

FACULTY OF ENGINEERING TECHNOLOGY  
Dynamical Systems, Signal Processing and Data Analytics (STADIUS)  
Group T Leuven Campus  
Andreas Vesaliusstraat 13 - box 2600  
3000 Leuven  
tel. + 32 16 30 10 30  
emedia@kuleuven.be



Ahmed Youssef Ali

**KU LEUVEN**

ARENBERG DOCTORAL SCHOOL  
FACULTY OF ENGINEERING TECHNOLOGY

# Machine learning for Human Health: From Activity and Vital-Sign Monitoring to Mortality Prediction

MACHINE LEARNING FOR HUMAN HEALTH: FROM ACTIVITY  
AND VITAL-SIGN MONITORING TO MORTALITY PREDICTION

Ahmed Youssef Ali

Dissertation presented in partial  
fulfilment of the requirements for the  
degree of Doctor of Engineering  
Technology

May 2021

May 2021

# **Machine learning for Human Health: From Activity and Vital-Sign Monitoring to Mortality Prediction**

**Ahmed Youssef Ali**

**Supervisors:**

Prof. dr. ir. Bart Vanrumste (Promotor)

Prof. dr. ir. Jean-Marie Aerts (Co-Promotor)

Prof. dr. Stijn Luca (Co-Promotor)

**Doctoral committee:**

Prof. dr. ir. Johan Suykens

Prof. dr. md. Pieter Vandervoort

Prof. dr. ir. Carolina Varon

Dr. ir. Roel Smolders

**Chairman:**

Prof. dr. ir. Johan Claes

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Engineering Technology (PhD)

May 2021

© 2021 KU Leuven – Faculty of Engineering Technology  
Uitgegeven in eigen beheer, Ahmed Youssef Ali, Andreas Vesaliusstraat 13, B-3000 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaandelijke schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

**KU LEUVEN**



# Machine learning for Human Health: From Activity and Vital-Sign Monitoring to Mortality Prediction

Dissertation presented to obtain  
the degree of Doctor Engineering  
Technology

by

**Ahmed Youssef Ali**

---

May 2021

Daar de proefschriften in de reeks van de faculteit industriële ingenieurswetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

# Committee

## **Supervisors:**

Prof. dr. ir. Bart Vanrumste (Promotor)	KU Leuven
Prof. dr. ir. Jean-Marie Aerts (Co-Promotor)	KU Leuven
Prof. dr. Stijn Luca (Co-Promotor)	Ghent University

## **Doctoral committee:**

Prof. dr. ir. Johan Suykens	KU Leuven
Prof. dr. md. Pieter Vandervoort	Hasselt University
Prof. dr. ir. Carolina Varon	KU Leuven
Dr. ir. Roel Smolders	KU Leuven

## **Chairman:**

Prof. dr. ir. Johan Claes	KU Leuven
---------------------------	-----------



# Acknowledgements

Acknowledgement has more than one meaning: the first is accepting the truth or existence of something, the second meaning is the common one regarding the start of a book. I will not stick to the second meaning only, but I will start with the first one.

Meaning 1: I accept the truth that everything that happened to me in my life, and more particularly during my academic and PhD life, is a perfectly well-designed fate (divine fore-ordainment) to whom I am eternally grateful, Alhamdulillah. Starting with the way I came to Leuven, obtaining my degree of AI, PhD project acceptance, starting and progressing in my PhD.

Meaning 2: I would love to show my sincere gratitude to everyone who sincerely contributed to this achievement. Firstly, thanks to my supervisors (professors Bart Vanrumste, Jean-Marie Aerts and Stijn Luca), with whom I enjoyed my PhD, as it was not only an educational and research experience but also an exciting and pleasant period in my life. I will always remember the amount of freedom and appreciation that I have received from them during my PhD. Thanks to my supervisory committee (Professors Johan Suykens and Pieter Vandervoort), I have learned a lot from their feedback and constructive comments. Thanks to my thesis assessors (Prof. Carolina Varon and Dr Roel Smolders) for their time, effort, and interest in assessing my thesis. Thanks to my teachers, directors and professors in Egypt, especially MG Prof. Khairy Elbarbary and Eng. Mustafa Sharaf, may they rest in peace. Thanks to my KU Leuven professors, I have learned a lot from my masters, namely professors Johan Suykens, Sabine Van Huffel and Marc Van Hulle. Thanks to all my colleagues, former and current, one by one with whom I spent a wonderful time during my PhD in my two labs, eMedia (Yiyuan, Ine, Hannelore, Kymeng, Bert, Benjamin, Karsten, Dimitri, Chetanya, Esmaeel, Mulu, and Jonas) and M3BIORES (Kathleen, Vere, Sara, Jasper, 2x Michiel, Joren, Ioana, Deborah, Meiqing, and the best office mate, Alberto). Special thanks to Inge Vanpoecke, who handled my administrative issues since day one in the most efficient way. Special thanks to dear my friends, especially Ahmed Abdelhamid and Mohammad Khodery. Special thanks to my sister in law, Monia, for her caring and sincere support since I came to Belgium. Thanks to my sister in law, Ghada, for her care and support in tough times.

Thanks to my big brother, Mohamed, who supports and guides me in all possible means, who inspired me to be an Engineer. Thanks to my brother, teacher, mentor, advisor, friend and the Pierian spring, aka Ali, to whom I am eternally grateful.

Thanks to my parents, my mother taught me to be a good man (I hope she isn't disappointed);  
My Father, a man of chivalry and honour, may he rest in peace.

# Summary

In this dissertation, we aim at handling the problem of uncertainty in medicine and health care. This problem is at the core of clinical practice as clinicians have to make decisions on a daily basis for the good of their patients. Uncertainty in medicine and healthcare has several taxonomies, one of them classify it into three categories: personal, conceptual and technical uncertainty. In this thesis, we focus on technical uncertainty which is data related. Therefore, technical uncertainty is handled as a data analysis problem in which analytical methods (e.g. machine learning) can be used. From a clinical perspective, technical uncertainty comprises three elements: probability, ambiguity and complexity. Probability element reflects the risk and imprecision about future outcomes of a specific decision or action from a temporal perspective. For ambiguity, it reflects the imprecise predictions, conflicting information and lack of evidence. Finally, complexity reflects the multiplicity of influential factors.

A possible solution to technical uncertainty in medicine is developing reliable clinical decision support systems (CDSS). The CDSS covers several categories; the most relevant one to our study is the diagnostic decision support system (DDSS). The main components of DDSS are medical data and analytical methods. Medical data can be collected using either conventional or wearable medical devices. Each option has its advantage; for example, conventional medical devices still provide the most accurate and gold-standard measurements. On the other hand, wearable medical devices provide continuous monitoring and ease of installation. For analytical methods, there is a broad spectrum of methods that can be used in the context of human-health applications. Our first candidate amongst these methods is machine learning given its high computational capabilities and efficient error performance. In this thesis, we use simple but powerful machine learning approaches. One approach is intended to be used for online and streaming analytics with wearable devices which is the localised learning approach of  $k$  nearest neighbours least squares support vector machine ( $k$ NN-LS-SVM). The other method is a linear hard margin approach to support vector machines which is used to engineer explainable features.

Developing a reliable decision support system faces many challenges; these challenges are data and model-based. In this thesis, we focus on some of these challenges: class-imbalance, non-uniform data distribution, ambiguity (confusion), continuously increasing data-size, model personalisation, black-box nature, online and streaming analytics. These challenges are linked to the aforementioned elements of technical certainty. Therefore, handling these challenges can

restrain the elements of technical uncertainty.

This dissertation's main objective is to investigate the possibility to limit and reduce the technical uncertainty of applying machine learning approaches in human health applications. Reducing and restricting this uncertainty can be achieved by developing machine learning algorithms that can enhance the reliability of human health applications' decision-making process.

Five human health applications are investigated in this dissertation: human activity recognition, thermal comfort prediction, vital signs prediction, intensive care unit (ICU) mortality prediction, and COVID-like patient mortality prediction at ICU. These applications cover the aforementioned challenges. The human activity recognition (HAR) application focuses on recognising daily life activities and postures based on accelerometer data. The application of thermal comfort prediction focuses on developing a machine learning model to predict the thermal comfort level based on a set of vital signs that can be easily measured using wearable sensors. The vital signs prediction application focuses on estimating the monitored vital signs' early warning score with a relatively high rate (every minute). Moreover, in this application, we develop a machine learning predictive model to predict the monitored vital signs for the upcoming 1, 2, and 3 hours ahead. For these three applications, all measurements are obtained using wearable sensors and the used machine learning algorithm is kNN-LS-SVM for classification and regression. Moreover, these three applications cover the challenges of class-imbalance, non-uniform data distribution, ambiguity, increasing data-size, model personalisation, online and streaming analytics.

The fourth application of ICU mortality prediction focuses on engineering explainable features extracted from the monitored vital signs during patients' ICU stay. These engineered features are meant to provide clinical insight to medical staff. In this application, the used machine learning algorithm is a linear hard margin approach to SVM's which used to control the complexity of the model and assure a linear separation of the data points in the input space. The last application focuses on ICU mortality prediction of a specific profile of patients compared to the previous application which investigates multiple profiles together. The fifth application's profile of patients is pulmonary disease patients with infection and/or pneumonia, which experiences similar symptoms of COVID-19 patients. Both fourth and fifth applications are based on ICU data obtained using conventional ICU devices. Moreover, both applications cover the challenge of black-box nature and the resulting lack of interpretability and explainability.

# Samenvatting

In dit proefschrift willen we het probleem van onzekerheid in de geneeskunde en de gezondheidszorg aanpakken. Dit probleem staat centraal in de klinische praktijk omdat klinici dagelijks beslissingen moeten nemen voor het welzijn van hun patiënten. Onzekerheid in de geneeskunde en de gezondheidszorg kent verschillende taxonomieën, waarvan er één de onzekerheid in drie categorieën indeelt: persoonlijke, conceptuele en technische onzekerheid. In dit proefschrift richten we ons op technische onzekerheid, die gerelateerd is aan gegevens. Daarom wordt technische onzekerheid behandeld als een data-analyse probleem waarbij analytische methoden (bijv. machine learning) kunnen worden gebruikt. Vanuit een klinisch perspectief bestaat technische onzekerheid uit drie elementen: waarschijnlijkheid, ambiguïteit en complexiteit. Het element waarschijnlijkheid weerspiegelt het risico en de onzekerheid over de toekomstige uitkomsten van een specifieke beslissing of handeling vanuit een temporele invalshoek. Meerduidigheid weerspiegelt de onnauwkeurige voorspellingen, tegenstrijdige informatie en gebrek aan bewijs. Complexiteit, ten slotte, weerspiegelt de veelheid van invloedrijke factoren.

Een mogelijke oplossing voor de technische onzekerheid in de geneeskunde is de ontwikkeling van betrouwbare klinische beslissingsondersteunende systemen (CDSS). CDSS omvat verschillende categorieën; de meest relevante voor onze studie is het diagnostic decision support system (DDSS). De belangrijkste onderdelen van DDSS zijn medische gegevens en analysemethoden. Medische gegevens kunnen worden verzameld met behulp van conventionele of draagbare medische apparatuur. Elke optie heeft zijn voordeel; zo levert conventionele medische apparatuur nog steeds de nauwkeurigste en meest gouden standaardmetingen op. Draagbare medische apparatuur daarentegen biedt continue monitoring en is gemakkelijk te installeren. Wat de analysemethoden betreft, is er een breed spectrum van methoden die kunnen worden gebruikt in het kader van toepassingen op het gebied van de menselijke gezondheid. Onze eerste kandidaat onder deze methoden is machinaal leren, gezien de hoge rekencapaciteiten en efficiënte foutprestaties. In dit proefschrift gebruiken we eenvoudige maar krachtige machine learning benaderingen. Eén benadering is bedoeld voor gebruik voor online en streaming analytics met draagbare apparaten en is de gelokaliseerde leerbenadering van  $k$  nearest neighbours least squares support vector machine (kNN-LS-SVM). De andere methode is een lineaire harde marge-benadering van support vector machines die wordt gebruikt om verklaarbare kenmerken te ontwikkelen.

Het ontwikkelen van een betrouwbaar beslissingsondersteunend systeem kent vele uitdagingen; deze uitdagingen zijn gebaseerd op gegevens en op modellen. In deze dissertatie richten we ons op enkele van deze uitdagingen: klassenonevenwichtigheid, niet-uniforme gegevensverdeling, ambiguïteit (verwarring), voortdurend toenemende gegevensgrootte, modelpersonalisatie, black-box karakter, online en streaming analytics. Deze uitdagingen houden verband met de hierboven genoemde elementen van technische zekerheid. Daarom kan het aanpakken van deze uitdagingen de elementen van technische onzekerheid beteugelen.

Het hoofddoel van dit proefschrift is het onderzoeken van de mogelijkheid om de technische onzekerheid van het toepassen van machine learning benaderingen in menselijke gezondheidstoepassingen te beperken en te verminderen. Het verminderen en inperken van deze onzekerheid kan worden bereikt door het ontwikkelen van machine learning algoritmen die de betrouwbaarheid van het besluitvormingsproces van humane gezondheidstoepassingen kunnen verhogen.

Vijf menselijke gezondheidstoepassingen worden onderzocht in dit proefschrift: herkenning van menselijke activiteit, voorspelling van thermisch comfort, voorspelling van vitale functies, voorspelling van sterfte op de intensive care unit (ICU), en voorspelling van sterfte van COVID-achtige patiënten op de ICU. Deze toepassingen bestrijken de bovengenoemde uitdagingen. De toepassing van menselijke activiteitsherkenning (HAR) richt zich op het herkennen van dagelijkse levensactiviteiten en houdingen op basis van versnellingsmetergegevens. De toepassing voor het voorspellen van thermisch comfort richt zich op het ontwikkelen van een machine-learning model om het thermisch comfortniveau te voorspellen op basis van een set vitale functies die eenvoudig gemeten kunnen worden met draagbare sensoren. De toepassing voor de voorspelling van vitale functies richt zich op het schatten van de vroegtijdige waarschuwingsscore van de gecontroleerde vitale functies met een relatief hoge snelheid (elke minuut). Bovendien ontwikkelen we in deze toepassing een machine learning voorspellingsmodel om de bewaakte vitale functies voor de komende 1, 2 en 3 uur te voorspellen. Voor deze drie toepassingen worden alle metingen verkregen met behulp van draagbare sensoren en het gebruikte machine-learning algoritme is kNN-LS-SVM voor classificatie en regressie. Bovendien hebben deze drie toepassingen betrekking op de uitdagingen van klasse-onbalans, niet-uniforme gegevensverdeling, ambiguïteit, toenemende gegevensgrootte, modelpersonalisatie, online en streaming analytics.

De vierde toepassing van ICU-sterftevoorspelling richt zich op de engineering van verklaarbare kenmerken uit de bewaakte vitale functies tijdens het ICU-verblijf van patiënten. Deze eigenschappen zijn bedoeld om klinisch inzicht te verschaffen aan medisch personeel. In deze toepassing is het gebruikte machine-learning algoritme een lineaire harde marge benadering van SVM's die gebruikt wordt om de complexiteit van het model te controleren en een lineaire scheiding van de gegevens te verzekeren.

# List of Abbreviations

<b>AI</b>	Artificial Intelligence
<b>AKI</b>	Acute Kidney Injury
<b>AOGs</b>	And-or-Graphs
<b>ARD</b>	automatic relevance determination
<b>ARDS</b>	Acute Respiratory Distress Syndrome
<b>AUC</b>	Area Under Curve
<b>BP</b>	Blood Pressure
<b>BPM</b>	Beat Per Minute
<b>BPM</b>	Breath Per Minute
<b>CDSS</b>	Clinical Decision Support System
<b>CNN</b>	Convolutional Neural Networks
<b>COVID</b>	Corona Virus Disease
<b>CPOE</b>	Computerised Provider Order Entry
<b>CPR</b>	Cardiopulmonary Resuscitation
<b>CRFs</b>	Conditional Random Fields
<b>DAE</b>	Deep Autoencoder
<b>DBM</b>	Data-Based Mechanistic
<b>DBP</b>	Diastolic Blood Pressure
<b>DDSS</b>	Diagnostic Decision Support System

<b>DNR</b>	Do-Not-Resuscitate
<b>EMR</b>	Electronic Medical Record
<b>ECG</b>	Electrocardiogram
<b>EEG</b>	Electroencephalogram
<b>EMG</b>	Electromyography
<b>EVS</b>	Excercise Vital Sign
<b>EWS</b>	Early Warning Score
<b>FN</b>	False Negative
<b>FoG</b>	Freezing of Gaits
<b>FP</b>	False Positive
<b>FT</b>	Fine Tuning
<b>HAR</b>	Human Activity Recognition
<b>HBNs</b>	Hierarchical Bayesian Networks
<b>HMM</b>	Hidden Markov Model
<b>HR</b>	Heart Rate
<b>IBL</b>	Instance-Based Learning
<b>ICU</b>	Intensive Care Unit
<b>kNN</b>	K Nearest Neighbours
<b>LOO</b>	Leave-One-Out
<b>LOSO</b>	Leave-One-Subject-Out
<b>LS</b>	Least Squares
<b>LSTM</b>	Long-Short Term Memory
<b>LSVM</b>	Localised Support Vector Machines
<b>MAE</b>	Mean Absolute Error
<b>MAPE</b>	Mean Absolute Percentage Error

<b>MLN</b>	Markov Logic Network
<b>MLP</b>	Multi Layer Perceptron
<b>MMD</b>	Maximum Mean Discrepancy
<b>MSE</b>	Mean Squared Error
<b>NEWS</b>	National Early Warning Score
<b>PA</b>	Physical Activity
<b>PAVS</b>	Physical Activity Vital Sign
<b>PMV</b>	Predicted-Mean-Vote
<b>PP</b>	Pulse Pressure
<b>PPG</b>	Photoplethysmograph
<b>PTT</b>	Pulse Transit Time
<b>PSVM</b>	Profile Support Vector Machine
<b>RBF</b>	Radial Basis Function
<b>RR</b>	Respiration Rate
<b>SBP</b>	Systolic Blood Pressure
<b>SNAP</b>	Speedy Nutrition and Physical Activity Assessment
<b>SpO<sub>2</sub></b>	Peripheral Oxygen Saturation
<b>SRL</b>	Statistical Relation Learning
<b>SVMs</b>	Support Vector Machines
<b>T</b>	Temperature
<b>TC</b>	Thermal Comfort
<b>TN</b>	True Negative
<b>TP</b>	True Positive
<b>TS</b>	Thermal Sensation
<b>TSV</b>	Thermal Sensation Vote
<b>XAI</b>	Explainable Artificial Intelligence



# Contents

Doctoral committee	i
Acknowledgements	iii
Abstract	v
Samenvatting	vii
List of Abbreviations	ix
List of Figures	xvii
List of Tables	xxiii
<b>1 General introduction</b>	<b>I</b>
1.1 Uncertainty in Medicine and Healthcare	2
1.1.1 Sources of Uncertainty	4
1.1.2 Technical Uncertainty	4
1.2 Data-Based Clinical Decision Support System	6
1.3 Decision Support Components	7
1.3.1 Medical Data Acquisition	7
1.3.2 Analytics	8
1.3.3 Data and Model-based Challenges	9
1.3.4 Machine Learning in Medicine	13
1.4 Objective	19
1.4.1 Sub-objectives	20
1.4.2 Research Questions	20
1.5 Human Health Applications	21
<b>2 A Localised Learning Approach For Human Activity Recognition</b>	<b>29</b>
2.1 Introduction	29
2.2 Related Work	31

2.2.1	State-of-the-art	31
2.2.2	Benchmark Study	32
2.3	Local learning of SVMs	32
2.3.1	Support vector machines	32
2.3.2	localised LS-SVMs and KNN-LS-SVMs	34
2.4	Simulation experiments	36
2.4.1	Class-imbalance	37
2.4.2	Ambiguity	41
2.5	Real-world Data	43
2.5.1	datasets	43
2.5.2	Classification performance	44
2.5.3	Time Performance	49
2.6	Discussion	49
2.7	Conclusion	50
<b>3</b>	<b>Towards Online Personalised-Monitoring of Thermal Sensation</b>	<b>59</b>
3.1	Introduction	60
3.2	Methods	62
3.2.1	Data Processing and Classification	62
3.2.2	Experiments and Experimental Setup	65
3.3	Results and Discussion	69
3.3.1	General Classification Models	69
3.3.2	Class Reduction	73
3.3.3	Personalised Classification Models	75
3.3.4	Streaming Algorithm Approach for Personalised Thermal Sensation Monitoring	77
3.4	Discussion	79
3.5	Conclusions	81
<b>4</b>	<b>Vital Signs Prediction Based on Continuous Monitoring</b>	<b>87</b>
4.1	Introduction	88
4.2	Data Generation	89
4.2.1	Study Design	89
4.2.2	Study Population	90
4.2.3	Measuring Device	90
4.3	Methods	90
4.3.1	High-Rate EWS Computation	90
4.3.2	Vital Signs Time-series Prediction	92
4.3.3	Local learning of SVMs	92
4.4	Results	97

4.4.1	High-Rate EWS Computation . . . . .	97
4.4.2	Vital Signs Time-series Prediction . . . . .	97
4.5	Discussion . . . . .	105
4.6	Conclusions . . . . .	108
<b>5</b>	<b>Feature Engineering for ICU Mortality Prediction</b>	<b>115</b>
5.1	Introduction . . . . .	116
5.2	Materials and Methods . . . . .	118
5.2.1	Data . . . . .	118
5.2.2	Hard-Margin SVM . . . . .	119
5.3	Feature-Engineering . . . . .	122
5.3.1	Evaluation . . . . .	122
5.3.2	Feature Extraction . . . . .	122
5.3.3	Feature Fine-Tuning . . . . .	126
5.4	Results . . . . .	128
5.5	Discussion . . . . .	130
5.6	Conclusions . . . . .	133
<b>6</b>	<b>Mortality Prediction for COVID-Like Patients at the ICU</b>	<b>139</b>
6.1	Introduction . . . . .	139
6.2	Data . . . . .	141
6.3	Methods . . . . .	142
6.3.1	Hard Margin SVM . . . . .	142
6.4	Feature Engineering . . . . .	144
6.4.1	Feature Extraction . . . . .	144
6.4.2	Feature Fine-Tuning . . . . .	146
6.5	Results . . . . .	148
6.6	Discussion . . . . .	150
6.7	Conclusions . . . . .	152
<b>7</b>	<b>General Discussion</b>	<b>155</b>
7.1	Question N <sup>o</sup> 1 . . . . .	156
7.1.1	Data-related Challenges . . . . .	156
7.1.2	Model-related Challenges . . . . .	160
7.2	Question N <sup>o</sup> 2 . . . . .	162
7.2.1	Model Interpretability and Explainability . . . . .	162
7.2.2	Error Performance and Model Interpretation . . . . .	164
7.3	Limitations . . . . .	164
7.3.1	Extra Hyperparameter . . . . .	165
7.3.2	Data-size . . . . .	165

7.3.3	Missing a Gold Standard	166
7.3.4	Lack of Data	166
7.4	Valorization	166
7.4.1	Human activity recognition	166
7.4.2	For Thermal Comfort Prediction	167
7.4.3	For Vital-Signs and Physical Activity	168
7.4.4	WearIT4health	169
8	<b>Conclusions and Future Work</b>	<b>175</b>

# List of Figures

I.1	Photograph of an original scroll from The ancient Egyptian papyrus <i>Edwin Smith Papyrus</i> , It is written right-to-left in Hieratic (the Egyptian cursive form of hieroglyphs). Rare Book Room, New York Academy of Medicine.[5] . . . . .	2
I.2	a) Research Map indicating the problem, solution, and applications. b) Data and model-related challenges from the clinical and data analysis perspectives and corresponding applications. . . . .	3
I.3	Three aspects of technical uncertainty: probability, ambiguity and complexity for the response to breast cancer treatment (from [14]). . . . .	5
I.4	Solution block from Figure 1.2a . . . . .	6
I.5	Solution components block from Figure 1.2a . . . . .	7
I.6	Solution material from Figure 1.2a . . . . .	7
I.7	Data and model-based challenges block from Figure 1.2b . . . . .	10
I.8	Two-dimensional binary classification problem with class imbalance ratio 1/9 between classes 1 and 2 respectively. . . . .	11
I.9	Two dimensional binary classification problem with one test point ( $\mathbf{x}$ ) experiencing ambiguity. Adapted from [51]. . . . .	11
I.10	Solution methods block from Figure 1.2a . . . . .	13
I.11	Two-dimensional nonlinear binary classification problem. a) The classification outcome of the global classifier of LS-SVM, b) The classification outcome of the localised classifier of kNN-LS-SVM. . . . .	15
I.12	Two-dimensional nonlinear binary classification problem with two new data points <b>a</b> and <b>b</b> . . . . .	16
I.13	Performance vs explainability of different learning algorithms (e.g. and-or-graphs (AOGs), support vector machines (SVMs), statistical relation learning (SRL), hierarchical Bayesian networks (HBNs), conditional random fields (CRFs), and Markov logic networks (MLNs))(from [67]). . . . .	18
I.14	Schematic representation of a two dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised. . . . .	19
I.15	Applications block from Figure 1.2a . . . . .	21

2.1	Schematic representation of a two dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised. . . . .	33
2.2	A flow chart illustrating the localised learning algorithm of KNN-LS-SVM. .	36
2.3	Two-dimensional datasets consisting of two classes with data generated from Gaussian distributions $X_+ \sim N(\mathbf{m}_+, \Sigma_+)$ and $X_- \sim N(\mathbf{m}_-, \Sigma_-)$ with $\mathbf{m}_+ = (\frac{1}{2}, \frac{1}{2})$ , $\mathbf{m}_- = (-\frac{1}{2}, -\frac{1}{2})$ and different covariance matrices: (a) $\Sigma_+ = \Sigma_- = 0.35I_2$ , (b) $\Sigma_+ = 2\Sigma_- = 0.70I_2$ and (c) $\Sigma_+ = \frac{1}{2}\Sigma_- = 0.35I_2$ . . . . .	38
2.4	(a) Averaged $F_1$ -scores of a KNN-LS-SVM, KNN, and an LS-SVM obtained from a 10-fold cross-validation experiment using the synthetic dataset shown in Figure 2.3a and using different percentages $p_+$ of instances from the positive class. (b) $F_1$ -scores of KNN-LS-SVM, KNN, and LS-SVM at different degrees of imbalance when $\Sigma_+ = 2\Sigma_- = 0.70I_2$ . (c) $F_1$ -scores of KNN-LS-SVM, KNN, and LS-SVM at different degrees of imbalance when $\Sigma_+ = \frac{1}{2}\Sigma_- = 0.35I_2$ and averaged over the runs of a 10-fold cross-validation experiment. The labels (*) indicated on top of the horizontal axis refer to the imbalance percentages where the difference in performance scores between KNN-LS-SVM and LS-SVM is statistically significant at the 0.05 level. . . . .	40
2.5	(a) Two-dimensional non-linearly separable dataset of two classes where $p_+ = 50\%$ . (b) $F_1$ -scores of the KNN-LS-SVM, KNN, and LS-SVM averaged over the runs of a 10-fold cross-validation test experiment. Performance scores are obtained by using the synthetic dataset describing confusion where the degree of imbalance is varied in the range {50%, 33%, 20%, 10%, 5%, 2.5%, 1.25%}. The labels (*) indicated on top of the horizontal axis refer to the imbalance percentages where the difference in performance scores between KNN-LS-SVM and LS-SVM is statistically significant at the 0.05 level. . . . .	42
2.6	$F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1 . . . . .	45
2.7	$F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Daphnet FoG . . . . .	46
2.8	$F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v2.0 . . . . .	47
2.9	$F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Skoda dataset . . . . .	47
3.1	A flow chart illustrating the algorithm of K-Nearest Neighbours Least Squares Support vector Machines (KNN-LS-SVM) classifier . . . . .	65
3.2	Photographic picture of the three climate-controlled rooms (from right to left, AC). . . . .	67

3.3	Sensor placement. (A) Ear channel for aural temperature measurement via the Cosinuss One, (B) upper arm where skin temperature and heat flux are measured with the gSKIN patch, (C) middle upper chest where skin temperature and heat flux are measured with the gSKIN patch, (D) lower chest where heart rate is measured with the Polar H7, (E) Scapula where skin temperature is measured with the shimmer, (F) mouth and nose where metabolic rate is measured via the MetaMAX-3B spiroergometer sensor. . . . .	68
3.4	Plots showing the climate chambers set-point temperatures programmed during the 55 min low activity phase (upper graph) and the 75 min high activity phase (lower graph). . . . .	69
3.5	Error performance of the developed general classification model with 54 input features for 7-classes classification problem. . . . .	71
3.6	Error performance of the developed general classification model with the selected 12 input features (reduced dimension input space) for 7-classes classification problem. . . . .	72
3.7	(a) A histogram of 7-class thermal sensation scale of ASHRAE system. (b) A histogram of 3-class thermal sensation of Configuration 1. (c) A histogram of 3-class thermal sensation of Configuration 2. (d) A histogram of 3-class of Configuration 3. . . . .	74
3.8	Schematic representation of the proposed streaming algorithm for online personalised thermal sensation monitoring. . . . .	77
3.9	Flowchart represents the main workflow of the streaming data acquisition and labeling process. . . . .	79
4.1	A scheme illustrating the allocation spots of SOMNOtouch-NIBP device and electrodes (adapted from <a href="http://www.somnomedics.eu">http://www.somnomedics.eu</a> ). . . . .	91
4.2	A flow chart illustrating the localised learning algorithm of KNN-LS-SVM for Regression. . . . .	95
4.3	The time-windows for feature extraction (red) comprised of two overlapped windows of 60 minutes each with an overlap of 50 minutes resulting in 70 minutes to predict the statistics of the target window (green) after N hours (e.g., 0,1 or 2 hours). . . . .	96
4.4	From the HR of a post-surgical patient: (a) The raw signal (BPM), (b) the denoised signal (BPM), (c) the maximum, (d) the mean, (e) the median, (f) the median based on moving median filter, and (g) the minimum values of the EWS component of HR for each one-minute segment. . . . .	98

4.5	Box-plots of the absolute errors of the Naive predictor at the upcoming hour of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming three hours (+1, +2, +3 hours) for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c) SpO <sub>2</sub> (%), (d) RR (BPM), and (e) PP (mmHg) for cardiology patients. . . . .	99
4.6	The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP, SpO <sub>2</sub> , RR, and PP for the upcoming one, two, and three hours (+1, +2, +3 hours) for cardiology patients. . . . .	99
4.7	The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Nieve predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO <sub>2</sub> , and (g,h) RR for cardiology patients. . . . .	100
4.8	Box-plots of the absolute errors of the Naive predictor at the upcoming hour of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming three hours (+1, +2, +3 hours) for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c) SpO <sub>2</sub> (%), (d) RR (BPM), and (e) PP (mmHg) for post-surgical patients. . . . .	101
4.9	The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP, SpO <sub>2</sub> , RR, and PP for the upcoming three hours (+1, +2, +3 hours) for post-surgical patients. . . . .	101
4.10	The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Nieve predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO <sub>2</sub> , and (g,h) RR for post-surgical patients. . . . .	102
4.11	Box-plots of the absolute errors of the Naive predictor of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming hour for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c) SpO <sub>2</sub> (%), (d) RR (BPM), and (e) PP (mmHg) for Dialysis patients. . . . .	103
4.12	The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP, SpO <sub>2</sub> , RR, and PP for the upcoming hour for dialysis patients. . . . .	103
4.13	The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Nieve predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO <sub>2</sub> , and (g,h) RR for Dialysis patients. . . . .	104
5.1	Distribution of the patient population and their reason for admission. The population was divided into age categories of 13-44, 45-64, 65-79 and >80 years of age. . . . .	119

5.2	Schematic representation of a two-dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised without tolerating any misclassifications (adapted from Reference [28]). . . . .	120
5.3	A flow chart illustrating the feature engineering methodology. . . . .	122
5.4	A flow chart illustrating the feature extraction process including the three feature categories (i.e. statistical, dynamic and physiological) and the sequence of the process marked by the evaluation steps. Also, in the process, the investigation is applied to the false negative patients only. . . . .	123
5.5	(a) Systolic blood pressure (BP), diastolic BP and pulse pressure (PP) of the last 150 observations (approximately the last nine days) of one false negative patient. The dashed window refers to the region where the systolic BP and diastolic BP measurements approach closely. (b) Systolic BP, diastolic BP and PP of the last 150 observations of another false negative patient. The mean value of the pulse pressure is 87.4 mmHg and median 88 mmHg. . . . .	125
5.6	(a) Respiration rate of a deceased patient with an obvious drop at specific observations below the normal range (12–20 BPM) (b) Oxygen saturation ( $SpO_2$ ) of another deceased patient that drops frequently (minimum 78% during the stay). . . . .	127
5.7	(a) $F_1$ -score, accuracy, sensitivity and precision of the classifier with all possible combinations of the three feature categories in addition to the fine-tuned combination. (b) The cumulative effect of the different fine-tuning stages on the classification accuracy, $F_1$ -score, precision, and sensitivity. WFT refers to 'without fine-tuning', FT $x$ refers to the $x$ th stage of fine-tuning as illustrated in the text. . . . .	130
6.1	Patients distribution based on the reason of hospitalisation . . . . .	142
6.2	Schematic representation of a two-dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised without tolerating any misclassifications (adapted from Reference [12]). . . . .	143
6.3	A scheme illustrates the last set of observations before discharge: time period for feature extraction (red) to predict mortality/survival at discharge (blue) given a specific prediction horizon (white). . . . .	145
6.4	Sensitivity and precision of models 1 and 2 for subpopulations I, II, and III based on 60 observations for feature extraction. . . . .	150
6.5	Sensitivity and precision of models 1 and 2 for subpopulations I, II, and III based on 30 observations for feature extraction. . . . .	150

6.6 Distribution of the patient population and their reason for admission. The population was divided into age categories of 13-44, 45-64, 65-79 and >80 years of age. . . . . 151

7.1 Distribution of the 25 participants' votes over their thermal sensation at the three environment temperatures of 24, 5, and 37° C. (from [4]) . . . . . 159

7.2 The wearable device ©WearIT4health to monitor hospitalised patients' vital signs in real time. . . . . 170

# List of Tables

2.1	Classification results of the KNN-LS-SVM, LS-SVM, <i>Ravi (2017)</i> , KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1. . . . .	44
2.2	Classification results of the KNN-LS-SVM, LS-SVM, <i>Ravi (2017)</i> , KNN, KNN-SVM, PSVM, and Stack-AE applied to Daphnet FoG. . . . .	45
2.3	Classification results of the KNN-LS-SVM, LS-SVM, <i>Ravi (2017)</i> , KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v2.0. . . . .	46
2.4	Classification results of the KNN-LS-SVM, LS-SVM, <i>Ravi (2017)</i> , KNN, KNN-SVM, PSVM, and Stack-AE applied to Skoda dataset. . . . .	48
2.5	The time performance of the classifiers KNN-LS-SVM, LS-SVM, KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1, Daphnet FoG, WISDM v2.0 and Skoda Datasets. The depicted results represent the consumed time in seconds to classify a single test point. . . . .	49
3.1	General physical information of the participants (test subjects). . . . .	66
3.2	Different temperature and relative humidity ranges that can be provided by the different Body and Mind (A, B, and C). . . . .	67
3.3	Overview of the 54 extracted features (CE = selected). . . . .	70
3.4	The normalized confusion matrix of Model I . . . . .	71
3.5	The normalized confusion matrix of Model II. . . . .	73
3.6	The error performance (precision, sensitivity, F1-score, and accuracy) of general LS-SVM model for the three different 3-class configurations. . . . .	75
3.7	The error performance (precision, sensitivity, F1-score, and accuracy) of the localized model KNN-LS-SVM model for the three different 3-class configurations based on LOSO testing approach. . . . .	76
3.8	The error performance (precision, sensitivity, F1-score, and accuracy) of the localized model KNN-LS-SVM model for the three different 3-class configurations based on LOO testing approach. (*) indicates the highest error performance value for each class in the different configurations. . . . .	76
4.1	Early Warning Scores system based on ZOL Hospital . . . . .	91

- 5.1 Feature Extraction results. . . . . 129
- 5.2 Feature Fine-tuning results. . . . . 129
  
- 6.1 Early Warning Scores system used by ZOL Hospital based on vital signs' values. 146
- 6.2 Three subpopulations with different combinations of pulmonary disease, infection, and pneumonia (M:mortality; S:survival). . . . . 148
- 6.3 The set of extracted features for each model/subpopulation combination. . . . 149
- 6.4 Precision, sensitivity and  $F_1$ -score of models 1 and 2 for subpopulations I, II, and III based on 60 observations for feature extraction. . . . . 149
- 6.5 Precision, sensitivity and  $F_1$ -score of models for subpopulations I, II, and III based on 30 observations for feature extraction. . . . . 149
  
- 7.1 The normalized confusion matrix of Model II. . . . . 158
- 7.2 Sensitivity and precision and  $F_1$ -score of recovery/survival for models 1 and 2 of subpopulations: Pulmonary+Infection (I), Pulmonary+Infection+Pneumonia (II), and Pulmonary+Pneumonia (III) based on 60 observations for feature extraction. . . . . 165

# Chapter 1

## General introduction

Decision-making is key in medical and clinical practice requiring adequate expertise from the medical specialists as well as relevant clinical information of the considered cases. Utilising analytical thinking in medicine appeared already since early ages as illustrated by the case where an ancient Egyptian physician drew inductive conclusions from observed facts while examining his patients [1]. In the ancient Egyptian papyrus (*Edwin Smith papyrus* Figure 1.1), an Egyptian physician documented a number of cases of bodily injuries. In some of these documented cases, the physician describes the steps of examination results, his diagnosis as well as suggested treatment. An example case showing these three steps, as quoted from the original English translation [2, 3] of the papyrus, is shown below:

1. *Examination: If thou examinest a man having a gaping wound in a vertebra of his neck, penetrating to the bone, (and) perforating a vertebra of his neck; if thou examinest that wound, (and) he shudders exceedingly, (and) he is unable to look at his two shoulders and his breast.*
2. *Diagnosis: Thou shouldst say concerning him: '(One having) a wound in his neck, penetrating to the bone, perforating a vertebra of his neck, (and) he suffers with stiffness in his neck. An ailment with which I will contend.*
3. *Treatment: Thou shouldst bind it with fresh meat the first day. Now afterward moor (him) at his mooring stakes until the period of his injury passes by.*

This papyrus is known as the oldest surgical text in history that contains the earliest known evidence of an inductive process in the history of the human mind [2]. It is estimated that this ancient Egyptian medical treatise is dating from c. 1600 BC [4].

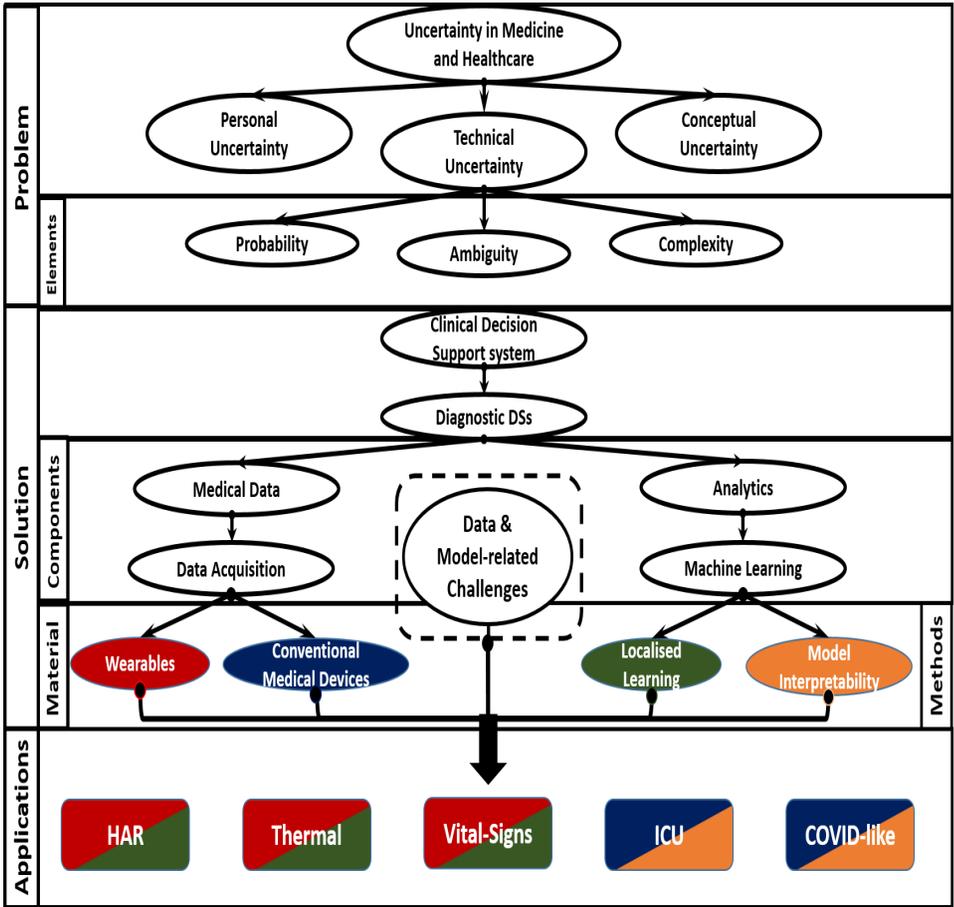


Figure 1.1: Photograph of an original scroll from The ancient Egyptian papyrus *Edwin Smith Papyrus*, It is written right-to-left in Hieratic (the Egyptian cursive form of hieroglyphs). Rare Book Room, New York Academy of Medicine.[5]

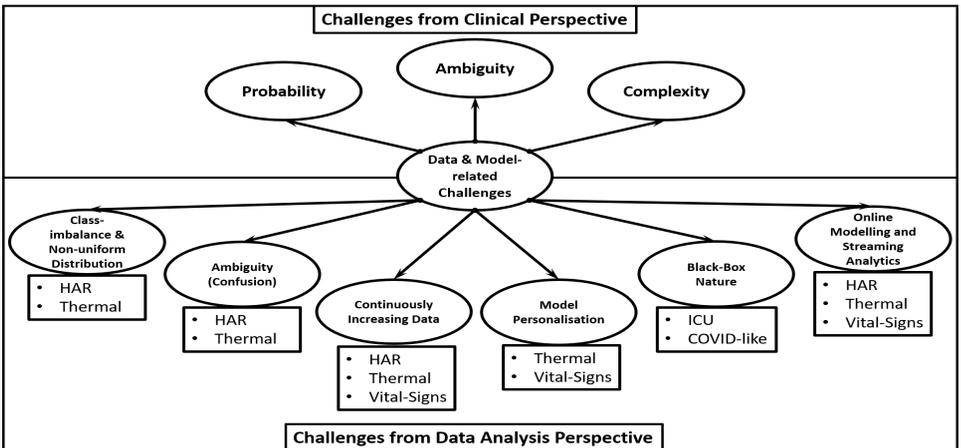
Nowadays, the (r)evolution in sensors and sensing-technology provides us with an enormous amount of medical data which can help to improve clinical decision-making. However, despite these huge evolutions since the era of the *Edwin Smith Papyrus*, healthcare outcomes today are still mostly probabilistic in nature resulting in uncertainty [6, 7, 8, 9, 10]. Restraining this medical and clinical uncertainty is needed since sound and accurate decision-making are crucial to clinical practice. In this thesis, as shown in Figure 1.2a, we investigate this problem (Problem block) by decomposing it into its elements and proposing potential solutions (Solution block) for these different elements. For contributing to this solution, we define possible challenges (Figure 1.2b) related to the different components and assign suitable methods to handle them. Finally, we assess our proposed methods by applying them to real-world human health applications (Applications block).

## 1.1 Uncertainty in Medicine and Healthcare

Uncertainty in medicine is a challenge faced by many medical specialists regardless of their experiences and the available information about the observed cases [11]. Three main sources of uncertainty in medicine are identified [12], namely: conceptual, personal and technical sources (Figure 1.2).



(a)



(b)

Figure 1.2: a) Research Map indicating the problem, solution, and applications. b) Data and model-related challenges from the clinical and data analysis perspectives and corresponding applications.

