ReliefF, PLS- and PCA-based.

Conclusions
The LSVM-based Classifier is deemed a potentially useful AI-based tool for aiding prediction of prognosis of patients with hepatitis. Future work is underway to provide AI-based metrics for monitoring the extent of patient outcome further to analgesic therapeutic strategies.
# AUSTRALIA

2 May

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:15 a.m.</td>
<td>Registration</td>
</tr>
<tr>
<td>8:30 a.m.</td>
<td>Introduction to MATLAB®</td>
</tr>
<tr>
<td></td>
<td>(Optional) session for people new to MATLAB</td>
</tr>
<tr>
<td></td>
<td>David Willingham, MathWorks</td>
</tr>
<tr>
<td>9:00 a.m.</td>
<td>Welcome and Opening</td>
</tr>
<tr>
<td></td>
<td>Wilco Ykelwerk, MathWorks</td>
</tr>
<tr>
<td>9:05 a.m.</td>
<td>Keynote: How to Build an Autonomous Anything</td>
</tr>
<tr>
<td></td>
<td>Mocha Kim, MathWorks</td>
</tr>
<tr>
<td>9:45 a.m.</td>
<td>What’s New in MATLAB® and Simulink® in R2017a</td>
</tr>
<tr>
<td></td>
<td>Daryl Ning, MathWorks</td>
</tr>
<tr>
<td>10:15 a.m.</td>
<td>Achieving Measurable Business Results Partnering with MathWorks: Practical Examples from Around the World</td>
</tr>
<tr>
<td></td>
<td>Branka Dijkstra, MathWorks</td>
</tr>
<tr>
<td>10:45 a.m.</td>
<td>Morning Tea, Partner Exhibition, and Networking</td>
</tr>
<tr>
<td>11:15 a.m.</td>
<td>A Machine learning-Based Speech Processing Solution for Facilitating Early Diagnosis of Parkinson’s Disease</td>
</tr>
<tr>
<td></td>
<td>Lucio Pavisi, Auckland Bioengineering Institute, The University of Auckland</td>
</tr>
<tr>
<td>11:45 a.m.</td>
<td>Developing and Deploying Analytics for IoT Systems</td>
</tr>
<tr>
<td></td>
<td>Daryl Ning, MathWorks</td>
</tr>
<tr>
<td>12:15 p.m.</td>
<td>Lunch, Partner Exhibition, and Networking</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Track A</th>
<th>Track B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:15 p.m.</td>
<td>Getting Business Insights with MATLAB and Big Data</td>
</tr>
<tr>
<td>David Willingham, MathWorks</td>
<td>Predictive Maintenance with MATLAB and Simulink</td>
</tr>
<tr>
<td></td>
<td>Predictive Maintenance with MATLAB and Simulink</td>
</tr>
<tr>
<td></td>
<td>Daryl Ning, MathWorks</td>
</tr>
<tr>
<td>1:45 p.m.</td>
<td>Integrating MATLAB Analytics into Enterprise Applications</td>
</tr>
<tr>
<td>David Willingham, MathWorks</td>
<td>Simulink as Your Enterprise Simulation Platform</td>
</tr>
<tr>
<td></td>
<td>Simulink as Your Enterprise Simulation Platform</td>
</tr>
<tr>
<td></td>
<td>Daryl Ning, MathWorks</td>
</tr>
<tr>
<td>2:15 p.m.</td>
<td>Afternoon Tea, Partner Exhibition, and Networking</td>
</tr>
<tr>
<td>2:30 p.m.</td>
<td>Developing Deep Learning Algorithms Using MATLAB</td>
</tr>
<tr>
<td>David Willingham, MathWorks</td>
<td>Using Computational Thinking to Foster Learning Curiosity</td>
</tr>
<tr>
<td></td>
<td>Using Computational Thinking to Foster Learning Curiosity</td>
</tr>
<tr>
<td></td>
<td>Bradley Norton, MathWorks</td>
</tr>
<tr>
<td>3:00 p.m.</td>
<td>End</td>
</tr>
</tbody>
</table>
A Minimum Viable Machine Learning-based Speech Processing Solution for Facilitating Early Diagnosis of Parkinson’s Disease
Luca Parisi\textsuperscript{1}, Marianne Lyne Manaog\textsuperscript{2}
\textsuperscript{1}Auckland Bioengineering Institute (ABI), University of Auckland
luca.parisi@eese.org
\textsuperscript{2}MedIntelligence\textsuperscript{®}, Auckland, New Zealand