### 1.1.1 Sources of Uncertainty

According to Beresford [12], sources of uncertainty are illustrated as follows:

- The personal source mainly focuses on the physician–patient relationship. Specifically, related to that knowledge obtained from patients themselves, which is based on their expectations and experiences. For instance, some physicians face the challenge to treat patients who are not able, for several reasons, to reveal their experience with treatment or their expectations from it [12]. The accuracy of this type of knowledge may affect the prognosis procedure. Another aspect of personal uncertainty is the attachment that can grow between the care provider and the patient, which may affect the decision–making process.
- The conceptual source of uncertainty is resulting from missing a standard measure for decision–making. In other words, the physician’s or the caretaker’s judgement in some situations is not based on standard or common measures. For instance, prioritising a patient over another having the same condition to be treated (e.g. ICU admission, medication provision) can be controversial in case of limited resources (e.g. COVID-19 pandemics). Another aspect of conceptual uncertainty is resulting from the difficulty to apply general criteria to specific situations [12].
- Finally, the technical source of uncertainty is the most apparent cause of medical uncertainty and this one is data–related. Data–related challenges that contribute to technical uncertainty have two main levels. One level concerns the data collection process, which can suffer quantitative and qualitative insufficiency. This insufficiency can be due to the limited amount of data and/or the quality of the collected data. In modern medicine, the presence of advanced monitoring technology can restrain data insufficiency since the amount of the collected medical data is enormous. Moreover, the new sensors technology assures high–quality and reliable measurements of the different parameters.

The other level concerns the rapidly growing data flow, which is difficult for any individual practitioner to process it analytically. Moreover, depending on the investigated case, it is also challenging to define which part(s) of the data are of more relevance than others [9, 12]. Therefore, the persistent challenge in the technical uncertainty is analytical, which is reflecting the capability to analyse and interpret the enormous amount of the collected medical data. Moreover, this analysis shall provide accurate and reliable decisions and/or predictions.

We will further focus in this thesis on technical uncertainty and its impact on the clinical decision–making process.

### 1.1.2 Technical Uncertainty

Elements of technical uncertainty emerge from the aspects of probability, ambiguity, and complexity [11, 14]. These three aspects comprise the difficulties that are faced by the analytical ap-

proaches of medical data and their outcome.

- Probability reflects the resulting risk from the indeterminacy of future outcomes [13]. In other words, this aspect concerns the temporal uncertainty due to the lack of a time-related predictive power.
- Ambiguity reflects the indecisiveness resulting from imprecise predictions, conflicting information, and lack of evidence.
- Complexity reflects the incomprehensibility resulting from the multiplicity of influential factors and variables or the lack of interpretability [11].

A concrete example of the three technical uncertainty aspects is illustrated in Figure 1.3 for the response to breast cancer treatment [14]. In this example, the three aspects of uncertainty, namely, probability, ambiguity and complexity are illustrated in the context of a treatment effect. The probability aspect reflects how likely the treatment will be successful. The ambiguity reflects the imprecision about the benefit for a patient of the treatment. In addition, it reflects the conflicting opinions of medical experts or insufficient information about the treatment benefit (effectiveness) and its side effects. Finally, the complexity aspect reflects the existence of multiple potential causal factors [14]. Another example is the diagnostic decision-making process and monitoring of ICU patients. For instance, monitoring intensive care unit (ICU) patients would benefit from an accurate prediction of the care outcome by predicting the behaviour of patients' monitored vital signs and general health status (probability). Moreover, the ambiguity aspect appears in the difficulty of making diagnostic decisions that leads to therapeutic strategies based on conflicting evidence or incomplete information which can be harmful to patients. Ultimately, the complexity aspect is present in the ICU decision-making process due to the influential factors' multiplicity.

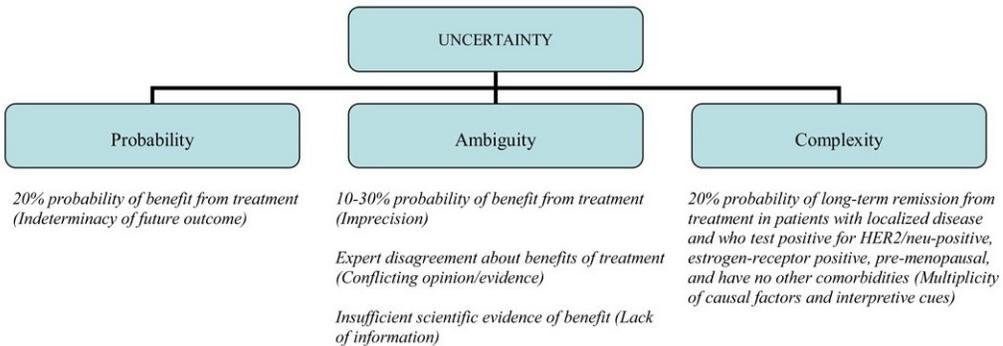


Figure 1.3: Three aspects of technical uncertainty: probability, ambiguity and complexity for the response to breast cancer treatment (from [14]).

Given the different aspects of technical uncertainty in medicine, we hypothesise that sufficient medical data and proper analytics can play an essential role in restraining the technical uncertainty. Especially that, technical uncertainty is data-related and can be tackled as a data analysis issue in the context of the medical domain. A practical solution to the problem of technical uncertainty is utilising clinical decision support systems whose components are medical data and analytical methods. Therefore, we propose restraining the technical uncertainty by providing proper analytical means, adequate and reliable analytical methodologies, to clinical decision support systems. By analytic methodologies, we mean the set of algorithms by which the monitored variables (e.g. vital signs) are analysed to extract useful information, define patterns and/or predict events to assess the health status of the monitored patients. In the following section, we will introduce the concept of clinical decision support systems and their components.

## 1.2 Data-Based Clinical Decision Support System

At the beginning of the information age (late 20th century), healthcare was expected to be revolutionised given the increasing amount of digital available medical data [15]. This expectation is partially legitimate, considering the ease of collecting medical data with advanced monitoring technologies and high storage capacities. However, the human mind alone cannot analyse this vast amount of data considering all potential factors reflected in these data. Hence, data is not enough to be the sole game-changer for healthcare services. In order to be actionable, data needs to be transformed into knowledge/insights. Advanced computer-based analytical methods are candidate approaches allowing to make these necessary transformations. Specifically, using and integrating different kinds of medical data (e.g. patients' information and clinical knowledge) is at the core of the computerised clinical decision support system (CDSS) in making clinical decisions [16].

Clinical decision support systems (CDSS) aim to enhance the clinical decisions by supporting human experts and broadening the investigation horizons [17]. Although the CDSS started in the 1970s, they faced several challenges regarding ethical, legal and explainability issues [18, 19, 20]. Nowadays, CDSs are linked to the patients' medical information collection and storing systems (e.g. electronic medical records (EMR)) and electronic entry systems of medical practitioner instructions for the treatment of patients (e.g. computerised provider order entry (CPOE)). Moreover, CDSs can be linked to wearable health technologies and biometric monitoring [21, 22] as sources of medical data.

CDSs can be categorised into knowledge-based and non-knowledge-based systems. The former is based on decision rules (IF-THEN), and the latter is based on data-based modelling algorithms [23].



Figure 1.4: Solution block from Figure 1.2a

Based on their functions, CDSs can be classified into several categories. Diagnostic decision support systems (DDSs), which are used for clinical diagnosis, is an important category of CDSs. DDSs are developed to provide a computerised consultation in the form of diagnosis or prediction, based on analysed medical data [24]. As indicated by Berner et al. [25] and Segal et al. [26], such systems face different challenges restraining their influence on the healthcare field. These challenges include, among others, negative perceptions of physicians and their biases against computational methods reliability, non-reliable accuracy, and poor system integration [25, 26, 16]. These challenges clearly indicate that there are still important issues to be solved before DDSS can become more influential in the field of medicine and healthcare (cf. objective section).

### 1.3 Decision Support Components

Two significant elements are necessary to develop a diagnostic decision support system: medical data acquisition and analytics.



Figure 1.5: Solution components block from Figure 1.2a

#### 1.3.1 Medical Data Acquisition

Medical data are collected either by conventional bedside medical devices at hospitals or clinics by specialised medical staff or, in a growing amount of cases, by wearable sensors attached to the body of the monitored patient.

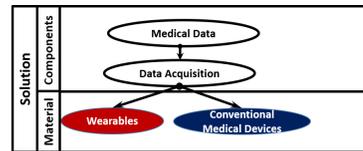


Figure 1.6: Solution material from Figure 1.2a

##### 1.3.1.1 Conventional Medical Devices

Conventional advanced bedside medical devices at hospitals and clinics are still considered the gold standard monitoring and measuring instruments because of their high accuracy and stability. However, these devices often suffer from being cumbersome and they need to be operated by trained specialised medical staff. Moreover, these devices are expensive and cannot be made accessible to all hospitalised patients. Therefore, they are often used for intermittent measurements of the different vital signs, except for critically-ill patients who are monitored continuously in intensive care units (ICU) using such devices. At ICUs, different vital signs of the admitted patients are continuously monitored

by bedside ICU equipment. However, the stored data in electronic medical records (EMR) are approved/validated typically at a much lower frequency (e.g. once every 1-2 hours) by nurses.

### 1.3.1.2 Wearable Medical Devices

These days we witness revolutionary improvements in sensor and communication technologies, leading to the development of more reliable wearable medical devices. It is envisioned that the continuous data stream generated by these wearable devices, in combination with artificial intelligence (AI), can revolutionise patient-care and health management in the future. In recent years, wearable technology has become more advanced in terms of design, ease of sensing, accuracy, power efficiency, and parameters to be monitored [27, 28]. This advancement makes the integration of wearable technology into healthcare applications a potential game-changer for health monitoring [29, 30, 31]. One significant advantage of using wearable technology in health monitoring is the continuous/high-rate monitoring of subjects. This continuous monitoring can contribute to preventive medicine by predicting several health-related risks at an early stage. In addition, it can provide a more complete picture of the health status of the subjects to the medical staff. The amount and the importance of measurable parameters using wearable technology are considered significant. In their study, King et al. [32] listed a set of these parameters/variables and the possible extracted features from them. They categorised them into three groups, namely physical activity variables, biometric variables, and environmental variables. The first comprises, among others, motions (e.g. walking, jogging running, falling, and step-count), postures (e.g. lying, sitting, and standing), sleeping duration, REM sleeping, and sleep latency by tracking heart rhythm with physical activity. The biometric group comprises variables, such as, heart rate, respiration rate, blood oxygen level, electrocardiogram (ECG), heart rate variability, blood pressure, glucose level, skin temperature, electroencephalogram (EEG) and electromyography (EMG). The third group comprises environmental variables such as environment temperature, humidity, and CO<sub>2</sub> concentration [32].

### 1.3.2 Analytics

The other key element that is necessary for a clinical decision support system is analytics. Several analytical methods are used to analyse medical data such as data-based mechanistic (DBM) modelling, statistical analysis and, in recent years, artificial intelligence (AI) and machine learning. Each of these analytical methods has its advantages. For instance, DBM modelling [33] concerns the mechanistic interpretation of the analysed systems, which can help to better understand the system. On the other hand, DBM handles a limited number of variables and data points for model development. AI approaches, especially machine learning, have shown a great potential in contributing to medical diagnosis and decision-making, because of their advanced computational capabilities for solving problems [34, 35, 36, 37, 38, 39]. In essence, both, AI and medicine, can be reduced to decision-making based on reasoning and inference. Hence, AI can work as a

complementary cognitive aid contributing to the decision-making process in medicine. In the review article of Yu et al. [40], a list of potential AI applications in medicine is discussed, including clinical practice applications (i.e. patient monitoring and patient risk stratification) [39, 40]. It is worth mentioning that AI and machine learning are not equivalent, as machine learning is only one branch amongst several emerging from AI. One of the first applications of AI in medicine was developing clinical decision support systems [41, 42]. The rule-based approach, a knowledge-based approach, is an AI branch that was used in the early days of clinical decision support systems [43, 44]. Rule-based approaches were utilised in e.g. ECG interpretation, disease diagnosis, treatment recommendation, and several other clinical applications. For instance, rule-based diagnosis follows the 'If (symptoms) Then (specific disease)' procedures. However, rule-based approaches require frequent human intervention as they incorporate expert judgement in addition to the limited flexibility of the decision rules to adapt to the newly obtained knowledge [45]. Another drawback of the rule-based approach is that its outcome is limited to human expert knowledge and hence, it cannot add any new clinical insight. More specifically, these reasoning rules translate the already existing medical knowledge into logical statements to assure a consistent decision. With the rise of applying AI in healthcare and in parallel with the evolution of machine learning methods, machine learning took the lead in AI applications in healthcare [46]. This was especially fuelled by the fact that machine learning methods have the power to handle complex medical problems and to identify data patterns which are too complex to be recognised/extracted by human observers [46, 47, 48]. These medical problems can be annotated with predefined labels, events, or real values such as diagnostic problems since specific outcomes are associated with obtained observations. This kind of problem requires 'supervised learning' (i.e. classification, regression) by which correct outcomes are assigned to new observations based on the learned model. On the other hand, some medical problems are without a defined outcome, such as partitioning patients of specific disease into groups based on similarity. This kind of problems requires unsupervised learning (e.g. clustering). Therefore, both supervised and unsupervised machine learning methods, enable discovering unrecognised patterns in medical data [47].

### 1.3.3 Data and Model-based Challenges

Many challenges result from the nature of modelling methods (i.e. machine learning algorithms) and medical data properties. In this section, we will elaborate more on these challenges. Important challenges resulting from the properties of medical data comprise class-imbalance and non-uniform data distribution of the events/classes in the input space, ambiguity (confusion), and continuously increasing data-size. Other important challenges result from machine learning algorithms such as model personalisation, black-box nature, online modelling and streaming analytics.

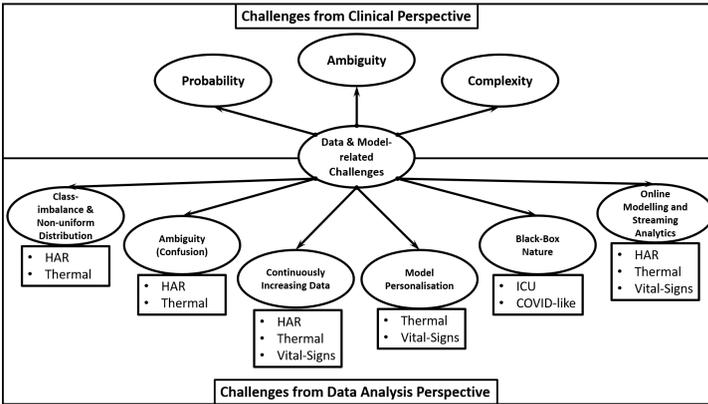


Figure 1.7: Data and model-based challenges block from Figure 1.2b

### 1.3.3.1 Class-imbalance and Non-uniform Data Distribution

Class-imbalance occurs when data points (observations) from some events are outnumbered by others. In the case of extreme imbalance, the challenge of rare events occurs (e.g. the detection of falls among several daily routine activities). Non-uniform distribution occurs when data-points of a specific class are scattered around the input space of the input features. Therefore, the model can consider some of these points as outliers. As shown in Figure 1.8, a binary classification two-dimensional problem with class-imbalance ratio of  $1/9$  is depicted. Moreover, few data points reflect non-uniform distribution by being scattered over the input space away from the centre of the class because of intra-class variance.

### 1.3.3.2 Ambiguity (Confusion)

Ambiguity or confusion from data analysis perspective occurs when an event is not distinguishable from another one. In the input space, where all data points are allocated representing the tuples of input variables' values, ambiguity could result from the overlap between the different classes or due to a nonlinear decision boundary between two closely spaced classes. The decision boundary is the boundary that partition the input space into two or more regions of the different classes. For instance, in human activity recognition, walking upstairs might be confused with walking downstairs depending on the accelerometer location. As shown in Figure 1.9, the test point ( $x$ ) is located close to the two classes and the pattern is nonlinear which is confusing to the decision making process.

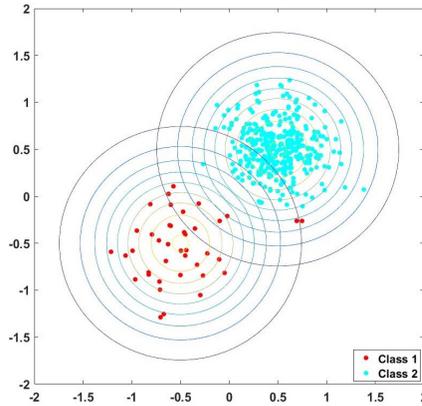


Figure 1.8: Two-dimensional binary classification problem with class imbalance ratio 1/9 between classes 1 and 2 respectively.

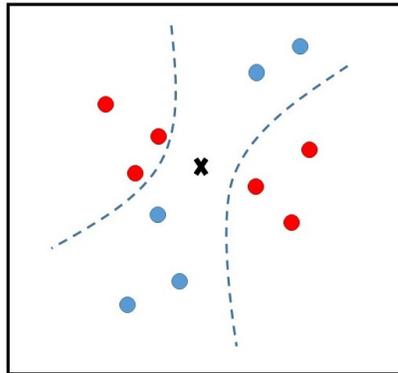


Figure 1.9: Two dimensional binary classification problem with one test point ( $x$ ) experiencing ambiguity. Adapted from [51].

### 1.3.3.3 Continuously Increasing Data-size

The large and continuously increasing data-size is another challenge that faces the usage of machine learning. The continuously increasing medical data may reflect different dynamics and behaviours of the monitored variables for the same subject or different subjects. Hence, continuous streaming can change the distribution of the data over the input space.

#### 1.3.3.4 Model Personalisation

Personalised medicine is that approach in which patients are stratified based on their disease sub-type, risk, prognosis, or treatment response using specialised diagnostic tests [52]. The concept behind this approach is to make medical decisions based on individual patient data instead of averages over a whole population [53]. From machine learning perspective, this approach requires personalised models that automatically take into account the individuality of each patient. Some machine learning approaches learn personalised models based on personal data only. Such an approach requires a sufficient amount of labelled data from that person which can be challenging. Moreover, considering only the personal data may lead to forgetting the previously obtained knowledge from other subjects' data. On the other hand, machine learning algorithms aim to generalise models which is a result of global learning via inductive inference [49]. By global learning, we refer to training the model using all available training data to identify a pattern that can be generalised over the training data in addition to the unseen data to be received. This aim of generalisation does not consider the local specificity of the different regions in the input space. Therefore, generalisation may lead to missing rare and minor events in terms of tolerance and robustness. In other words, to generalise a global model, it is required to tolerate some error to avoid overfitting and to obtain a robust model against outliers. From model personalisation perspective, global models disregard the individuality of the different data sources' profiles (e.g. subjects). Hence, there is a need for a machine learning model that can learn from other persons' data and the personal data in a balanced way that considers the individuality of each person's data.

#### 1.3.3.5 Black-box Nature

Machine learning algorithms are mostly black-box models. Black-box models are data-based models that estimate input/output relationships without taking into account the internal underlying mechanisms of the system. Black-box models can be contrasted with white-box models that are developed based on the system's knowledge and thus have physical meaning. For instance, black-box classifiers are developed by defining and optimising the function that associates input features to output labels, regardless of the complexity, explainability and interpretability of that function (model).

Briefly, model explainability is about how the model works, and model interpretability is a domain-specific notion. Therefore, the black-box nature compromises model interpretability and explainability. This compromise is not commendable in medicine, as obtaining clinical interpretation and insight into human health problems is crucial, especially for decision-making situations.

#### 1.3.3.6 Online Modelling and Streaming Analytics

Another challenge of machine learning algorithms is related to the need for real-time model adaption. For several health care applications, the stream of collected data is continuously increas-

ing in time which requires streaming analytics. In streaming analytics, an enormous amount of data in-motion flows over networks or clouds and is continuously collected, transmitted, analysed, and the resulting analysis outcome is provided in real-time. Most machine learning algorithms are basically global offline algorithms. In other words, they are trained by all available training data (global) and then applied to new unseen data without adapting to newly measured information (offline). Therefore, streaming new data points cannot be considered in the training process unless the model is retrained or adapted to the new data points. Both options are expensive and therefore enough new data points are required for an efficient update of the model. The other option is to continuously add new data points to the training set and to develop a model that is continuously updating (online) with a minimal computational cost.

### 1.3.4 Machine Learning in Medicine

The aforementioned challenges require applying adequate and reliable machine learning algorithms. These algorithms need to be simple to implement, interpretable and efficient in terms of performance and computation. An efficient performance considers a reliable error performance (e.g. accurate predictions). For the higher mentioned challenges of class-imbalance and non-uniform data distribution, ambiguity, continuously increasing data-size, model personalisation and streaming analytics, localised learning approaches of machine learning [51, 54] can be an adequate analytic methodology, as will be shown later in this chapter. For the challenge of the black-box nature (lack of interpretability and explainability), a feature engineering approach in combination with a linear hard margin approach to support vector machines can be a useful methodology. This approach can provide (partial) interpretability besides a high performance. Both machine learning approaches are elaborated in detail within the upcoming chapters, but meanwhile a brief introduction to them is presented in this chapter.

#### 1.3.4.1 Localised Learning

As mentioned previously, machine learning is used extensively and efficiently in several applications, including medicine. Many machine learning approaches and algorithms are used in medicine, from simple decision trees up to advanced deep neural networks. The majority of well-known moderate to powerful supervised machine learning algorithms are based on inductive inference which leads to global learning [49, 55, 56, 57]. This approach establishes a generalised hypothesis (model) from specific examples (training data) to assign a value/label to new data points based on the learned models. In other words, inductive

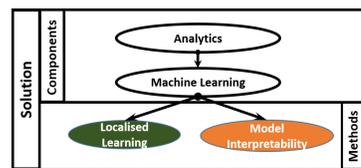
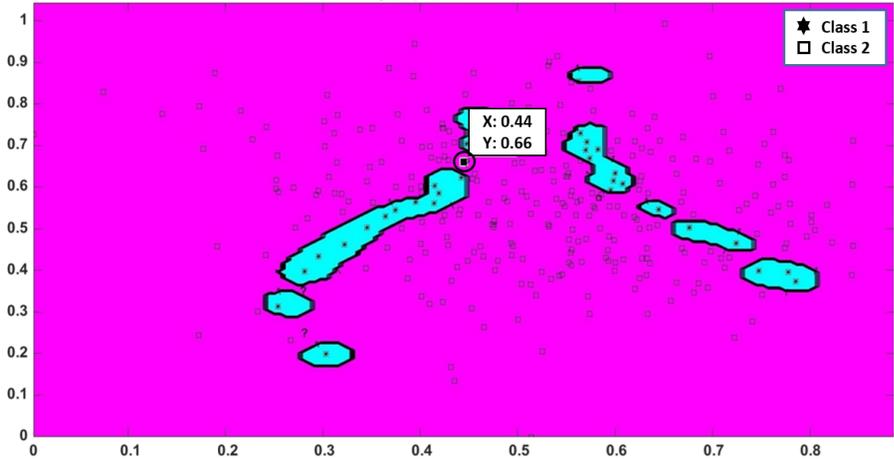


Figure 1.10: Solution methods block from Figure 1.2a

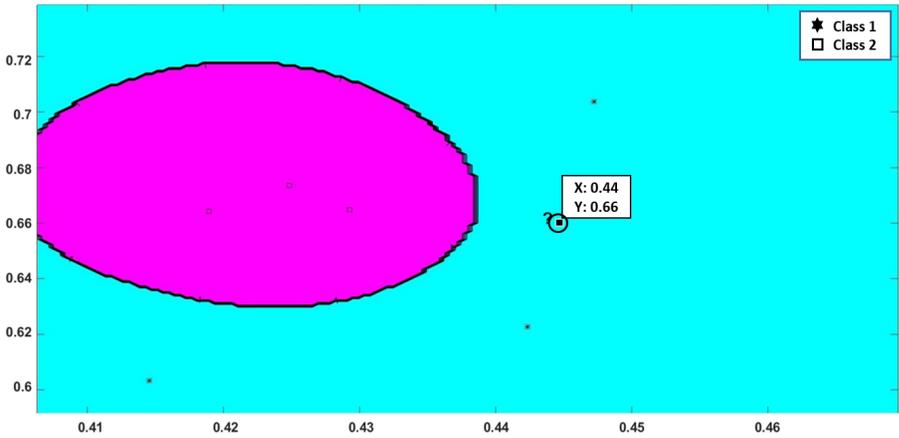
inference aims to define a general theory from the observed examples (specific-to-general). This inference has the advantage of providing general patterns that control the behaviour of the observed examples [49, 57, 58]. However, the local properties of the observed examples are prone to be compromised to obtain generalisation. Based on Bottou et al. [51], in case of uneven distribution of patterns in the input space, a proper local adjustment of the model capacity is needed to enhance the performance. In other words, if the feature space is not providing the possibility to find a solution considering both local and global properties of the observations, then the model may tolerate minority data points as an error. In Figure 1.11, a two-dimensional nonlinear binary classification problem is depicted. In Figure 1.11a, the global classifier of LS-SVM identifies the pattern of class 1 within the discontinuous blue region, although the original pattern is supposed to be continuous. This discontinuity results from the class imbalance with ratio 1/9 between classes 1 and 2 and non-uniform data distribution. We have chosen a point with approximate coordinates (0.44, 0.66) as a test point that is misclassified with the global model, as shown in Figure 1.11a. This point is chosen to be tested by a localised model (i.e. kNN-LS-SVM) to show the influence of the localised algorithm on enhancing the classification performance. As shown in Figure 1.11b, a close-up to the test point and its vicinity, the test point is classified correctly as class 1 by being included in the blue pattern with the other class 1 instances.

In addition to the local learning challenge, another challenge that global learning algorithms face, especially offline models, is the analysis of both time-series data and continuously growing data. In the case of time-series data, online modelling is required to provide a real-time analysis (e.g. time-series prediction) [59]. For instance, a real-time time-series predictive model cannot be implemented without considering the new measured values or observations. Referring to the medical domain, human vital-signs are dynamic as they are continuously changing for several factors (e.g. activity, health status), which requires a real-time update of the model. For continuously growing data, as mentioned in sections 1.3.1.3 and 1.3.1.6, data distribution can be affected in the input space, which requires updating the model as well. To develop a global model that is capable of predicting the outcome of new observations efficiently, this model has to be trained with an enormous amount of data that is assumed to comprise all possible dynamics of the system, which is difficult and expensive. On the other hand, conventional online modelling approaches are time-dependent and prone to forget old observations in favour of most recent observations. These analytical requirements regarding time-series and continuously growing data also apply when using wearable technology for medical monitoring and real-time time-series prediction. Therefore, we propose localised learning algorithms for wearable-based applications.

The analytical challenges of locality, time-series data, and continuously growing data require a real-time modelling approach to handle them. Such an approach is challenging in case of global learning as real-time modelling requires either retraining the model or applying model adaption techniques (e.g. incremental, active, and transfer learning) [60, 61, 62, 63, 64]. These approaches face difficulties that may cause lack of efficiency for our medical applications. These difficulties include the catastrophic forgetting that may occur by forgetting the obtained knowledge from



(a)



(b)

Figure 1.11: Two-dimensional nonlinear binary classification problem. a) The classification outcome of the global classifier of LS-SVM, b) The classification outcome of the localised classifier of kNN-LS-SVM.

previously collected data [65]. Another difficulty is applying gradual adaption with new data points, avoiding complete retraining of the model. In addition, limiting the number of training instances to control the computational complexity [65, 66]. These difficulties are still a matter of current research.

Another approach of reasoning that can handle the challenges of locality, time-series data, and continuously growing data is transductive inference. This approach uses the new observation(s) to build a dedicated model that is adapted to the local properties of new observations [56]. However, the transductive inference may experience the property of "laziness" as each model is only built after receiving the observations which is the case for  $k$ -nearest neighbours (kNN). KNN algorithm is a simple machine learning method that can be used for both classification and regression. The algorithm briefly assigns a value or label to the data points based on the majority vote of nearest  $k$  neighbours. As shown in Figure 1.12, for data point **a**, it is assigned the label of class 2 based on its nearest neighbours' labels. On the other hand, local learning models such as kNN as an example of transductive inference, are less suited to deal with nonlinear patterns that result from overlapping classes as shown in Figure 1.12 for point **b**. Therefore, a more efficient transductive inference approaches can be achieved by developing hybrid algorithms. These algorithms integrate a local learning algorithm (e.g.  $k$ -nearest neighbours) into a powerful global learning algorithm (e.g. support vector machines) resulting in localised learning algorithm [51, 54].

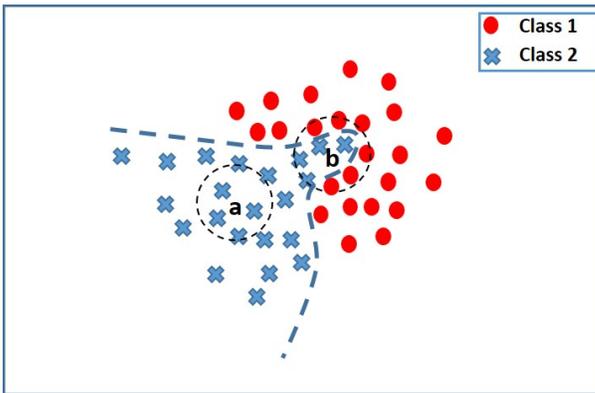


Figure 1.12: Two-dimensional nonlinear binary classification problem with two new data points **a** and **b**.

Local algorithms can provide the advantages of adapting the model to the local properties. On the other hand, the global algorithm shall provide the capability to handle complex and nonlinear patterns as shown in Figure 1.11. Hence, considering the local properties could potentially contribute to solving the challenges of ambiguity, class-imbalance, and data nonuniform distribution [51]. Moreover, the challenge of model personalisation can be handled by training

the model with the most similar observations from either the same subject or a similar subject or a similar sub-population of patients. Therefore, considering the locality can provide model personalisation, which reflects the individuality of each patient in the medical application. Furthermore, the challenges of streaming analytics and online modelling in medical applications can be handled by computationally efficient hybrid transductive algorithms (e.g. kNN-LS-SVM). Moreover, time performance is also crucial to be considered to address both streaming analytics and online modelling challenges. Optimising the time performance can be achieved by minimising the size of the local training set, such that the error performance is acceptable, as the computational complexity of the global learning algorithm is a function of the training set size. For instance, as shown in Figure 1.11b, the required number to train a model for a new data point can be only 7 data points.

Based on these elements, we propose to select transductive inference approaches for a subset of the aforementioned challenges. These challenges are namely [class-imbalance and nonuniform data distribution](#), [ambiguity](#), [increasing data-size](#), [model personalisation](#), [streaming analytics and online modelling](#) (Section 1.3.3).

#### 1.3.4.2 Model Interpretability and Explainability

In the previous section, we referred to all challenges identified in Section 1.3.3 except for the [black-box nature](#) (1.3.3.5). Therefore, the final challenge that we want to address here is the lack of interpretability and explainability of the developed models as a result of the black-box nature. Recently, this topic has received more and more attention. In their article, Gunning et al. [67] are elaborating on the concept of explainable artificial intelligence (XAI) and how to optimise the trade-off between explainability and the error performance as shown in Figure 1.13. For instance, decision trees are considered one of the most explainable methods, but with less optimal accuracy. On the other hand, deep learning shows excellent error performance but is least explainable. The importance of interpretability and explainability is strongly linked to the application domain. In medicine and healthcare, medical staff must understand the reasoning behind data-based models in order to be confident while using them. Both interpretability and explainability can be full or partial. For example, partial explainability may address a specific portion of an input signal that impacts the predicted outcome. On the other hand, full explainability may indicate which exact attribute(s) of this portion has this influence. Similarly, partial interpretability may indicate the order of the model and weights of each input variable. However, full interpretability reflects the meaning of the model's order and weights within the domain context.

Interpretability of models results from the characteristics of the machine learning method itself. For instance, as shown in Figure 1.14, a linear hard margin support vector machine (SVM) model has the property of providing a linear model that maximises the margin between the different classes within a feature space that is limited to the provided input variables [50]. Therefore, for a performant linear hard margin SVM model, the input features are supposed to provide linearly separable data points in the original input space where the margin between the different

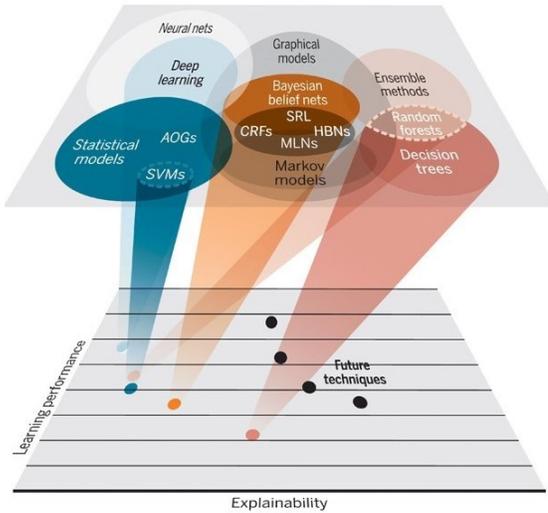


Figure 1.13: Performance vs explainability of different learning algorithms (e.g. and-or-graphs (AOGs), support vector machines (SVMs), statistical relation learning (SRL), hierarchical Bayesian networks (HBNs), conditional random fields (CRFs), and Markov logic networks (MLNs))(from [67]).

classes is maximised. By further investigation, we can identify the range of values for each feature that reflects the margin between different classes. In other words, we can determine at which value(s) within the range of values for each feature the label flips from one class to another. Moreover, calculating the distances between the decision boundary and different data points reflect the confidence about the decision and the status of what these data points represent. For instance, in mortality prediction problems for hospitalised patients, the decision boundary distinguishes between mortality and survival subjects. Hence, the distance of a data point (subject’s attributes) from the decision boundary indicates the subject’s status and the severity of his/her health status.

Moreover, model explainability can be achieved by either forward or backward methods. A forward way can be achieved by engineering explainable features and integrating them into the machine learning algorithm [68]. Engineering informative and meaningful features (e.g. average heart rate per hour) can explain to some extent the basis on which the model decision is made. On the other hand, the backward way can be achieved by using techniques that backtrack the best performant model and associate that performance to the most influencing features. A successful example of such a method is the technique of heatmapping [69] and its application to deep neural networks.

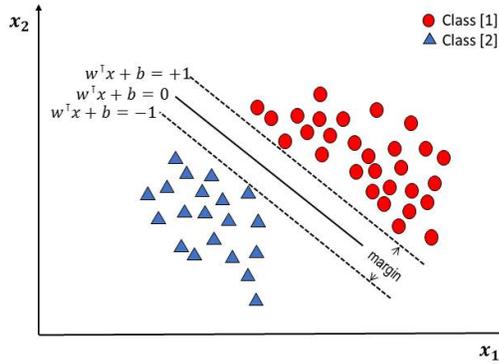


Figure 1.14: Schematic representation of a two dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised.

## 1.4 Objective

After decomposing the [technical uncertainty \(1.1.2\)](#) problem to its essential elements (i.e. complexity, probability, and ambiguity) [1.3](#), we can link it to the aforementioned [challenges](#). These challenges are class-imbalance, non-uniform data distribution, ambiguity, increasing data-size, model personalisation, black-box nature, streaming analytics and online modelling.

Starting with probability, which reflects the resulting risk from the indeterminacy of future outcomes, its essential challenge is related to time-based prediction accuracy. This accuracy is directly influenced by the challenges of [class-imbalance and non-uniform data distribution](#) and [ambiguity](#).

For ambiguity, which reflects the indecisiveness resulting from imprecise predictions, lack of evidence and conflicting information, its essential challenges are linked to relation-based prediction accuracy, data sufficiency and engineered features' efficiency. The relation-based prediction (e.g. classification) is a time-independent prediction and its accuracy is influenced by [class-imbalance and non-uniform data distribution](#) and [ambiguity](#). Data sufficiency, which is required for strong evidence, is directly linked to the amount of data and its challenges [increasing data-size](#) and [streaming analytics and online modelling](#). Moreover, sufficient data shall include several sources of data (e.g. multiple subjects for human health applications) which imposes the challenge of [model personalisation](#). Engineered features' efficiency is needed to synthesise the different information pieces and resolve the conflict between them, which is linked to model explainability ([black-box nature](#)).

For complexity, which results from the multiplicity of influential factors and variables or the lack of interpretability, its essential challenges are related to the difficulty of considering many influential factors and model interpretability. Therefore, we proposed machine learning algorithms (e.g. SVM's) to handle the computational problem as they can handle a large number

of input variables. On the other hand, model interpretability is linked to the challenge of **black-box nature**.

Motivated by the problem of technical uncertainty in medicine and healthcare, we aim in this thesis at investigating the possibility to limit and reduce the technical uncertainty of applying machine learning approaches in human health applications. Reducing and restricting this uncertainty can be achieved by developing machine learning algorithms that can enhance the reliability of the decision-making process in human health applications.

Therefore, the main objective of this thesis is to develop performant machine learning algorithms and approaches that contribute to handling the aforementioned **challenges** in the context of human health applications. Such algorithms can be the core of reliable diagnostic decision support tools that can reduce and limit the technical uncertainty in the medical domain. In Section 1.5, we illustrate the human health application cases that will be investigated in this thesis.

#### 1.4.1 Sub-objectives

In the light of the main objective of this thesis, we specify four sub-objectives that are addressing more specific machine learning solutions to the main aforementioned challenges. More specifically, we aim at:

1. Developing machine learning algorithms for tackling the challenges of ambiguity, class-imbalance, non-uniform data distribution, and continuously increasing data-size (Chapters 2, 3, and 4).
2. Proposing a machine learning algorithm to the challenges of streaming analytics and online modelling and their requirements (Chapters 2, 3, and 4).
3. Proposing a machine learning algorithm to the challenge of model personalisation (Chapters 3, and 4).
4. Providing an approach that satisfies both partial interpretability and explainability for black-box modelling (Chapters 5, and 6).

#### 1.4.2 Research Questions

After defining the main objective and sub-objectives, two research questions are raised regarding the aforementioned challenges. These questions are stated below:

- Question N°1

Is the proposed localised learning algorithm of kNN-LS-SVM able to handle class-imbalance, non-uniform data distribution, ambiguity, continuously increasing data, online modelling, streaming analytics and model personalisation?

- Question N°2

To what extent can a feature engineering procedure integrated with a linear hard margin approach of SVM provide interpretability and explainability while providing an acceptable error performance for prediction?

## 1.5 Human Health Applications

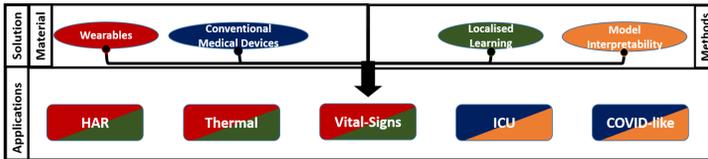


Figure 1.15: Applications block from Figure 1.2a

In this section, we list the set of human health applications investigated through the up-coming chapters. They cover a spectrum of human health applications that goes from healthy subjects over hospitalised patients at general wards to critically ill patients admitted to the ICU.

1. The first application (Chapter 2) focuses on human activity recognition (HAR) based on daily life physical activities and postures monitored using wearable accelerometer sensors. In this application we introduce kNN-LS-SVM algorithm in addition to investigating the challenges related to [class-imbalance and non-uniform data distribution](#), [ambiguity](#), [increasing data-size](#) in the context of HAR.
2. The second application (Chapter 3) focuses on human well-being and thermal sensation and degree of comfort recognition using easily measurable variables by wearable sensors. The main focus of this application is on developing personalised thermal comfort classification models in the context of an automated climate control platform. This application is based on wearable sensors for all vital signs considered in this study. Moreover this application is investigating the challenges related to [class-imbalance and nonuniform data distribution](#), [ambiguity](#), [increasing data-size](#) and [model personalisation](#) in the context of thermal comfort prediction.
3. The third application (Chapter 5) concerns hospitalised patients and their continuously monitored vital signs by wearable sensors to develop a high-rate (every minute) early warning score (EWS) and predictive models by which the future vital sign values are predicted. These hospitalised patients are admitted to general wards namely cardiology, post-operative, and dialysis wards. All three applications are based on wearable sensors and handle the challenges related to [ambiguity](#), [increasing data-size](#), [model personalisation](#), [streaming analytics and online modelling](#).

4. The fourth application (Chapter 6) concerns critically ill patients who are admitted to the ICU for different hospitalisation reasons (e.g. cardiovascular, renal or pulmonary diseases). This application focuses on the mortality prediction of ICU patients based on gold standard measurements at the ICU using standard ICU monitoring devices. The investigated data in this application is discrete data with the rate of a single approved measurement every one to two hours. This application investigates the challenge of **black-box nature** and possible model interpretability and explainability for ICU patients.
5. Finally, the fifth application (Chapter 7) is inspired by the recent situation of COVID-19 pandemic and the resulting under-capacity experienced at intensive care units. Therefore, the developed predictive models and engineered features of the fourth application are applied to patients with a similar profile as COVID-19 patients (COVID-like patients) as a basis for later applying to actual COVID-19 patients. The main objective of this application is to predict the mortality and survival of COVID-like patients to support medical decisions regarding the predicted outcome. This application investigates the challenge of **black-box nature** and possible model interpretability and explainability for pulmonary ICU patients.

# Bibliography

- [1] Stiefel M. et al.. The Edwin Smith Papyrus: The Birth of Analytical Thinking in Medicine and Otolaryngology. *Laryngoscope*. 116(2) 2009. <https://doi.org/10.1097/01.mlg.0000191461.08542.a3>.
- [2] Breasted J. H.. The Edwin Smith Surgical Papyrus. Chicago, Ill. :The University of Chicago Press, 1930.
- [3] Hughes J. T.. The Edwin Smith Surgical Papyrus: an analysis of the first case reports of spinal cord injuries. *Paraplegia*. 26(2), 1988, 71–82. doi:10.1038/sc.1988.15.PMID: 3045730.
- [4] Edwin Smith Papyrus, Encyclopedia Britannica. <https://www.britannica.com/topic/Edwin-Smith-papyrus>
- [5] The US National Library of Medicine Turning the Pages Information System (TTPI).
- [6] Gillett G.. Clinical medicine and the quest for certainty. *Social Science and Medicine*, 58(4), 2004, 727–738.
- [7] Kaplan RM, Ganiats TG, Frosch DL.. Diagnostic and treatment decisions in US healthcare. *J. Health Psychol.*, 9, 2004, 29–40.
- [8] Kemm, John R.. "Evidence based must not equal judgment free.". *Bmj*, 328(7446), 2004, 1018.
- [9] Kaplan RM, Frosch DL. Decision making in medicine and health care. *Annu Rev Clin Psychol*, 2005, 525–556. doi:10.1146/annurev.clinpsy.1.102803.144118
- [10] Ghosh A. K. Understanding medical uncertainty: a primer for physicians. *JAPI*, 52, 2004, 739–42.
- [11] Kangmoon K., and Young-Mee L.. Understanding uncertainty in medicine: concepts and implications in medical education. *Korean journal of medical education*, 30(3), 2018, 181–188.

- [12] Beresford EB. Uncertainty and the shaping of medical decisions. *Hastings Cent Rep.*, 21(4), 1991, 6–11.
- [13] Savulescu J. Treatment limitation decisions under uncertainty: the value of subsequent euthanasia. *Bioethics*. 8(1), 1994, 49–73.
- [14] Han P.K., Klein W.M. and Arora N.K.. Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*, 31(6), 2011, 828–838.
- [15] Carlile S., and Sefton A. J.. Healthcare and the information age: implications for medical education. *Medical Journal of Australia*, 168(7), 1998, 340–343.
- [16] Sutton Reed T., et al.. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3(1), 2020, 1–10.
- [17] Osheroff J., et al. Improving Outcomes with Clinical Decision Support: An Implementers Guide. *HIMSS Publishing*, 2012.
- [18] Sim I. et al.. Clinical decision support systems for the practice of evidence-based medicine. *J. Am. Med Inf. Assoc. Jamia.*, 8, 2001, 527–534.
- [19] De Dombal F.. Computers, diagnoses and patients with acute abdominal pain. *Arch. Emerg. Med.*, 9, 1992, 267–270.
- [20] Shortliffe E. H. and Buchanan B. G.. A model of inexact reasoning in medicine. *Math. Biosci.*, 379, 1975, 233–262.
- [21] Middleton B., Sittig D. F. and Wright A.. Clinical decision support: a 25 year retrospective and a 25 year vision. *Yearb. Med. Inform.*, 25(S 01), 2016, 103–116.
- [22] Dias D.. Wearable health devices vital sign monitoring, systems and technologies. 2018. <https://doi.org/10.3390/s18082414> (2018).
- [23] Berner E. S. (Ed.). *Clinical Decision Support Systems*. Springer, New York, NY, 2007.
- [24] Berner E. *Clinical Decision Support Systems: Theory and Practice* 3rd edition. 2016 <https://doi.org/10.1007/978-0-387-38319-4>.
- [25] Berner E. S.. Diagnostic decision support systems: why aren't they used more and what can we do about it? *AMIA Annual Symposium Proceedings*, 2006, 1167–1168.
- [26] Segal, M. M. et al.. Experience with integrating diagnostic decision support software with electronic health records: benefits versus risks of information sharing. *EGEMs Gener. Evid. Methods Improv. Patient Outcomes*, 5(23), 2017.

- [27] Ghaffari R., Schlatka B. L., Balooch G., Huang Y., Rogers J. A.. Reinventing biointegrated devices. *Mater Today*, 16(5), 2013, 156-157.
- [28] MC10 Inc. MC10 Reshaping electronics. 2014, <http://www.mc10inc.com/>.
- [29] Godfrey A., Conway R., Meagher D., OLaighin G.. Direct measurement of human movement by accelerometry. *Med Eng Phy*, 3(10), 2008,1364-1386.
- [30] Cheung V. H., Gray L., Karunanithi M.. Review of accelerometry for determining daily activity among elderly patients. *Arch Phys Med Rehabil*, 92(6), 2011, 998-1014.
- [31] Patel S., Park H., Bonato P., Chan L., Rodgers M.. A review of wearable sensors and systems with application in rehabilitation. *J Neuroeng Rehabil*, 2012, 9-21.
- [32] King R. C., Villeneuve E., White R. J., Sherratt R. S., Holderbaum W., Harwin W. S.. Application of data fusion techniques and technologies for wearable health monitoring. *Medical engineering and physics*, 42, 2017, 1-12.
- [33] Young P.. Data-Based Mechanistic Modeling of Engineering Systems. *Journal of Vibration and Control*, 4(1), 1998,5-28. <https://doi:10.1177/107754639800400102>.
- [34] Murdoch T. B., Detsky A. S.. The inevitable application of big data to health care. *JAMA*, 309, 2013, 1351-1352.
- [35] Kolker E., Özdemir V., Kolker E.. How Healthcare can refocus on its Super-Customers (Patients, n=1) and Customers (Doctors and Nurses) by Leveraging Lessons from Amazon, Uber, and Watson. *OMICS*, 20, 2016, 329-33.
- [36] Dilsizian SE, Siegel EL.. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr Cardiol Rep*, 16, 2014, 441.
- [37] Patel V. L., Shortliffe E. H., Stefanelli M. et al.. The coming of age of artificial intelligence in medicine. *Artif Intell Med*, 46, 2009, 5-17.
- [38] Jha S., Topol E. J.. Adapting to Artificial Intelligence: radiologists and pathologists as information specialists. *JAMA*, 316, 2016, 2353-2354.
- [39] Jiang Fei, et al.. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017, 230-243.
- [40] Yu Kun-Hsing, Andrew L. B., and Isaac S. K.. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10), 2018, 719-731.
- [41] Check Hayden E.. The automated lab. *Nature*, 516, 2014 ,131-132.

- [42] Miller R. A.. Medical diagnostic decision support systems past, present, and future: a threaded bibliography and brief commentary. *J. Am. Med. Inform. Assoc.*, **1**, 1994, 8–27.
- [43] Shortliffe E.. Computer-Based Medical Consultations. *MYCIN*, Elsevier, New York, 2, 2012.
- [44] Szolovits P., Patil R. S. and Schwartz W. B.. Artificial intelligence in medical diagnosis. *Ann. Intern. Med.*, **108**, 1988, 80–87.
- [45] Berner E. S. et al.. Performance of four computer-based diagnostic systems. *N. Engl. J. Med.*, **330**, 1994, 1792–1796.
- [46] Deo R. C.. Machine learning in medicine. *Circulation* **132**, 2015, 1920–1930.
- [47] Yu K. H. and Snyder M.. Omics profiling in precision oncology. *Mol. Cell. Proteomics*, **15**, 2016, 2525–2536.
- [48] Roberts K. et al.. Biomedical informatics advancing the national health agenda: the AMIA 2015 year in review in clinical and consumer informatics. *J. Am. Med. Inform. Assoc.*, **24**, 2017, 185–190.
- [49] Abu Mostafa Y.S., Malik M. I., Hsuan Tien L.. Learning from Data. *AMLBook: New York, NY, USA*, 2012.
- [50] Boser B. E., Guyon I. M., Vapnik V. N.. A training algorithm for optimal margin classifiers. *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA*, 1992, 144–152.
- [51] Vapnik V. and Bottou. Local algorithms for pattern recognition and dependencies estimation. *Neural Computation*, **5**, 1993, 893–909.
- [52] Sobradillo P., Pozo F., Agustí A. P4 medicine: the future around the corner. *Arch Bronconeumol*, **2011**, 35–40. <https://doi.org/10.1016/j.arbres.2010.09.009>.
- [53] Fröhlich H., Balling R., Beerenwinkel N. et al. From hype to reality: data science enabling personalized medicine. *BMC Med* **16**, 150 2018. <https://doi.org/10.1186/s12916-018-1122-7>.
- [54] Amer A.. Localised least squares support vector machines with application to weather forecasting. *Masters thesis, KU Leuven*, 2016.
- [55] Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., and Vandewalle J.. Least Squares Support Vector Machines. *Singapore: World Scientific Publishing Co.*, 2002.
- [56] Pang S., Kasabov N.. Inductive vs transductive inference, global vs local models: SVM, TSVM, and SVMT for gene expression classification problems. *IEEE International Joint Conference on Neural Networks*, **2**, 2004, 1197–1202.

- [57] Mitchell M.T.. Machine Learning, *MacGraw-Hill*, 1997
- [58] Vapnik V.. Statistical Learning Theory. *John Wiley and Sons, Inc*, 1998.
- [59] Fontenla-Romero Ó., Guijarro-Berdiñas B., Martínez-Rego D., Pérez-Sánchez B., Peteiro-Barral D.. Online machine learning. In *Efficiency and Scalability Methods for Computational Intellect. IGI Global*, 2013, 27–54.
- [60] Yang J., Yan R. and Hauptmann A. G.. Adapting SVM Classifiers to Data with Shifted Distributions. *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, Omaha, NE, 2007, 69–76.
- [61] Syed N., Liu H., and Sung K.. Incremental learning with support vector machines. In *In Workshop on Support Vector Machines, at the IJCAI*, 1999.
- [62] Tong S. and Koller D.. Support vector machine active learning with applications to text classification. In *Proceedings of ICML*, 2000, 999–1006.
- [63] Wang H., Fan W., Yu P. S., and Han J.. Mining concept-drifting data streams using ensemble classifiers. In *Proceedings of SIGKDD*, 2003, 226–235.
- [64] Wu P. and Dietterich T. G.. Improving SVM accuracy by training on auxiliary data sources. In *Proceedings of ICML*, 2004.
- [65] Yue W., et al.. Large scale incremental learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
- [66] Shmelkov K., Schmid C., Alahari K.. Incremental learning of object detectors without catastrophic forgetting. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017, 3400–3409.
- [67] Gunning, D. et al.. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)*, nd Web 2, 2017, 2.
- [68] Rudin C.. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 2019, 206–15.
- [69] Samek W., Montavon G., Vedaldi A., Hansen L. K., Müller K. R.. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. *Springer LNCS*, 11700, 2019.



## Chapter 2

# A Localised Learning Approach Applied to Human Activity Recognition<sup>1</sup>

### Abstract

The recognition of human physical activities and postures based on sensor data has received much research attention in several human health and biomedical engineering applications. In this study, the challenges of class-imbalance and ambiguity (or confusion) are discussed that frequently arise in data from human activity recognition (HAR) systems. In order to reduce the influence of imbalance and ambiguity in HAR problems, a novel hybrid localised learning approach of K-nearest neighbours least-squares support vector machine (KNN-LS-SVM) is proposed. The classifier is applied to different synthetic and real-world datasets where imbalance and ambiguity are present. In this study, it is novel to apply a hybrid localised learning algorithm to the HAR problem. When compared to different global and local approaches, higher classification performances could be obtained by using the proposed localised learning approach. Furthermore, the computational effort could be reduced in an online learning mode.

### 2.1 Introduction

Recognising human physical activities automatically via soft computing techniques is at the core of human activity recognition (HAR) studies. Human physical activities can be recognised by using computer vision techniques through the analysis of images and videos or by exploring sensory data that are obtained by wearable or portable sensors [1]. The importance of HAR systems is illustrated by the various amount of applications where they are used, e.g. medical monitoring [2], healthcare [3], military training, and sports [4]. Moreover, the use of HAR algorithms is enhanced by rapid advancements in sensor technology that enable to monitor people in their daily environments, e.g. by using smart-phones [5] or wearables [6]. In this study, we

---

<sup>1</sup><https://doi.org/10.1109/MIS.2020.2964738>

will focus on the problem of activity classification based on accelerometer data acquired through an inertial sensor-based HAR system.

For most of the real-world problems (e.g., HAR), data is not evenly distributed in the input space [7] which is a challenge to global learning algorithms in general (e.g., SVM, conventional and deep neural networks). In particular, we will discuss the problems of *class-imbalance* and *ambiguity* that frequently arise in data obtained from HAR systems and how they can influence the performance of a classifier. Class-imbalance occurs when instances from some activities are outnumbered by others. In case of extreme imbalance, the problem of rare events occurs (e.g. the detection of falls among several daily activities [8]). Ambiguity or confusion occurs when an activity is not clearly distinguishable from another one (e.g. eating might be confused with brushing teeth due to arm motion [9]). From input space perspective, ambiguity (confusion) could be a result of the overlap between the different classes or due to a highly nonlinear decision boundary between two closely spaced classes.

In order to reduce the influence of these problems (i.e. class-imbalance and ambiguity) on HAR systems, we introduce a localised learning approach. This approach is leading to a novel hybrid algorithm which is obtained from integrating the  $K$ -nearest neighbours (KNN) algorithm into a least-squares support vector machine (LS-SVM) algorithm, namely KNN-LS-SVM. In this approach, a classification model is built for each test example using only the training examples located in the vicinity of the test example. This novel localised approach of an LS-SVM algorithm is then applied to HAR problems. Which is the first time to apply a hybrid localised learning algorithm to the HAR problem.

In addition to handling the problems of class imbalance and ambiguity, we will show that the proposed kNN-LS-SVM has other advantages as well. These advantages include simplicity of implementation that can lead to a computational advantage compared to other classifiers (e.g., deep learning neural networks). Moreover it can deal with non-linearity due to the use of the LS-SVM method as opposed to a standard kNN. Our results indicate that the KNN-LS-SVM can be a suitable approach for the HAR application especially when applied to online problems and streaming analytics for which the data size is continuously increasing.

This article is structured as follows. In Section 2, an overview of related work on classification techniques for HAR systems and local learning is given. Section 3, gives a gentle introduction to LS-SVM and KNN-LS-SVM. Performances of LS-SVM and KNN-LS-SVM are compared in Section 4 using synthetic datasets. In Section 5, KNN-LS-SVM, as well as various global and local classifiers, are applied to four real-world datasets and their performances are compared to those of a benchmark study of a deep-learning approach. The obtained results are discussed in section 6. Finally, a conclusion is presented in Sections 7.

## 2.2 Related Work

In this section, the state-of-the-art of the HAR problem and localised learning algorithms is introduced. Moreover, the benchmark study is briefly introduced.

### 2.2.1 State-of-the-art

A variety of classification algorithms have been applied to the problem of HAR, such as decision trees [10, 11], Naïve Bayes [10], Bayesian Networks [4], KNN [10, 11], convolutional neural networks [12], support vector machines (SVMs) [13, 3], and hidden Markov models (HMMs) [14]. Furthermore, deep learning approaches have recently gained much research attention [15, 16] and have been applied for HAR using low-power wearable devices [17, 18]. Moreover, transfer learning applied to deep neural networks for HAR application has recently received some attention in order to transfer models between different subjects. This approach is presented by *Renjie et al.* in their study [19] by applying the maximum mean discrepancy (MMD) algorithm to a two-layer convolutional neural network. However uncontrolled environment and online application of such an approach is still a challenge. Developing HAR models to be compatible with wearable systems is an important approach that is introduced by Cheng et al. in their work [20] namely InnoHAR model. This model is developed by concatenating convolution kernels of different scales and splicing with max-pooling layers. An important challenge that they are willing to tackle in their future work is class-imbalance in real-life human activities. For the purpose of real-time online data stream processing of HAR, the recent study by Amin et al. [21] developed a HAR model based on visual sensory data. The developed model is an optimised convolutional neural network (CNN) based model in which, deep features are extracted via a pre-trained CNN. The extracted features are fed to a deep autoencoder (DAE) to learn the temporal behaviour of the signal and finally, the classification is done via a quadratic SVM.

From the literature review, there are some challenges that need to be handled. These challenges are class-imbalance [20], and applicability to real-world online modelling [19]. These challenges are motivating us to introduce a machine learning approach that can provide a high error performance regardless of the balance degree between the available instances of different activities. In addition, a low computational and temporal cost are desired in order to be suitable for its application in an online mode. Ultimately, we aim at introducing an approach that is compatible with streaming data analysis in which the modelling complexity is not affected by the continuously increasing size of the dataset.

Localised learning algorithms have been limited studied for HAR problems. A general framework for local learning was introduced in [22], where it was demonstrated that a localised approach might be very efficient to deal with the problems of imbalance and ambiguity. Among the most common local learning approaches are KNN algorithms which have been studied intensively in the context of HAR problems, e.g. [6, 23]. Zhang et al. [24] introduced a KNN-SVM algorithm that combined a KNN with an SVM for a visual object recognition problem. In [25]

a more integrated framework, called localised support vector machine (LSVM), was introduced, which incorporates the neighbourhood information directly into SVM learning. The use of such localised approaches of SVMs, however, has not yet been studied in the context of HAR problems, nor its influence on the classification performance when imbalance or ambiguity is present.

The design of the localised hybrid algorithm KNN-LS-SVM aims to reduce the influence of imbalance and ambiguity in HAR problems. The choice of an LS-SVM classifier to be localised rather than a standard SVM is inspired by its computational advantage of solving a set of linear equations instead of solving the quadratic programming problem of standard SVM [26]. Moreover, an LS-SVM is considered as a very efficient global machine learning technique in many fields [27]. Based on synthetic and real-world data, the difference in performance between a KNN-LS-SVM and its alternative LS-SVM is illustrated for different degrees of class-imbalance and ambiguity.

### 2.2.2 Benchmark Study

In their study [18], *Ravi et al.* developed their model based on extracting shallow features in addition to deep learnt features via a CNN for HAR. The raw data from 3-axes accelerometer are segmented into time-windows with specific widths. From the extracted segments, deep learnt features and shallow features are extracted in parallel. Deep learnt features are extracted from the spectrogram of the segmented signal via the temporal convolutional layer. All extracted features, deep learnt and shallow features, are combined and fed to a fully connected layer whose output is propagated to the soft-max layer to be classified. The proposed approach in [18] outperformed other deep learning and ensemble approaches (MLP, J48 and logistic regression) [17, 28, 29] when applied to a set of published datasets [30, 31, 32, 33]. Because of its high performance and recency, the approach of [18] will be the benchmark of our study.

## 2.3 Local learning of SVMs

In this section, we start by reviewing the main concepts behind SVMs and localised approaches for SVMs. We will proceed by introducing our hybrid KNN-LS-SVM algorithm.

### 2.3.1 Support vector machines

SVMs are originally presented as binary classifiers, that assign each data instance  $\mathbf{x} \in \mathbb{R}^d$  to one of two classes described by a class label  $y \in \{-1, 1\}$  based on the decision boundary that maximises the margin  $2/\|w\|_2$  between the two classes as shown in Figure 1 [34]. Generally, a feature map  $\phi : \mathbb{R}^d \mapsto \mathbb{R}^p$  is used to transform the geometric boundary between the two classes to a linear boundary  $L : \mathbf{w}^T \phi(\mathbf{x}) + b = 0$  in feature space, for some weight vector  $\mathbf{w} \in \mathbb{R}^{p \times 1}$  and  $b \in \mathbb{R}$ . The class of each instance can then be found by  $y = \text{sgn}(w^T \phi(x) + b)$ , where  $\text{sgn}$  refers to the sign function.

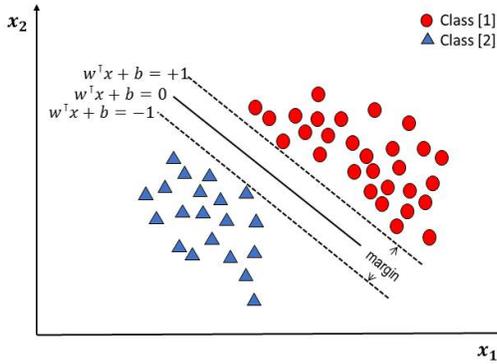


Figure 2.1: Schematic representation of a two dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised.

The estimation of the boundary  $L$  is performed based on a set of training examples  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) with corresponding class labels  $y_i \in \{-1, 1\}$ . An optimal boundary is found by maximising the margin that is defined as the smallest distances between  $L$  and any of the training instances. In particular, one is interested in constants  $\mathbf{w}$  and  $b$  that minimise a *loss-function*:

$$\min_{\mathbf{w}, b; \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N \xi_i, \quad (2.0)$$

and are subject to:

$$y_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, 2, \dots, N.$$

The constant  $C$  in (2.0) denotes the *penalty term* that is used to penalise missclassification through the slack variables  $\xi_i$  in the optimisation process.

The so-called *kernel-trick* avoids the explicit introduction of a feature map  $\phi$  and implicitly allows to use feature spaces of infinite dimensionality. A commonly used kernel is given by the Gaussian kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_0^2}\right),$$

where  $\sigma_0$  denotes the *kernel bandwidth*. Both  $\sigma_0$  and  $C$  can be optimised as hyper-parameters in a cross-validation experiment.

LS-SVMs are obtained by using a least-squares error loss function [26]:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2, \quad (2.1)$$

such that

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) = 1 - e_i, \quad i = 1, 2, \dots, N.$$

This optimisation procedure introduces errors  $e_i$  such that  $1 - e_i$  is proportional to the signed distance of  $\mathbf{x}_i$  from the decision boundary. In fact, the non-negative slack variable constraint is removed and the solution of the optimisation problem can be obtained by a set of linear equations, reducing computational effort [26].

### 2.3.2 localised LS-SVMs and KNN-LS-SVMs

In many HAR problems data are not evenly distributed in the input space. The presence of underrepresented data and severe class distribution skews affects the performance of learning algorithms that underly the HAR system [35]. Furthermore, the quality of a classifier further decreases when patterns are ambiguous, i.e., when they are not clearly belonging to one class or the other (i.e. ambiguity). Local learning approaches try to overcome such problems by building models that fit the data in the local neighbourhood around a test example and by locally adjusting to the properties of the data [22].

A well-known example of a local learning method is given by the KNN algorithm [36]. While nearest neighbours classifiers are very natural local learning methods, they suffer from the problem of high variance in the case of limited sampling. The use of a localised SVM can overcome such disadvantage as they often perform better than other classification methods in the neighbourhood consisting of a small number of examples ( $k \ll N$ ) [24].

Furthermore, the complexity of global SVMs rapidly grows as the size of training instances increases. Besides, determining the right hyperparameters (kernel width and penalty term) of these models in a cross-validation experiment is computationally expensive. Local SVMs attempt to overcome these disadvantages by building small SVM models based on data in the local neighbourhood around a test example. This computational advantage is of particular importance in an online learning mode where one is interested to cheaply update the HAR model with the additional knowledge of a new data point. When using a global model, the model has to be recomputed from scratch, while for a local model only the training instances in the vicinity of the test examples matter.

While global SVMs consider the same weight for all training instances in the optimisation process (2.1), local learning approaches allow that the training samples near a test point are more influential than others. localised approaches of SVMs [25, 37] are based on weighting functions  $\lambda(\mathbf{x}_s, \mathbf{x}_i)$  that express the similarity between the features vectors of the  $i$ -th data point  $\mathbf{x}_i$  and a test instance  $\mathbf{x}_s$ . For an LS-SVM, this leads to the following cost function:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i^2, \quad (2.2)$$

such that

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) = 1 - e_i, \quad i = 1, 2, \dots, N.$$

Weighted least-squares support vector machines [43] use a similar approach, but here a different weighting function can be used for any given test point  $\mathbf{x}_s$ . In [37] the use of continuous similarity functions were studied including the Gaussian similarity criterion given by:

$$\lambda(\mathbf{x}_s, \mathbf{x}_i) = \exp\left(-\frac{\|\phi(\mathbf{x}_s) - \phi(\mathbf{x}_i)\|_2^2}{h^2}\right),$$

where  $\|\cdot\|_2$  denotes the Euclidean norm and  $h$  denotes a bandwidth parameter to be tuned. In this work we will study a binary valued similarity criterion:

$$\lambda(\mathbf{x}_s, \mathbf{x}_i) = \begin{cases} 1 & \text{if } \|\phi(\mathbf{x}_s) - \phi(\mathbf{x}_i)\|_2 \leq r_s \\ 0 & \text{otherwise,} \end{cases} \quad (2.3)$$

where  $r_s$  is the  $K$ -th smallest distance among  $\{\|\phi(\mathbf{x}_s) - \phi(\mathbf{x}_j)\|_2; 1 \leq j \leq N\}$  which has to be tuned as a hyperparameter. This formulation leads to the hybrid KNN-LS-SVM method that we will apply on HAR problems. In particular a classification model is built for each test example using only the training examples located in the vicinity of the test example [39]. In contrast to the localised LS-SVM proposed in [37], a KNN-LS-SVM has the additional advantage of sparseness. Indeed, for an LS-SVM or the localised version that uses a continuous similarity function all input data is required to construct the separating hyperplane [43]. This can be seen by solving the optimisation problem (2.1). Using the method of the Lagrangian multipliers, we find:

$$\mathcal{L}(w, b, e; \alpha) = \frac{1}{2}\|w\|_2^2 + \frac{1}{2}\gamma \sum_{i=1}^N \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i^2 - \sum_{i=1}^N \alpha_i (y_i[\mathbf{w}^\top \phi(\mathbf{x}_i) + b] - 1 + e_i), \quad (2.4)$$

where  $\alpha_i$  are the *Lagrangian* multipliers. The optimality conditions are found by setting the first order partial derivatives to zero:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^N \alpha_i y_i \phi(\mathbf{x}_i), \\ \frac{\partial \mathcal{L}}{\partial b} = 0 &\Rightarrow \sum_{i=1}^N \alpha_i y_i = 0, \\ \frac{\partial \mathcal{L}}{\partial e} = 0 &\Rightarrow \alpha_i = \gamma \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i, \\ \frac{\partial \mathcal{L}}{\partial \alpha} = 0 &\Rightarrow y_i (w^\top \phi(\mathbf{x}_i) - b) = 1 - e_i, \forall 1 \leq i \leq N. \end{aligned}$$

From the third condition, it is clear that the support values  $\alpha_i$  are weighted by the similarity function and are zero when  $\lambda(\mathbf{x}_s, \mathbf{x}_i) = 0$ . Thus, for a KNN-LS-SVM the sparseness charac-

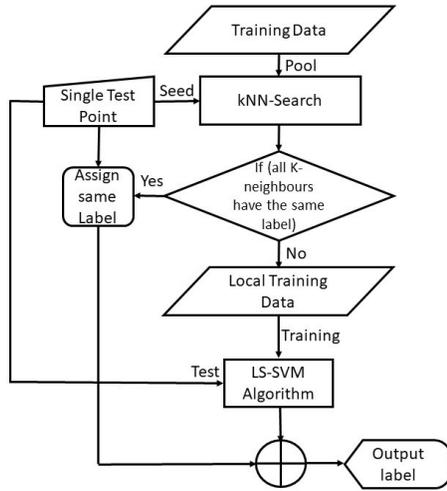


Figure 2.2: A flow chart illustrating the localised learning algorithm of KNN-LS-SVM.

teristic is returned to the LS-SVM. In an online learning mode, this sparseness will result in a computational advantage compared to LS-SVM, as we will show in Section 2.5.

As shown in Figure 2.2, the algorithm of KNN-LS-SVM is implemented as follows:

1. Given a test example  $\mathbf{x}_s$ , compute distances to all training examples and pick the nearest  $K$  neighbours;
2. If all  $K$  neighbours would have the same label, assign the same label to  $\mathbf{x}_s$ .
3. Else, train the LS-SVM model with the  $K$  nearest neighbours.
4. Use the resulting classifier to label  $\mathbf{x}_s$ .

The parameter  $K$  and the distance metric (e.g. Euclidean, Mahalanobis or Chebyshev) are additional hyperparameters next to the kernel width  $\sigma_0$  and the penalty term  $\gamma$  that are optimized in a cross validation approach.

## 2.4 Simulation experiments

The objective in this section is to present the problems of class-imbalance and ambiguity, that frequently arise in HAR problems, with controlled synthetic datasets. The generated synthetic datasets are two dimensional in order to simplify and visualise the problems which is not possible with high dimensional real-world datasets.

In this section, the performance of our KNN-LS-SVM method is compared with that of a global LS-SVM using three synthetic datasets. Two challenges are presented where global

classifiers perform suboptimally: class-imbalance and ambiguity (or confusion). This is the first time to illustrate the problems of class-imbalance and ambiguity with synthetic data in the context of localised learning. Results are presented using the  $F_1$  score that is defined as the harmonic average of precision and recall:

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

In a two-class setting, recall and precision are defined as follows:

1. Recall is the ratio of instances that are correctly classified as positive to all positive instances;
2. Precision is the ratio of the instances that are correctly classified as positive to all instances classified as positive.

In case of multi-class setting, recall and precision are calculated based on one-vs-all approach (one class is positive and all other classes are negative).

The training of a local learning model requires the training of a local model for each individual test point. The selection of the hyperparameters of these local models is based on a cross-validation experiment where the accuracy (i.e. the ratio of correctly classified instances) is maximised. Accuracy-based model selection can handle the different distributions of the classes that can be present in a local region [32, 31]. In contrast with the global learning algorithms where the  $F_1$ -score is recommended for hyperparameters selection in case of class-imbalance, accuracy provides a balanced performance to the local learning algorithms as the majority/minority ratios changes in the local scale.

### 2.4.1 Class-imbalance

Learning from imbalanced data is still a focus of intense research, treating the problem of skewed class-distributions [40, 41]. It occurs when representatives of some classes appear much more frequently which poses a difficulty for learning algorithms, as they will be biased towards the majority group. In this section we study the use of a local learning method to deal with such imbalance.

A synthetic dataset is constructed consisting of data generated from two planar Gaussian distributions  $X_+$  and  $X_-$ , that respectively represent a positive and a negative class, see Figure 2.3a. The distributions are centred at respectively  $\mathbf{m}_+ = (\frac{1}{2}, \frac{1}{2})$  and  $\mathbf{m}_- = (-\frac{1}{2}, -\frac{1}{2})$ , with identical isotropic covariance matrices  $\Sigma_+ = \Sigma_- = 0.35I_2$ , where  $I_2$  denotes the identity matrix in  $\mathbb{R}^{2 \times 2}$ . Experiments were performed where a number of  $N = 400$  instances were simulated and the percentage  $p_+$  of instances in the positive class  $X_+$  varied in the range  $p_+ \in \{50\%, 25\%, 12.5\%, 5, 2.5\%, 1.25\%\}$  with a number of instances  $\{200, 100, 50, 20, 10, 5\}$  respectively. Both, an LS-SVM and a KNN-LS-SVM were trained with a Gaussian kernel. The models depend on hyperparameters (kernel width, penalty term and number of neighbours), the

value of which are estimated in a 10-fold cross-validation experiment, where 80% of the simulated instances were used for training and 20% for testing. Figure 2.4a shows the  $F_1$ -scores averaged over the folds as a function of the imbalance percentage.

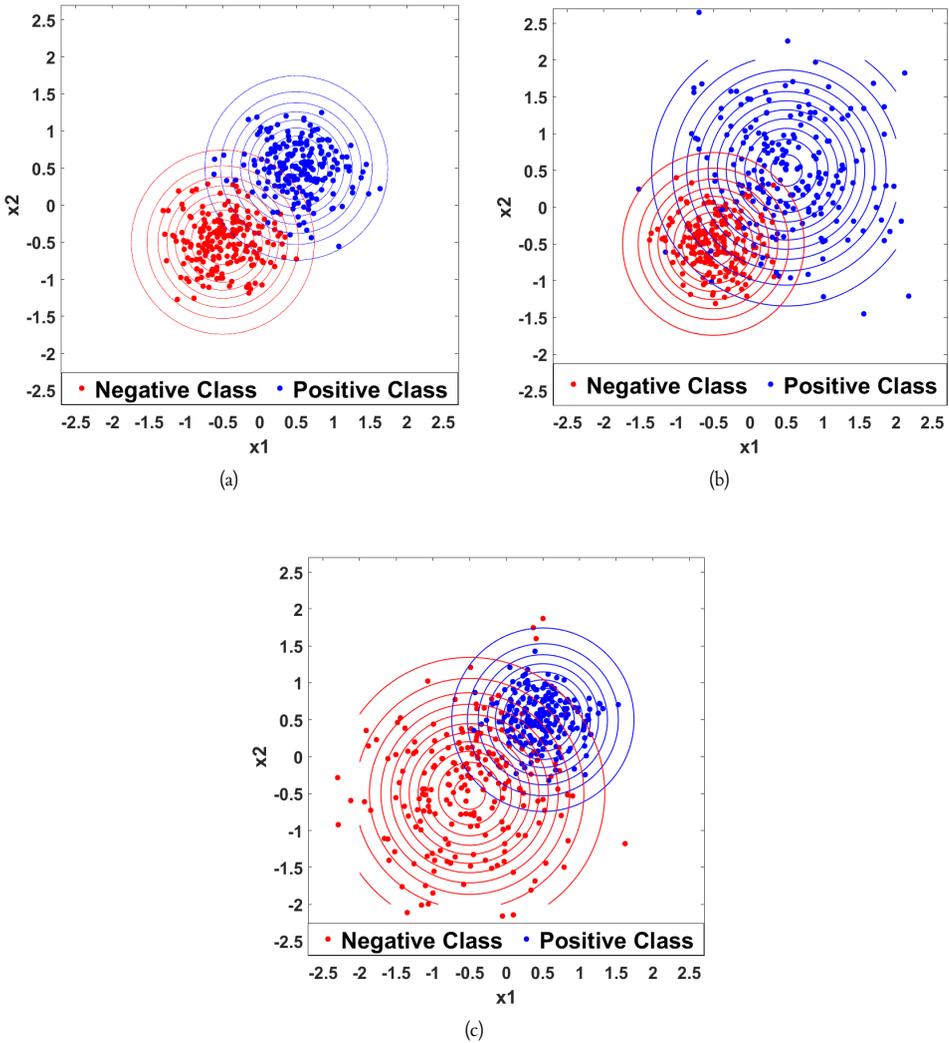
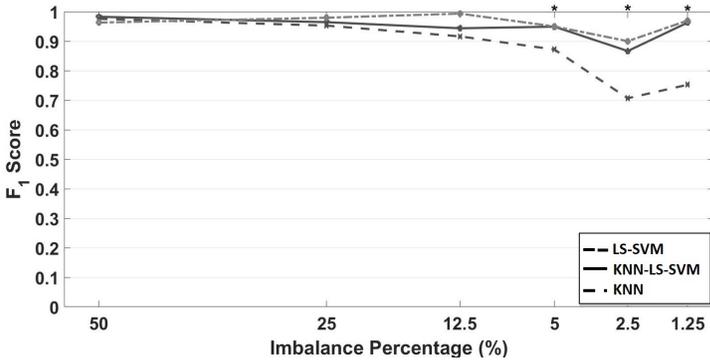


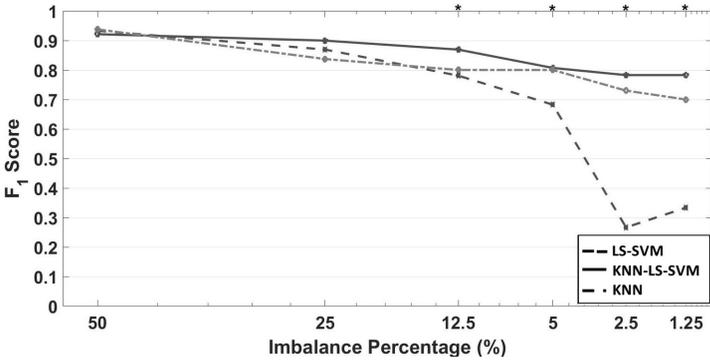
Figure 2.3: Two-dimensional datasets consisting of two classes with data generated from Gaussian distributions  $X_+ \sim N(\mathbf{m}_+, \Sigma_+)$  and  $X_- \sim N(\mathbf{m}_-, \Sigma_-)$  with  $\mathbf{m}_+ = (\frac{1}{2}, \frac{1}{2})$ ,  $\mathbf{m}_- = (-\frac{1}{2}, -\frac{1}{2})$  and different covariance matrices: (a)  $\Sigma_+ = \Sigma_- = 0.35I_2$ , (b)  $\Sigma_+ = 2\Sigma_- = 0.70I_2$  and (c)  $\Sigma_+ = \frac{1}{2}\Sigma_- = 0.35I_2$ .

Clearly, the classifiers perform equally well when classes are balanced. However when there is class-imbalance the localised KNN-LS-SVM outperforms the global LS-SVM. Furthermore, the difference in performance tends to increase with an increasing degree of class-imbalance. At the percentages 50%, 25%, 12.5%, 5%, 2.5%, and 1.25%, the use of a KNN-LS-SVM results in a mean increasing difference in  $F_1$ -score of respectively 0.71%, 1.17%, 2.74%, 7.74%, 16.00% and 19.68%. A statistical comparison of the  $F_1$  scores using a paired t-test resulted in one-sided p-values: 0.4200, 0.4000, 0.2000, 0.0219, 0.0278, and 0.0276 respectively which show that the differences corresponding to an imbalance percentage at and above 5% were statistically significant with a significance level of 0.05.

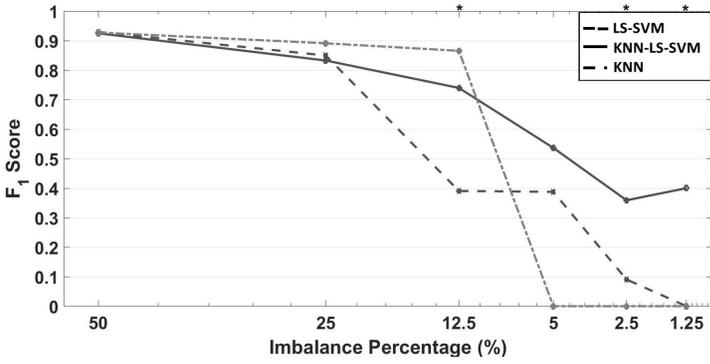
The performance in case of class-imbalance can be further studied by changing the intra-class variance while fixing the inter-class variance leading to overlapping classes as shown in Figures 2.3b and 2.3b. Figures 2.4a, 2.4b, and 2.4c show respectively the difference in  $F_1$ -scores when  $\Sigma_+ = 2\Sigma_- = 0.70I_2$  (i.e. the variances of the positive class are as twice as large than the variances of the negative class) and when  $\Sigma_+ = \frac{1}{2}\Sigma_- = 0.35I_2$  (i.e. the variances of the positive class are as half as large than the variances of the negative class). Clearly, the overall performance of both classifier decreases, when compared to the case where  $\Sigma_+ = \Sigma_-$ . The localised method, however, still outperforms the global LS-SVM for lower percentages  $p_+$ . In case,  $\Sigma_+ = 2\Sigma_-$ , there is a difference in  $F_1$ -score of  $-1.17\%$ ,  $3.30\%$ ,  $8.82\%$ ,  $12.45\%$ ,  $51.59\%$ , and  $44.93\%$  at the percentages 50%, 25%, 12.5%, 5%, 2.5%, and 1.25% respectively. A paired t-test showed that the differences were significant at the 0.05 level for imbalance percentages above 25% (with one-sided p-values 0.1310, 0.1300, 0.0236, 0.0296, 0.0204, and 0.0495 at the percentages 50%, 25%, 12.5%, 5%, 2.5%, and 1.25% respectively). In case,  $\Sigma_+ = \frac{1}{2}\Sigma_-$ , the performance curves of both classifiers remarkably decrease due to the overlap between the two classes. However, the KNN-LS-SVM still outperforms the LS-SVM with a difference in  $F_1$ -score of 0.48%, 1.72%, 34.92%, 14.91%, 26.85%, and 40.00% at the percentages 50%, 25%, 12.5%, 5%, 2.5%, and 1.25% respectively. The one-sided p-values of a paired t-test are 0.3500, 0.2445, 0.0034, 0.1498, 0.0160, and 0.0088 at the different imbalance percentages respectively. Thus, only at imbalance percentages 12.5%, 2.5% and 1.25% a significant difference at the 0.05 level was found. The significance is missed at the imbalance percentage of 5%, however the outperformance of the KNN-LS-SVM still present. Conclusively, the performance of the classifiers is not only influenced by the class-distribution imbalance but also by the intra-class variance and besides the inter-class variance in the input space. Moreover, as shown in Figures 2.4a, 2.4b, and 2.4c, the kNN classifier shows a similar or a better performance than both kNN-LS-SVM and LS-SVM especially for the cases of non-extreme class imbalance. This can be interpreted in light of the importance of locality since kNN considers the local properties of the data points. However, kNN is not expected to handle complex nonlinear and overlapping patterns.



(a)



(b)



(c)

Figure 2.4: (a) Averaged  $F_1$ -scores of a KNN-LS-SVM, KNN, and an LS-SVM obtained from a 10-fold cross-validation experiment using the synthetic dataset shown in Figure 2.3a and using different percentages  $p_+$  of instances from the positive class. (b)  $F_1$ -scores of KNN-LS-SVM, KNN, and LS-SVM at different degrees of imbalance when  $\Sigma_+ = 2\Sigma_- = 0.70I_2$ . (c)  $F_1$ -scores of KNN-LS-SVM, KNN, and LS-SVM at different degrees of imbalance when  $\Sigma_+ = \frac{1}{2}\Sigma_- = 0.35I_2$  and averaged over the runs of a 10-fold cross-validation experiment. The labels (\*) indicated on top of the horizontal axis refer to the imbalance percentages where the difference in performance scores between KNN-LS-SVM and LS-SVM is statistically significant at the 0.05 level.

### 2.4.2 Ambiguity

Ambiguity (or confusion) arises when regions exist in data space that are occupied by more than one class or when classes are very closely spaced [42]. In such cases, a global classifier will fit highly nonlinear boundaries that can become very complex on input space. A local learning algorithm attempts to locally adjust the complexity of the boundary to the properties of the data in each area of the input space.

In this section, we will study a simulated experiment where ambiguity occurs proportionally with the class-imbalance due to the discontinuity of the different classes' patterns and that is inspired from a general representation of Bottou & Vapnik [22], see Figure 2.5a. The data of 1000 instances is generated by mapping two variables  $x_1$  and  $x_2$  that are distributed according to a standard normal distribution  $N(0, 1)$  to an univariate score  $z = \sin(x_1) * \sin(x_2) + x_1$ . By setting specific ranges on the distribution of  $z$  different degrees of imbalance can be achieved. Where the percentage  $p_+$  of instances in the positive class  $X_+$  varied in the range  $p_+ \in \{50\%, 33\%, 20\%, 10\%, 5\%, 2.5\%, 1.25\%\}$  with number of instances  $\{500, 333, 200, 100, 50, 25, 12\}$  respectively. To simulate the set of  $p_+$  of instances the range on  $z$  was chosen as:

$$0.5 < |z| < b, \text{ such that } P(0.5 < |z| < b) = p_+.$$

Hence, by varying  $b$ , the width of the positive class pattern and the number of the positive instances varies proportionally. Figure 2.5a shows an example of the boundary between the classes for  $p_+ = 50\%$ . The positive class is scattered in two stripes and surrounded by negative observations on both sides.

Figure 2.5b shows a comparison of the  $F_1$ -scores between an LS-SVM and a KNN-LS-SVM algorithm applied on the data set. For each percentage  $p_+$ , a 10-fold cross-validation experiment, similar as in Section 2.4.1, was performed. At the percentages 50%, 33%, 20% and 10%, the performance of both local and global classifiers is more or less the same. However, the use of a KNN-LS-SVM at the imbalance percentages of 5%, 2.5% and 1.25% lead to an increase in  $F_1$ -score of 10.86, 19.28 and 20.05% respectively. A paired t-test showed that these differences were all significant at a 0.05 levels (with one-sided p-values below 0.05) with one-sided p-values of 0.0245, 0.0150 and 0.0100 for  $p_+ \in \{5\%, 2.5\%, 1.25\%\}$  respectively. Clearly, the  $F_1$ -score rapidly decreases as the imbalance percentage increases. For higher imbalance percentages, the number of available instances from the positive class decreases making it hard for both classifiers to model the complex boundary between the classes. Ultimately, as shown in Figures 2.4 and 2.5, the KNN-LS-SVM can handle overlapping and nonlinearity problems together with class imbalance more efficiently than standard KNN.

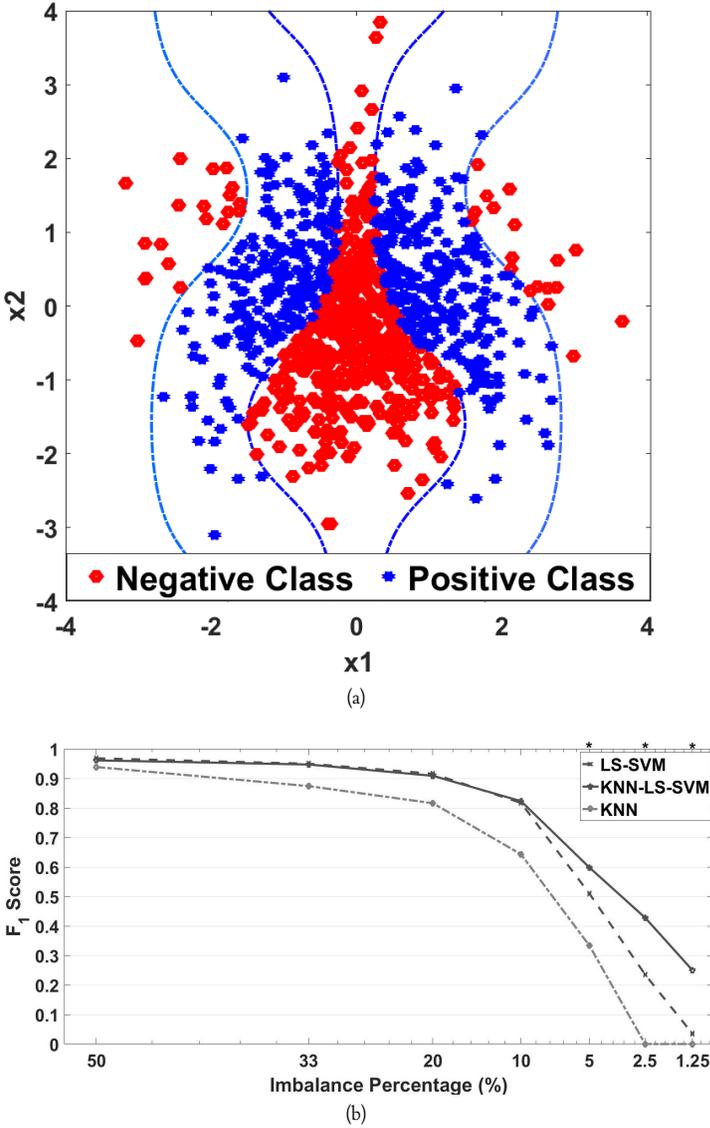


Figure 2.5: (a) Two-dimensional non-linearly separable dataset of two classes where  $p_+ = 50\%$ . (b)  $F_1$ -scores of the KNN-LS-SVM, KNN, and LS-SVM averaged over the runs of a 10-fold cross-validation test experiment. Performance scores are obtained by using the synthetic dataset describing confusion where the degree of imbalance is varied in the range  $\{50\%, 33\%, 20\%, 10\%, 5\%, 2.5\%, 1.25\%\}$ . The labels (\*) indicated on top of the horizontal axis refer to the imbalance percentages where the difference in performance scores between KNN-LS-SVM and LS-SVM is statistically significant at the 0.05 level.

## 2.5 Real-world Data

As illustrated in section IV, class-imbalance is considered a challenge to global learning algorithms under some conditions such as the imbalance ratio, the overlap between different classes, inter-class and intra-class variances, and ambiguity.