Recent advances in speech processing have facilitated the assessment of prognosis of patients with Parkinson’s Disease (PD); however, there is currently no objective method to provide an early diagnosis of PD. The clinical potential of the following Artificial Intelligence classifiers was assessed: a four-dimensional self-organising map (SOM), a Lagrangian Support Vector Machine (LSVM) and a multi-layer perceptron (MLP) with ten sigmoid hidden and Softmax output neurons. Dysphonia measures and the “Unified Parkinson’s Disease Rating Scale” (UPDRS) were obtained from the University of California-Irvine (UCI) database. Deploying the ReliefF, the partial least square- and the MLP-based feature selection algorithms, UPDRS, the mean pitch, the standard deviation of the period were ranked as the three most important factors towards the diagnosis of PD. The MATLAB toolboxes “rftool” and “rptool” were used for fitting and pattern recognition purposes respectively, also enabling a clear visualisation of the classification outcomes. Findings indicate test classification accuracies of 84.52\%, 96.27\% and 100\% for the SOM, LSVM and MLP respectively. Having negligible cross-entropy and mean squared errors (0.039 and $1.91 \times 10^{-14}$ respectively) and the maximum area under the receiving operating characteristic (ROC) curve, the MLP is deemed a minimum viable solution for aiding early diagnosis of PD.
A Minimum Viable Machine Learning-based Speech Processing Solution for Facilitating Early Diagnosis of Parkinson’s Disease

Luca Parisi, Auckland Bioengineering Institute (ABI), The University of Auckland
Marianne Lyne Manaog, MedIntellego®, Auckland, New Zealand

Presentation Outline

- Problem Definition and Motivation
- The Disease of Interest: Parkinson’s Disease
- High-level Objectives
- Value Proposition – A Minimum Viable AI Solution
- Findings
- Discussion
- Conclusion

![A Scheme of a Multi-Layer Perceptron (MLP)-based learning classifier](image)

Problem Definition and Motivation

motor problems → PD leads to untreatable motor symptoms [1]

Parkinson’s Disease (PD) → no proven means of slowing progression

loss of postural reflexes → dopamine replacement therapy leads to frequent complications [2]

There is no objective method for early diagnosis of PD.

Main Features to Diagnose PD

Methods:
ReliefF-, partial least square-, Multi-layer Perceptron-based algorithms.

Results:
Three main diagnostic features: UPDRS (main clinical score of PD), mean pitch, standard deviation of the period.
Methodology

Artificial Intelligence-based classifiers:

- self-organising maps,
- Lagrangian Support Vector Machine and
- multi-layer perceptron.

Artificial Neural Networks to Aid Diagnosis of PD

Multi-layer Perceptron (MLP)
Machine Learning (ML) to Aid Diagnosis of PD

Adapted from: Cross Validated, 2014

![Diagram of Machine Learning]

Lagrangian Support Vector Machine

Adapted from: Merguander and Mucicand, 2001

![Diagram of Lagrangian Support Vector Machine]

Overly Sophisticated Attempts of Diagnosing PD

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Journal</th>
<th>PWP</th>
<th>Controls</th>
<th>Algorithms</th>
<th>Accuracy</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanharani et al.</td>
<td>2014</td>
<td>Comput Methods</td>
<td>23/48</td>
<td>8/20</td>
<td>One FP, four FSs, three LCs</td>
<td>100%</td>
<td>Five PMs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Programs Biomed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alam and</td>
<td>2014</td>
<td>J Am Stat</td>
<td>48/48</td>
<td>29/20</td>
<td>Two LCs</td>
<td>93.30%</td>
<td>Two PMs</td>
</tr>
<tr>
<td>Almazaydeh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2014</td>
<td>PLoS ONE</td>
<td>23/48</td>
<td>8/20</td>
<td>Two FSs, three LCs</td>
<td>91.80%</td>
<td>Three PMs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behrooz and 2016</td>
<td>2016</td>
<td>Int J Telemed</td>
<td>20/48</td>
<td>20/20</td>
<td>Two FSs, four LCs</td>
<td>87.50%</td>
<td>Three PMs</td>
</tr>
<tr>
<td>Sami</td>
<td></td>
<td>Appl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al.</td>
<td>2016</td>
<td>BioMed Eng Online</td>
<td>48/48</td>
<td>29/20</td>
<td>Two FIs, two LCs</td>
<td>87.80%</td>
<td>Two PMs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FP = feature pre-processing; FS = feature selection; LC = learning classifier; PM = performance measure
The importance of Deploying MATLAB

MATLAB toolboxes “nftool” and “nprtool” → Reduced development time and easy-to-use visualisation tools.

Results

- Classification accuracies: 84.52% for self-organising maps, 98.27% for Lagrangian Support Vector Machine and 100% for Multi-layer Perceptron.
- Mean squared error (MSE): $1.91 \times 10^{-14}$.
- Area under the Receiving Operating Characteristic curve (AUC): 1.

Conclusion

- Assessment of clinical potential of Artificial Intelligence-based classifiers;
- Feature selection of parameters for early diagnosis of PD;
- Simple method for early diagnosis of PD: Multi-layer perceptron;
- Applications in a clinical setting.

Acknowledgements

Thank you

Any questions