In this section, four real-world datasets are used to compare the performances of a KNN-LS-SVM to an LS-SVM [43], KNN, KNN-SVM [24], Profile SVM (PSVM) [25], Stacked Autoencoders (Stack-AE) [44]. Moreover, we compare our results with the results of a recent study of HAR that is based on the use of deep learning techniques [18]. Through the rest of the study, we will refer to the proposed algorithm of the benchmark study [18] as *Ravi (2017)*. Finally, the time performance of a KNN-LS-SVM is compared to those of the already mentioned classifiers except for *Ravi (2017)* due to the dedicated platform used in their study [18].

### 2.5.1 datasets

**WISDM v1.1** The first dataset that we will study has been used to evaluate a system that uses phone-based accelerometers to perform HAR [32]. Several activities were recorded with different frequencies of occurrence: walking (38.6%), jogging (31.2%), walking upstairs (11.2%), walking downstairs (9.1%), sitting (5.5%), and standing (4.4%). Activities of 36 subjects were recorded using an impeded accelerometer of a smartphone with a sampling rate of 20 Hz and that was located in the front pocket. In this way, a total number of 1,048,576 samples were acquired within approximately 14.56 recording hours. In a preprocessing phase, features were extracted as will be discussed later using a non-overlapping sliding window of 10 seconds.

**Daphnet FoG** This dataset contains annotated readings of 3 accelerometers attached to Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks [31]. Since freezing of gait occurs rarely compared to other movement activities, the data is very imbalanced. Only 1/9 of all recorded instances corresponded to the freezing-class. Sensors were attached to the shank (just above the ankle) and the thigh (just above the knee) using an elasticised strap and Velcro. A third sensor was attached to the lower back via a belt. The number of patients in this study is 10. The sampling rate of the accelerometers recordings was 64 Hz and the total number of the acquired samples is 1,917,887 within approximately 8.32 recording hours. The features were extracted as will be discussed later from non-overlapping sliding windows of length 4 seconds.

**WISDM v2.0** This dataset is used to evaluate a system that uses phone-based accelerometers to perform HAR [33]. Several activities were recorded with different frequencies of occurrence: walking (42.1%), jogging (14.7%), sitting (22.3%), standing (9.7%), Lying down (9.3%), and stairs (1.9%). Activities of 563 subjects were recorded using an impeded accelerometer of a smartphone with a sampling rate of 20 Hz and that was located in the front pocket. In this way, a total number of examples 2,980,765 were acquired. In a preprocessing phase, features were extracted as will be discussed later using a non-overlapping sliding window of 10 seconds.

**Skoda** This dataset contains 10 manipulative gestures (classes) performed in a car maintenance scenario [30]. They are a subset of the 46 activities performed in the factory in one of the quality control checkpoints. Data is collected from one subject, with a sampling rate of 98 Hz. For comparison purpose, the accelerometer signals from one node are used (Node 16). The total number of samples is approximately 705,440 samples. The ten classes will be shown in *Classification Performance* section.

### 2.5.2 Classification performance

We compare the classification performances of the KNN-LS-SVM to the LS-SVM, KNN, KNN-SVM, PSVM, Stacked Autoencoders (Stack-AE) and *Ravi (2017)*. For error performance evaluation, precision, recall and  $F_1$  – score, are presented for the recognition of the activities present in the studied datasets. To make a consistent comparison with the recent study of HAR that is based on a deep learning approach, we use the same set of features of method *Ravi (2017)* [18] to train the KNN-LS-SVM, LS-SVM, KNN, KNN-SVM, PSVM, and Stack-AE. Several features are used: interquartile range, amplitude kurtosis, root mean square, variance, mean, standard deviation, skewness, minimum, median, maximum, mean-cross, and zero-cross. All these features are used similar to the benchmark studies for a consistent comparison.

Tables 2.1, 2.2, 2.3, and 2.4 show the recall/precision scores of the KNN-LS-SVM, LS-SVM, KNN, KNN-SVM, PSVM, Stack-AE and *Ravi (2017)* applied to the datasets WISDM v1.1, Daphnet FoG, WISDM v2.0 and Skoda respectively. Figures 2.6, 2.7, 2.8, and 2.9 show the  $F_1$ -scores of the classifiers applied to the real-world datasets.

Table 2.1: Classification results of the KNN-LS-SVM, LS-SVM, *Ravi (2017)*, KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1.

		Walk	Jog	Sit	Stand	Walk Up	Walk Down
<b>KNN-LS-SVM</b>	Recall	99.36	96.70	97.97	99.04	93.88	95.34
	Precision	96.99	99.73	99.18	98.10	95.47	96.41
<b>LS-SVM</b>	Recall	99.51	99.87	96.25	96.19	94.22	90.39
	Precision	98.68	99.87	96.65	96.65	93.55	96.06
<b>Ravi (2017)</b>	Recall	99.37	99.40	98.56	97.25	95.13	95.90
	Precision	99.37	99.64	97.85	98.15	95.52	94.44
<b>KNN</b>	Recall	100	100	100	95.2	96.7	100
	Precision	98.9	100	96.77	100	100	100
<b>KNN-SVM</b>	Recall	82.7	89.12	86.40	95.23	80.36	74.00
	Precision	83.1	96.32	95.00	100	65.22	70.83
<b>PSVM</b>	Recall	78.05	96.15	96.30	100	55.56	56.41
	Precision	79.60	96.77	100	90.05	47.62	64.71
<b>Stack-AE</b>	Recall	98.50	98.70	90.30	89.50	87.30	74.30
	Precision	96.60	99.40	87.50	89.50	82.80	92.9

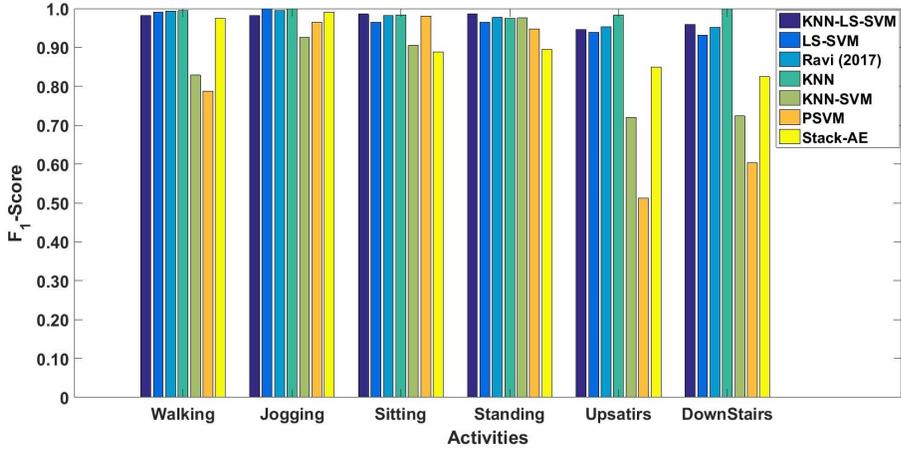


Figure 2.6:  $F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1

Table 2.2: Classification results of the KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Daphnet FoG.

		Non-Freezing	Freezing
<b>KNN-LS-SVM</b>	Recall	97.79	72.92
	Precision	97.18	77.55
<b>LS-SVM</b>	Recall	98.31	62.66
	Precision	95.96	79.89
<b>Ravi (2017)</b>	Recall	98.15	59.92
	Precision	97.40	67.89
<b>KNN</b>	Recall	98.88	82.85
	Precision	98.34	82.32
<b>KNN-SVM</b>	Recall	97.80	60.00
	Precision	96.23	72.00
<b>PSVM</b>	Recall	91.94	54.72
	Precision	94.63	43.88
<b>Stack-AE</b>	Recall	97.00	70.12
	Precision	96.6	72.93

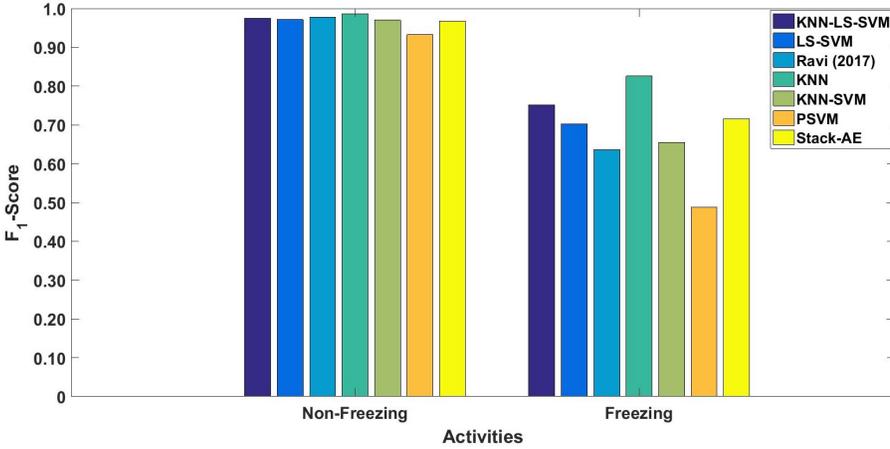


Figure 2.7:  $F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Daphnet FoG

Table 2.3: Classification results of the KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v2.0.

		Walk	Jog	Sit	Stand	Lying Down	Stairs
<b>KNN-LS-SVM</b>	Recall	97.00	98.00	91.79	78.00	85.82	95.00
	Precision	96.60	97.41	86.62	92.00	88.80	97.96
<b>LS-SVM</b>	Recall	97.94	94.39	89.00	63.23	82.48	88.90
	Precision	95.96	98.54	80.59	92.45	83.09	88.90
<b>Ravi (2017)</b>	Recall	97.19	97.73	89.28	82.11	85.80	76.98
	Precision	97.17	98.01	87.32	82.05	88.65	85.00
<b>KNN</b>	Recall	96.53	92.83	82.46	66.20	42.45	61.76
	Precision	91.84	96.92	69.80	75.81	75.64	70.00
<b>KNN-SVM</b>	Recall	97.96	97.29	84.79	74.63	86.99	96.15
	Precision	96.64	98.17	88.81	79.37	76.98	96.17
<b>PSVM</b>	Recall	83.54	71.36	51.00	32.12	76.19	32.00
	Precision	90.88	56.72	72.11	70.97	34.78	15.69
<b>Stack-AE</b>	Recall	95.43	95.75	76.01	70.31	80.39	29.63
	Precision	92.79	96.21	79.22	73.77	70.29	100

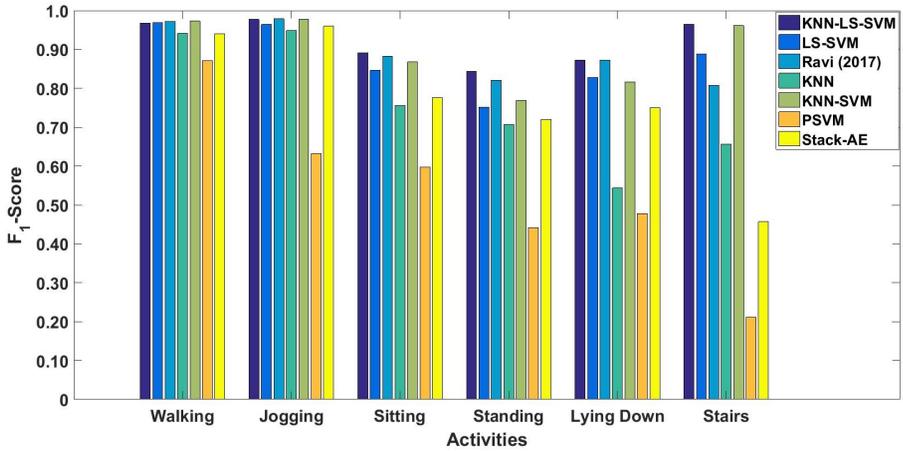


Figure 2.8:  $F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v2.0

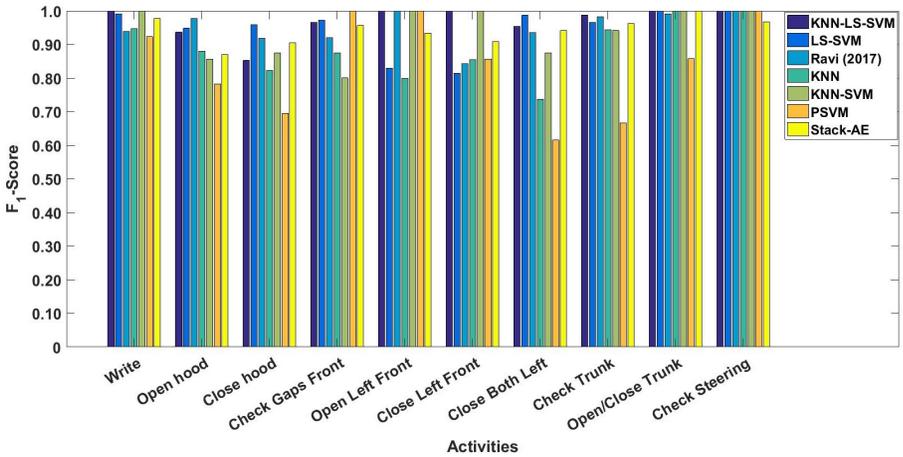


Figure 2.9:  $F_1$ -Scores of the classifiers KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Skoda dataset

Table 2.4: Classification results of the KNN-LS-SVM, LS-SVM, Ravi (2017), KNN, KNN-SVM, PSVM, and Stack-AE applied to Skoda dataset.

<i>KNN-LS-SVM</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	100	93.62	88.90	100	100
	Precision	100	93.62	81.84	93.33	100
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	100	91.30	97.50	100	100
	Precision	100	100	100	100	100
<i>LS-SVM</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	100	96.97	95.08	94.59	89.47
	Precision	98.33	92.75	96.67	100	77.27
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	73.33	97.37	96.55	100	100
	Precision	91.67	100	96.55	100	100
<i>Ravi (2017)</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	91.34	97.78	94.44	92.79	100
	Precision	96.67	97.78	89.47	91.15	100
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	80.00	94.20	97.59	98.04	100
	Precision	88.89	92.86	98.78	100	100
<i>KNN</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	90.00	91.67	82.35	87.50	80.00
	Precision	100	84.62	82.35	87.50	80.00
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	85.1	77.80	94.44	100	100
	Precision	85.71	70.00	94.44	100	100
<i>KNN-SVM</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	100	100	87.5	66.67	100
	Precision	100	75.00	87.50	100	100
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	100	77.78	100	100	100
	Precision	100	100	88.90	100	100
<i>PSVM</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	100	100	80.00	100	100
	Precision	85.71	64.29	61.54	100	100
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	75.00	44.44	50.00	90.00	100
	Precision	100	100	100	81.82	100
<i>Stack-AE</i>		Write	Open hood	Close hood	Check gaps Front	Open Left Front
	Recall	100	86.96	88.89	91.67	100
	Precision	95.65	86.96	92.31	100	87.50
		Close Left Front	Close Both Left	Check Trunk Gaps	Open and Close Trunk	Check Steering
	Recall	83.33	94.12	96.30	100	100
	Precision	100	94.12	96.30	100	93.75

### 2.5.3 Time Performance

To compare the time performance between an LS-SVM and KNN-LS-SVM, we make use of a non-dedicated platform (i.e. System Type x64-based PC, Processor Intel(R) Core(TM) i7-6820HQ CPU @ 2.70GHz, 2701 MHz, 4 Core(s), 8 Logical Processor(s), Installed Physical Memory (RAM) 8,00 GB). We compare the elapsed time that is required to run the algorithms in an online mode. In an online mode instances are evaluated one by one. For global classifiers, we measure the test time of one instance as in practice one model is trained and applied for all test instances. However, for the localised models (i.e. KNN, KNN-SVM and KNN-LS-SVM), the measured time includes both training and testing time. For PSVM, local models are trained offline by the training data points of each profile that result from partitioning the training set. Hence, the measured time for PSVM is test time only. The elapsed test times for each classifier applied to each dataset are depicted in Table 2.5.

Table 2.5: The time performance of the classifiers KNN-LS-SVM, LS-SVM, KNN, KNN-SVM, PSVM, and Stack-AE applied to WISDM v1.1, Daphnet FoG, WISDM v2.0 and Skoda Datasets. The depicted results represent the consumed time in seconds to classify a single test point.

	WISDM v1.1	Daphnet FoG	WISDM v2.0	Skoda
KNN-LS-SVM	<b>0.0126</b>	<b>0.0014</b>	<b>0.0089</b>	<b>0.0035</b>
LS-SVM	0.0321	0.0062	0.0887	0.2283
KNN	25.32	34.00	19.77	16.822
KNN-SVM	19.78	14.017	20.10	12.88
PSVM	0.0290	0.0032	0.0111	0.0036
Stack-AE	0.0166	0.0171	0.0205	0.0248

## 2.6 Discussion

Applying the various classifiers on WISDM v1.1, we obtain the results shown in Table 2.1 and Figure 2.6. Notice that class-imbalance is not dominant in this dataset, except for *Sitting* and *Standing* activities (5.5 and 4.4% respectively). The best performance is that of the KNN classifier for all activities except for *Standing*. The main drawback of the KNN classifier is the temporal complexity to get the optimum hyperparameters (i.e. K-number and distance metric). To obtain such good results, as shown in Table 2.5, it takes approximately 25 seconds for one test point which is 2500 times the required time by KNN-LS-SVM. Moreover, KNN-LS-SVM provides the best performance for the minority classes of *Sitting* and *Standing*. We can notice here that KNN-LS-SVM is providing a performance that compromises between the superior performance of KNN with an expensive temporal complexity and LS-SVM with an acceptable error and time performance. The algorithm of KNN-LS-SVM is relying on KNN, but both the number  $K$  and the distance metric are globally optimised independently of the test set. For this dataset, we can

claim that the error performance of KNN-LS-SVM is comparable in case of balanced activities and better for unbalanced activities. Moreover, KNN is not applicable to the online application as 25 seconds to classify a single point is more than the window size of 10 seconds that is used in this dataset.

For Daphnet FoG dataset, the class-imbalance is the dominant characteristic with a class-imbalance ratio of 1 to 9. As shown in Figure 2.7 and both Table 2.1 and Table 2.5, KNN provides the best error performance and worst time performance due to hyperparameter optimisation. The second best error performance is the one of KNN-LS-SVM with best time performance which is  $4 * 10^{-4}$  times that of KNN. The average run-time of KNN to classify one test point is 34 seconds which is not applicable for online classification as the window size for this dataset is only 4 seconds.

For WISDM v2.0, KNN-LS-SVM approximately provides the best error and time performance over all activities especially the extremely minor activity of Stairs (1.9%). The only competitive classifier is the KNN-SVM. However, its temporal and computational complexity is much higher than that of the KNN-LS-SVM. The superiority of the KNN-SVM and KNN-LS-SVM which are both based on an RBF kernel can be due to the presence of strong nonlinearity and overlapping classes in this dataset.

By applying the various classifiers to the Skoda dataset with 10 classes, KNN-LS-SVM provides the best performance for 6 classes. This dataset does not suffer class-imbalance such that KNN-LS-SVM is competing with the other global and local classifiers in the error performance. However, it outperforms the other classifiers in the time performance.

Ultimately, from Table 2.5, it is obvious that KNN-LS-SVM provides the best time performance over all real-world datasets compared to the other classifiers which are implemented on the same platform while at the same time providing robustness against class imbalance.

## 2.7 Conclusion

In this paper, we discussed the problems of class-imbalance and ambiguity that frequently arise in data obtained from HAR systems. A novel hybrid localised learning approach of KNN-LS-SVM is proposed to tackle these problems. Moreover, for the first time, these problems are analysed with synthetic datasets in the context of localised learning with a detailed illustration of the KNN-LS-SVM algorithm. In contrast to the already existing literature on HAR that mainly focuses on the use of global learning methods, we applied for the first time a hybrid localised learning algorithm to the problem of HAR. Furthermore, we compared the performance of the KNN-LS-SVM with other global and local learning techniques and the benchmark study of [18].

A localised method has the advantage to locally adjust the complexity of the decision boundary to the properties of the data in each area of the input space. The choice of LS-SVM instead of a standard SVM to be localised was motivated by the relatively computational simplicity of the LS-SVM compared to SVM [43]. This choice was further supported by the increased time

performance in an online mode such that the KNN-LS-SVM has much potential to be suitable for online and streaming analytics problems in which the data size is continuously increasing.

Experiments using the synthetic data showed that the local classifier (i.e. KNN-LS-SVM) can be more robust against class-imbalance and ambiguity compared to a global classifier (i.e. LS-SVM). This was also confirmed by our experiments on the real-world data sets where the highest difference in the performance was obtained when the class-distribution was highly skewed.

The proposed algorithm is applicable to HAR applications as human health monitoring, e.g. fall detection, independent living of elderly, freezing of gait detection for Parkinson's patients, among others [45, 46, 31]. Experiments with real-world data illustrated the potential of the use of the localized approach for online and streaming analytics problems in HAR applications especially when applied to middle-sized data sets.

Moreover, the proposed localised approach has application potential to medical diagnostic problems that can suffer from class-imbalance problems, e.g. abnormality detection via screening, cancerous cells detection and Hyperthyroid diagnosis [47, 48, 49].

Finally, the proposed localised algorithm outperformed the benchmark global models under the following data-based conditions:

- The data sets contained up to 15,000 data points.
- The imbalance percentages  $p_+$  where the KNN-LS-SVM outperformed the LS-SVM ranged from 1.25% to 12.5%.
- Comparable performances were found for percentages  $p_+$  given by 25% and 50%.
- The Fisher discriminant ratio that describes the overlap between classes could take values up to 4.08.

The Fisher discriminant ratio is defined as:

$$f = \frac{(m_+ - m_-)^2}{(\sigma_+^2 + \sigma_-^2)},$$

where  $m_+, m_-$  are the means of the positive and negative classes respectively. And  $\sigma_+, \sigma_-$  are the standard deviations of the positive and negative classes respectively.

In future research, we plan to apply and validate the method further in the context of real-time activity tracking of hospitalised patients.



# Bibliography

- [1] Bulling A., Blanke U., and Schiele B.. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 1, 2014, 1-33.
- [2] Clifton L., Clifton D., Pimentel M., Watkinson P., and Tarassenko L.. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. *IEEE Journal of Biomedical and Health Informatics*, 18, 2014,722-730.
- [3] Luca S., Vuegen L., Van hamme H., Karsmakers P., and Vanrumste B.. Decision support systems for home monitoring applications: Classification of activities of daily living and epileptic seizures. *London: IET, Institution of Engineering and Technology, ch. 13, 2016, 271-291.*
- [4] Lara O. and Labrador M.. A mobile platform for real time human activity recognition. *in Proc. IEEE Conference on Consumer Communications and Networks, 2012.*
- [5] Morales J. and Akopian D..Physical activity recognition by smartphones, a survey. *Bio-cybernetics and Biomedical Engineering*, 37(3), 2017,388-400.
- [6] Attal F. ,Mohammed S., Dedabrishvili M., Chamroukhi F., Oukhellou L., and Amira Y..Physical human activity recognition using wearable sensors. *Sensors*, 15(12), 2015,314-338.
- [7] Vapnik V. and Bottou. Local algorithms for pattern recognition and dependencies estimation. *Neural Computation*, 5, 1993,893-909.
- [8] Baldewijns G., Debarde G., Mertes G., Vanrumste B., and Croonenborghs T.. Bridging the gap between real-life data and simulated data by providing a highly realistic fall dataset for evaluating camera-based fall detection algorithms. *Healthcare Technology Letters*, 3(1),2016, 6-11.
- [9] Lara O. and Miguel A. L.. A survey on human activity recognition using wearable sensors.*IEEE Communications Surveys and Tutorials*, 15, 2013, 1192-1209.

- [10] Jatoba L., Grosmann U., Kunze C., Ottenbacher J., and Stork W.. Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, **2008**, 5250–5253.
- [11] Muarer U., Smailagic A., Siewiorek D., and Deisher M., Activity recognition and monitoring using multiple sensors on different body positions. in *Proc. International Workshop on Wearable and Implantable Body Sensor Networks, Washington, DC, USA*, **2006**.
- [12] Ming Z., Nguyen L. T., Bo Y., Ole J., Jiang Z., Pang W., and Joy Z.. Convolutional neural networks for human activity recognition using mobile sensors. in *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services*, **6**, **2014**.
- [13] He Z.-Y. and Jin L.-W.. Activity recognition from acceleration data using ar model representation and svm. in *International Conference on Machine Learning and Cybernetics*, **2008**, 2245–2250.
- [14] Zhu C. and Sheng W.. Human daily activity recognition in robot assisted living using multi-sensor fusion. in *IEEE International Conference on Robotics and Automation*, **2009**, 2154–2159.
- [15] Miotto R., Wang F., Wang S., Jiang X., and Dudley J. T., Deep learning for health-care:review, opportunities and challenges. *Briefings in Bioinformatics*, **2017**.
- [16] Wu D., Sharma N., and Blumenstein M.. Recent advances in video-based human action recognition using deep learning: A review. in *International Conference on Neural Networks*, **2017**, 2865–2872.
- [17] Ravi D., Wong C., Lo B., and Yang G.-Z.. Deep learning for human activity recognition:A resource efficient implementation on low-power devices. in *Proc. of IEEE 13th Int. Conf. Wearable Implantable Body Sensor Network*, **2016**, 71–76.
- [18] Ravi D., Wong C., Lo B., and Yang G.-Z.. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, **21**, **2017**, 56–64.
- [19] Renjie D. et al.. Empirical study and improvement on deep transfer learning for human activity recognition. *Sensors*, **2019**.
- [20] Cheng X. et al.. Innohar: a deep neural network for complex human activity recognition. **2019**, 9893–9902.
- [21] AminULLAH et al.. Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments. *Future Generation Computer Systems*, **2019**, 386–397.

- [22] Bottou L. and Vapnik V.. Local learning algorithms. *Neural computation*, 4, 1992, 888-900.
- [23] Paul P. and George T.. An effective approach for human activity recognition on smart-phone. in *IEEE International Conference on Engineering and Technology (ICETECH)*, 2015, 1-3.
- [24] Zhang H., Berg A., Maire M., and Malik J.. SVM - KNN: discriminative nearest neighbor for visual object recognition. in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, 1-3.
- [25] Cheng H., Tan P.-N., and Jin R.. Localized support vector machine and its efficient algorithm. in *Proceedings of the SIAM International Conference on Data Mining*, 2007.
- [26] Suykens J. and Vandewalle J.. Least squares support vector machine classifiers. *Neural processing letters*, 9, 1999, 293-300.
- [27] Huang Y. and Ma H.. Application of least square support vector machine in electronic engineering based on principal component analysis. in *Advances in Mechanical and Electronic Engineering*, D. Jin and S. Lin, Eds. Berlin, Heidelberg: Springer, 2012.
- [28] Alsheikh M. A., Selim A., Niyato D., Doyle L., Lin S., and Tan H.-P.. Deep activity recognition models with triaxial accelerometers. in *AAAI Workshop: Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, 2016.
- [29] Catal C., Tufekci S., Pirmir E., and Kocabag G.. On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing*, 37, 2015, 1018-1022.
- [30] Zappi P. et al.. Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. 2007.
- [31] Bachlin M., Plotnik M., Roggen D., Maidan I., Hausdorff J., Giladi N., and Troster G.. Wearable assistant for Parkinsons disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 2010, 436-446.
- [32] Kwapisz J., Weiss G., and Moore S.. Activity recognition using cell phone accelerometers. in *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*. 2010, 10-18.
- [33] Lockhart J. W. et al.. Design considerations for the WISDM smart phone-based sensor mining architecture. 2011, 25-33.
- [34] Schölkopf B. et al.. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 2001, 1443-1471.

- [35] Haibo H. and Garcia E.. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9),2009, 1263-1284.
- [36] Webb A. and Copsey K.. Statistical Pattern recognition, 3rd edition. *Wiley, UK*, 2011.
- [37] Karevan Z., Feng Y., and Suykens J.. Moving least squares support vector machines for weather temperature prediction. in *Proceedings of European Symposium on Artificial Neural Networks*, 2017.
- [38] Suykens J.A.K., De Brabanter J., Lukas L., Vandewalle J.. Weighted least squares support vector machines : robustness and sparse approximation. *Neurocomputing, Special issue on fundamental and information processing aspects of neurocomputing*, 48, 2002, 85-105.
- [39] Amer A.. Localised least squares support vector machines with application to weather forecasting. *Masters thesis, KU Leuven*, 2016.
- [40] He H. and Ma Y.. Imbalanced Learning: Foundations, Algorithms, and Applications. 1st ed. *Wiley-IEEE Press, New York*, 2013.
- [41] Krawczyk B.. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221-232.
- [42] Trappenberg T. and Back A.. A classification scheme for applications with ambiguous data. in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Net-works, IJCNN Neural Computing: New Challenges and Perspectives for the New Millen-nium*, 6, 2000, 296-301.
- [43] Suykens J. et al., Least Squares Support Vector Machines. *Singapore: World Scientific Publishing Co.*, 2002.
- [44] Charte D. et al.. A practical tutorial on autoencoders for nonlinear feature fusion taxonomy, models, software and guidelines. *Information Fusion*, 2018, 78-96.
- [45] Choi Y., Ralhan A., and Ko S. A study on machine learning algorithms for fall detection and movement classification. in *IEEE Information Science and Applications (ICISA)*, 2011, 1-8.
- [46] Ni Q. et al.. The elderlys independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors*, 15(5), 2015, 312-362.
- [47] Tataru C., Yi D., Shenoyas A., and Ma A.. Deep learning for abnormality detection in chest x-ray images. 2017.
- [48] Kourou K. et al. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13, 2015, 8-17.

- [49] Maysanjaya I. M. D. et al.. A comparison of classification methods on diagnosis of thyroid diseases. in *Intelligent Technology and Its Applications (ISITIA) International Seminar on. IEEE*, 2015, 89-92.



## Chapter 3

# Towards Online Personalised-Monitoring of Human Thermal Sensation Using Machine Learning Approach<sup>1</sup>

### Abstract

Thermal comfort and sensation are important aspects of building design and indoor climate control, as modern man spends most of the day indoors. Conventional indoor climate design and control approaches are based on static thermal comfort/sensation models that view the building occupants as passive recipients of their thermal environment. To overcome the disadvantages of static models, adaptive thermal comfort models aim to provide opportunity for personalised climate control and thermal comfort enhancement. Recent advances in wearable technologies contributed to new possibilities in controlling and monitoring health conditions and human well-being in daily life. The generated streaming data generated from wearable sensors are providing a unique opportunity to develop a real-time monitor of an individuals thermal state. The main goal of this work is to introduce a personalised adaptive model to predict individuals thermal sensation based on non-intrusive and easily measured variables, which could be obtained from already available wearable sensors. In this paper, a personalised classification model for individual thermal sensation with a reduced-dimension input-space, including 12 features extracted from easily measured variables, which are obtained from wearable sensors, was developed using least-squares support vector machine algorithm. The developed classification model predicted the individuals thermal sensation with an overall average accuracy of 86%. Additionally, we introduced the main framework of streaming algorithm for personalised classification model to predict an individuals thermal sensation based on streaming data obtained from wearable sensors.

---

<sup>1</sup><https://doi.org/10.3390/app9163303>

### 3.1 Introduction

Thermal comfort (TC) is an ergonomic aspect determining satisfaction with the surrounding environment and is defined as that condition of mind which expresses satisfaction with the thermal environment and is assessed by subjective evaluation [1]. The effect of thermal environments on occupants might also be assessed in terms of thermal sensation (TS), which can be defined as a conscious feeling commonly graded into the categories cold, cool, slightly cool, neutral, slightly warm, warm, and hot [1]. Thermal sensation and thermal comfort are both subjective judgements, however, thermal sensation is related to the perception of ones thermal state, and thermal comfort is related to the evaluation of this perception [2]. In other words, TS expresses the perception of the occupants, while TC assesses this perception, taking into account physiological and psychological factors [3]. The assessment of thermal sensation has been regarded as more reliable and as such is often used to estimate thermal comfort [4]. Thermal sensation is the result of the body psycho-physical reaction to certain thermal stimuli related to indoor conditions [5]. Human thermal sensation mainly depends on the human body temperature (core body temperature), which is a function of sets of comfort factors [5, 6]. These comfort factors include indoor environmental factors, such as mean air temperature around the body, relative air velocity around the body, humidity, and mean radiant temperature of the environment to the body [6]. Additionally, some personal (individual-related) factors, namely metabolic rate or internal heat production in the body, which vary with the activity level and clothing thermo-physical properties (such as clothing insulation and vapor clothing resistance), are included. It should be mentioned that the individual thermal perception is deepening, as well, on psychological factors, expectations and short/long-term experience, which directly affect individuals perceptions, time of exposure, perceived control, and environmental stimulation [7]. The most considered way to have an accurate assessment of TS is to ask the individuals directly about their thermal sensation perception [5, 6]. The thermal-sensation-vote (TSV) is one of the most used concepts to address the opinion of individuals concerning TS. That is, individuals express their vote to rate their TS when they are exposed to given thermal conditions, by using a scale from cold to hot, with a predefined number of points. Thermal sensation mathematical models are developed in order to overcome the difficulties of direct enquiry of subjects. The development of such models is mostly dependent on statistical approaches by correlating experimental conditions (i.e. environmental and personal variables) data to thermal sensation votes obtained from human subjects [4, 6]. The recent intensive review work of Enescu (2019), explored the most important contributions to model and predict thermal sensation (TS) under both steady-state and transient conditions. It is shown that the most used models to assess TS of the human body with respect to the environment have been developed starting from Fangers predicted-mean-vote (PMV) empirical model [5] for steady-state conditions and from the Gagge model [8] for transient conditions. Since then, numerous models are developed to assess and predict TS (e.g., [9, 10, 11, 12, 13, 14, 15, 16]). Most of the aforementioned models (e.g., PMV) are static in the sense that they predict the average vote

of a large group of people based on the seven-point thermal sensation scale, instead of individual thermal comfort, they only describe the overall thermal sensation of multiple occupants in a shared thermal environment. To overcome the disadvantages of static models, adaptive thermal comfort models aim to provide insights in increasing opportunities for personal and responsive control, thermal comfort enhancement, energy consumption reduction and climatically responsive and environmentally responsible building design [17, 18]. The idea behind adaptive models is that occupants and individuals are no longer regarded as passive recipients of the thermal environment but rather, play an active role in creating their own thermal preferences [18]. Many adaptive thermal comfort models are developed based on regression analysis (e.g., [18, 19, 20]). Besides regression analysis, thermal sensation prediction can also be seen as a classification problem where various classification algorithms can be implemented [17]. In their work [21], Lee et al., proposed a method for learning personalised thermal preference profiles by formulating a combined classification and inference problem with 5-cluster models. Moreover, the thermal preference of a new user is inferred by a mixture of sub-models for each cluster, where clusters are used to group occupants with similar thermal preferences. Recently, a number of research works (e.g., [22, 23, 24, 25, 26]) have demonstrated the possibility of using machine learning techniques, such as support vector machine (SVM), to assess and predict human thermal sensation. It can be concluded, based on the published work (see the recent literature review [17] by Lu et al.), that classification-based models have performed as well as regression models. Different related works investigated the problem of thermal sensation and comfort prediction via machine learning algorithms. Ghahramani et al. [22] applied the hidden Markov model (HMM) technique to the thermal comfort prediction problem with three levels of thermal comfort. There is a main issue in the used dataset in this study is the class imbalance, which is not tackled by the proposed methodology. In their study, Ghahramani et al. did not discuss the problem of streaming analytics and model personalisation. In order to develop personalised models, Jiang et al. [28] applied support vector machines classifiers to the personal data of each subject to predict the thermal sensation level for the same subject. The obtained results are promising, however, their approach requires a sufficient number of data-points to obtain an acceptable performance, which is not applicable to our dataset (9 data-points per subject). The very recent study of Lu et al. [17] proposed a personalised model, however, the study strictly investigated two subjects and developed a dedicated model for each subject. In comparison with many relevant studies, our study is tackling several challenges at the same time. These issues are feature reduction, streaming, and online modeling compatibility and model personalisation. The latter issue is tackled in a novel way by considering both personal and nonpersonal data relying on the similarity either inter or intra-subjects. In general, it can be stated that it is a real modeling challenge to correlate the physiological variables with information concerning global and local sensation [5]. Recent advances in mobile technologies in healthcare, in particular, wearable technologies (m-health) and smart clothing, have positively contributed to new possibilities in controlling and monitoring health conditions and human well-being in daily life applications. The wearable sensing technologies

and their generated streaming data are providing a unique opportunity to understand the users behaviour and to predict future needs [29]. The generated streaming data is unique due to the personal nature of the wearable devices. However, the generated streaming data forms a challenge related to the need for personalised adaptive models that can handle newly arrived personal data. The main goal of this work is to introduce a personalised adaptive modeling algorithm to predict an individuals thermal sensation based on non-intrusive and easily measured variables, which could be obtained from already available wearable sensors.

## 3.2 Methods

### 3.2.1 Data Processing and Classification

Thermal sensation prediction based on wearable sensors can be considered as a classification machine learning problem, the input of which is the set of extracted features from the measured variables and the output is the subjects feedback with the standard thermal sensation labels. Several machine learning techniques can be used for such a problem. Support vector machines (SVMs) is one of the efficient classification techniques used in different relevant studies [21, 22, 23, 25, 26]. In this study, the least squares support vector machine (LS-SVM) is proposed to be used for general models as it is as powerful as standard SVMs, but, it has less computational cost [30]. Most, if not all, relevant studies of thermal sensation prediction rely on global general models. Global models are models that are trained using the whole available training dataset with a uniform weight (i.e. all training points are equally contributing to the training process). However, global models are not that efficient for online classification and streaming analytics applications in which a stream of new data is collected from subjects via wearable technology, especially when aiming at personalised models. Hence, for this purpose, we suggest a localized version of LS-SVM, namely K-Nearest Neighbours (KNN)-LS-SVM [31] to be compatible with the wearable sensors for online and streaming analytics. The classification problem of thermal sensation is a multiclass classification problem, the input variables of this problem from which features are extracted are: aural temperature  $T_{er}$ , average skin temperature  $\bar{T}_{sk} = \frac{1}{3}[T_{scap} + T_{ch} + T_{arm}]$ , ambient temperature  $T_a$ , chest skin temperature  $T_{ch}$ , heart rate  $H_R$ , average heat flux from the skin  $\bar{q}_{sk}$ , temperature gradient between core and skin  $\Delta T_{sk} = T_{er} - \bar{T}_{sk}$ , age, gender, body mass index  $BMI$ , metabolic rate  $M_r$ . As some variables are time measurements of the different parameters, the process of feature extraction is applied to a specified time window from the recordings, namely the last five minutes preceding the sensation labeling by the test subjects. The extracted features from time-variant variables are: minimum ( $min$ ), maximum ( $max$ ), variance ( $var$ ), energy, time-derivative ( $\frac{d}{dt}$ ), root mean square ( $rms$ ). Target labels are the seven classes of the standard thermal comfort sensation scores: Cold (-3), Cool (-2), Slightly Cool (-1), Neutral (0), Slightly Warm (1), Warm (2), and Hot (3).

### 3.2.1.1 Support Vector Machines (SVMs)

SVMs are originally presented as binary classifiers, that assign each data instance  $\mathbf{x} \in \mathbb{R}^d$  to one of two classes described by a class label  $y \in \{-1, 1\}$  based on the decision boundary that maximises the margin  $2/\|w\|_2$  between the two classes as shown in Figure 1 [30]. Generally, a feature map  $\phi: \mathbb{R}^d \mapsto \mathbb{R}^p$  is used to transform the geometric boundary between the two classes to a linear boundary  $L: \mathbf{w}^\top \phi(\mathbf{x}) + b = 0$  in feature space, for some weight vector  $\mathbf{w} \in \mathbb{R}^{p \times 1}$  and  $b \in \mathbb{R}$ . The class of each instance can then be found by  $y = \text{sgn}(\mathbf{w}^\top \phi(\mathbf{x}) + b)$ , where  $\text{sgn}$  refers to the sign function. The estimation of the boundary  $L$  is performed based on a set of training examples  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) with corresponding class labels  $y_i \in \{-1, 1\}$ . An optimal boundary is found by maximising the margin that is defined as the smallest distances between  $L$  and any of the training instances. In particular, one is interested in constants  $\mathbf{w}$  and  $b$  that minimise a *loss-function*:

$$\min_{\mathbf{w}, b; \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N \xi_i, \quad (3.0)$$

and are subject to:

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, 2, \dots, N.$$

The constant  $C$  in (3.0) denotes the *penalty term* that is used to penalise missclassification through the slack variables  $\xi_i$  in the optimisation process.

The so-called *kernel-trick* avoids the explicit introduction of a feature map  $\phi$  and implicitly allows to use feature spaces of infinite dimensionality. A commonly used kernel is given by the Gaussian kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_0^2}\right),$$

where  $\sigma_0$  denotes the *kernel bandwidth*. Both  $\sigma_0$  and  $C$  can be optimised as hyper-parameters in a cross-validation experiment.

### 3.2.1.2 Least Squares Support Vector Machine (LS-SVM)

LS-SVMs are obtained by using a least-squares error loss function [30]:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2, \quad (3.1)$$

such that

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) = 1 - e_i, \quad i = 1, 2, \dots, N.$$

This optimisation procedure introduces errors  $e_i$  such that  $1 - e_i$  is proportional to the signed distance of  $\mathbf{x}_i$  from the decision boundary. In fact, the non-negative slack variable constraint is

removed and the solution of the optimisation problem can be obtained by a set of linear equations, reducing computational effort [30].

### 3.2.1.3 kNN-LS-SVM

While global SVMs consider the same weight for all training instances in the optimization process, local learning approaches allow for training samples near a test point to be more influential than others. Localized approaches of SVMs [31] are based on weighting functions  $\lambda(\mathbf{x}_s, \mathbf{x}_i)$  that express the similarity between the features vectors of the  $i$ th data point  $\mathbf{x}_i$  and the test instance  $\mathbf{x}_s$ . For an LS-SVM, this leads to the following cost function:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i^2, \quad (3.2)$$

such that

$$y_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) = 1 - e_i, \quad i = 1, 2, \dots, N.$$

For KNN-LS-SVM a binary valued similarity criterion:

$$\lambda(\mathbf{x}_s, \mathbf{x}_i) = \begin{cases} 1 & \text{if } \|\phi(\mathbf{x}_s) - \phi(\mathbf{x}_i)\|_2 \leq r_s \\ 0 & \text{otherwise,} \end{cases}$$

where  $r_s$  is the  $K$ -th smallest distance among  $\{\|\phi(\mathbf{x}_s) - \phi(\mathbf{x}_j)\|; 1 \leq j \leq N\}$ . This formulation leads to the hybrid KNN-LS-SVM method [31]. In practice, implementing the hybrid classifier of KNN-LS-SVM, as shown in Figure 3.1, starts with receiving an unlabelled new test point  $x_s$  and finding the nearest  $K$  points from the training set in the feature space. Based on the nearest  $K$  points, an LS-SVM model is trained only with the new subset, hence, for each test point a dedicated model is trained. The advantage of this localised approach is that it can enhance the classification performance in case of class imbalance, in addition to the computational and temporal efficiency especially for online modelling and streaming analytics. For more detail concerning localised learning, reference [30] includes a detailed explanation of the algorithms.

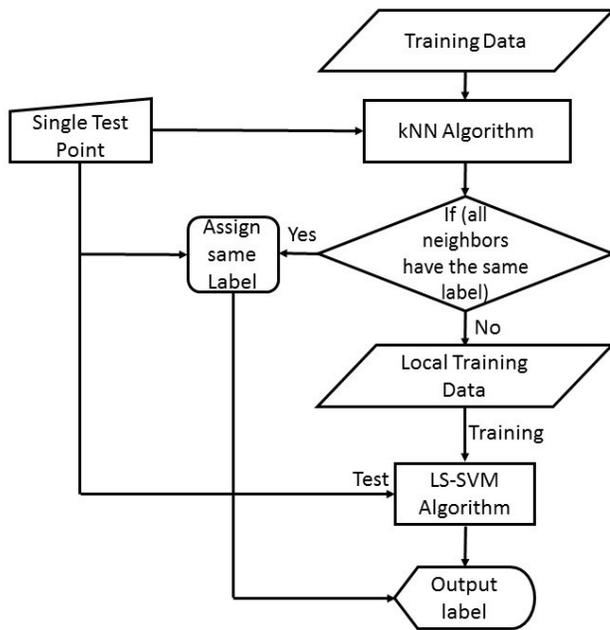


Figure 3.1: A flow chart illustrating the algorithm of K-Nearest Neighbours Least Squares Support vector Machines (KNN-LS-SVM) classifier

### 3.2.2 Experiments and Experimental Setup

#### 3.2.2.1 Test Subjects

In total 25 healthy participants (6 females and 19 males), between the age of 25 and 35 (average age  $26 \pm 4.2$ ) years, with average weight and height of  $70.90 \pm 12.70$  kg and  $1.74 \pm 0.10$  m, respectively, volunteered to perform the aforementioned experimental protocol. Detailed physical information about the test subjects is shown in Table 3.1.

Table 3.1: General physical information of the participants (test subjects).

Subject	Gender	Height (m)	Weight (kg)	Age (year)
P1	M	1.69	59	23
P2	M	1.77	75	20
P3	M	1.82	73	29
P4	F	1.61	53	31
P5	M	1.86	88	21
P6	F	1.57	50	22
P7	M	1.73	86	33
P8	M	1.81	67	21
P9	M	1.86	92	36
P10	M	1.65	62	31
P11	F	1.7	61	23
P12	M	1.86	80	23
P13	M	1.82	86	27
P14	F	1.6	51	22
P15	M	1.7	58	29
P16	M	1.75	74	26
P17	F	1.68	76	26
P18	M	1.8	74	29
P19	M	1.78	79	29
P20	M	1.83	81	22
P21	M	1.78	78	28
P22	M	1.81	69	22
P23	F	1.57	49	26
P24	M	1.75	68	24
P25	M	1.78	83	28

### 3.2.2.2 Climate Chambers

During the course of this study, three (Rooms A, B and C) climate-controlled chambers designed and built to investigate the dynamic mental and physiological responses of humans to specific indoor climate conditions were used. Figure 2 shows a photographic picture of the three climate rooms, namely, A, B and C. The Body and Mind Rooms are experimental facilities at the M3-BIORES laboratory (Division of Animal and Human Health Engineering, KU Leuven). The three rooms are dimensionally identical; however, each room is designed to provide different ranges of climate conditions as shown in Table 3.2.

Table 3.2: Different temperature and relative humidity ranges that can be provided by the different Body and Mind (A, B, and C).

Room	Air temperature range ( $^{\circ}\text{C}$ )	Relative humidity range (%)
A	+23 - +37	50 - 80
B	+10 - +25	50 - 80
C	-5 - +10	40 - 60

The three rooms are equipped with axial fans to simulate wind velocities between 2.5 and 50  $\text{km h}^{-1}$ .



Figure 3.2: Photographic picture of the three climate-controlled rooms (from right to left, AC).

### 3.2.2.3 Measurements and Gold Standards

During the course of the experiments, participants heart rate  $H_R$ , metabolic rate  $M_r$ , average skin temperature  $T_{sk}$ , heat flux  $q_{sk}$  between the skin and the ambient air, core body temperature  $T_c$  represented by the aural temperature  $T_{er}$  were measured continuously. The heart rate of each participants was monitored using the Polar H7 ECG strap that is placed under the chest. For Polar H7 ECG strap, the ECG sampling rate is 128 Hz. The metabolic rate as metabolic equivalent tasks ( $METS$ ) of each participant was calculated based on indirect calorimetry using MetaMAX 3B spiroergometer sensor. The average skin temperature was calculated based on measurements from three body-places, namely, scapula  $T_{scap}$ , chest  $T_{ch}$  and arm  $T_{arm}$  (Figure 3). The skin temperature measurements were performed using one Shimmer temperature sensor and two  $gSKIN$  bodyTEMP patches. Two heat flux  $gSKIN$  patches were placed on both the chest and the left arm (Figure 3.3). The skin temperatures and heat flux measurements were acquired at sampling frequency of 1 Hz. Core body temperature was estimated based on au-

ral temperature measure measurements, which was performed using in-ear wireless (Bluetooth) temperature sensors (Cosinuss One) with a sampling rate of 1 Hz. At the end of each applied temperature level during the course of both experimental phases, a thermal sensation questionnaire, based on ASHRAE 7-points thermal scale, was performed for each test subject.

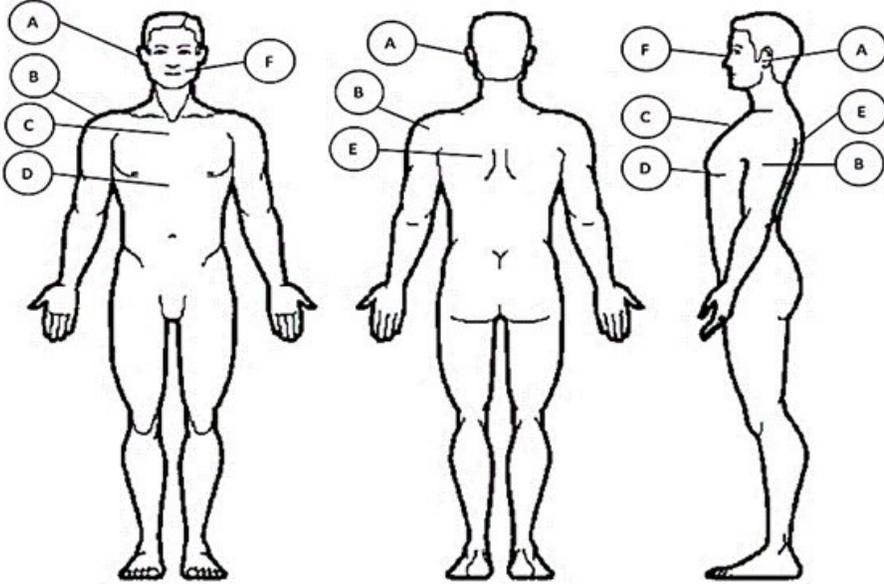


Figure 3.3: Sensor placement. (A) Ear channel for aural temperature measurement via the Cosinuss One, (B) upper arm where skin temperature and heat flux are measured with the gSKIN patch, (C) middle upper chest where skin temperature and heat flux are measured with the gSKIN patch, (D) lower chest where heart rate is measured with the Polar H7, (E) Scapula where skin temperature is measured with the shimmer, (F) mouth and nose where metabolic rate is measured via the MetaMAX-3B spirometry sensor.

#### 3.2.2.4 Experimental Protocol

The experimental protocol used in the present study was designed in such way to investigate the subjects thermal and physiological responses to three different temperature (low, normal and high) under two levels of physical activities (low and high). The three predefined temperatures (low =  $5^{\circ}\text{C}$ , normal =  $24^{\circ}\text{C}$  and high =  $37^{\circ}\text{C}$ ) were chosen based on the thermal-comfort-chart of the ASHRAE-55 [32] and the effects on health according to the Wind Chill Chart for cold exposure (National Weather Service of the US) and for hot temperatures exposure according to [33]. The conducted experiments were consisted of two phases (Figure 3.4, upper graph), namely, low activity and high activity phases. During the first experimental phase, low activity phase,

the test subjects (while being seated = low activity) were exposed, during 55 minutes, to three levels of temperatures in the following order: normal, low, high and normal again (Figure 3.4). During the high activity phase, the test subjects was exposed to a 15 minutes of light physical stress (80W of cycling on a fastened racing bicycle). During the course (75 minutes) of the active phase, each test subject was exposed to the predefined three temperature levels (Figure 4, lower graph). During each temperature level, starting from the normal level (24 °C), the test subjects are performed 15 minutes of cycling (with 80 W power) and followed 4 minutes of resting (seated). During the course of conducted experiments, the clothing insulation factor ( $Col$ ) was kept constant at  $Col = 0.34$ , which accounted for a cotton short and t-shirt as a standard clothing for all test subjects. The experimental protocol was approved by the SMEC (Sociaal-Maatschappelijke Ethische Commissie), on the 16 January 2019 with number G-2018 12 1464.

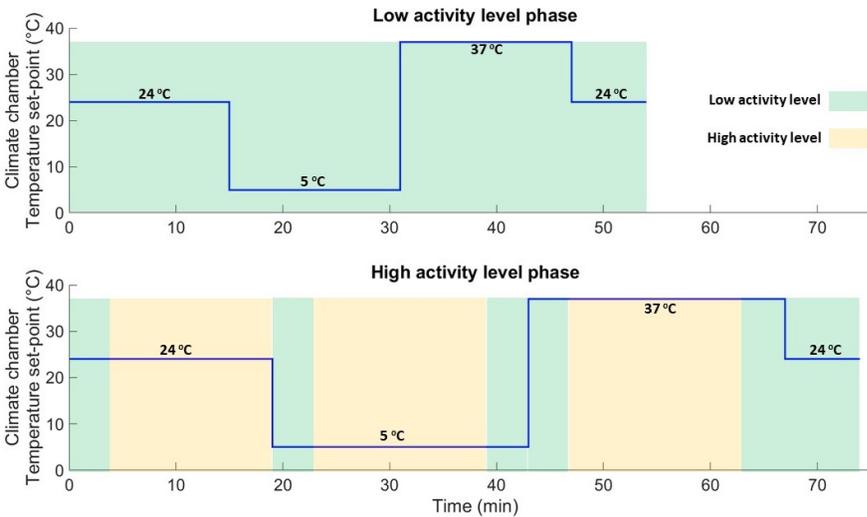


Figure 3.4: Plots showing the climate chambers set-point temperatures programmed during the 55 min low activity phase (upper graph) and the 75 min high activity phase (lower graph).

### 3.3 Results and Discussion

#### 3.3.1 General Classification Models

In this section, classification models are developed globally, in other words the classification models are trained using all available training dataset with the same weight (i.e. all training data-points are contributing equally to the training process). The whole dataset ( $N$ -subjects) are divided, based on leave-one-subject-out approach (LOSO), into  $N-1$  subjects for training and 1 subject

for testing.

### 3.3.1.1 Developing general model using all extracted features for 7-classes problem (Model I)

Initially, in this stage of developing a general classification model to predict thermal sensation, in total 54 features are used to form the input space of the classification model for the 7-classes classification problem. The extracted features are meant to be simple and basic features that are not computationally expensive and represent the basic characteristics of segmented time windows. A feature space including the mean value of the measured input variables, namely,  $T_{er}$ ,  $H_R$ ,  $\bar{q}_{sk}$ ,  $\Delta\bar{T}$  and  $M_r$ . Additionally, other features are extracted by computing the variance, min, max, root mean squares (*RMS*), energy ( $E = \frac{1}{N} \sum_{n=1}^N x_n^2$ , where  $N$  is the number of samples of variable  $x_n$ ) and first derivative ( $\frac{dx}{dt}$ ) of the aforementioned measured variables as shown in Table 3.3. The age, gender, body-mass-index (*BMI*) and ambient temperature ( $T_\infty$ ) are also included in the feature spaces. The output confusion matrix is computed for each subject based on LOSO testing approach. The averaged normalised confusion matrix over all test subjects is shown in Table 3.4 where the value of each cell ( $i, j$ ) represents the number of times (as percentage %) that class  $j$  is classified as class  $i$ . Given that the optimal situation is 100% for  $i = j$ . From the resulted confusion matrix (Table 3.4) the overall accuracy of the developed classifier (Model I) is calculated to be 51%. In Table 3.4, there is the prediction result noted as Else, which represents the case that the classifier could not assign the test point to any of the presented classes. The error performance of the developed general model is depicted in Figure 3.5.

Table 3.3: Overview of the 54 extracted features (CE = selected).

	Variance	Mean	Min	Max	RMS	$E$	$\frac{d}{dt}$
$T_{er}$	x	x	x	x	x	-	x
$H_r$	x	x	x	x	x	-	x
$\bar{q}_{sk}$	x	x	x	x	x	x	x
$\bar{T}_{sk}$	x	x	x	x	x	-	x
$\Delta\bar{T}$	x	x	x	x	x	-	x
$M_r$	x	x	x	x	x	x	x
$T_{arm}$	x	x	x	x	x	-	x
$T_{scap}$	x	x	x	x	x	-	x

Table 3.4: The normalized confusion matrix of Model I

		Actual Label (j)						
		Cold %	Cool %	Slightly Cool %	Neutral %	Slightly Warm %	Warm %	Hot %
Predicted Label (i)	Else	0	0	3.30	2.80	4.20	8.00	6.70
	Cold	0	0	0	0	0	0	0
	Cool	33.3	52.3	23.4	1.40	0	0	0
	Slightly Cool	66.7	42.9	40.0	12.7	2.10	0	0
	Neutral	0	4.80	33.3	60.6	27.1	0	0
	Slightly Warm	0	0	0	19.7	56.3	32.0	13.3
	Warm	0	0	0	1.40	10.4	44.0	40.0
	Hot	0	0	0	1.40	0	16.0	40.0

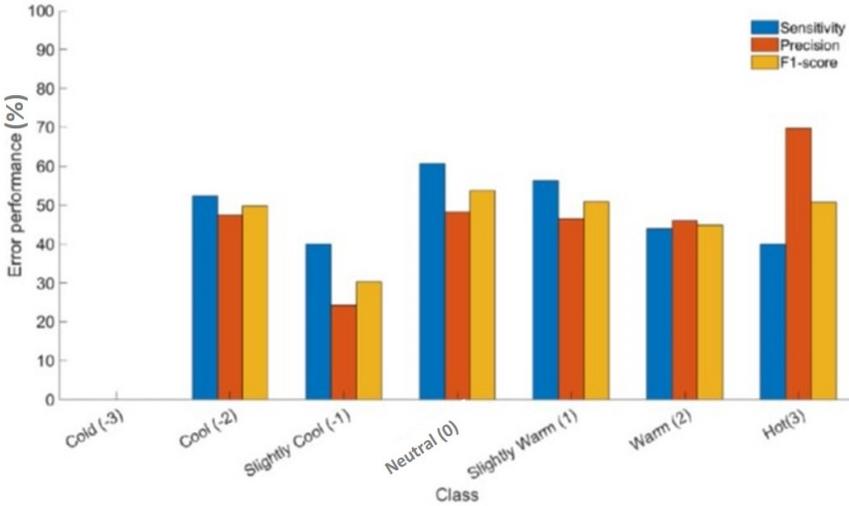


Figure 3.5: Error performance of the developed general classification model with 54 input features for 7-classes classification problem.

3.3.1.2 Developing a General Model for 7-Classes Classification Problem with Dimension Reduction (Model II)

As shown in Table 3.3, the input space of Model I included all extracted features (54 features) that were obtained from the measured variables. However, for the sake of the main objective of the present work, the computational cost of the developed algorithm should be low enough to be compatible with wearable technology and online modeling. Hence, a feature selection pro-

cedure was employed to obtain the most reduced-dimension input space for the classification model yet with the best error performance. Feature selection here is based on evaluating all possible feature combinations and selecting the combination with best error performance. The used feature selection procedure resulted in a reduced input space of only 12 features with optimal feature combination. The selected features comprise: gender, age,  $H_R$ ,  $T_{er}$ ,  $T_{sk}$ ,  $\Delta\bar{T}$ ,  $\bar{q}_{sk}$ ,  $RMS(H_R, T_c, T_{sk}, \dot{q})$ , and  $\frac{d\bar{q}_{sk}}{dt}$  (time-derivative of average heat flux). The feature selection step reduced the input space from 54 features to only 12, which effectively reduced the computational costs of the classification algorithm during online implementation. The reduced dimension input space, including the selected 12 features, was used to develop a general classification model for the 7-classes classification problem to predict the thermal sensation of all test subjects. The resulted classification confusion matrix for the developed general model using the reduced-dimension input space is shown in Table 3.5. The results showed an overall accuracy of the developed classification model of 57% with an improvement of 6% compared to the results of model I. The overall error performance (sensitivity, precision and  $F_1$ -score) results are shown in Figure 3.6.

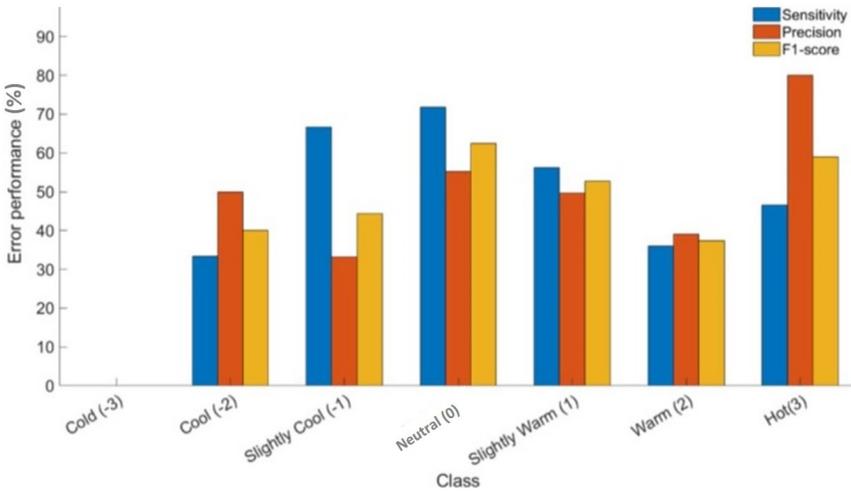


Figure 3.6: Error performance of the developed general classification model with the selected 12 input features (reduced dimension input space) for 7-classes classification problem.

Table 3.5: The normalized confusion matrix of Model II.

		Actual Label (j)						
		Cold %	Cool %	Slightly Cool %	Neutral %	Slightly Warm %	Warm %	Hot %
Predicted Label (i)	Else	0	4.79	6.65	0	10.40	16.03	0
	Cold	0	0	0	0	0	0	0
	Cool	33.20	33.33	0	0	0	0	0
	Slightly Cool	66.80	57.08	66.69	8.45	2.10	0	0
	Neutral	0	4.79	26.66	71.81	22.90	4.03	0
	Slightly Warm	0	0	0	18.32	56.25	31.96	6.71
	Warm	0	0	0	1.42	8.35	35.99	46.65
	Hot	0	0	0	0	0	12.00	46.65

### 3.3.2 Class Reduction

From the confusion matrix in Table 3.5, it can be seen that the confusion is mostly observed between the adjacent classes. The main reason of such interclass confusion is that the features are not able to discriminate completely between these adjacent classes. For instance, the actual neutral class (0) is confused with 8.45% and 18.32% with slightly cool (-1) and slightly warm (1) classes, respectively. Hence, it is more convenient to reduce the seven thermal sensation classes into three classes representing thermal comfort (comfortable, uncomfortably cool, and uncomfortably warm). The class reduction is done based on three criteria, namely, maximum confusion, acceptable class imbalance, and avoiding overlap between classes. As mentioned earlier, the maximum confusion is observed between the adjacent classes (see Table 3.5). However, it is not possible to merge all adjacent confused classes due to the overlap. For example, the Slightly-Warm class is confused with the Neutral class by 22.9%, on the other hand, the Warm class is confused with the Slightly-warm by 31.96%. Hence, in order to merge the Slightly-warm class with the Neutral it should not be merged with Warm and vice versa. Therefore, merging must avoid any overlap between different classes. Another criterion is the class imbalance, as shown in Figure 7a and Table 3.5, where Cold is not recognised by the classifier due to the relatively very low number of instances labeled as Cold compared to the other classes. For an acceptable class imbalance, it is meant to consider the already existing class imbalance between the whole states that the frequency of a state occurrence is reducing by moving far from the Neutral state, as shown in Figure 7a.

Finally, it is necessary to avoid any overlap between the reduced classes by assigning each state to only one class. As there are different possibilities to obtain the new three classes, it is found that three configurations are the closest to the thermal comfort levels, considering the earlier mentioned criteria. Based on these criteria the seven classes were reduced into three classes with three different configurations as follows:

- **Configuration 1** Merging the states of Cold (-3) and Cool (-2) into Class 1 (27 instances),

merging Slightly cool (-1), Neutral (0), and Slightly warm (1) into Class 2 (149 instances), and merging Warm (2) and Hot (3) into Class 3 (40 instances) (Figure 7b).

- **Configuration 2** Merging the states of Cold, Cool and Slightly-cool into Class 1 (57 instances), Neutral as Class 2 (71), and merging Slightly-warm, warm and Hot into Class 3 (88 instances) (Figure 7c).
- **Configuration 3** Merging the states of Cold, Cool and Slightly-cool into Class 1 (57 instances), merging Neutral, and Slightly-warm into Class 2 (119 instances), and merging warm and Hot into Class 3 (40 instances) (Figure 7d).

As shown in Figure 3.7, each configuration has a different class distribution (i.e. number of instances per class).

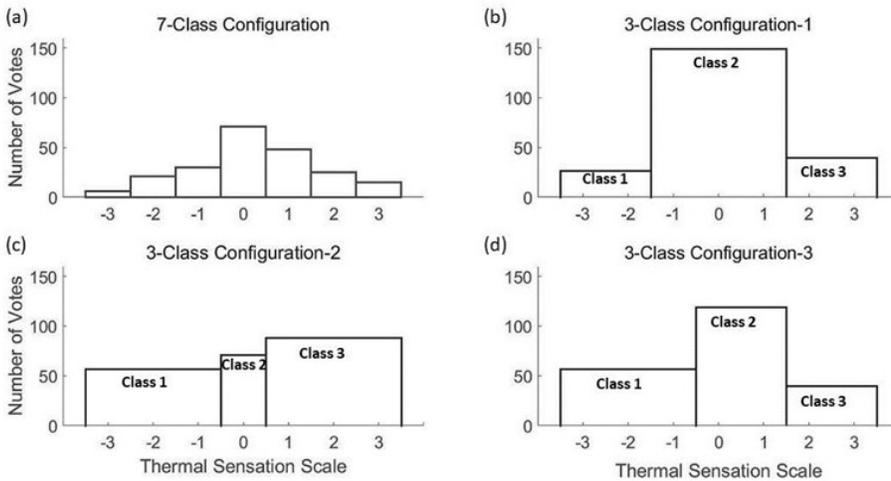


Figure 3.7: (a) A histogram of 7-class thermal sensation scale of ASHRAE system. (b) A histogram of 3-class thermal sensation of Configuration 1. (c) A histogram of 3-class thermal sensation of Configuration 2. (d) A histogram of 3-class of Configuration 3.

### 3.3.2.1 Developing General Models with the Selected Features for 3-Class Problem with Different Class Configurations (Model III)

Class Configurations (Model III) The error performance results of the developed classification model (Model III), based on the 12 selected features, for the three labelling configurations (Conf. 1, Conf. 2 and Conf. 3) are shown in Table 3.6. Comparing the three configurations is not consistent, as for each configuration, the number of datapoints change, which influences the performance especially for such small size dataset.

Table 3.6: The error performance (precision, sensitivity,  $F_1$ -score, and accuracy) of general LS-SVM model for the three different 3-class configurations.

Configurations	Classes	Precision	Sensitivity	$F_1$ -score	Accuracy
Conf. 1	Class 1	0.53	0.37	0.44	0.81
	Class 2	0.83	0.89	0.86	
	Class 3	0.79	0.75	0.77	
Conf. 2	Class 1	0.88	0.88	0.88	0.81
	Class 2	0.75	0.66	0.70	
	Class 3	0.82	0.89	0.86	
Conf. 3	Class 1	0.88	0.88	0.88	0.85
	Class 2	0.88	0.91	0.89	
	Class 3	0.88	0.78	0.83	

### 3.3.3 Personalised Classification Models

In order to develop online-personalised models, it is necessary to consider two main challenges, first the developed model should be able to handle the new, personal, data in the training set. Additionally, the developed model should be adapted to the new personal data without any bias to the majority of the old (non-personal) data. Different approaches are used to handle these challenges such as incremental learning methods [34], which work on adapting and retuning the parameters of the general model based on the newly collected data. Another approach is the localised learning, which is based on developing a local model for each test point or subset of the test set [35]. In the present paper, the KNN-LS-SVM localised learning approach is used because of its simplicity and efficiency. Two techniques were used to test the localised models, the first based on LOSO testing approach, and the second approach was based on leave-one-out (LOO) testing approach.

#### 3.3.3.1 Developing Personalised Models Using the Selected 12 Features and Different Class-Configurations Based on LOSO Testing Approach

As explained earlier, to develop a personalised classification model the new personal data were not considered in the training set to compare the performance with the global model. In other words, the new subject (the subject data that left out of the training set) is completely unknown to the model, which simulates the case when the model is dealing with an unknown test subject. The used localised learning approach of KNN-LS-SVM searches for the most similar (based on the similarity criterion, see (3)) training points to the new test point (from the new subject) in the input space by which a local model is developed to classify this test point. The resulted error performance (precision, sensitivity,  $F_1$ -score, and accuracy) of the KNN-LS-SVM classifier based on LOSO testing approach and  $K = 5$  is presented in Table 3.7.

Table 3.7: The error performance (precision, sensitivity, F1-score, and accuracy) of the localized model KNN-LS-SVM model for the three different 3-class configurations based on LOSO testing approach.

Configurations	Classes	Precision	Sensitivity	F1-score	Accuracy
Conf. 1	Class 1	0.47	0.36	0.41	0.83
	Class 2	0.84	0.90	0.87	
	Class 3	0.83	0.71	0.77	
Conf. 2	Class 1	0.84	0.95	0.89	0.81
	Class 2	0.74	0.68	0.70	
	Class 3	0.84	0.83	0.83	
Conf. 3	Class 1	0.87	0.94	0.90	0.85
	Class 2	0.88	0.89	0.88	
	Class 3	0.86	0.74	0.79	

### 3.3.3.2 Developing Personalised Models Using the Selected 12 Features and Different Class-Configurations Based on Leave-One-Out (LOO) Approach

In contrast with the first approach, for each subject one data-point is tested and the rest of the same subject datapoints are integrated with the training data. This approach mimics online personalised streaming modelling, since the new streaming personal data is considered in the training dataset and a dedicated classifier is developed online for each new test data-point. The obtained error performance of the KNN-LS-SVM classifier based on LOO testing approach and  $K = 5$  is depicted in Table 3.8.

Table 3.8: The error performance (precision, sensitivity, F1-score, and accuracy) of the localized model KNN-LS-SVM model for the three different 3-class configurations based on LOO testing approach. (\*) indicates the highest error performance value for each class in the different configurations.

Configurations	Classes	Precision	Sensitivity	F1-score	Accuracy
Conf. 1	Class 1	0.75	0.56	0.64	0.86
	Class 2	0.87	0.93*	0.90*	
	Class 3	0.86	0.78	0.82	
Conf. 2	Class 1	0.84*	0.95*	0.89*	0.79
	Class 2	0.71	0.62	0.66	
	Class 3	0.81	0.83*	0.82	
Conf. 3	Class 1	0.84*	0.91	0.87	0.87*
	Class 2	0.89*	0.88	0.88	
	Class 3	0.89*	0.80	0.84*	

For the proposed personalised models, the first approach of LOSO is mimicking the case that the model is applied to an unknown subject to predict individuals thermal sensation level based on the measured variables. The localised model is searching for the most similar (nearest) training points to each test point, of this subject, that to train the classification model for each test point. This approach could be useful in case of having a large amount of data with a diversity of subjects especially in the absence of streaming data from new subjects. The second approach of LOO mimics the case of having a prior knowledge about the test subject through personally labelled data. The localised model in this approach is also searching for the most similar training points, which may include this subject personal data. This approach can be efficient in the presence of streaming personal data that is labeled by the test subject.

### 3.3.4 Streaming Algorithm Approach for Personalised Thermal Sensation Monitoring

In this paper, we introduce the main framework of streaming algorithm for personalised classification model to predict individuals thermal sensation based on streaming data obtained from wearable sensors. The main framework of the proposed streaming algorithm approach is depicted in Figure 3.8.

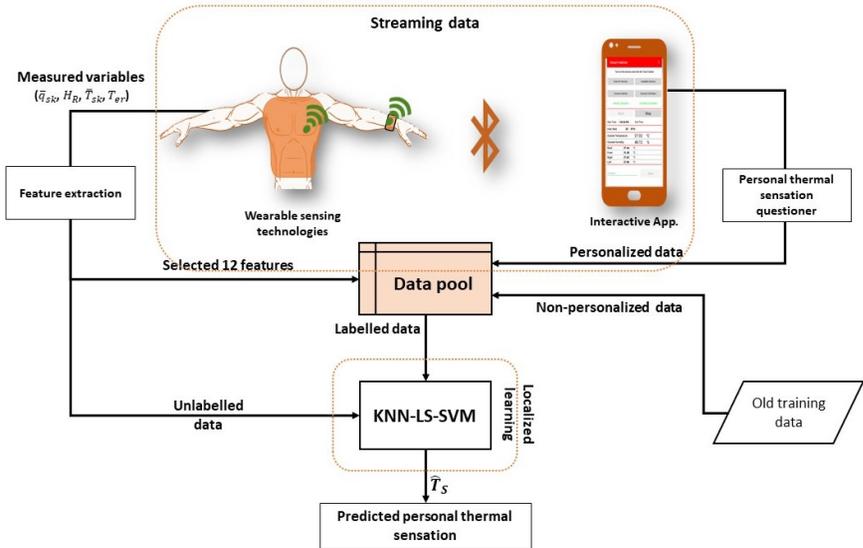


Figure 3.8: Schematic representation of the proposed streaming algorithm for online personalised thermal sensation monitoring.

The main components of the proposed algorithm (Figure 3.8) are explained in the following:

- *Streaming Data*

The availability of the real-time sensors data, from the wearable technologies, has given the possibility of streaming data, which processed via the proposed online streaming algorithm to adapt and personalise the classifier model. The streaming data includes:

- I Wearable sensor data, which consists of the continuously measured variables, namely, individuals heart rate, skin heat flux, skin temperature, ambient temperature and aural temperature.
- II Data obtained from the interactive mobile App., which consists of personal data, namely, age and gender. Additionally, the individuals thermal sensation vote is to be obtained via mobile application-based questioner.

The workflow procedures of streaming data acquisition and labelling are depicted by the flowchart shown in Figure 3.8.

- *Feature Extraction*

As shown earlier, the selected 12 features are extracted from the continuously measured variables, namely,  $H_R$ ,  $T_{er}$ ,  $T_{sk}$ ,  $\Delta\bar{T}$ ,  $\bar{q}_{sk}$ ,  $RMS(H_R, T_c, T_{sk}, \dot{q})$ , and  $\frac{d\bar{q}_{sk}}{dt}$ . Other personal features, namely, age and gender are to be obtained via the interactive mobile application from individual users.

- *Labeled Data*

All training data must be labelled, either the old training data or the new personal data. Personal data is labelled manually via the questionnaire provided by the mobile application.

- *Unlabeled Data*

Unlabelled data is the new data points to be labelled by the classifier, these unlabelled data points include the extracted features from the measured variables.

- *Localised Learning Algorithm*

The localised learning algorithm (i.e. KNN-LS-SVM) is the classifier that receives the unlabelled data points and train a dedicated model with the K nearest training points in order to label the unlabelled ones. The output of this process is a predicted label of personal thermal sensation ( $\hat{T}_s$ ).

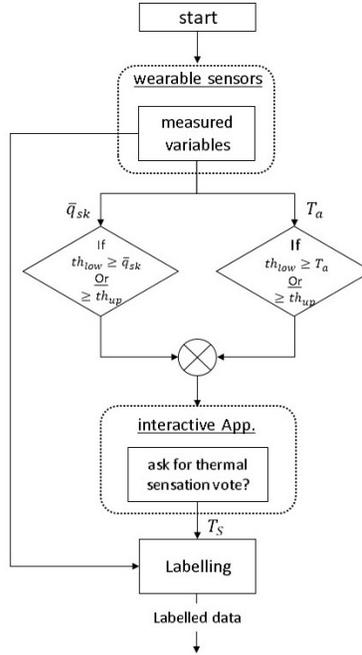


Figure 3.9: Flowchart represents the main workflow of the streaming data acquisition and labeling process.

### 3.4 Discussion

The main advantage of the proposed classification model in the present work, in comparison with other proposed models in recent studies (e.g., [17, 22, 25, 28]), is its capability to handle the requirements for adaptive personalisation and online streaming modelling. Moreover, the proposed model is reduced-dimension, with the minimum possible number of features, which makes it computationally suitable for smart wearable technologies. The main results and findings of the present study is compared with recent studies that treat the prediction of the thermal sensation/comfort as a classification problem using machine-learning techniques. In their study [22], Ghahramani et al. used HMM classification technique, in which three classes of thermal comfort, namely, comfortable, uncomfortably cool and uncomfortably warm are used. An important point to be considered in the work of Ghahramani et al. [22] is the class imbalance in their used experimental data between the positive class (comfortable), which represents 81% of the data and the negative class (uncomfortable), which represents only 19% of the data. Therefore, using the classification accuracy (reported 82.8%) is considered misleading in this case. Hence,

it is much more suitable in their case to compare the precision and sensitivity of this model and our general model (Model III Conf. 3). The reported results [22] of Ghahramani et al. showed a precision of 93.3% and sensitivity of 56.22% without clarifying the precision and sensitivity of the uncomfortable states of warm and cool. On the other hand, our results of (Model III Conf. 3), which is the closest to the compared approach, show a precision of 88% for all classes and sensitivity of 88%, 91%, and 78% for Class 1, Class 2, and Class 3, respectively. These results show more balance between precision and sensitivity for each class. Moreover, personalisation and streaming algorithm compatibility is missing in their study. Another relevant study [28], by Jiang et al., attempted to develop a personalised classification model, as for each subject, a classification model is trained with 50% of that subject data and tested with the rest. The reported result of this study [28] showed an average accuracy over all subjects of 89.82%. However, there is no clarification of the class distribution; hence, it is not clear whether the accuracy is efficient enough for evaluation. Moreover, it is not consistent to compare our final personalised model with that model as the latter is learned with seven classes; however, the former is learned with three classes. In another comparable study [25] to our present work, Farhan et al., predicted individual thermal comfort using machine learning classifier. In their study Farhan et al., used publicly available dataset from which a balanced number of each class is chosen to train and test the classification model. Their developed classification model is trained with three classes that represent the three thermal comfort states of uncomfortably cool, neutral and uncomfortably warm divided based on predefined comfort thresholds. In contrast to our proposed classification model, the proposed classifiers in [25] do not consider model personalisation or streaming online modelling. The best-obtained results amongst their developed models are of the SVM classifier as follows: precision of (76.92, 62.8, and 94.2%) and sensitivities of (67.5, 89.8, and 75.7%) for classes -1, 0, and 1 respectively. On the other hand, our obtained results of (Model III conf. 3) are precision of (88, 88, and 88%) and sensitivities of (88, 91, and 78%) of classes 1, 2, and 3 respectively. It is observed that the precision of our developed classification model are more consistent for all classes, and the sensitivities are higher in total. Ultimately, their approach [25] does not consider personalising the model or streaming online modelling. In another recent study [17], a personal model is discussed; however, it is strictly applied to two subjects (male and female), unlike the case in the present study where we test the model on 25 test subjects. After comparing our methodology and results with number of relevant and comparable studies, it is obvious that the presented study tackled number of classification and modeling challenges unlike many of the aforementioned relevant works. These challenges included the feature selection and dimension reduction, considering new streaming personal data into the training set with keeping the model complexity, rigidity against the problem of class imbalance, and ultimately personalising the classification model using easily measured variable obtained from wearable sensors.

### 3.5 Conclusions

In this present paper, 25 participants are subjected to three different environmental temperatures, namely 5 °C (cold), 20 °C (moderate) and 37 °C (hot) at two different activity levels, namely, at low level (rest) and high level (cycling at 80W power). Metabolic rate, heart rate, average skin temperature (from three different body locations), heat flux and aural temperature are measured continuously during the course of the experiments. The thermal sensation votes are collected from each test subject based on ASHRAE 7-points questioner. A general classification model based on LS-SVM technique is developed to predict the individuals thermal sensation. A localised learning algorithm based on KNN-LS-SVM approach is used to develop a personalised classification model to predict the individuals thermal sensation for 3-classes classification model. The developed classification model has the advantage of using a reduced-dimension input-space, which is suitable for wearable applications and online streaming algorithm. The developed personalised model showed an overall accuracy result of 86%. Additionally, we introduced the main framework of streaming algorithm based on the developed personalised classification model to predict individuals thermal sensation based on streaming data obtained from wearable sensors. In the present work, we believe that it is the first time to utilise the localised learning approach in the thermal state classification problem. One of the main advantages of the proposed approach, in this paper, that it is suitable for streaming algorithm and online modelling as the computational cost is not influenced by increasing the number of data-points. However, the newly obtained data-points is to be considered to develop the online model, which is the main advantage of the KNN-LSSVM. Furthermore, the localised learning approach enables personalisation of the classification model by considering either the personally labelled data-points or the most similar data-points of other persons. On the other hand, number of limitations, concerning the developed model, should be acknowledged here. One important limitation to the developed classification model is regarding to the data size, as the number of data-points per person and in total are generally limited. Moreover, the 7-classes labeling is unbalanced, which made the class reduction is necessary to enhance the overall prediction performance during the course of this study. Otherwise, this study would be extended to be applied to a 7-classes classification problem. The data balance and data size can be enhanced by asking for more frequent votes during the experiment and considering more than three environment temperature levels. Finally, another limitation regards the proposed KNN-LSSVM modeling approach, in which an extra hyperparameter (i.e.  $k$ ) is to be optimised, which adds an extra computational cost to the overall streaming algorithm.



# Bibliography

- [1] ASHRAE. Thermal Environmental Conditions for Human Occupancy; American Society of Heating, Refrigeration and Air Conditioning Engineers. *Inc.: Atlanta, GA, USA, 2004.*
- [2] ISO-10551. Ergonomics of the Thermal EnvironmenAssessment of the Influence of the Thermal Environment Using Subjective Judgement Scales. *ISO: Brussels, Belgium, 1995.*
- [3] Fanger P.O.. Thermal Comfort: Analysis and Applications In Environmental Engineering, 1st ed.*Danish Technical Press: Lyngby, Denmark, 1970.*
- [4] Koelblen B., Psikuta A., Bogdan, A., Annaheim, S., Rossi R.M.. Thermal sensation models: A systematic comparison. *Indoor Air, 2017, 27, 680-689.*
- [5] Enescu D.. Models and Indicators to Assess Thermal Sensation Under Steady-state and Transient Conditions. *Energies, 2019, 12, 841.*
- [6] Kenneth K.C., Parsons C.. Human Thermal Environments: The Effects of Hot, Moderate and Cold Environments on Human Health, Comfort and Performance. *3rd ed.; CRC Press: Boca Raton, FL, USA, 2014.*
- [7] Nikolopoulou M., Steemers K.. Thermal comfort and psychological adaptation as a guide for designing urban spaces. *Energy Build, 2003, 35, 95-101.*
- [8] Gagge A.P., Stolwijk J.A.J., Nishi Y.. An Effective Temperature Scale Based on a Simple Model of Human Physiological Regulatory Response. *ASHRAE Trans., 1971, 77, 247-262.*
- [9] Takada S., Matsumoto S., Matsushita T.. Prediction of whole-body thermal sensation in the non-steady state based on skin temperature. *Build. Environ., 2013, 68, 123-133.*
- [10] Fiala D.. Dynamic Simulation of Human Heat Transfer and Thermal Comfort. *De Montfort University: Leicester, UK, 1998.*
- [11] Lomas K.J., Fiala D., Stohrer M.. First principles modeling of thermal sensation responses in steady-state and transient conditions. *ASHRAE Trans., 2003, 109, 179-186.*

- [12] Zhang H.. Human Thermal Sensation and Comfort in Transient and Non-Uniform Thermal Environments. *University of California, Berkeley: Berkeley, CA, USA*, 2003.
- [13] Guan Y., Hosni M.H., Jones B.W., Giolda T.P.. Investigation of Human Thermal Comfort Under Highly Transient Conditions for Automotive Applications-Part 2: Thermal Sensation Modeling. *ASHRAE Trans.*, 2003, 109, 898-907.
- [14] Guan Y., Hosni M.H., Jones B.W., Giolda T.P.. Investigation of Human Thermal Comfort Under Highly Transient Conditions for Automotive Applications-Part 1: Experimental Design and Human Subject Testing Implementation. *ASHRAE Trans.*, 2003, 109, 885-897.
- [15] Nilsson H.O., Holmer I.. Comfort climate evaluation with thermal manikin methods and computer simulation models. *Indoor Air*, 2003, 13, 28-37.
- [16] Kingma, B.R.M., Schellen L., Frijns A.J.H., Lichtenbelt W.D.V.. Thermal sensation: A mathematical model based on neurophysiology. *Indoor Air*, 2012, 22, 253-262.
- [17] Lu S., Wang W., Wang S., Hameen E.C.. Thermal Comfort-Based Personalised Models with Non-Intrusive Sensing Technique in Office Buildings. *Appl. Sci.*, 2019, 9, 1768.
- [18] De Dear R., Brager G.S.. Developing an adaptive model of thermal comfort and preference. *ASHRAE Trans.*, 1998, 104, 145-167.
- [19] Rijal H.B., Humphreys M., Tuohy P., Nicol F.. Development of adaptive algorithms for the operation of windows, fans and doors to predict thermal comfort and energy use in Pakistani buildings. *ASHRAE Trans.*, 2008, 114, 555-573.
- [20] Humphreys M.. Outdoor temperatures and comfort indoors. *Batim. Int. Build. Res. Pract.*, 1978, 6, 92.
- [21] Lee S., Bilonis I., Karava P., Tzempelikos A.. A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings. *Build. Environ.*, 2017, 118, 323-343.
- [22] Ghahramani A., Castro G., Karvigh S.A., Becerik-Gerber B.. Towards unsupervised learning of thermal comfort using infrared thermography. *Appl. Energy*, 2018, 211, 41-49.
- [23] Kim J., Zhou Y., Schiavon S., Raftery P., Brager G.. Personal comfort models: Predicting individuals thermal preference using occupant heating and cooling behavior and machine learning. *Build. Environ.*, 2018, 129, 96-106.
- [24] Chaudhuri T., Soh Y.C., Li H., Xie L.. Machine learning based prediction of thermal comfort in buildings of equatorial Singapore. *In Proceedings of the IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), Singapore*, 2326 July 2017, 72-77.

- [25] Farhan A.A., Pattipati K., Wang B., Luh P.. Predicting individual thermal comfort using machine learning algorithms. In *Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden, 2428 August 2015*, 708–713.
- [26] Dai C., Zhang H., Arens E., Lian Z.. Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions. *Build. Environ.*, **2017**, *114*, 1–10.
- [27] Huang C.C., Yang R., Newman M.W.. The potential and challenges of inferring thermal comfort at home using commodity sensors. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 711 September 2015*, 1089–1100.
- [28] Jiang L., Yao R.. Modelling personal thermal sensations using C-Support Vector Classification (C-SVC) algorithm. *Build. Environ.*, **2016**, *99*, 98–106.
- [29] Hussain S., Kang B.H., Lee S.. A Wearable Device-Based Personalized Big Data Analysis Model. In *Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014*, 8867, 236–242.
- [30] Suykens, J.A.K., Van Gestel T., and de Brabanter J.. Least Squares Support Vector Machine. *World Scientific: Singapore, 2002*.
- [31] Amer A.Y.A.. Localized Least Squares Support Vector Machines with Application to Weather Forecasting. *KU Leuven: Leuven, Belgium, 2016*.
- [32] Ashrae, Ashrae Standard 55; American Society of Heating, Refrigerating and Air-Conditioning Engineers. *Inc.: Atlanta, GA, USA, 2017*.
- [33] Dewhirst M.W., Vigiante B.L., Lora-Michiels M., Hanson M., Hoopes P.J.. Basic principles of thermal dosimetry and thermal thresholds for tissue damage from hyperthermia. *Int. J. Hyperth*, **2003**, *19*, 267–294.
- [34] Losing V., Hammer B., Wersing H.. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, **2018**, *275*, 1261–1274.
- [35] Bottou L., Vepnik V.. Local learning algorithms. *Neural Comput.*, **1992**, *4*, 888–900.



## Chapter 4

# Vital Signs Prediction Based on Continuous Monitoring of Hospitalised Patients Using Wearable Technology

### Abstract

In this prospective, interventional, international study, we investigate continuous monitoring of hospitalised patients vital signs using wearable technology as a basis for real-time early warning scores (EWS) estimation and vital signs time-series prediction. The collected continuous monitored vital signs are heart rate, blood pressure, respiration rate, and oxygen saturation of a heterogeneous patient population hospitalised in cardiology, post-surgical and dialysis wards. Two aspects are elaborated in this study. The first is the high-rate (every minute) estimation of the statistical values (e.g., minimum and mean) of the vital signs components of the EWS for one-minute segments in contrast with the conventional routine of 2 to 3 times per day. The second aspect explores the use of a hybrid machine learning algorithm of kNN-LS-SVM for predicting future values of monitored vital signs. It is demonstrated that a real-time implementation of EWS in clinical practice is possible. Furthermore, we showed a promising prediction performance of vital signs compared to the most recent state of the art of a boosted approach of LSTM. The reported mean absolute percentage errors of predicting one-hour averaged heart rate are 4.1, 4.5, and 5 % for the upcoming one, two and three hours respectively for cardiology patients. The obtained results in this study show the potential of using wearable technology to continuously monitor the vital signs of hospitalised patients as the real-time estimation of EWS in addition to a reliable prediction of the future values of these vital signs is presented. Ultimately, both approaches of high-rate EWS computation and vital signs time-series prediction is promising to provide efficient cost-utility, ease of mobility and portability, streaming analytics, and early warning for vital signs deterioration.

## 4.1 Introduction

Monitoring of vital signs of hospitalised patients is of paramount importance to deliver timely and adequate care. Numerous studies that focus on analysing vital signs, hypothesise that many adverse events are preceded with a disruption in the vital signs [1, 2, 3, 4, 5, 6]. Monitoring is conventionally achieved via expensive and cumbersome devices [7]. In addition, these conventional monitoring devices have limited mobility and portability. In hospitals, early warning score (EWS) systems are used to indicate deterioration of the vital signs heart rate, respiration rate, systolic blood pressure, oxygen saturation, and temperature [9]. One limitation of this EWS is that it evaluates the current instantaneous measurement of the vital sign, but provides no past trends or future predictions of vital signs. Another limitation is the low frequency of observations in clinical practice which is typically between two to three times a day [10]. This relatively low frequency results from monitoring vital signs with cumbersome devices in combination with manual recording of the EWS by the nurses (e.g., respiration rate).

Therefore, continuous monitoring of the vital signs of hospitalised patients using wearable technology is expected to overcome the limitation of the conventional low-rate measurement of EWS at the hospital. Furthermore, continuous monitoring of vital signs provides the medical staff a more complete picture and clinical insight into the patients' health status and progression [8]. Motivated by these challenges, we aim in this paper to develop a model for predicting future values of vital signs in addition to continuously monitoring of EWS using wearable technology. For critical care patients, predicting any adverse events based on vital signs analysis is investigated exhaustively in several studies such as sepsis prediction [11] and mortality prediction [12]. However, for general wards, we hypothesise that predicting the future value of vital signs can provide early detection of any deterioration of the patients' health state. Moreover, due to the difficulty to obtain a real-time annotation/labelling of the monitored vital signs, we have to predict values instead of labels or scores. An important aspect in this regard is the recording frequency since it influences the magnitude of the prediction horizon. Recently Shiyu Liu et al. [13] proposed a generative boosting approach of long-short term memory (LSTM) deep neural networks to predict the vital signs values for specific prediction horizons (up to three hours ahead). Their dataset included demographic data and vital signs of 177 medical patients (non-specific ward) at regular intervals of 5 minutes over 24 hours. Moreover, they used a mutual information-based clustering algorithm to select a more representative dataset to train the generative model. To the authors knowledge this study describes one of the most performant algorithms in the state of the art for predicting vital signs. More specifically, they predicted heart rate and systolic blood pressure 20 minutes in advance, with a mean absolute percentage error of 7.41 and 6.17 %, respectively.

In our study, we aim at investigating two aspects of monitoring vital signs and EWS using wearable technology. Firstly, we aim at computing and monitoring the vital signs components of the EWS at a high-rate (i.e. every minute). Secondly, we aim at predicting at least one-hour ahead statistical attributes of the different vital signs (i.e. minimum, maximum, mean). The latter

approach is achieved by using the localised learning approach of KNN-LS-SVM. The KNN-LS-SVM approach is chosen to provide an online prediction with more personalised characteristics of the model [14]. To the authors knowledge, it is the first time to apply a localised learning algorithm (i.e. kNN-LS-SVM) to predict vital-signs time series. It is worth mentioning that kNN-LS-SVM for regression is proposed before in the studies of [15, 16, 17] as a continuation of the fundamental study of local learning algorithms [18].

This article is arranged as follows; after the introduction section, an overview of the EAGLE study is introduced in section 2. Then, the third section is about the methods that are going to be applied to analyse the vital signs for the two proposed approaches. In section 4, the results of the two proposed approaches are illustrated. Next we discuss our findings in section 5. Finally, the conclusion of this study is presented in section 6.

## 4.2 Data Generation

The EWS is a scoring system implemented in many care centers worldwide, which helps to prevent clinical deterioration of the patients at general wards, via early recognition of disease worsening (i.e. change in one or more vital parameters) [9, 19, 20, 21, 22]. In detail, essential vital parameters (heart rate, respiration rate, blood pressure, temperature, oxygen saturation, and neurological responsiveness) are recorded multiple times a day. However, the classic EWS system has its limitations as mentioned earlier in the introduction.

The EWS scoring system has already been proven to be an effective tool in reducing clinical deterioration, reducing the admission to intensive care units and thus overall reducing mortality [19, 23, 25]. However, as mentioned above the EWS is measured in clinical practice at a rather low frequency. Therefore, estimation of the EWS score via continuously monitored parameters is expected to further increase patient survival.

The data in this work were generated in the framework of the The new gold standard: the Early Warning Score algorithm (i.e. EAGLE) study, which was part of the Interreg EMR project WearIT4health on the development of wearables for hospitalised patients. The objectives of the EAGLE study are: (1) to collect continuously monitored vital parameter data using wearable sensors of patients admitted at the general ward. (2) To develop an algorithm that can early identify clinical deterioration and can optimise the application of the conventional EWS system, and (3) to explore the possibility to predict the measures of the vital signs and EWS for specific prediction horizons.

### 4.2.1 Study Design

The EAGLE study is an international, multicentre, prospective, and interventional study that includes the following study sites: Centre Hospitalier Universitaire de Liège (CHU) at Liège, Belgium. Ziekenhuis Oost-Limburg (ZOL) at Genk, Belgium and Academisch Ziekenhuis Maastricht (aZM) at Maastricht, The Netherlands. The data collection is done according to the study

protocol in accordance with GCP and ICH guidelines. The study was approved by the ethical commission of CHU (B707201836800) and aZM.

### 4.2.2 Study Population

Patients of the participating hospitals were selected taking into account the following exclusion criteria: i) eligible patients should not be younger than 18 years old, ii) eligible patients should not be recruited from high-intensity units (i.e. intensive care units, coronary care units and emergency rooms), iii) eligible patients should not participate in other studies that might influence the study results (e.g. experimental medication that could affect the heart rate), iv) eligible patients should not be hospitalised for a 1 day clinic stay, v) eligible patients should not have infectious diseases.

This resulted in the following target populations: 52 cardiology patients at ZOL and 21 at CHU (on average 23 hours x 2 days), 10 post-surgical patients at aZM (on average 23 hours x 1 day) and 7 dialysis patients at CHU (on average 4 hours x 2 sessions).

The used measuring device in this study is SOMNOtouch-NIBP (Figure 4.1) <http://www.somnomedics.eu> of the CE and ISO certified company of SOMNOmedics GmbH. The SOMNOtouch-NIBP device is capable of measuring several physiological parameters needed for the estimation of the EWS score in a continuous, non-invasive manner.

### 4.2.3 Measuring Device

The SOMNOtouch-NIBP device consists of the following sensors:

1. Accelerometer ( $x$ ,  $y$  and  $z$ -axis) ( $m/s^2$ ) capable of detecting motion and the position of the body
2. Pulse oximeter to measure the oxygen saturation (%)
3. Photoplethysmograph (PPG) which is used in combination with the ECG to derive cuffless, non-invasive blood pressure using pulse transit time (PTT) technique.
4. 3-channel ECG from which the heart rate and respiration rate can be derived.
5. Intercostal electromyography (EMG) electrodes to estimate the respiration based on muscle Movement.

## 4.3 Methods

### 4.3.1 High-Rate EWS Computation

Based on the continuous measurement of the vital signs (heart rate, blood pressure, oxygen saturation  $SpO_2$  and respiration rate) the EWS component of each vital sign is computed every

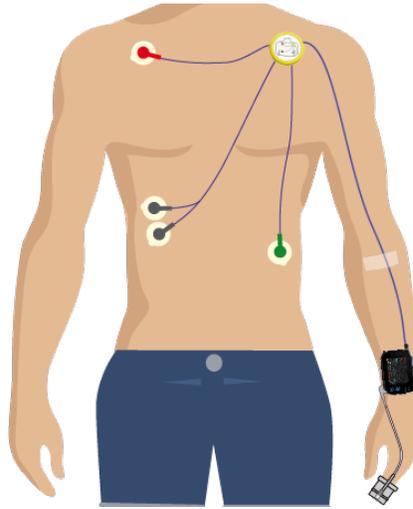


Figure 4.1: A scheme illustrating the allocation spots of SOMNOtouch-NIBP device and electrodes (adapted from <http://www.somnomedics.eu>).

minute using a wearable device. However, the used device does not measure body temperature, hence, the EWS component of temperature is excluded from this study. In order to avoid the instantaneous calculation of EWS components of each vital sign, the statistics of the vital signs are calculated within each minute (i.e. mean, median, minimum and maximum) and based on them the EWS component of each vital sign is computed. The EWS component of each vital sign is computed based on the standard ranges of the EWS used by the hospital of ZOL as listed in Table 4.1.

Table 4.1: Early Warning Scores system based on ZOL Hospital

SCORE	3	2	1	0	1	2	3
Temperature (°C)		<35.1	35.1-36.5	36.6-37.5	>37.5		
Heart Rate (BPM)		<40	40-50	51-100	101-110	111-130	>130
Respiration Rate (BPM)		<9		9-14	15-20	21-30	>30
Oxygen Saturation (%)	<91	91-93	94-95	>95			
Systolic Blood Pressure (mmHg)	<70	70-80	81-100	101-180	180-200	>200	

Due to motion artefacts and estimation noise (i.e. HR estimation from ECG), the recorded signals have to be denoised. The denoising filter that is used for this approach is a fourth order Butterworth low pass zero-phase filter with cut-off frequency 0.03125 HZ. The signal is denoised for every minute (60 samples) and the early warning score component of the clean signal's statistics is computed.

### 4.3.2 Vital Signs Time-series Prediction

In this approach, we aim at providing early detection of vital signs' behaviour based on time series prediction. In contrast with conventional EWS monitoring, the early detection in our approach is based on predicting the future measures of each vital sign for specific prediction horizons, based on the historical measurements of these vital signs. Moreover, we aim to develop a predictive approach that is suitable for online prediction and model personalisation. The online model is required to adapt the prediction given new recorded measurements, and the personalisation is required to consider the individuality of each subject. In order to develop a model with these characteristics, we suggest using a localised learning approach that can handle continuously increasing recorded measurements in addition to online predictions. The chosen localised learning approach is k-nearest neighbours least-squares support vector machines (kNN-LS-SVM) [15, 14] that has shown an acceptable performance in different studies tackling the problems of streaming analytics, online prediction and model personalisation [14, 24].

### 4.3.3 Local learning of SVMs

In this section, we start by reviewing the main concepts behind SVMs and localised learning approaches for SVMs. Many localised learning algorithms are developed; these algorithms can be divided into two categories. The first category is multiple prototype method which mainly relies on partitioning the input space prior to training the models by either clustering (e.g., k-means) or *Voronoi* partitioning. Therefore, these algorithms provide offline models which do not consider the new streams of data points [15, 27, 28, 29]. A well-known algorithm of that category is Profile SVM (PSVM) [27]. For the second category, some localised learning algorithms rely on weighing functions that train models online. These algorithms are called instance-based learning algorithms (IBL) as locality refers to the training instances in the vicinity of the new test instance. The weighting function can be a square kernel that provides uniform weight to specific neighbouring instances and excluding all other instances such as k-nearest neighbours. Moreover, the weighting function can be a smooth kernel which gives decaying weights to all instance with the distance from the test instance. The disadvantage of the smooth kernel function (e.g., Gaussian and Cosine similarity measure) [18, 30, 31] is that all training instances or an indefinite number of them are included in the training process but with different weights. On the other hand, the square kernel function (i.e. kNN) provides a controlled computational complexity regardless of the incremental data size in contrast with the other weighing functions. However, the disadvantage of it is the crucial influence of the k number. Therefore, kNN-LS-SVM [15, 14, 32] is chosen for its advantages of handling streaming data and providing online modelling given the fixed computational complexity. Hence, We will proceed by introducing the hybrid KNN-LS-SVM algorithm for regression.

### 4.3.3.1 Support vector machines

SVMs are originally presented as binary classifiers, that assign each data instance  $\mathbf{x} \in \mathbb{R}^d$  to one of two classes described by a class label  $y \in \{-1, 1\}$  based on the decision boundary that maximises the margin  $2/\|\mathbf{w}\|_2$  between the two classes. Generally, a feature map  $\phi : \mathbb{R}^d \mapsto \mathbb{R}^p$  is used to transform the geometric boundary between the two classes to a linear boundary  $L : \mathbf{w}^\top \phi(\mathbf{x}) + b = 0$  in feature space, for some weight vector  $\mathbf{w} \in \mathbb{R}^{p \times 1}$  and  $b \in \mathbb{R}$ . The class of each instance can then be found by  $y = \text{sgn}(w^\top \phi(x) + b)$ , where  $\text{sgn}$  refers to the sign function.

Similar to the classification problems, regression models are obtained via estimating the boundary  $L$  based on a set of training examples  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) with corresponding output values  $y_i \in \mathbb{R}$ . In particular, one is interested in parameters  $\mathbf{w}$  and  $b$  that minimise a *loss-function*:

$$\min_{\mathbf{w}, b; \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N (\xi_i + \xi_i^*), \quad (4.0)$$

and are subject to:

$$\begin{aligned} y_i - \mathbf{w}^\top \phi(\mathbf{x}_i) - b &\leq \epsilon + \xi_i, & i = 1, 2, \dots, N, \\ \mathbf{w}^\top \phi(\mathbf{x}_i) - y_i + b &\leq \epsilon + \xi_i^*, & i = 1, 2, \dots, N, \\ \xi_i, \xi_i^* &\geq 0, & i = 1, 2, \dots, N. \end{aligned}$$

The constant  $C$  in (4.0) denotes the *penalty term* that is used to penalise estimation error through the slack variables  $\xi_i$  and  $\xi_i^*$  outside Vapnik  $\epsilon$ -sensitivity loss function in the optimisation process.

The so-called *kernel-trick* avoids the explicit introduction of a feature map  $\phi$  and implicitly allows to use feature spaces of infinite dimensionality. A commonly used kernel is given by the Gaussian kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_0^2}\right),$$

where  $\sigma_0$  denotes the *kernel bandwidth*. Both  $\sigma_0$  and  $C$  can be optimised as hyper-parameters in a cross-validation experiment.

LS-SVM's are obtained by using a least-squares error loss function [33]:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2, \quad (4.1)$$

such that

$$y_i = \mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i, \quad i = 1, 2, \dots, N.$$

Where  $\gamma$  is the regularisation constant for LS-SVM. The optimisation procedure introduces errors  $e_i$  such that  $1 - e_i$  is proportional to the signed distance of  $\mathbf{x}_i$  from the decision boundary. In fact, the non-negative slack variable constraint is removed and the solution of the optimisation

problem can be obtained by a set of linear equations, reducing computational effort [33].

#### 4.3.3.2 KNN-LS-SVM Regressor

Local learning approaches build models that fit the data in the local neighbourhood around a test example and by locally adjusting the model parameters to the properties of the data [18].

While global SVMs consider the same weight for all training instances in the optimisation process (4.1), local learning approaches allow that the training samples near a test point are more influential than others. Localised learning approaches of SVMs [14] are based on weighting functions  $\lambda(\mathbf{x}_s, \mathbf{x}_i)$  that express the similarity between the features vectors of the  $i$ -th data point  $\mathbf{x}_i$  and a test instance  $\mathbf{x}_s$ . For an LS-SVM, this leads to the following cost function:

$$\min_{\mathbf{w}, b; e} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i^2, \quad (4.2)$$

such that

$$y_i = \mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i, \quad i = 1, 2, \dots, N.$$

Weighted least-squares support vector machines [34] use a similar approach, but here a different weighting function can be used for any given test point  $\mathbf{x}_s$ . In this work we will study a binary valued similarity criterion:

$$\lambda(\mathbf{x}_s, \mathbf{x}_i) = \begin{cases} 1 & \text{if } \|x_s - x_i\|_2 \leq r_s \\ 0 & \text{otherwise,} \end{cases}$$

where  $r_s$  is the  $K$ -th smallest distance among  $\{\|x_s - x_i\|; 1 \leq i \leq N\}$ . This formulation leads to the hybrid KNN-LS-SVM method that we will apply on the time-series prediction approach. In particular a regression model is built for each test example using only the training examples located in the vicinity of the test example [15].

KNN-LS-SVM has the additional advantage of sparseness. Indeed, for an LS-SVM or the localised version that uses a continuous similarity function all input data is required to construct the separating hyperplane [34]. This can be seen by solving the optimisation problem (4.1). Using the method of the Lagrangian multipliers, we find:

$$\mathcal{L}(w, b, e; \alpha) = \frac{1}{2} \|w\|_2^2 + \frac{1}{2} \gamma \sum_{i=1}^N \lambda(\mathbf{x}_s, \mathbf{x}_i) e_i^2 - \sum_{i=1}^N \alpha_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i - y_i),$$

where  $\alpha_i$  are the *Lagrangian* multipliers. Thus, for a KNN-LS-SVM the sparseness characteristic

is returned to the LS-SVM. In an online learning mode, this sparseness will result in a computational advantage compared to LS-SVM.

As shown in Figure 4.2, the algorithm of KNN-LS-SVM is implemented as follows:

1. Given a test example  $\mathbf{x}_s$ , compute distances to all training examples and pick the nearest  $K$  neighbours;
2. Train the LS-SVM model with the  $K$  nearest neighbours.
3. Use the resulting regressor to estimate the output of  $\mathbf{x}_s$ .

The parameter  $K$  and the distance metric (e.g. Euclidean, Mahalanobis or Chebyshev) are additional hyperparameters next to the kernel width  $\sigma_0$  and the penalty term  $\gamma$  that are optimised in a cross-validation approach [14]. One challenge does face finding the nearest neighbour in continuously increasing data pool is the search complexity. However, several advanced search algorithms are developed to reduce this complexity [35].

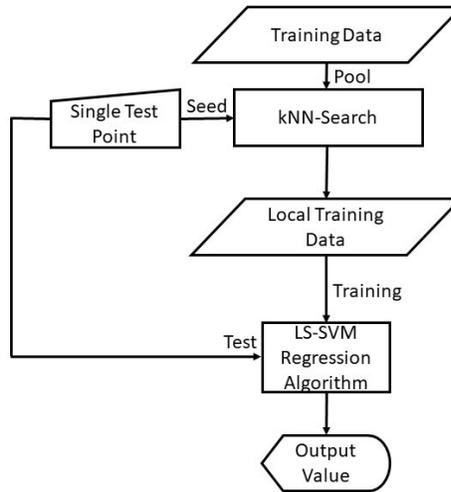


Figure 4.2: A flow chart illustrating the localised learning algorithm of KNN-LS-SVM for Regression.

### 4.3.3.3 Prediction-approach Design

In order to develop a predictive model several considerations are taken into account that can predict statistics of the vital signs for a specific prediction horizon. Firstly, the time-series prediction problem is formulated as a regression problem whose input comprised of the extracted features from time windows of the measurements of the vital signs ( $t$ -hour windows). After signal preprocessing, the unified sampling rate for all signals is set to 1 HZ. Hence, it would be more useful to

predict the statistics of a time window of a specific size instead of exact samples. After discussions with medical experts, we concluded that predicting the statistics of the vital signs measurements for a prediction horizon of 1 hour is of high clinical relevance for the different profiles of patients that we are targeting in this study (i.e. cardiology, dialysis, and post-surgical). Hence, the output of the regression problem is represented by the statistical values (i.e. minimum, maximum, and mean) of the time windows representing the upcoming three consecutive hours (+1, +2, and +3 hours) from the end of the feature-extraction period to test the prediction power of our model. For dialysis patients, the prediction will be restricted to the upcoming hour (+1 hour) as they are only hospitalised during the dialysis sessions (3-4 hours for each session). Furthermore, the next step to define the regression problem is to set the window size, the number of windows, and the overlap percentage between windows from which the features are extracted to be used as input. After testing different window sizes and different overlapping percentages, we found that two overlapping windows of window-size one hour (3600 samples) with an overlap of 50 minutes (3000 Observations) can provide the best possible prediction for the upcoming hour. For practical reasons, this approach is designed to provide a prediction every 10 minutes based on the previous 70 minutes recordings as shown in Figure 4.3. Ultimately, for the train/test division, the leave-one-instance-out approach is used as the training set includes the data instances of all patients except for one data instance from one patient. Hence, by using kNN-LS-SVM, the nearest points can be from the same subject or similar subjects.

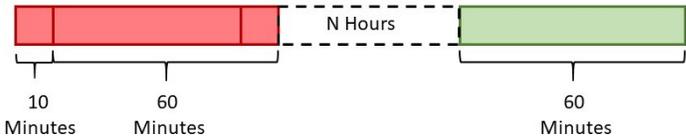


Figure 4.3: The time-windows for feature extraction (red) comprised of two overlapped windows of 60 minutes each with an overlap of 50 minutes resulting in 70 minutes to predict the statistics of the target window (green) after N hours (e.g., 0,1 or 2 hours).

As we have three profiles of patients, namely cardiology, post-surgical, and dialysis patients, it is found to be efficient to test the predictive models on the different profiles individually. However, the main characteristics of the predictive models for each profile will be the same except for the profile of dialysis as the number of observations is relatively low compared to the other wards. Moreover, the extracted features for all profiles are the same, namely minimum, mean, median, maximum, standard deviation, variance and energy from the denoised signal and its first derivative forming 11 features in total excluding energy of the first derivative. The main difference that is imposed on the predictive model characteristics for dialysis patients is the number of nearest neighbours to train the models locally with. For both cardiology and post-surgical

patients, the number of nearest neighbours is 25 data instances, on the other hand, the number of nearest neighbours is 15 for dialysis patients.

Features are extracted from the vital-signs of heart and respiration rate, systolic, diastolic, mean arterial blood pressure, oxygen saturation and pulse pressure (7 variables). These features are extracted from the time-windows of 1-hour (3600 observations), resulting in 77 (11x7) dimensions. As shown in Figure 4.3, the input of the prediction model of kNN-LS-SVM is of two windows, 1-hour each, resulting in 144 (2x77) input dimensions.

## 4.4 Results

In this section, the results of implementing and testing the proposed approaches of high-rate EWS computation and vital signs time-series prediction are presented. For the high-rate EWS computation approach, the outcome of each implementation stage is illustrated. On the other hand, the vital-signs prediction approach is tested on the different profiles of patients using the leave-one-instance-out test procedure. Both approaches are applied to 52 cardiology patients at ZOL and 21 at CHU (on average 23 hours x 2 days), 7 post-surgical patients at aZM (on average 23 hours x 1 day) and 5 dialysis patients at CHU (on average 4 hours x 2 sessions). For post-surgical and Dialysis patients, 3 and 2 patients are excluded respectively due to low quality measurements.

### 4.4.1 High-Rate EWS Computation

For the first approach, the different statistical values of the monitored vital signs (i.e. maximum, minimum, mean, and median) are computed and then the vital sign score is calculated based on these statistical values. Hence, based on the method elaborated in Section 3.1, the original signal is segmented into non-overlapping 1-minute segments. Each segment is denoised and then the statistical values are computed for the denoised signals (i.e. maximum, minimum, mean, and median). From these statistical values, the vital signs scores are estimated based on the depicted ranges in Table 4.1. As shown in Figure 4.4, the results of the different stages of the aforementioned method applied to the heart rate (HR) of a post-surgical patient for approximately 6.5 hours of monitoring are depicted.

As mentioned earlier, the recordings of HR in Figure 4.4 are from a post-surgical patient for the afternoon period. Based on specialists, patients during this period rest after a physiotherapy session in the morning except for the periods of going to the toilet or eating on the table. Therefore, we expect that for those patients, the noise can be due to poor conductivity or local muscular motions.

### 4.4.2 Vital Signs Time-series Prediction

For vital signs prediction, the results will be illustrated for the three profiles of patients of cardiology, post-surgical, and dialysis ward. The targeted vital signs are heart rate (HR), systolic

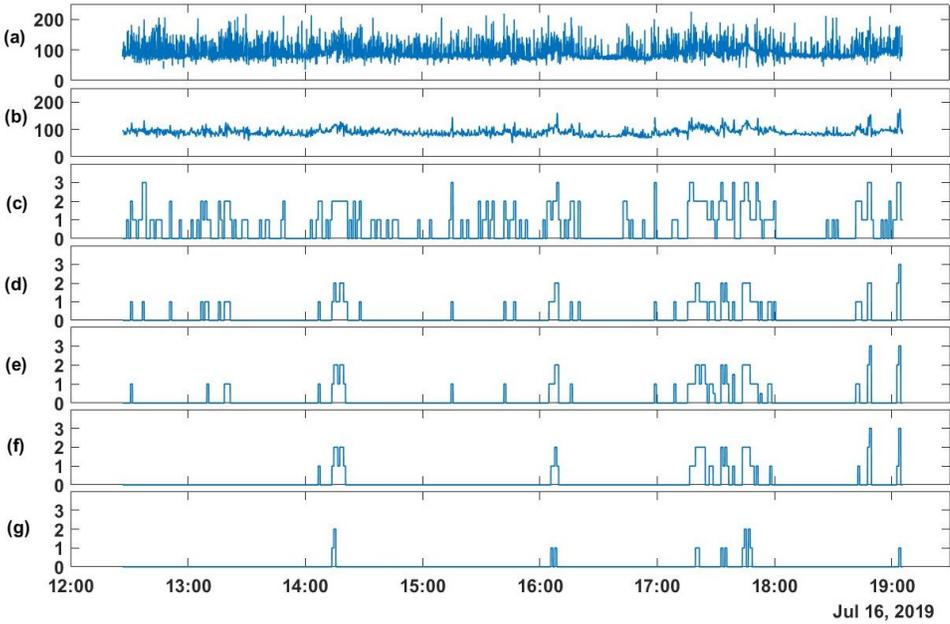


Figure 4.4: From the HR of a post-surgical patient: (a) The raw signal (BPM), (b) the denoised signal (BPM), (c) the maximum, (d) the mean, (e) the median, (f) the median based on moving median filter, and (g) the minimum values of the EWS component of HR for each one-minute segment.

blood pressure (SBP), oxygen saturation ( $SpO_2$ ), respiration rate (RR), and pulse pressure (PP) that is derived from the systolic (SBP) and diastolic blood pressure (DBP). The predicted values are the statistical values (minimum, maximum and mean) of these vital signs for specified future time windows. The upcoming results are based on the leave-one-instance-out test approach.

#### 4.4.2.1 Cardiology and post-surgical Patients

For Cardiology and post-surgical patients, the same regression models of kNN-LS-SVM are applied with the same number of k-nearest points (25 data instances). The choice of the number k is optimised based on a cross-validation procedure based on the error performance as explained in [14]. To evaluate the influence of the proposed algorithm, a naive predictor is proposed to be compared with. The naive predictor is assigning the previously observed mean value to the predicted one. The prediction performance of a naive predictor (NaiveMean) in addition to the prediction results for 1-hour, 2-hours and 3-hours ahead are evaluated using the absolute error as shown in Figure 4.5. These results are for the vital signs HR (5.a), SBP (5.b),  $SpO_2$  (5.c), RR (5.d), and PP (5.e). Furthermore, the mean absolute percentage error (MAPE) of the prediction results are shown in Figure 4.6 for cardiology patients. Regarding  $SpO_2$ , its values are normally

skewed to be within a range of 20% between 80-100%. Therefore, the worst expected value of MAPE is 20%.

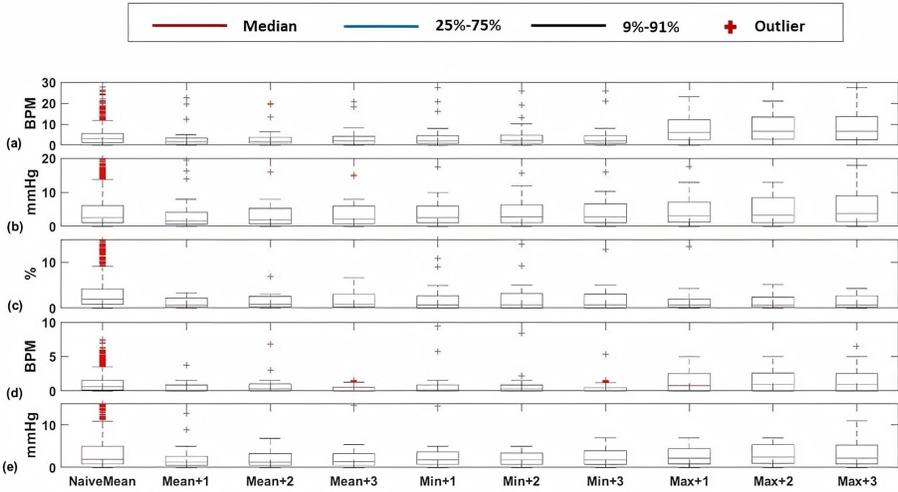


Figure 4.5: Box-plots of the absolute errors of the Naive predictor at the upcoming hour of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming three hours (+1, +2, +3 hours) for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c) SpO<sub>2</sub> (%), (d) RR (BPM), and (e) PP (mmHg) for cardiology patients.

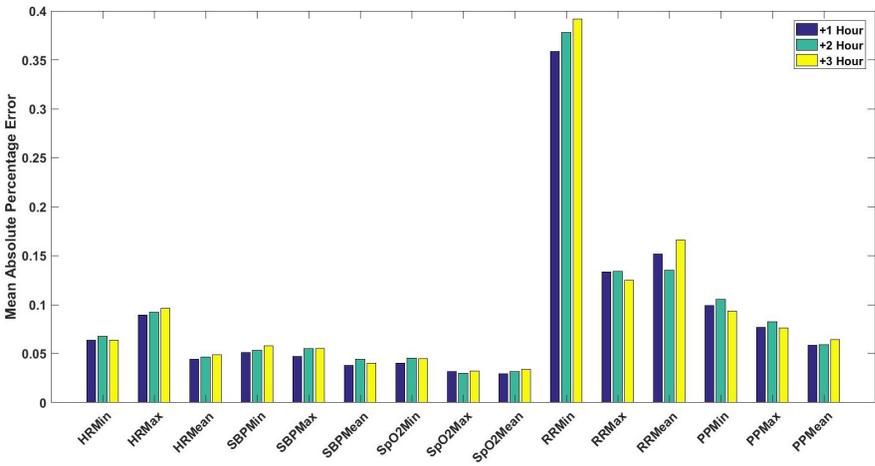


Figure 4.6: The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP, SpO<sub>2</sub>, RR, and PP for the upcoming one, two, and three hours (+1, +2, +3 hours) for cardiology patients.

The next stage is to calculate the EWS for both predicted and actual measures of the vital signs (i.e. HR, SBP, SpO<sub>2</sub>, and RR). In Figure 4.7(a, c, e and g), the normalised histograms of absolute error between the predicted components of EWS and the actual components of EWS for the four vital signs are depicted. Similarly, normalised histograms of EWS error for naive predictor are shown in Figure 4.7 (b, d, f, and h).

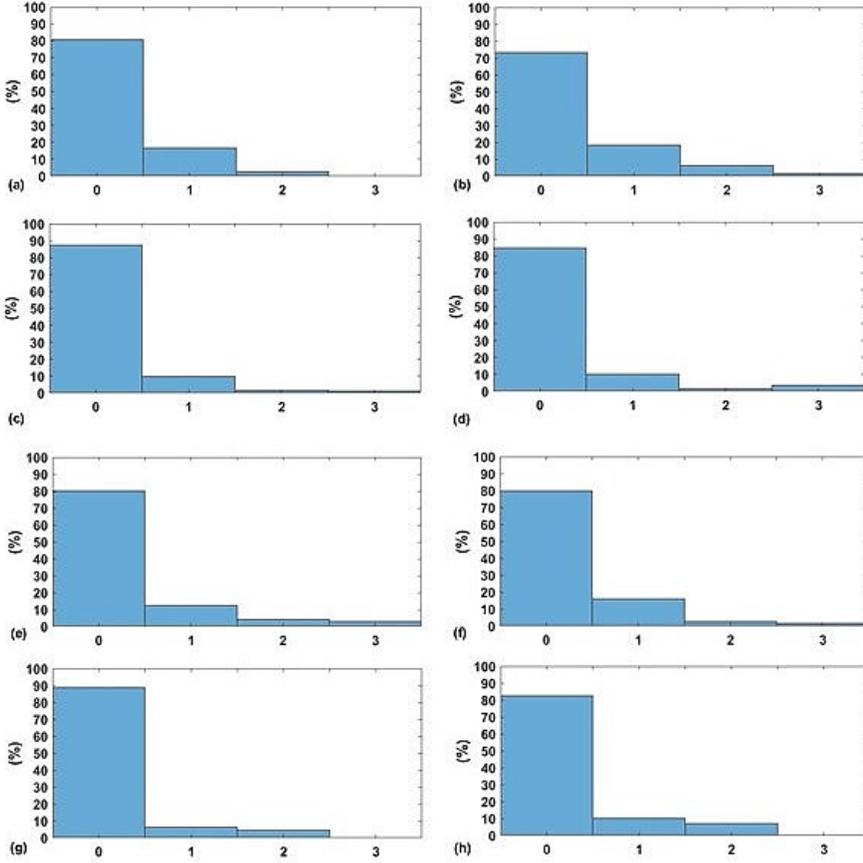


Figure 4.7: The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Naive predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO<sub>2</sub>, and (g,h) RR for cardiology patients.

Similar to cardiology patients, both absolute error and MAPE results of the vital signs HR (8.a), SBP (8.b), SpO<sub>2</sub> (8.c), RR (8.d), and PP (8.e) for post-surgical patients are depicted in Figures 4.8 and 4.9. Moreover, the normalised histograms of the EWS components absolute error of HR, SBP, SpO<sub>2</sub>, and RR respectively are shown in Figure 4.10 (a, c, e, and g). Moreover, normalised

histograms of EWS error for naive predictor are shown in Figure 4.10 (b, d, f, and h).

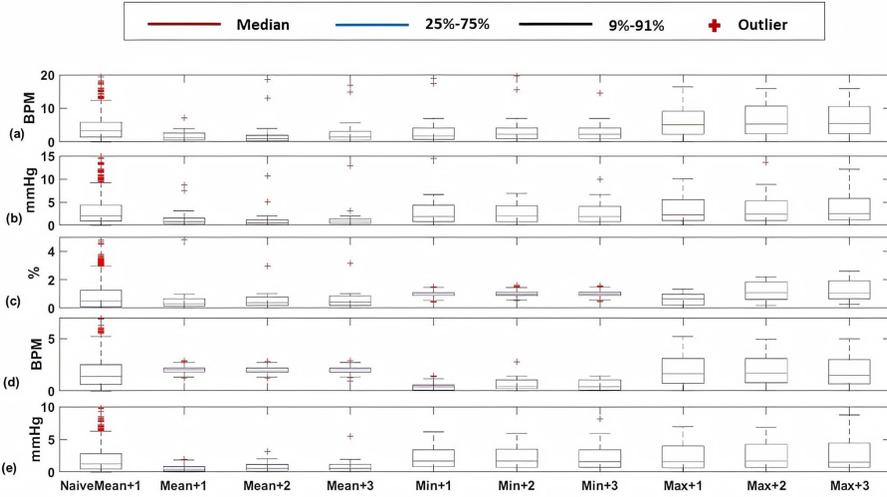


Figure 4.8: Box-plots of the absolute errors of the Naive predictor at the upcoming hour of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming three hours (+1, +2, +3 hours) for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c)  $SpO_2$  (%), (d) RR (BPM), and (e) PP (mmHg) for post-surgical patients.

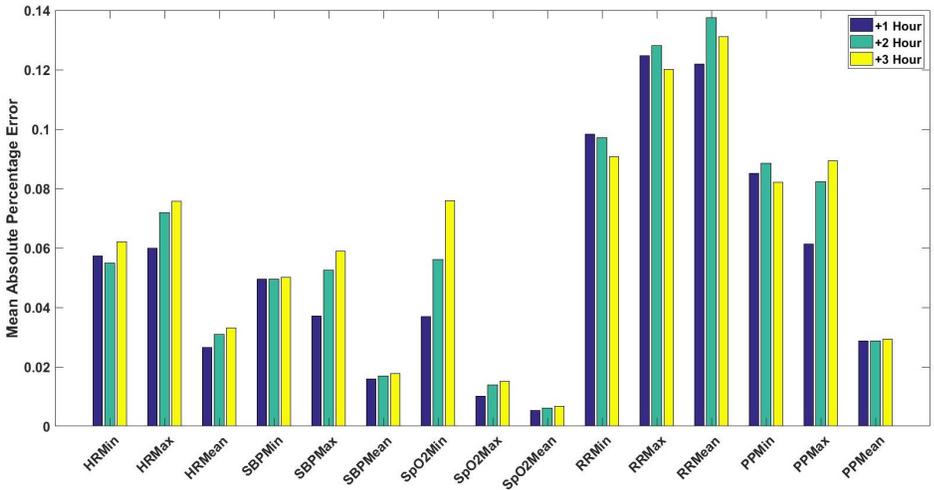


Figure 4.9: The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP,  $SpO_2$ , RR, and PP for the upcoming three hours (+1, +2, +3 hours) for post-surgical patients.

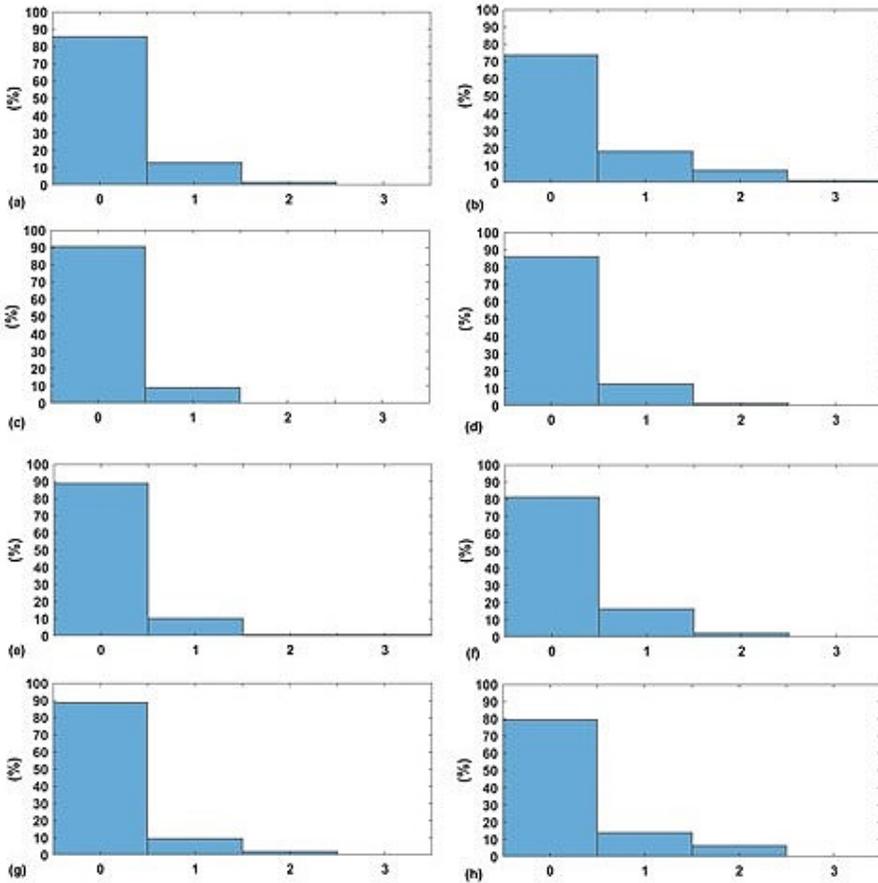


Figure 4.10: The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Nieve predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO<sub>2</sub>, and (g,h) RR for post-surgical patients.

#### 4.4.2.2 Dialysis Patients

For dialysis patients, the predictive models are slightly different from the previous models due to the disease characteristics of these patients. As dialysis patients at haemodialysis are regularly in-hospital (i.e. at least 3 days per week), with a duration of four hours per session, the prediction is only applied for the next hour instead of three hours. Furthermore, it is found that the optimal k-number of the nearest neighbours is 15 data instances based on the cross-validation procedure. The prediction error performance, absolute error and MAPE, for the statistical values of the vital signs for the upcoming hour are depicted in Figures 4.11 and 4.12 respectively. Ultimately, the

normalised histograms for the vital signs components of EWS absolute error are shown in Figure 4.13 (a, c, e, and g). Finally, normalised histograms of EWS error for naive predictor are shown in Figure 4.13 (b, d, f, and h).

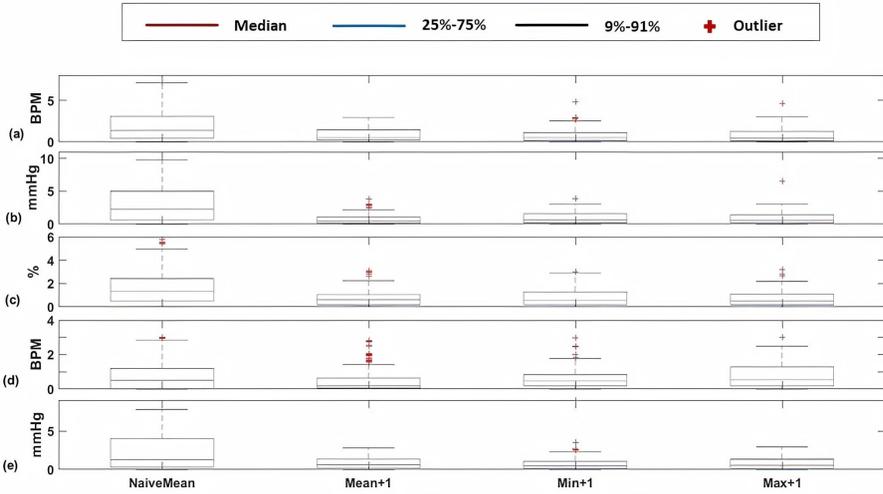


Figure 4.11: Box-plots of the absolute errors of the Naive predictor of the mean value and the proposed algorithm of kNN-LS-SVM of the mean, minimum, and maximum values at the upcoming hour for the vital signs (a) HR (BPM), (b) SBP (mmHg), (c) SpO<sub>2</sub> (%), (d) RR (BPM), and (e) PP (mmHg) for Dialysis patients.

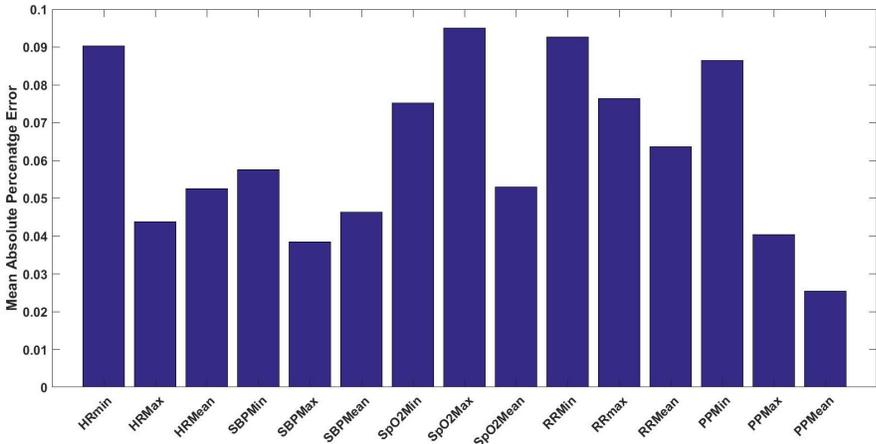


Figure 4.12: The mean absolute percentage error (MAPE) of the predicted statistical values (i.e. minimum, maximum, and mean) for the vital signs of HR, SBP, SpO<sub>2</sub>, RR, and PP for the upcoming hour for dialysis patients.

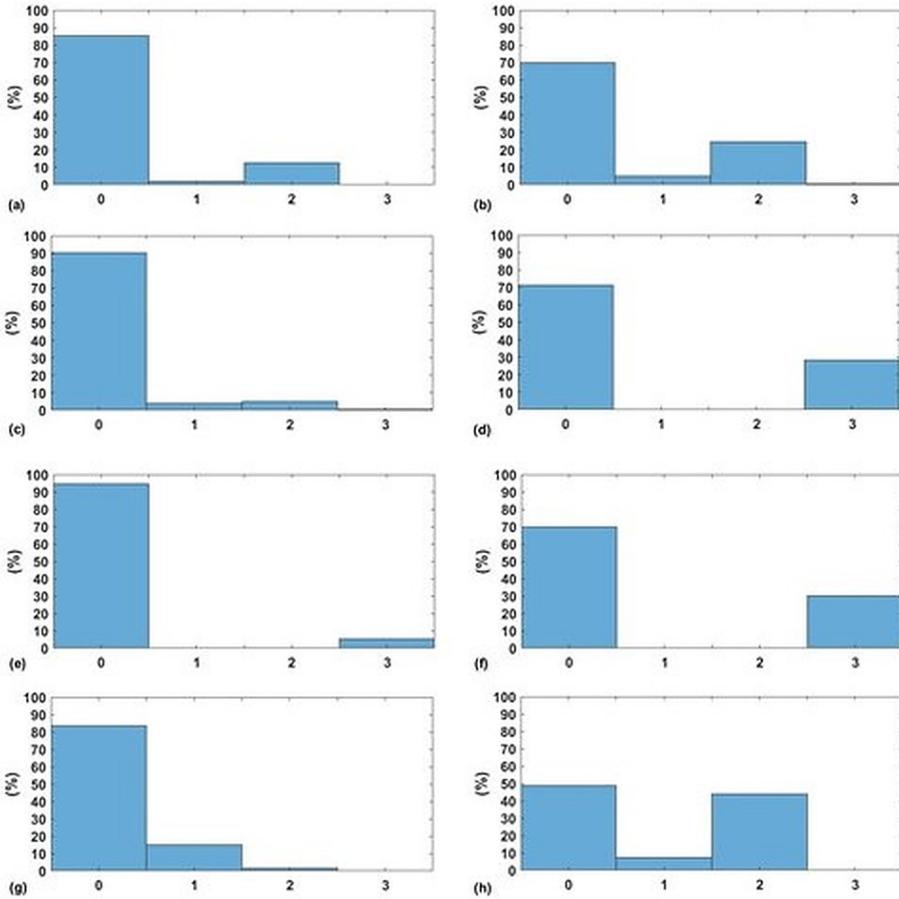


Figure 4.13: The normalised histogram of EWS components absolute Error for both kNN-LS-SVM and Nieve predictors respectively of (a,b) HR, (c,d) SBP, (e,f) SpO<sub>2</sub>, and (g,h) RR for Dialysis patients.

After showing the results of the three profiles of patients it is noticed that for both Cardiology and Post-surgical patients, the significance was achieved at  $\alpha$  of 0.01 resulting in  $p$ -values approximately zeros for all vital signs. The significance is achieved for dialysis patients at  $\alpha$  of 0.05 with  $p$ -values 0.05, 0.06, 0.064, 0.042, and 0.033 for HR, SBP, SpO<sub>2</sub>, RR, and PP respectively.

## 4.5 Discussion

Several studies investigated vital sign data collected from continuously monitored hospitalised patients, more in particular, intensive care unit (ICU) patients. For this purpose, publicly available datasets such as MIMIC have been used of which different versions are available (MIMIC, MIMIC II, and MIMIC III) [36, 37]. However, continuously monitoring vital signs requires expensive cumbersome devices at the ICU [38, 1]. Many of these studies target the early detection of vital signs deterioration based on novelty detection approaches [39, 40, 41, 42, 43]. In our study, monitoring is performed with medically approved wearable technology. Such devices have the advantage to be relatively cheaper and they allow a mobile and portable monitoring approach. Moreover, this study assesses the vital signs by time-series prediction and real-time estimation of EWS components from each vital sign. It is worthy to mention that the used EWS standard thresholds are of the hospital of ZOL, however, our algorithms can be adapted to any EWS standards of any hospital.

As shown in section 4, the first approach provides a high-rate real-time estimation of the EWS components obtained from the vital signs' data. (HR, SBP, SpO<sub>2</sub>, and RR). Real-time EWS components are obtained by estimating vital signs' scores every 60 seconds after signal preprocessing. Such real-time vital signs assessments are already possible for critically ill patients at ICUs that are monitored by expensive equipment and are restrained to their bed. For general-ward patients however, such real-time EWS components estimation is not performed in clinical practice due to the restrictions of, among others, manual and infrequent measurements by nurses. Here, we demonstrate that frequent measurements using wearables for general-ward patients allow real-time estimation of EWS as well. To illustrate the implementation of the first approach, a representative example of a post-surgical patient is used. As shown in Figure 4.4, the raw signal (4.a) of HR is noisy due to motion artefacts and possible conductivity issues. Hence, it is important to denoise the raw signal on segments of one-minute each to discover the underlying trend and deviations from it. For that purpose, a Butterworth low pass filter zero-phase (fourth order) is used to denoise the signals with cut-off frequencies between 0.03 and 0.04 Hz for the different signals. The signal shown in (4.b) is a sequence of denoised non-overlapping segments (one-minute each). After obtaining a clean signal, the EWS can be easily calculated given the hospital standard thresholds for each vital sign as shown in Table 4.1. For practical reasons, we find that instead of providing sample-by-sample EWS we can provide the EWS components of the statistical values of the vital sign within each segment. Therefore, within each segment we provide the maximum, mean, median, and minimum observed vital sign score as shown in Figure 4.4. We notice from these observed vital signs scores that the variance within the same segment is reasonable as the difference between the minimum and maximum scores for the same segment is mostly unity, taking into account that the EWS ranges are continuous without guard-intervals. Moreover, both the mean and the median are very similar which indicates the absence of outliers and consistency within each segment. Furthermore, in Figure (4.f), we show the

median EWS component of HR based on moving median filter. This filter is implemented by extracting the median value within a moving window of one minute width and shifting with one sample. Compared to Figure (4.e), the moving median filter is missing several details especially those that lead to unity EWS component of HR.

For the time-series prediction approach, the results in section 5 show that the prediction of vital signs based on historical values of these vital signs is feasible considering the balance between the feature-extraction horizon (70 minutes) and the prediction horizon (1-3 hours). As mentioned earlier, there are three profiles of patients that are monitored and analysed.

Firstly, Figure 4.5 shows significant difference between a naive predictor, that just predicts the next hour value similar to the most previous measure, and the kNN-LS-SVM predictor. This significance is approved by the paired t-test for all vital signs for different profiles. For the two profiles cardiology and post-surgical patients, the parameters of the kNN-LS-SVM are set to the same values. In particular the number of nearest neighbours is set to  $k = 25$ . In Figure 4.5 and 4.6, the absolute error and MAPE for cardiology patients are illustrated respectively. It is observed that the extreme values (minimum and maximum), especially maximum, have higher errors than the mean. This is due to the high variability of the extreme values from one window to another especially when the window size is one hour. Moreover, the parameters with high average amplitude (e.g., 124 mmHg for SBP) have higher absolute error than that of RR which has low average amplitude (approximately 14 BPM), in contrast with results of MAPE since the opposite is observed. Furthermore, it is observed that the maximum value of HR has the largest absolute error (mean, median, and standard deviation) which may indicate the high fluctuation of the extreme values of HR. On the other hand, the highest MAPE is observed by the minimum value of RR as the minimum values can be approximately 5 BPM, hence having an absolute error of 2 leads to 40% MAPE. Another important observation is that the error performance evolution from one prediction horizon to another is at maximum 1.8, 2.5% for absolute error (mean) and MAPE respectively. In Figure 4.7, the normalised histograms of EWS components absolute error of the vital signs HR, SBP, SpO<sub>2</sub>, and RR are depicted. These histograms show the dominance of zero error with minimum 80% (HR and SpO<sub>2</sub>) and maximum 88% (RR), but the unity error of maximum 14% (HR) and minimum 7% (RR). The rest of the possible absolute error values (2 and 3) are minority with maximum 5% (SpO<sub>2</sub>).

For post-surgical patients, as shown in Figures 4.8 and 4.9, we notice that the fluctuation of absolute error of blood pressure parameters (SBP and PP) in addition to HR is low compared to cardiology patients. Moreover, the MAPE for all vital signs is less than that of cardiology patients. In Figure 4.10, the normalised histograms of EWS components absolute error show a higher dominance of zero error with minimum 87% (HR) and maximum 90% (SBP). However, the unity error of maximum 14% (HR) and minimum 10% (RR). The rest of the possible absolute error values (2) are minority with maximum 2% (SpO<sub>2</sub>) and now error of 3 is observed.

For dialysis patients, some features about the model and the approach are modified to meet the characteristics of this profile of patients. As mentioned earlier, this profile of patients is nor-

mally hospitalised and monitored during the dialysis sessions (3-4 hours/session for haemodialysis). Hence, it is found that one-hour prediction is sufficient for that profile, also the optimal number of nearest points is 15 data-points for the kNN-LS-SVM regression model. As shown in Figures 4.11 and 4.12, the overall error performance for +1 hour prediction is comparable to that of the other profiles of patients either for absolute error or MAPE. One remark is that, in contrast with the other profiles, the prediction error for the minimum value for HR and SBP is higher than that of the maximum. That may indicate that extremely low values of these vital signs for those patients are less predictable than the other values. One general observation for the different profiles of patients is that the error performance is not systematically degrading with increasing the prediction horizon. A possible interpretation is regarding the temporal resolution provided by the window size. Hence, increasing the window-size can eliminate this observation, but this will lead to decrease our data-size which affects the performance. Therefore, we keep the window-size of one hour which provides the best error performance together with keeping a proper data size. In Figure 4.13, the normalised histograms of EWS components absolute error show a maximum zero error of 94% ( $SpO_2$ ) and minimum 83% (RR). The unity error is observed, in contrast with the other profiles, less than that of 2 (HR, SBP and  $SpO_2$ ) and 3 ( $SpO_2$ ). This can be due to the relatively smaller size of data compared to the other profiles, hence, the distribution became not skewed normal as expected. Moreover, the errors of 2 and 3 are mainly observed in the vital signs of  $SpO_2$  and RR. This can be interpreted as a result of the narrow EWS ranges for both vital signs as shown in Table 4.1.

From the obtained histograms, we have noticed that the observed EWS error has in few cases the value of 3 especially for the vital signs of SBP and  $SpO_2$ . By investigating the instances that have an EWS error of 3 for SBP, we found that this error of 3 is due to noise. This conclusion is drawn from observing abrupt changes in the EWS from 0 to 3 or vice versa. Moreover, there are only 10 minutes between two consecutive observations representing windows of one hour each, with 50 minutes overlap. Therefore, we conclude that these extreme EWS errors are more due to artefacts and their associated noise than physiological behaviour. On the other hand, the EWS error of 3 for  $SpO_2$  is either due to the small range of  $SpO_2$  between scores 0 and 3 (90-96%) as shown in Table 4.1 or due to noise.

A general observation regarding histograms (Figures 4.7, 4.10, 4.13) of the Naive predictor EWS error is that the performance for both cardiology and post-surgical patients is worse than our proposed model but not radically worse. This is due to the high percentage (> 80%) of the observed 0 EWS for all vital signs along the period of recording. On the other hand, the histogram of the Naive predictor for dialysis patients is much worse than that of our proposed model. This can be because of the more dynamic behaviour of the vital signs during the dialysis sessions. Hence, the variability from one observation to another can mislead the Naive predictor.

Ultimately, it is noticed that the locally selected instances for the three profiles of patients are not all from the same patient. Moreover, more than 50% of the nearest neighbours are from different subjects. Which shows the advantage of using our localised learning approach that

can provide an accurate online performance using same-subject data and together with the most similar instances of other subjects. Therefore, the concept of model personalisation is extended to include similar subjects in addition to same-subject data.

After showing the results for each of the considered profile of the patients, we discuss our predictive model's error performance in the light of the study of Shiyu et al. [13] which is the most relevant one according to the authors. However, the differences between the two studies have to be considered regarding the prediction horizon and the nature of the predicted values (i.e. exact or statistical). In their study, Shiyu et al. show that their best results of the boosting generative LSTM approach for vital signs prediction are 7.41% for HR and 6.17% for SBP of MAPE. These prediction results are for 20 minutes prediction horizon only. On the other hand, our prediction results are not for exact values but for statistical values of time windows. However, for comparison purposes, the mean values of the future time windows are used considering the average prediction errors of the three prediction horizons (i.e. +1, +2, +3 hours). For HR, the MAPE's are 4.5, 3.1, and 6.3% for cardiology, post-surgical, and dialysis patients respectively. For SBP, the MAPE's on average are 4.4, 1.8, and 4.6% for cardiology, post-surgical, and dialysis patients respectively. Based on these results, we claim that our approach is promising given the prediction horizons and the error performance. Moreover, a more recent study [44] is presenting an LSTM based predictive model to predict the vital signs of hospitalised patients. However, the used time windows are relatively small since the best results obtained with 1-minute windows and the number of admissions needed for the best performance is 2500 admissions. The developed deep network requires 22 hours for training and the best error performance obtained is 81%. Ultimately, as mentioned earlier, the vital sign of body temperature, one of the five main EWS components, is missing in this study as the used wearable device does not provide it. However, our resulting outcome of continuous monitoring and time-series prediction of EWS components of the other vital signs is still informative. These components can be provided to the medical staff in an aggregated form to give the complete EWS once the body temperature is available. Moreover, the body temperature of general wards patients is not as dynamic as the other vital signs (e.g., HR). Hence, we do not expect any difficulty to predict it once its values are provided.

## 4.6 Conclusions

In this study, we proposed two approaches for high-rate EWS computation and time-series prediction based on vital signs measured on hospitalised patients using a wearable device in a cardiology, post-surgical, and dialysis ward. The first approach is the estimation of the high-rate (every minute) scores of the statistical values of the measured vital signs of HR, SBP, RR, and SpO<sub>2</sub>, for each one-minute segment, based on the depicted thresholds in Table 4.1. On the other hand, the second approach comprises predictive models by which the future values of monitored vital signs in addition to the pulse pressure (PP) are predicted. This approach is designed to provide a prediction result every ten minutes. The used technique is the hybrid machine learning

algorithm of kNN-LS-SVM. The predicted values are the statistical values (i.e. minimum, maximum, and mean) of the future time-windows within specified prediction horizons. The used prediction horizons for both cardiology and post-surgical patients are 1-, 2-, 3-hours ahead (+1, +2, +3 hours). For dialysis patients, the prediction horizon is only the upcoming hour due to the relatively short stay. The prediction performance is evaluated based on the error metrics of the absolute error and mean absolute percentage error (MAPE), and the followed test procedure is that of leave-one-out. The prediction error performance shows outperforming results compared to a naive predictor as well as to the best performing and most recent state-of-art [13]. Hence, we conclude that our prediction approach can provide an acceptable prediction performance that can add a predictive insight to the medical staff monitoring the health status of the monitored patients. Furthermore, the prediction approach can handle both online prediction and streaming analytics as a main feature of the used method of kNN-LS-SVM in addition to model personalisation. For the high-rate estimation of EWS, the proposed approach shows the possibility to provide an online estimation of the EWS based on the real-time signal preprocessing and vital sign score computation. Ultimately, continuous vital signs monitoring using wearable technologies can provide a real-time estimation of the EWS and time-series prediction using a localised learning algorithm. In this way, the combination of wearables and machine learning can contribute to a more accurate monitoring of patients in hospital settings.

For future work, we would suggest investigating the patients taking into consideration their status during their hospitalisation stay. For instance, we would expect clinicians to give a label to the analysed patients regarding their likelihood of deterioration. In this case, we would test our models specifically on those critical cases to evaluate the efficiency of our models with such cases.



# Bibliography

- [1] Brekke I. J., et al.. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PLoS one*, 14.1,2019.
- [2] Kause J., Smith G., Prytherch D., Parr M., Flabouris A., Hillman K.. A comparison of Antecedents to Cardiac Arrests, Deaths and EMERGENCY Intensive care Admissions in Australia and New Zealand, and the United Kingdomthe ACADEMIA study. *Resuscitation*, 2004, 62(3), 275-82.
- [3] Buist M., Bernard S., Nguyen T. V., Moore G., Anderson J.. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation*, 2004, 62(2), 137-41.
- [4] Hillman K. M., Bristow P. J., Chey T., Daffurn K., Jacques T., Norman S. L., et al.. Antecedents to hospital deaths. *Intern Med J*, 2001, 31(6), 343-348.
- [5] Henriksen D. P., Brabrand M., Lassen A. T.. Prognosis and risk factors for deterioration in patients admitted to a medical emergency department. *PLoS One*, 2014, 9(4).
- [6] Barfod C., Lauritzen M. M. P., Danker J. K., Sölétormos G., Forberg J. L., Berlac P. A., et al.. Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency departmenta prospective cohort study. *Scand J Trauma Resusc Emerg Med*, 2012, 20(1).
- [7] Javanbakht M., Mashayekhi A., Trevor M., et al.. Cost utility analysis of continuous and intermittent versus intermittent vital signs monitoring in patients admitted to surgical wards. *J Med Econ.*, 2020, 1-9. doi:10.1080/13696998.2020.1747474.
- [8] Downey C. L., Chapman S., Randell R., Brown J. M., Jayne D. G.. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. *Int J Nurs. Stud.*, 84, 2018, 19-27. doi:10.1016/j.ijnurstu.2018.04.013
- [9] Fang A., Lim W. and Balakrishnan T.. Early warning score validation methodologies and performance metrics: a systematic review. *BMC Med Inform Decis Mak*, 2020, 20(111).<https://doi.org/10.1186/s12911-020-01144-8>.

- [10] Petersen O. A., Kristian A., and Lars S. R.. Frequency of early warning score assessment and clinical deterioration in hospitalized patients: A randomized trial. *Resuscitation*, **2016**, 43191-96.
- [11] Reyna M. A., et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, **2019**.
- [12] Youssef Ali Amer A., et al. Feature Engineering for ICU Mortality Prediction Based on Hourly to Bi-Hourly Measurements. *Applied Sciences*, **2019**, 9(17), PP 3525.
- [13] Liu S., Yao J. and Motani M., Early Prediction of Vital Signs Using Generative Boosting via LSTM Networks. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, **2019**, 437-444.
- [14] Youssef Ali Amer A., Aerts JM, Vanrumste B, Luca S. A Localised Learning Approach Applied to Human Activity Recognition. *IEEE Intelligent Systems*, **2020**.
- [15] Youssef Ali Amer A.. Localised least squares support vector machines with application to weather forecasting, *Masters thesis, KU Leuven*, **2016**.
- [16] Zifang H., and Shyu M.-L.. k-NN based LS-SVM framework for long-term time series prediction. *IEEE International Conference on Information Reuse and Integration*, **2010**.
- [17] Zifang H., and Shyu M.-L.. Long-term time series prediction using k-NN based LS-SVM framework with multi-value integration. *Recent Trends in Information Reuse and Integration*. Springer, Vienna. **2012**, 191-209.
- [18] Bottou L. and Vapnik V.. Local Learning Algorithms. *Neural computation*, **1992**, 4, 888-900.
- [19] Moon A., Cosgrove J. F., Lea D., Fairs A., Cressey D. M.. An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR. *Resuscitation*, **2011**, 82(2), 150-154.
- [20] Jarvis S., Kovacs C., Briggs J., Meredith P., Schmidt P. E., Featherstone P. I., et al. Aggregate National Early Warning Score (NEWS) values are more important than high scores for a single vital signs parameter for discriminating the risk of adverse outcomes. *Resuscitation*, **2015**, 87, 75-80.
- [21] Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al.. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*. **2007**, 33(4), 667-679.
- [22] Subbe C. P., Duller B., Bellomo R.. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. *Crit Care*, **2017**, 21(1), 52.

- [23] Smith M. E., Chiovaro J. C., O'Neil M., Kansagara D., Quinones A. R., Freeman M., et al.. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Annals of the American Thoracic Society*, **2014**, *11(9)*, 1454–1465.
- [24] Youssef A., Youssef Ali Amer A., Caballero, N. and Aerts, J.M.. Towards Online Personalized-Monitoring of Human Thermal Sensation Using Machine Learning Approach. *Applied Sciences*, **2019**, *9(16)*, p.3303.
- [25] Smith M. E. B., Chiovaro J. C., O'Neil M., Kansagara D., Quinones A., Freeman M., et al.. Early Warning System Scores: A Systematic Review. *VA Evidence-based Synthesis Program Reports, Washington (DC)*, **2014**.
- [26] Pasti L. et al.. Optimization of signal denoising in discrete wavelet transform. *Chemometrics and intelligent laboratory systems*, *48(1)*, **1999**, 21–34.
- [27] Cheng H., Tan P.-N., and Jin R.. Localized support vector machine and its efficient algorithm. in *Proceedings of the SLAM International Conference on Data Mining*, **2007**.
- [28] Bischl B., Schiffner J., and Weihs C.. Benchmarking local classification methods. *Computational Statistics*, *28(6)*, **2013**.
- [29] Cheng H., Tan P., and Jin R.. Efficient algorithm for localized support vector machine. *IEEE Transactions on Knowledge and Data Engineering*, *22(4)*, **2010**.
- [30] Yang H., Huang K., King I. and Lyu M. R.. Localized support vector regression for time series prediction. *Neurocomputing*, *72(10-12)*, **2009**.
- [31] Karevan Z., Feng Y., and Suykens J.. Moving least squares support vector machines for weather temperature prediction. in *Proceedings of European Symposium on Artificial Neural Networks*, **2017**.
- [32] Berg A., Maire M., and Malik J.. SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *2*, **2006**, 2126–2136.
- [33] Suykens J. A. K., Vandewalle J.. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.*, **1999**, *9*, 293–300
- [34] Suykens J. A. K., De Brabanter J., and Lukas L., and Vandewalle J.. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, **2002**, *48(1)*, 85–105.
- [35] Cayton L.. Fast nearest neighbor retrieval for bregman divergences. *Proceedings of the 25th International Conference on Machine Learning*, **2008**, 112–119. doi:10.1145/1390156.1390171. ISBN 9781605582054.

- [36] Saeed M., Villarroel M., Reisner A. T., Clifford G., Lehman L. W., Moody G., Heldt T., Kyaw T. H., Moody B., Mark R. G.. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit. Care Med.* **2011**, *39*, 952–960.
- [37] Johnson A. E. W., Pollard T. J., Shen L., Li-wei H. L., Feng M., Ghassemi M., Moody B., Szolovits P., Celi L. A., Mark R. G.. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035.
- [38] Lehman L. W., Adams R. P., Mayaud L., et al.. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE Journal Biomed. Health Inform.*, *19(3)*, **2015**, 1068–1076. doi:10.1109/JBHI.2014.2330827
- [39] Reyes-García J., Galeana-Zapién H., Galaviz-Mosqueda A. and Torres-Huitzil A.. Evaluation of the Impact of Data Uncertainty on the Prediction of Physiological Patient Deterioration. *IEEE Access*, *6*, **2018**, 38595–38606. doi:10.1109/ACCESS.2018.2853701.
- [40] Pimentel M. A. F., Clifton D. A., Clifton L., Watkinson P. J. and Tarassenko L.. Modelling physiological deterioration in post-surgical patient vital-sign data. *Med. Biol. Eng. Comput.*, *51*, **2013**, 869–877.
- [41] Clifton L., Clifton D. A., Watkinson P. J. and Tarassenko L.. Identification of patient deterioration in vital-sign data using one-class support vector machines. *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, **2011**, 125–131.
- [42] Clifton L., Clifton D. A., Zhang Y., Watkinson P., Tarassenko L. and Yin H.. Probabilistic novelty detection with support vector machines. *IEEE Trans. Rel.*, *63(2)*, **2014**, 455–467.
- [43] Luca S.E., Pimentel M.A F., Watkinson P.J., and Clifton D.A.. Point process models for novelty detection on spatial point patterns and their extremes. *Computational Statistics and Data Analysis*, *125*, **2018**, 86–10.
- [44] Da Silva D. B., Schimidt D., Da Costa C. A., Da Rosa R. R., Eskofier B.. DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. *Expert Systems with Applications*. **2020 Aug 27**, 113905.

## Chapter 5

# Feature Engineering for ICU Mortality Prediction Based on Hourly to Bi-Hourly Measurements<sup>1</sup>

### Abstract

Mortality prediction for intensive care unit (ICU) patients is a challenging problem that requires extracting discriminative and informative features. This study presents a proof of concept for exploring features that can provide clinical insight. Through a feature engineering approach, it is attempted to improve ICU mortality prediction in field conditions with low frequently measured data (i.e. hourly to bi-hourly). Features are explored by investigating the vital signs measurements of ICU patients, labelled with mortality or survival at discharge. The vital signs of interest in this study are heart and respiration rate, oxygen saturation and blood pressure. The latter comprises systolic, diastolic and mean arterial pressure. In the feature exploration process, it is aimed to extract simple and interpretable features that can provide clinical insight. For this purpose, a classifier is required that maximises the margin between the two classes (i.e. survival and mortality) with minimum tolerance to misclassification errors. Moreover, it preferably has to provide a linear decision surface in the original feature space without mapping to an unlimited dimensionality feature space. Therefore, a linear hard margin support vector machine (SVM) classifier is suggested. The extracted features are grouped in three categories: statistical, dynamic and physiological. Each category plays an important role in enhancing classification error performance. After extracting several features within the three categories, a manual feature fine-tuning is applied to consider only the most efficient features. The final classification, considering mortality as the positive class, resulted in an accuracy of 91.56%, sensitivity of 90.59%, precision of 86.52% and  $F_1$ -score of 88.50%. The obtained results show that the proposed feature engineering approach and the extracted features are valid to be considered and further enhanced for the

---

<sup>1</sup><https://doi.org/10.3390/app9173525>

mortality prediction purpose. Moreover, the proposed feature engineering approach moved the modelling methodology from black-box modelling to grey-box modelling in combination with the powerful classifier of SVMs.

## 5.1 Introduction

Intensive care unit (ICU) patients are admitted because of an acute critical illness or because of the high need for intensive continuous monitoring. In addition, critical ICU patients are prone to rapid deterioration, resulting in a possibly fatal outcome when not monitored closely. Hence, the main challenge at the ICU is to reduce the morbidity of the admitted patients and prevent mortality which has a high likelihood due to severe illness [1]. Mortality prevention requires an intensive monitoring of vital signs, such as heart and respiration rate, oxygen saturation, non-invasive or arterial blood pressure, and so forth, that can capture clinical deterioration earlier and thus improve patient outcome. In the past, multiple scoring systems have been developed (e.g., Acute Physiology, Age, Chronic Health Evaluation II, Simplified Acute Physiology Score, Sequential Organ Failure Assessment) to provide insights and even predictions regarding ICU patient mortality [2]. However, these scoring systems are population-based and often use summarised nongranular data. This calls for the need for an in-depth investigation of vital signs and associated indicators preceding any deterioration using granular continuous data. This investigation can be handled by time-series analytics to understand the behaviour and interaction of different signals.

Most of the ICU mortality prediction studies focus on developing powerful mortality prediction models [3, 5, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16] in which the higher priority is to provide an accurate label or score about the admitted patients' status. One drawback of such an objective is paying less attention to features simplicity and interpretability, which is the case with deep learning approaches [7, 8, 9, 10, 11, 12, 13]. The key approach in these studies is black-box modelling focusing mainly on predictive model error performance, regardless the interpretability of the features. Hence, the useful information that can be provided to the medical staff is strictly the prediction output. Moreover, a considerable number of relevant studies focus on investigating the continuously recorded vital signs of ICU patients in order to predict the mortality risk of those patients [3, 5, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. A frequently used database in these studies is the medical information mart for intensive care (MIMIC) in its three releases (MIMIC, MIMIC II and III) with different versions [17, 18]. These databases provide a diverse and very large population of ICU patients and contain high temporal resolution data including lab results, electronic documentation and bedside monitor trends and waveforms. In contrast, another approach that is used in investigating critically ill patients in the ICU is mechanistic modelling [19, 20]. Mechanistic modelling is used to describe the system from mathematical and physical dynamics perspective. The main focus of mechanistic modelling is on the system dynamics, the interaction between the different variables and the way they interact from a system perspective taking into account

biological and physiological laws [21]. A mechanistic modelling approach is used in investigating biological systems by developing mathematical models [22, 23, 24].

The main focus in this presented study is to engineer features that can provide clinical insight by which the medical staff is guided through the different parameters. However, prediction accuracy is used in this study to assess the relevancy of the extracted features to the mortality events. Moreover, the dataset in our study is a low frequently measured data (i.e. hourly to bi-hourly) as it is a daily-life dataset that is not generated for research purpose. Moreover, the set of variables, parameters and the investigated population here is limited compared to the ones provided by the MIMIC databases. In the light of the given approaches (Black-box predictive models and mechanistic models) and reviewed studies, our study stands between the two approaches (i.e. pure black-box modelling and mechanistic modelling), as the main focus of the study is to achieve an efficient and informative set of physiologically meaningful features (mechanistic aspect) by means of enhancing the predictive model error performance (black-box aspect) that could be representable for European ICU departments.

From an analytical perspective, the series of recordings for each vital sign is considered a time-series that is sampled by a specific sampling rate. During ICU monitoring, different vital signs are measured and recorded simultaneously, in which the simultaneity facilitates studying correlation, interaction and behaviour between and within the different vital signs. Moreover, the time-series of recorded vital signs enable extraction of different features (typically statistical and dynamic) within segmented time windows, showing the dynamic behaviour of the recorded sign.

Many features can be extracted within consecutive or overlapping time windows for different vital signs, either individually or in combination. This option provides a large number of dimensions that have to be evaluated and adjusted to inform the decision making of the algorithm, which requires an exhaustive investigation. However, such an investigation including a large number of numerical features is not an easy task for medical experts. Due to the high dimensionality issue, it is required to conduct such an investigation via a computational algorithm. In order to cope with these challenges, a simple and powerful classifier is used to explore the features. Ideally, this classifier should handle the problem of classification intuitively with the optimal margin hypothesis [25] which maximises the separability between the different classes. Moreover, the classifier should be capable of dealing with high dimensional data efficiently.

The proposed classifier for this purpose is the linear hard margin support vector machine (SVM) classifier which represents the simplest version of the powerful SVMs. The reason for using SVMs that it is relying on the maximum margin hypothesis. For linear hard margin SVM, it restrictively works efficiently once the input features provide linearly separable data points. With this property, it is feasible to extract features that may have a medical interpretation or physiological ground as the classifier would deal with the features as they are presented in the input space. In other words, it is required to have an acceptable performance only if the data points in the presented feature space are linearly separable with minimum misclassification error [25, 26]. This error intolerance (or minimum tolerance) ensures that the introduced features provide a clear

separation between the different classes (i.e. mortality and survival). Moreover, utilising such a linear classifier controls the dimensionality of the solution as it would only find a solution in the introduced dimensions. In other words, using a more sophisticated classifier (e.g., Radial Basis Function (RBF) SVM) would find a solution in an uncontrolled dimensionality, for instance, RBF SVM reaches infinite dimensionality due to the characteristics of the Gaussian kernel [26].

In this study, the problem is presented as integration between time-series prediction and classification. This integration is obtained by extracting features from the time-series and considering the dynamic behaviour of the time-series to construct the input space of the model. On the other hand, the output of the model is represented by the labels mortality/survival. The prediction is obtained by predicting the state (label) after the final record (last moment at ICU) on average 1.5 days ahead. The final record is the record preceding the patients death (mortality label) or transfer to a lower care ward (survival label).

The objective of this study is to present a proof of concept for exploring features that can provide clinical insight through a feature engineering approach in order to improve the ICU mortality prediction in field conditions with low frequently measured data. The feature engineering approach is based on the hypothesis that utilising the linear hard margin SVM would provide a controllable and interpretable feature extraction approach.

This paper is arranged as follows: After the introduction, the second section of materials and methods comprises data description and an introduction to linear hard margin support vector machines. The third section includes the feature engineering process and results. The fourth section includes the discussion and the final section gives the conclusion.

## 5.2 Materials and Methods

### 5.2.1 Data

Data used for testing and evaluating the features were collected at the hospital Ziekenhuis Oost-Limburg (Genk, Belgium) during the period of 2015–2017. In detail, data were collected from patients hospitalised at the ICU and coronary care unit who were at these wards for at least ten days. Data consisted of vital parameters which were recorded continuously by Philips Intellivue monitors (Philips Electronics Nederland B.V., Amsterdam, The Netherlands), that recorded continuously and was annotated on average hourly to bi-hourly by the nursing staff. The recorded data was extracted from the electronic medical record for a total of 447 different patients, three of them readmitted to the unit again, hence, in total 450 recorded admissions annotated with either mortality or survival by discharge. The age of the patients was 65 ( $\pm 16$ ) years old, 305 of patients were males and 142 were females. The average duration of stay at the ICU is 20.96 days with a minimum of 10 days, maximum of 97 days, median of 30 days and IQR of 20–53 days of ICU stay. The vital parameter data consisted of the heart rate, the respiration rate, oxygen saturation, arterial blood pressure (ABP), non-invasive blood pressure (if ABP was not measured) and body temperature (not frequently). The patient population of the study has different reasons for ICU

admission as shown in Figure 5.1. The local Ethical Committee was notified and approval was obtained (19/0023R).

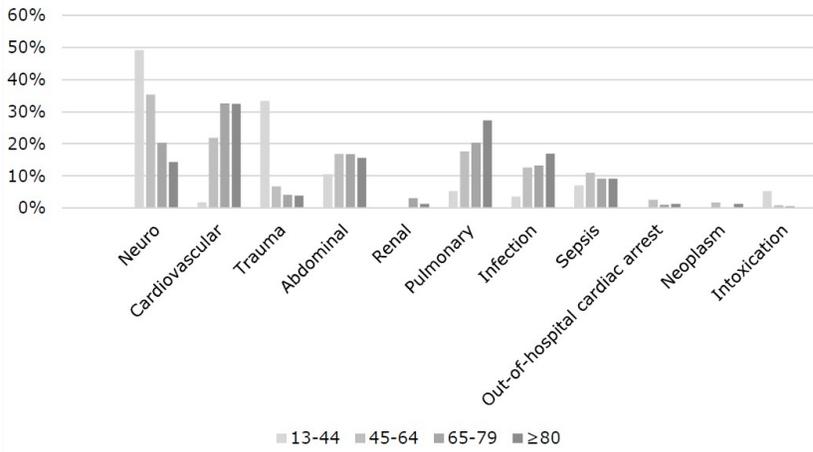


Figure 5.1: Distribution of the patient population and their reason for admission. The population was divided into age categories of 13-44, 45-64, 65-79 and >80 years of age.

### 5.2.2 Hard-Margin SVM

SVMs are originally presented as binary classifiers, that assign each data instance  $\mathbf{x} \in \mathbb{R}^d$  to one of two classes described by a class label  $y \in \{-1, 1\}$  based on the decision boundary that maximises the margin  $2/\|w\|_2$  between the two classes as shown in Figure 5.2. The margin is determined by the distance between the decision boundary and the closest data point from each class [27, 28, 25, 26].

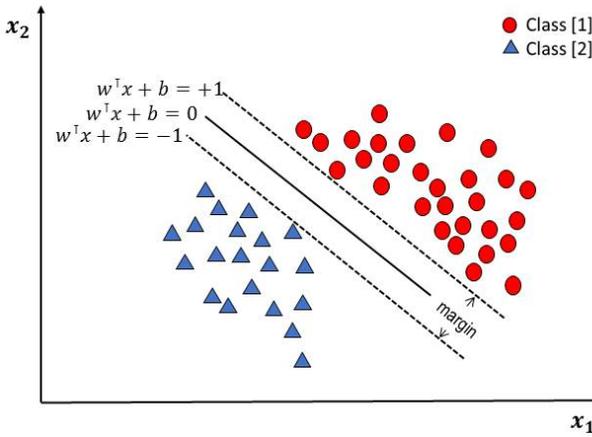


Figure 5.2: Schematic representation of a two-dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised without tolerating any misclassifications (adapted from Reference [28]).

Generally, a feature map  $\phi : \mathbb{R}^d \mapsto \mathbb{R}^p$ , where  $d$  is the number of input space dimensions and  $p$  is the number of feature space dimensions, is used to transform the geometric boundary between the two classes to a linear boundary  $L : \mathbf{w}^\top \phi(\mathbf{x}) + b = 0$  in feature space, for some weight vector  $\mathbf{w} \in \mathbb{R}^{p \times 1}$  and  $b \in \mathbb{R}$ . The class of each instance can then be found by  $y = \text{sgn}(w^\top \phi(x) + b)$ , where  $\text{sgn}$  refers to the sign function.

The estimation of the boundary  $L$  is performed based on a set of training examples  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) with corresponding class labels  $y_i \in \{-1, 1\}$ , where  $N$  is the number of data points. An optimal boundary is found by maximising the margin that is defined as the smallest distances between  $L$  and any of the training instances. In particular, one is interested in constants  $\mathbf{w}$  and  $b$  that minimise a *loss-function* [28]:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^\top \mathbf{w},$$

and are subject to:

$$y_i(\mathbf{w}^\top \phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, N.$$

By applying the lagrangian to the problem we get

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|_2^2 - \left( \sum_{i=1}^N \alpha_i (y_i [\mathbf{w}^\top \phi(x_i) + b] - 1) \right),$$

where  $\alpha_i \geq 0$  are the Lagrangian multipliers for  $i^{th}$  data point. By solving the optimisation

problem

$$\max_{\alpha} \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \alpha),$$

the following optimisation conditions are obtained:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 &\longrightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \phi(x_i), \\ \frac{\partial \mathcal{L}}{\partial b} = 0 &\longrightarrow \sum_{i=1}^N \alpha_i y_i = 0, \\ \frac{\partial \mathcal{L}}{\partial \alpha} = 0 &\longrightarrow y_i (\mathbf{w}^\top \phi(x_i) - b) = 1, \end{aligned}$$

The resulting classifier in both primal space and dual space are

$$\begin{aligned} f(\mathbf{x}) &= \text{sgn}(\mathbf{w}^\top \phi(x) + b), \\ f(\mathbf{x}) &= \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i \phi(x_i)^\top \phi(\mathbf{x}) + b\right). \end{aligned}$$

The dot product  $\phi(x_i)^\top \phi(\mathbf{x})$  is computationally expensive, hence, it is replaced with the *kernel* function  $k(x_i, \mathbf{x})$ , this replacement is known as the *kernel* trick. With the *kernel* trick, there is no need to execute the step of feature map as it is implicitly done by the *kernel* function. Hence, the dual space classifier with the *kernel* trick is

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i k(x_i, \mathbf{x}) + b\right).$$

For practical reasons, we suggest to obtain the linear hard margin SVM from the standard SVM formula that tolerate misclassification errors [29]

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N \xi_i,$$

subject to:

$$y_i (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, 2, \dots, N.$$

where the constant  $C$  denotes the *penalty term* that is used to penalise misclassification through the slack variables  $\xi_i$  in the optimisation process. The linear hard margin SVM can be obtained via penalising the error extremely by giving  $C$  a very high value (e.g.,  $10^{10}$ ). With this trick, we can get a solution with misclassified instances to be investigated through the feature engineering phase.

### 5.3 Feature-Engineering

The process of feature engineering is implemented in an interactive way between extracting new features and the classifier error performance as shown in Figure 5.3. This process is executed in three phases: feature-extraction, evaluation and feature fine-tuning. This process has a closed-loop nature as shown in Figure 5.3, since the three phases influence each other. The proposed three categories of features are statistical features, dynamic features, physiological features. The following sections describe the different feature engineering phases and the extracted features per category.

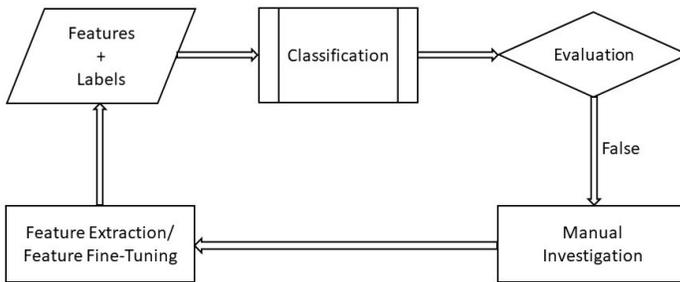


Figure 5.3: A flow chart illustrating the feature engineering methodology.

#### 5.3.1 Evaluation

The engineered features are evaluated by feeding them into a linear hard-margin SVM classifier to predict mortality or survival of a subject. For this purpose, a leave-one-out procedure is used to produce a confusion matrix showing the true positives (TP), the true negatives (TN), the false positives (FP) and the false negatives (FN). The positive class is the mortality state and the negative class is the survival one. Using these numbers, different error performance metrics are calculated (i.e. sensitivity, precision, accuracy and  $F_1$ -score). Furthermore, we evaluate the features by looking at the effect on the number of true positives and true negatives when they are added to the model.

#### 5.3.2 Feature Extraction

Firstly, the period of analysis is within the last 84 observations which represent on average five days before the patients discharge. The first 60 observations (3.5 days on average) out of 84 are considered for feature extraction to predict mortality/survival 24 observations ahead (1.5 days on average) at discharge (i.e. after observation 84). This period is determined after different test trials with different periods and is found to be the most efficient and informative period based on the classification performance. Moreover, this average period of 3.5 days agrees with the experience

of clinical experts in the field. This agreement is based on the fact that there is no standard at the moment that refers to a minimum or maximum of observations to use, in order to provide the best of the care. As it is a human/medical judgement which made based on a combination of patient-specific prognosis and trends, clinical expertise and experience and often corresponds to 3–4 days. The scheme of the feature extraction process is shown in Figure 5.4. Three categories of features are extracted, as described below.

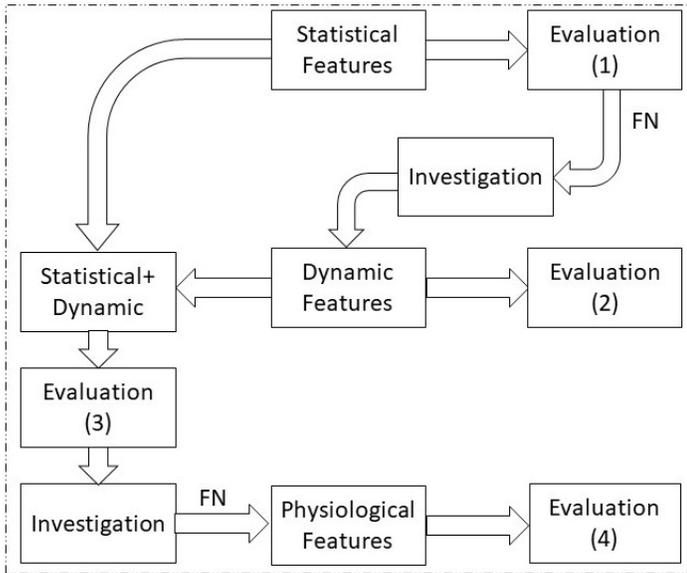


Figure 5.4: A flow chart illustrating the feature extraction process including the three feature categories (i.e. statistical, dynamic and physiological) and the sequence of the process marked by the evaluation steps. Also, in the process, the investigation is applied to the false negative patients only.

5.3.2.1 Statistical Features

The first category of features to be extracted is the set of statistical features which represent the basic characteristics of each time-series within segmented, non-overlapping time windows: *minimum, maximum, mean, median, standard deviation, variance, and energy.*

Statistical features are extracted within windows whose sizes are defined by the number of observations and not by a specific time period due to the nonuniform sampling rate (hourly to bi-hourly) as mentioned before. Extraction is based on the raw measurements of the vital signs and their first derivatives as well as the calculated standard early warning scores (EWS) of these measurements based on ZOL hospital standards. A weak point about statistical features is the static nature of these features as they do not reveal the dynamic behaviour of the time-series. Therefore, another category of features is required to be explored, namely dynamic features.

### 5.3.2.2 Dynamic Features

The extracted dynamic features are *Pearson correlation coefficients*, *crossing-the-mean count*, *outlier-occurrence count*, and *outlier indicator*. *Correlation coefficient* is computed between each pair of vital signs within each window. For this feature, it is necessary to be applied to the *z-score* of the vital signs. *Crossing-the-mean count* of a vital sign is determined by counting the number of times that the recorded vital sign crosses its mean value within each window. This feature indicates the abrupt changes in the vital sign from one observation to another. *Outlier-occurrence count* is computed by counting the number of outliers detected within each window. An outlier is detected by the statistical definition: any point outside the range  $\mu \pm 3\sigma$  for a normally distributed variable is an outlier. For this feature, it is not expected to work with the vital sign of oxygen saturation ( $SpO_2$ ) as it is negatively skewed, however, it will be tested as a feature to prove the concept. Finally, the *outlier indicator* is determined by the difference between the mean and the median of the records within each window.

### 5.3.2.3 Physiological Features

In order to enhance the classification performance, a manual investigation of the misclassified instances (based on the statistical and dynamic features) is required. The investigation is focusing on the false negative patients (i.e. deceased patients classified as survived) as the main objective is relevant to a reliable mortality prediction which is inversely proportional to the false negative count. This manual investigation is based on the measured physiological vital signs and uses physiological process knowledge resulting in physiological features. The different physiological features are described hereafter. By investigating the time-series of false negative patients, a consistent behaviour is noticed within the period of interest, in which the systolic blood pressure (SBP) approaches the diastolic blood pressure (DBP) as shown in Figure 5.5a.

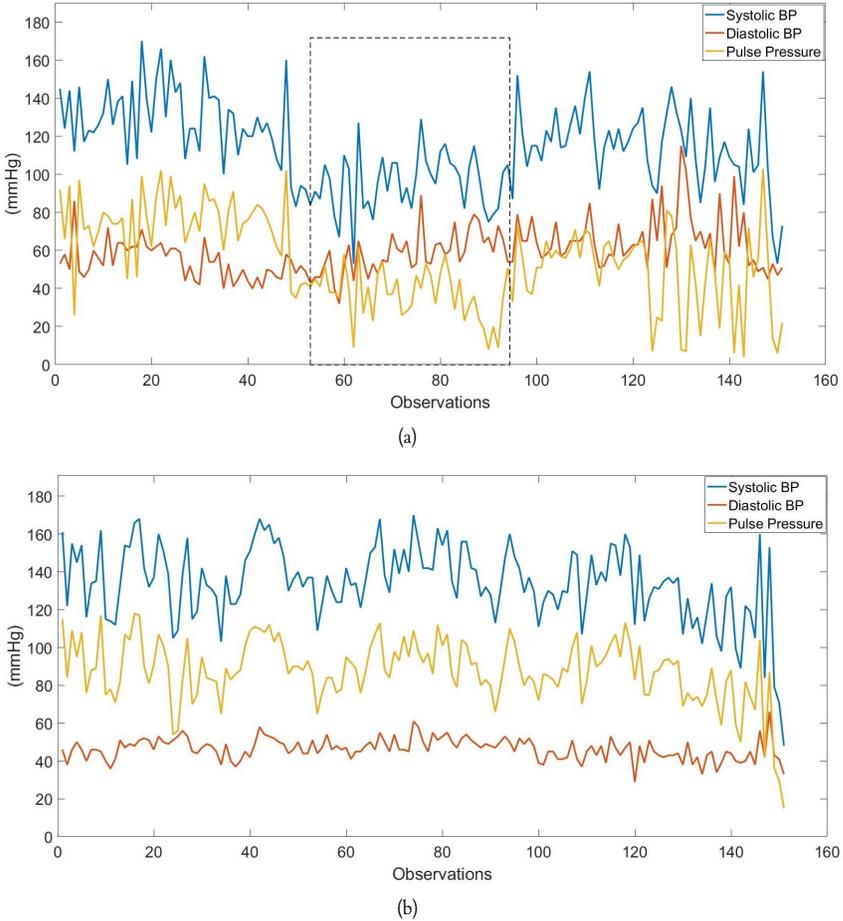


Figure 5.5: (a) Systolic blood pressure (BP), diastolic BP and pulse pressure (PP) of the last 150 observations (approximately the last nine days) of one false negative patient. The dashed window refers to the region where the systolic BP and diastolic BP measurements approach closely. (b) Systolic BP, diastolic BP and PP of the last 150 observations of another false negative patient. The mean value of the pulse pressure is 87.4 mmHg and median 88 mmHg.

It is found that the difference (SBP-DBP) within certain measurement periods is smaller than 20 mmHg. A relevant observation that is noticed with other false negative patients is that this difference is relatively high (i.e. greater than 60 mmHg) during certain measurement periods as shown in Figure 5.5b. This difference between SBP and DBP is also known as the pulse pressure (*PP*) and varies normally in a range between 40–60 mmHg [30, 31]. As the *PP* is a linear combination between two vital signs, it can be considered as a new variable from which both statistical and dynamic features can be extracted. By reviewing medical literature focusing on

$PP$  and its effect on the mortality prediction (e.g., References [32, 33]), our finding is partially consistent with their conclusion.

By further investigating the data, another behaviour is noticed with false negative patients, namely a frequent drop in respiration rate ( $RR$ ) as shown in Figure 5.6a. Due to this behaviour, a new feature is proposed to represent this drop and the count of its occurrence. This feature is defined as the number of times the  $RR$  drops below a specific threshold within each window and is further referred as *low- $RR$  count*. For this feature, two parameters are selected: the threshold and the window size. Both of them are searched exhaustively by maximising the classification performance by considering the new feature. The best-found combination is a threshold of 5 bpm and a window size of 60 observations.

Another observation in some false negative patients' vital signs is a physiological feature related to a frequent drop of oxygen saturation  $SpO_2$  as shown in Figure 5.6b. Similar to *low- $RR$  count*, this feature is defined as the number of times the  $SpO_2$  drops below a specific threshold within each window. Moreover, the threshold and window size combination affects the influence of the feature on the performance. The best-found combination is a threshold of 77% and a window size of 60. This feature is further referred to *low- $SpO_2$  count*.

Both, *low- $SpO_2$  count* and *low- $RR$  count* created only an added value to the classification performance after the fine-tuning step.

Finally, a physiological feature that is imported directly from the patients' medical record is their positive and negative diagnosis with cardiovascular diseases ( $CVD$ ). By considering this feature exclusively in the input space, no single positive class is recognised. However, by adding this feature to the optimal combination of features, a remarkable enhancement is achieved as will be discussed later.

### 5.3.3 Feature Fine-Tuning

After defining three different categories of features, it is necessary to fine-tune the proposed features in order to obtain the most efficient combination and representation of them. As will be shown in Section 5.4, the error performance can drop after combining features from different categories. One interpretation of this drop is that some features are strictly efficient for a group of patients and confusing for the rest. In order to limit this effect a fine-tuning step is performed.

The feature fine-tuning phase is based on the selection of vital signs instead of the selection of dimensions which is in contrast with existing automatic and conventional feature-selection techniques. Indeed, the rows of the input matrix of our data correspond with the different subjects in the study and contain the different features calculated on multiple windows (e.g., the statistical feature of mean is extracted from  $m$  vital signs within  $n$  time-windows resulting in  $mn$  columns for each subject). Conventional feature-selection techniques [14] select the columns of the matrix that are most representative for the study. However, in this way feature values within a specific time-window can be excluded leading to features that are hard to interpret. For this reason, we propose a backward selection approach where a feature (corresponding to multiple columns

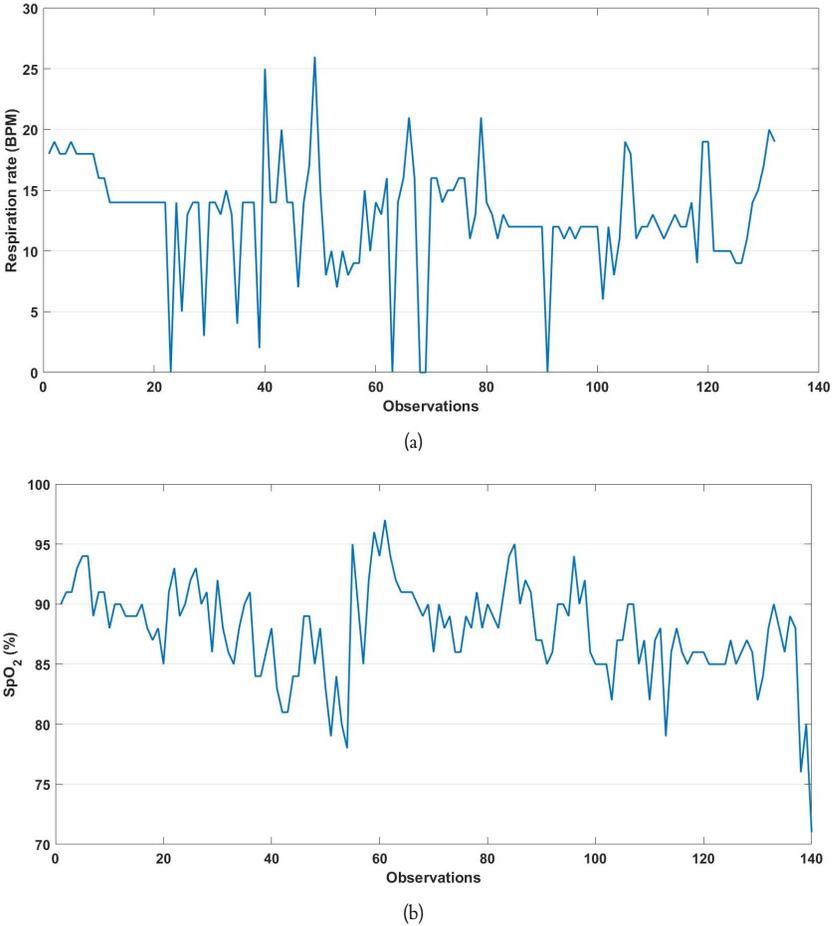


Figure 5.6: (a) Respiration rate of a deceased patient with an obvious drop at specific observations below the normal range (12–20 BPM) (b) Oxygen saturation ( $SpO_2$ ) of another deceased patient that drops frequently (minimum 78% during the stay).

in the input matrix) can be excluded from the set of features. Moreover, prior knowledge is used in order to reduce the randomness in the selection process of the features. For instance, we will exclude the statistical and dynamic features of the  $HR$  guided by the prior knowledge that the heart is a main actuator in the control system of a human body that responds to different excitations (e.g., medication), not only critical events [34]. The effect on the performance score of this selection will be discussed in Section 5.4.

The procedure of feature fine-tuning that we propose in this work starts with exploring whether statistical and dynamic features are providing high performance when extracted from all vital signs or strictly from a subset of these vital signs. Moreover, we assess the effect on the

classification performance of using aggregate features which are calculated on a group of vital signs together rather than on individual vital signs. Furthermore, feature values can be presented as either real or absolute. This procedure is applied exhaustively to the statistical and dynamic features and is assessed by the error performance. The resulting fine-tuning (FT) steps are as follows:

1. FT1: For *HR* extracted features, it is found that excluding both statistical and dynamic features enhances the error performance.
2. FT2: The *correlation coefficients* feature is found more efficient when presented in both real and absolute values.
3. FT3: *Outlier-occurrence count*, is found most efficient when applied to *SBP*, *MAP*, *RR* and *PP* excluding *DBP* and *SpO<sub>2</sub>*. Moreover, the *outlier-occurrence count* is found more efficient when presented in an aggregate form instead of individually except for the vital sign *SBP*.
4. FT4: The *correlation coefficients* feature is providing the best performance when computed only between *HR* and *SBP*. Together with considering the features *low-SpO<sub>2</sub> count* and *low-RR count* the classification performance is improved.
5. FT5: *crossing-the-mean count* is found more efficient when applied only to *SBP* and *RR* and represented in the aggregate form.
6. FT6: The dynamic feature of *outlier indicator* is more efficient when applied only to *SBP* and *DBP*.
7. FT7: Ultimately, considering the physiological feature of *CVD* enhanced the performance.

## 5.4 Results

The obtained results based on the previously mentioned evaluation metrics for each category and for each fine-tuning step are explained below.

Starting with the statistical features, the resulting classification output is 83 TP's, 148 TN's, 87 FN's and 132 FP's. This result is fixed over the different test trials score-wise and patient-wise. In other words, the correctly classified patients are fixed over the different test trials because of using the linear hard margin SVM.

For dynamic features, the resulting classification output considering only the dynamic features is 32 TP's, 247 TN's, 138 FN and 33 FP's. Despite the remarkable reduction in the number of TP's, 18 new TP's are recognised by the dynamic features that are not recognised by the statistical features, in addition to 116 new TN's. This result is again fixed over the different test

Table 5.1: Feature Extraction results.

Feature Combination	Results							
	TP	TN	FN	FP	Sensitivity (%)	Precision (%)	F1-Score	Accuracy (%)
Statistical (Stat)	83	148	87	132	48.88	38.60	43.14	51.33
Dynamic (Dyn)	32	247	138	33	18.82	49.23	27.23	62.00
Stat+Dyn	85	159	85	121	50.00	41.26	45.21	54.22
Physiological (Phy)	45	222	125	58	26.47	43.69	32.97	59.33
Phy+Stat+Dyn	83	118	87	162	48.88	33.88	40.02	44.67

trials score-wise and patient-wise. With both statistical and dynamic features, the classifier performance is improved slightly compared to only statistical features with 2.8% increment in the accuracy. As the resulting classification output after combining both categories is 85 TP's, 159 TN's, 85 FN and 121 FP's. Despite the weak performance at this stage, the correctly classified instances are fixed with each test trial. This means that extracted features at this level are able to discriminate clearly between the correctly classified patients.

For physiological features, namely the *PP*, the resulting classification output with exclusively the extracted statistical and dynamic features of *PP* is 45 TP's, 222 TN's, 125 FN's and 58 FP's. It is important to note that the investigated FN's at the earlier stage are correctly classified by the *PP* extracted features. However, adding the *PP* extracted features to both statistical and dynamic features provided the following results: 83 TP's, 118 TN's, 87 FN's and 162 FP's. The classification output of the different feature-categories combinations are shown in Figure 5.7a. Moreover, feature extraction results are combined and depicted in Table 5.1.

Before showing the results of the fine-tuning phase, we present the results of using the feature selection and ranking technique of automatic relevance determination (ARD) [28] based on backward selection method. The classification output of the ARD selected dimensions is 92 TP's, 218 TN's, 78 FN's and 62 FP's.

For the fine-tuning phase, the results are depicted in Table 5.2 and Figure 5.7b in a cumulative way.

Table 5.2: Feature Fine-tuning results.

Cumulative Fine-Tuning Steps	Results							
	TP	TN	FN	FP	Sensitivity (%)	Precision (%)	F1-Score	Accuracy (%)
ARD	92	218	78	62	54.12	59.74	56.80	68.89
FT1	99	164	71	116	58.23	46.04	51.42	58.44
+FT2	101	179	69	101	59.41	50.00	54.30	59.41
+FT3	106	185	64	95	62.35	52.74	57.14	64.67
+FT4	129	219	41	61	75.88	67.89	71.66	82.67
+FT5	143	243	27	37	84.11	79.44	81.70	85.78
+FT6	147	251	23	29	86.47	83.52	84.97	88.44
+FT7	154	256	16	24	90.59	86.52	88.50	91.56

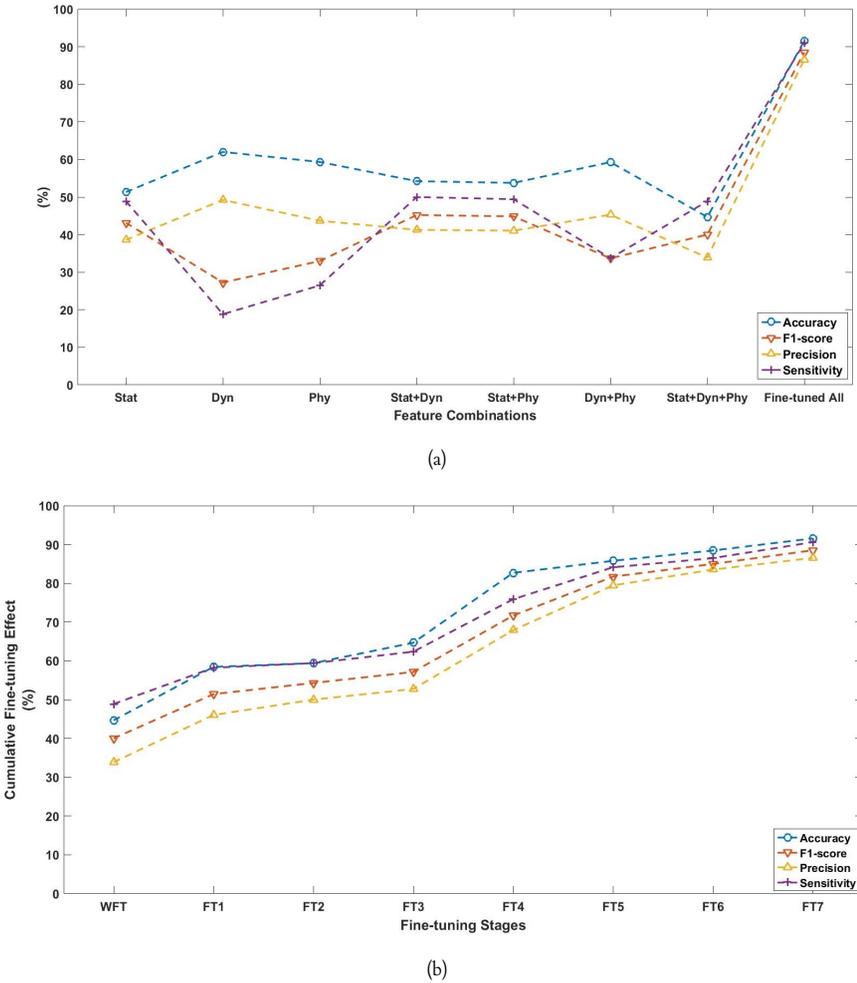


Figure 5.7: (a)  $F_1$ -score, accuracy, sensitivity and precision of the classifier with all possible combinations of the three feature categories in addition to the fine-tuned combination. (b) The cumulative effect of the different fine-tuning stages on the classification accuracy,  $F_1$ -score, precision, and sensitivity. WFT refers to 'without fine-tuning',  $FT_x$  refers to the  $x$ th stage of fine-tuning as illustrated in the text.

### 5.5 Discussion

Many studies are using the area under receiver operating characteristics curve (AUC) as an evaluation metric. In this study, we prefer to use the confusion matrix for evaluation and direct quantification of error metrics of concern (e.g., sensitivity, precision). However, the calculated AUC for our optimised classifier is 0.91 for comparison purposes. This result, when compared

to several recent studies is satisfactory. For instance, a recent study focusing on a special profile of ICU patients reported an AUC of 0.70 using a developed novel mortality prediction SOFA-RV [35]. Another study [12] that evaluates the Super ICU Learner Algorithm (SICULA) and its predictive power applied to MIMIC II database reported an AUC of 0.88 on average under specific conditions and 0.94 on average when applied to an external validation set with calibration. The study of Luo Y. et al. [11] reported an AUC of 0.848. Luo Y. et al. proposed an unsupervised feature learning algorithm that extracts features automatically from the clinical multivariate time-series. Luo Y. et al. applied their algorithm to the MIMIC-II [17] dataset with a prediction horizon extending to 30 days. The study in Reference [8] that developed a convolutional neural network (CNN) as a deep learning approach to predict mortality risk at ICU reported, as the highest performance, an AUC of 0.87, a precision of 0.7443 and a recall of 0.8188. The developed model used the variables of heart and respiration rate, systolic and diastolic blood pressure obtained from the MIMIC-III dataset [18]. Landon et al. [8] referred to the difficulties and limitations of using electronic medical report (EMR) data, similar to our dataset, for the purpose of mortality prediction at ICU. Nemati et al. in their study [36] of sepsis early prediction, which is a lead cause of morbidity and mortality of ICU patients, developed a machine learning model that reported an AUC of 0.83–0.85 for a prediction horizon of 12 down to 4 h prior to clinical recognition. Nemati et al. used an EMR data with high-resolution vital signs time-series obtained from the MIMIC-III dataset [18]. Two medical studies [32, 33] reported an observed relevance between the low pulse pressure and mortality risk. Which is consistent with our finding of considering the pulse pressure as an independent variable from which both statistical and dynamic features can be extracted to inform mortality prediction. Moreover, the medical study in Reference [37] concludes the relevance between the widened (high) pulse pressure and the mortality risk for a special profile of critically ill patients. This conclusion as well is consistent with our finding, as we referred to the statistical and dynamic features of the pulse pressure which will indicate either abnormally high or low levels of pulse pressure. It is important to note that each study has different conditions, different objectives, different datasets, parameters and variables and predictive models.

At the feature extraction phase, the variation of results with different categories shows that a set of features can be efficient with a group of patients (i.e. correctly classified) but the same set of features can be inefficient or confusing to another group of patients (i.e. misclassified). For instance, statistical features classified correctly 83 TPs and 148 TNs, on the other hand, dynamic features classified correctly 32 TPs and 247 TNs. Considering the patient identity, it is found that dynamic features correctly classified 18 TPs and 116 TNs that the statistical features misclassified. The same observation is noticed with *PP* extracted features (45 TPs and 222 TNs) and those features extracted from both *SBP* and *DBP* together (72 TPs and 199 TNs). The difference in this situation is that *PP* is a result of a linear combination between *SBP* and *DBP*, however, *PP* extracted features correctly classified 14 TPs and 58 TNs that are misclassified by *SBP/DBP* extracted features. Hence, the influence of features should be evaluated on a subject-basis in addition to error metrics. Another observation is that the physiological features of *low-RR* and

*low-SPO<sub>2</sub> count* do not correctly classify any true positive patient despite their physiological basis when presented as the only input features. However, their contribution is significant when combined with the consistent set of features as shown at the feature fine-tuning phase. Therefore, excluding a feature has to be done after that it has been tested in combination with different groups of features especially if the extracted feature has a physiological basis.

At the fine-tuning phase, we have to note that this process is based on feature-vector-level not dimension-level as a single extracted feature may include multiple dimensions (e.g., the mean within each window for a specific vital sign). Which is in contrast with conventional feature selection techniques that rely on selecting the most relevant dimensions regardless of the interpretation of the selected dimensions. The initial modification is excluding both statistical and dynamic features extracted from *HR* in order to enhance the performance. This modification is required as many of the cardiovascular patients in this study have common cardiac diseased behaviour, which confuses the classifier. Moreover, the heart acts as one of the main actuators in the human control system responding to different types of excitations. Hence, *HR* disturbances might not be sufficient to predict mortality, leading to a high false alarm rate. Ultimately, considering the cardiovascular patients specifically, HR statistical characteristics, as well as their HR dynamic features are both technically confusing to mortality prediction. Moreover, the enhancement of detecting more TP's by presenting some dynamic features in an aggregate form can be interpreted by the fact that the concurrence of vital signs deterioration is partially a sufficient mortality indicator but not a necessary one. In other words, total deterioration implies mortality but not vice versa. Introducing the *correlation coefficients* feature with absolute values in addition to real values provides an improvement. Both absolute and real values help the linear classifier to distinguish between the instances based on the correlation strength and correlation sign respectively. Restricting the *crossing-the-mean count* to *SBP* and *RR* caused an improvement. Thus, observation-to-observation variability of both vital signs even for a relatively low sampling rate (i.e. 0.5–1 sample/hour) is more informative than the other vital signs for resting patients such as ICU patients.

As the main objective of this study is to engineer feature that can provide clinical insight about mortality prediction, it is important to refer to the decision tree classifiers. As one of the decision trees advantages is model interpretability in terms of the input attributes. However, some shortages are present in decision trees in contrast with SVMs that supported the choice of the latter. These shortages are mainly the greedy nature of the algorithm, local optimisation, prone to overfitting and expensive computational cost compared to linear hard margin SVM in which there are no hyperparameters to optimise. Moreover, we based our study on the optimal margin hypothesis which is not provided by decision trees in contrast with SVMs. For comparison reasons, a decision tree analysis is applied to the final set of features. A CART algorithm decision tree (MATLAB 2017) is used with the following settings: the splitting criterion of *gdi*, minimum parent size of 368, minimum leaf size of 184, maximum splits of 450 and pruning based on classification error criterion. The classification output of the optimised decision tree is as fol-

lows: sensitivity of 41.2% precision of 42.42%, F1-score 41.80% and accuracy of 52.22%. It is obvious that the results are poor compared to the results of linear hard margin SVM. The poor performance is quite expected because of the conceptual differences between the two classification techniques (i.e. Decision trees and SVMs). It is possible that if the whole feature engineering process is designed based on the decision tree classifier properties, the results can be better. Model development, feature extraction and fine-tuning are implemented on observation-basis instead of time-basis (hourly/daily). We hypothesise that observation-basis are more realistic as the events (observations) within a specific time period are more informative than time period regardless of the number of observations. Ideally, the number of observations is fixed along a specific period for all patients and uniformly distributed as well which is not the case with our dataset. However, for a proof of concept, we evaluate the classification performance based on extracting the same features on time-basis. Time-basis is implemented by considering the last 7 days before discharge, considering the first 5 days for feature extraction to predict mortality 2 days ahead. These periods are defined based on the observation-basis analysis. By extracting statistical, dynamic and physiological features without fine-tuning, the output classification performance is 88 TPs, 163 TNs, 82 FNs and 117 FPs. In comparison with the classification performance on observation-basis (83 TPs, 118 TNs, 87 FNs and 162 FPs) the error performance is higher. However, by following the same feature fine-tuning steps the final classification output (82 TPs, 160 TNs, 88 FNs and 120 FPs) is dropped compared to that obtained by an observation-based approach (154 TPs, 256 TNs, 16 FNs and 24 FPs). This drop can be interpreted by the fact that the fine-tuning phase is a manual crafting of the feature combination which is sensitive to the features setup (i.e. observation-basis or time-basis).

## 5.6 Conclusions

In this study, we proposed a proof of concept for a feature engineering approach to explore features that can provide clinical insight in order to enhance the mortality prediction of ICU patients using the machine learning algorithm of linear hard margin SVM. The optimal combination of features that provided the best classification performance comprises the following features:

1. Statistical features of the raw physiological variables, their first derivative of *SBP*, *DBP*, *MAP*, *RR*, *SpO<sub>2</sub>* and *PP*. Moreover, the statistical features extracted from the EWS of *SBP*, *RR* and *SpO<sub>2</sub>*. A window size of 15 observations.
2. Real and absolute values of *correlation coefficients* between *HR* and *SBP* in a window size of 30 observations.
3. *Outlier-occurrence count* of *SBP*, *MAP*, *RR* and *PP*. represented in an aggregate form except for the *SBP* represented individually as well. A window size of 60 observations.
4. *crossing-the-mean count* of *SBP* and *RR*, it is presented in the aggregate form. A window size of 60 observations.

5. *Outlier indicator of SBP and DBP*. A window size of 60 observations.
6. *Low-SpO<sub>2</sub> count less than 77% and low-RR count less than 5 BPM*. A window size of 60 observations.

The proposed approach allows moving from black-box to grey-box modelling, starting from a powerful black-box technique such as SVMs. Moreover, in this case study, low frequently measured vital signs (hourly to bi-hourly) enabled us to extract efficient features for the purpose of relatively long term analysis.

From a feature engineering perspective, some features or variables are individually unable to distinguish between the two classes (i.e. mortality and survival). However, by combining such features in suitable feature combinations, their use becomes beneficial. Furthermore, combining different efficient features might cause a drop in performance. Therefore, a feature fine-tuning phase is essential in order to synthesise efficient feature-combination.

From the medical perspective, we can conclude that the heart rate as an individual variable can be confusing to predict the mortality. This conclusion is supported by improving the error performance by excluding the heart rate features. Moreover, we can recommend paying more attention to the pulse pressure explicitly, either high or low level, since both levels are found associated with the mortality of a group of patients. Watching the pulse pressure requires implicitly to consider the diastolic blood pressure which is excluded from the EWS standards. Finally, we conclude that different profiles of patients require a different set of features to handle the mortality prediction efficiently.

For future work, we propose to test the developed model with the extracted features along the stay of the ICU patients. In other words, we can scan the complete period of stay with the moving window of 60 observations for feature extraction to predict the mortality-risk 24 observations ahead. Despite the fact that along the stay the patients will be labelled as survival, the medical doctors may label any upcoming events with possible mortality-risk.

# Bibliography

- [1] Braber A., van Zanten A.R. Unravelling post-ICU mortality: Predictors and causes of death. *Eur. J. Anaesthesiol.* **2010**, *27*, 486–490.
- [2] Goldhill D.R., McNarry A.F., Mandersloot G., McGinley A. A physiologicallybased early warning score for ward patients: The association between score and outcome. *Anaesthesia* **2005**, *60*, 547–553.
- [3] Lokhandwala S., McCague N., Chahin A., Escobar B., Feng M., Ghassemi M.M., Stone D.J., Celi L.A. One-year mortality after recovery from critical illness: A retrospective cohort study. *PLoS ONE* **2018**, *13*, e0197226.
- [4] Celi L.A., Galvin S., Davidzon G., Lee J., Scott D., Mark R. A database-driven decision support system: Customized mortality prediction. *J. Pers. Med.* **2012**, *2*, 138–148.
- [5] Celi L.A., Tang R.J., Villaroel M., Davidzon G.A., Lester W.T., Chueh H.C. A clinical database-driven approach to decision support: Predicting mortality among patients with acute kidney injury. *J. Healthc. Eng.* **2011**, *2*, 97–110.
- [6] Johnson A.E.W., Mark R.G. Real-time mortality prediction in the Intensive Care Unit. *AMIA Ann. Symp. Proc.* **2017**, *2017*, 994–1003.
- [7] Alves T., Laender A., Veloso A., Ziviani N. Dynamic Prediction of ICU Mortality Risk Using Domain Adaptation. *IEEE Int. Conf. Big Data* **2018**, 1328–1336, doi:10.1109/BigData.2018.8621927.
- [8] Landon B., Aditya P., Izzatbir S., Clayton B. Real Time Mortality Risk Prediction: A Convolutional Neural Network Approach. *Int. Conf. Health Inf.* **2018**, 463–470, doi:10.5220/0006596204630470.
- [9] Zhu Y., Fan X., Wu J., Liu X., Shi J., Wang C. Predicting ICU Mortality by Supervised Bidirectional LSTM Networks. In Proceedings of the IJCAI 2018 Joint Workshop on Artificial Intelligence in Health (AIH 2018), Stockholm, Sweden, 13–19 July 2018, 49–60.

- [10] Johnson A.E., Pollard T.J., Mark R.G. Reproducibility in critical care: A mortality prediction case study. *Mach. Learn. Healthc. Conf.* **2017**, 2017, 361–376.
- [11] Luo Y., Xin Y., Joshi R., Celi L., Szolovits P. Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
- [12] Pirracchio R., Petersen M.L., Carone M., Rigon M.R., Chevret S., van der Laan M.J. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *Lancet Respir. Med.* **2015**, 3, 42–52.
- [13] Mayaud L., Lai P.S., Clifford G.D., Tarassenko L., Celi L.A., Annane D. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Crit. Care Med.* **2013**, 4, 954–962.
- [14] Verplancke T., Van Looy S., Benoit D., Vansteelandt S., Depuydt P., De Turck F., Decruyenaere J. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC Med. Inform. Dec. Mak.* **2008**, 8, 56–63.
- [15] Kim S., Kim W., Park R.W. A Comparison of Intensive Care Unit Mortality Prediction Models through the Use of Data Mining Techniques. *Healthc. Inform. Res.* **2011**, 17, 232–243.
- [16] Vieira S.M., Mendonça L.F., Farinha G.J., Sousa J.M. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. *Appl. Soft Comput.*, **2013**, 13, 3494–3504.
- [17] Saeed M., Villarroel M., Reisner A.T., Clifford G., Lehman L.W., Moody G., Heldt T., Kyaw T.H., Moody B., Mark R.G. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit. Care Med.* **2011**, 39, 952–960.
- [18] Johnson A.E.W., Pollard T.J., Shen L., Li-wei H.L., Feng M., Ghassemi M., Moody B., Szolovits P., Celi L.A., Mark R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, 3, 160035.
- [19] Aerts J.M., Haddad W.M., An G., Vodovotz Y. From data patterns to mechanistic models in acute critical illness. *J. Crit. Care* **2014**, 29, 604–610.
- [20] Young P.C. *Recursive Estimation and Time-Series Analysis: An Introduction*, Springer Science and Business Media: Berlin, Germany, 2012.
- [21] Vodovotz Y., Csete M., Bartels J., Chang S., An G. Translational systems biology of inflammation. *PLoS Comput. Biol.* **2008**, 4, e1000014.

- [22] Kumar R., Clermont G., Vodovotz Y., Chow C.C. The dynamics of acute inflammation. *J. Theor. Biol.* **2004**, *230*, 145–155.
- [23] Reynolds A., Rubin J., Clermont G., Day J., Vodovotz Y., Ermentrout G.B. A reduced mathematical model of the acute inflammatory response: I. Derivation of model and analysis of anti-inflammation. *J. Theor. Biol.* **2006**, *242*, 220–236.
- [24] Day J., Rubin J., Vodovotz Y., Chow C.C., Reynolds A., Clermont G. A reduced mathematical model of the acute inflammatory response II. Capturing scenarios of repeated endotoxin administration. *J. Theor. Biol.* **2006**, *242*, 237–256.
- [25] Boser B.E., Guyon I.M., Vapnik V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992, 144–152.
- [26] Cortes C., Vapnik V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297.
- [27] Suykens J., Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300
- [28] Suykens J.A.K., Van Gestel T., De Brabanter J., Vandewalle J.. *Least Squares Support Vector Machines*, World Scientific Publishing Co.: Singapore, 2002.
- [29] Abu-Mostafa Y.S., Malik M.-I., Hsuan-Tien L. *Learning from Data*, AMLBook: New York, NY, USA, **2012**.
- [30] Homan T.D., Cichowski E. *Physiology, Pulse Pressure*, StatPearls [Internet]. StatPearls Publishing **2018**.
- [31] Stergiopoulos N., Segers P., Westerhof N. Use of pulse pressure method for estimating total arterial compliance in vivo. *Am. J. Physiol.-Heart Circ. Physiol.* **1999**, *276*, 424–428.
- [32] Yildiran T., Koc M., Bozkurt A., Sahin D.Y., Unal I., Acarturk E. Low pulse pressure as a predictor of death in patients with mild to advanced heart failure. *Texas Heart Inst. J.* **2010**, *37*, 284–290.
- [33] Voors A.A., Petrie C.J., Petrie M.C., Charlesworth A., Hillege H.L., Zijlstra F., McMurray J.J., van Veldhuisen D.J. Low pulse pressure is independently related to elevated natriuretic peptides and increased mortality in advanced chronic heart failure. *Eur. Heart J.* **2005**, *26*, 1759–1764.
- [34] Grodins Fred S. *Control Theory and Biological Systems*, Columbia University Press: New York, NY, USA, 1963.

- [35] Akin S., Caliskan K., Soliman O.I., Muslem R., Guven G., Van Thiel R.J., Struijs A., Gommers D., Zijlstra F., Bakker J., et al. A novel mortality risk score predicting intensive care mortality in cardiogenic shock patients treated with veno-arterial extracorporeal membrane oxygenation. *Eur. Heart J.* **2018**, *39*, 5690.
- [36] Nemati S., Holder A., Razmi F., Stanley M.D., Clifford G.D., Buchman T.G. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit. Care Med.* **2018**, *46*, 547–553.
- [37] Al-Khalisy H., Nikiforov I., Jhajj M., Kodali N., Cheriyaath, P. A widened pulse pressure: A potential valuable prognostic indicator of mortality in patients with sepsis. *J. Community Hosp. Intern. Med. Perspect.* **2015**, *5*, 29426.

## Chapter 6

# Mortality Prediction for COVID-Like Patients at the ICU

### Abstract

The recent hit of COVID-19 pandemic that shocked the healthcare systems around the globe stresses the necessity to provide intelligent monitoring solutions. These intelligent monitoring solutions are needed to optimise the available resources at hospitals, more specifically at intensive care units (ICUs). During the first wave of the pandemic, ICUs in hospitals experienced under-capacity due to high admission rates. In this study, we introduce mortality prediction models for COVID-like patients admitted to ICU. COVID-like patients are pulmonary patients who are diagnosed with a respiratory infection and/or pneumonia. The developed models in this study are based on a linear hard margin approach to support vector machines. Moreover, the input features of these models are a subset of engineered features for ICU patients who are admitted for different reasons (e.g., cardiovascular, neurological diseases). The developed models are built considering a one-day ahead prediction horizon to predict mortality or survival at discharge. Two major groups of models are developed, these two groups are based on 5 and 2-day time-windows for feature extraction, respectively. The best and balanced 5-day model has an error performance of 89.29% sensitivity, 89.29% precision, and 0.8929  $F_1$ -score. On the other hand, the best performing 2-day model has an error performance of 88.73% sensitivity, 90.84% precision, and 0.8959  $F_1$ -score. Based on the obtained results, the engineered features for the developed models in this study can support mortality prediction at ICU for COVID-like patients.

### 6.1 Introduction

This study is introduced in light of the ongoing COVID-19 pandemic that hits the whole world since December 2019. A significant challenge that the hospitals all over the world are facing is the high rate of admission of coronavirus patients to intensive care units (ICU's) [1]. This high

rate may result in under-capacity of the ICU beds. Hence, continuous mortality risk assessment of the ICU patients is valuable for supporting medical staff in making decisions regarding ICU patients' treatments and follow up.

In this study, we aim at developing mortality prediction models of intensive care unit patients who are admitted because of pulmonary diseases together with infection and/or pneumonia. The aforementioned profile of patients is hypothesised to be similar to COVID-19 patients. This investigation is implemented prior to being applied to data obtained from COVID-19 patients admitted to ICU.

Mortality prediction in this study is implemented based on a feature engineering approach for ICU mortality prediction that was introduced in a previous study [2]. This approach resulted in a set of features that are extracted from vital signs of heart and respiration rates, systolic, diastolic, mean arterial blood pressure, and oxygen saturation. These extracted features provided an acceptable performance (i.e. accuracy of 91.56%, sensitivity of 90.59%, precision of 86.52% and  $F_1$ -score of 88.50%) of mortality prediction for 1-2 days ahead [2]. One of the main conclusions of that study was that each profile of patients requires a specific set of features for an efficient prediction performance.

Due to the sudden hit of the pandemic, it was necessary to find a similar profile of patients to that of coronavirus patients. The already existing data set that used in our previous study [2] and its extended version have different profiles of patients (e.g., cardiovascular, neurology, pulmonary, infection, sepsis, and abdominal diseases). We selected pulmonary and infection patients, either in the presence or absence of pneumonia as COVID-like patients because of the presence of similar symptoms.

The candidate profiles are suggested based on understanding the pathophysiology of COVID-19 as it is a very recent disease. Therefore, the amount of available and conclusive research about the virus is limited and little is known about the pathophysiology of COVID-19. The main conclusions of the relevant studies are summarised as follows. Firstly, a brief review paper reported fever, cough and dyspnoea as main symptoms and acute respiratory distress syndrome (ARDS), acute kidney injury (AKI) and myocardial injury as leading causes of mortality in COVID-19 patients [3, 4]. Another study specifically investigated patients with COVID-19 that were admitted to the ICU. The 36 investigated patients had a mean age of 66 years, and 61% of them suffered from ARDS. Other common complications among these patients were arrhythmia (44%), shock (31%) and acute cardiac injury (22%) [5]. Furthermore, in their study [6], Klok et al. observed a high incidence of thrombotic complications and acute pulmonary embolism, which influence blood oxygen levels, respiration rate and blood pressure. On top of that, mechanical ventilation, both invasive and non-invasive, was a standard treatment administered to patients with COVID-19 as the infection of the lungs and the subsequent immune reaction impede airflow to the alveoli [5, 7]. Finally, accurate overall mortality estimations for COVID-19 were not yet available at the time because of the large number of non-reported cases [8]. Based on these studies, the common pathophysiology between the different observed cases is acute respiratory problems resulting

from a respiratory infection.

The main objective of this study is to develop mortality prediction models for COVID-like ICU patients. These models are based on the engineered features from our previous study [2]. This article is arranged as follows: After the introduction, the second section of data in which we describe the data. In the third section, we introduce the linear hard margin support vector machines and feature engineering. The fourth section includes the results. The fifth section consists of the discussion, and the final section gives the conclusion.

## 6.2 Data

The used data for this research was gathered from monitoring 801 ICU patients from the Belgian hospital ZOL (Ziekenhuis Oost-Limburg) as part of the PRINCESS (Prolonged Intensive Care Stay Characterisation) study. This retrospective study aims to create a clinical profile of patients hospitalised in the intensive care unit. The patients were distributed over 3 hospital units: ICU 1, ICU 2, and the intensive heart care unit and they were all admitted between 2015 and 2019. For privacy reasons, all data had been anonymised before the investigation and the data collection was approved by the medical ethical committee of ZOL.

The analysed data comprises the vital parameters of the patients and general/patient-specific information retrieved from the electronic medical record (EMR). The vital parameters were recorded continuously by Philips Intellivue monitors (Philips Electronics Nederland B.V., Amsterdam, The Netherlands), and were annotated on average hourly to bi-hourly by the nursing staff. The measured vital parameters are heart rate (HR) in beats per minute (bpm), respiration rate (RR) in breaths per minute (bpm), oxygen saturation ( $SpO_2$ ) in percentages, systolic and diastolic arterial blood pressure, (SBP and DBP) in mmHg, oxygen administered to the patient ( $O_2$ ) in litres per minute (L/min), the temperature of body or bladder (T) in degrees Celsius ( $^{\circ}C$ ) and central venous pressure (CVP) in  $cmH_2O$ . Arterial blood pressure can be measured in two different ways: invasively and non-invasively. Upon check-up, these measurements are read out by the medical staff and entered manually into the patients EMR. The time and date of the measurement are saved as well. The general/patient-specific information retrieved from the EMR includes the gender and age of the patient, smoking history, the reason for admission to the ICU as well as comorbidities, like diabetes, hypoxia, hypertension, cardiovascular diseases, etc. Finally, every patient has a label indicating his/her outcome at the end of the ICU stay: 0 for survival and 1 for mortality. The age of the patients was 65 ( $\pm 16$ ) years old. The distribution of patients based on their reasons of hospitalisation in this dataset is displayed in Figure 6.1.

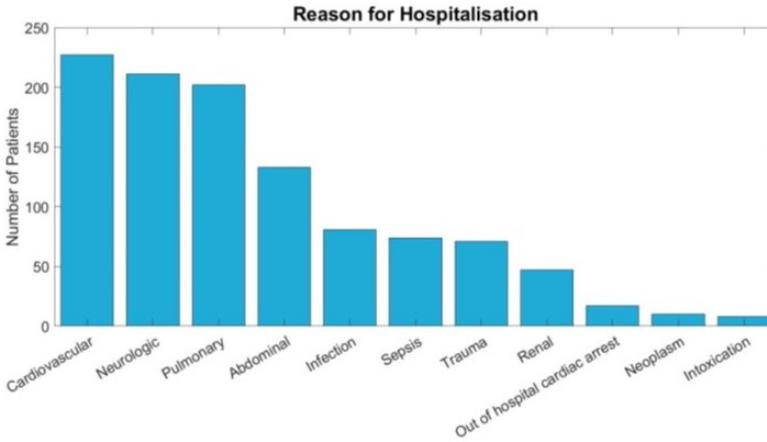


Figure 6.1: Patients distribution based on the reason of hospitalisation

## 6.3 Methods

As mentioned earlier, the followed methodology to develop the mortality prediction models is similar to the one we developed in our previous study [2]. Subsequently, we shortly explain the developed approach, starting with linear hard margin SVM and then the engineered features.

### 6.3.1 Hard Margin SVM

SVMs are originally presented as binary classifiers, that assign each data instance  $\mathbf{x} \in \mathbb{R}^d$  to one of two classes described by a class label  $y \in \{-1, 1\}$  based on the decision boundary that maximises the margin  $2/\|w\|_2$  between the two classes as shown in Figure 6.2. The margin is determined by the distance between the decision boundary and the closest data point from each class [11, 12, 9, 10].

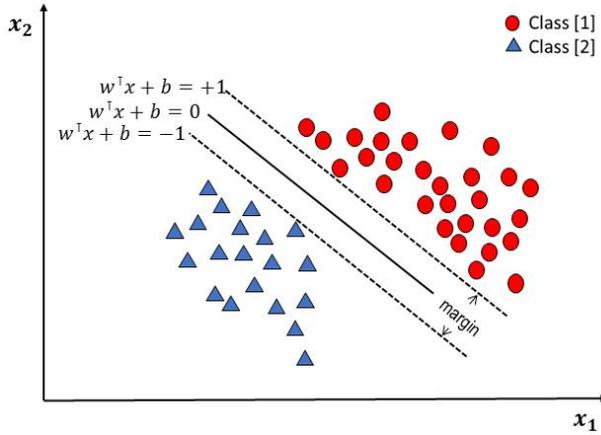


Figure 6.2: Schematic representation of a two-dimensional dataset consisting of two linearly separable classes. The dotted lines indicate the boundaries where the margin is maximised without tolerating any misclassifications (adapted from Reference [12]).

Generally, a feature map  $\phi : \mathbb{R}^d \mapsto \mathbb{R}^p$ , where  $d$  is the number of input space dimensions and  $p$  is the number of feature space dimensions, is used to transform the geometric boundary between the two classes to a linear boundary  $L : \mathbf{w}^\top \phi(\mathbf{x}) + b = 0$  in feature space, for some weight vector  $\mathbf{w} \in \mathbb{R}^{p \times 1}$  and  $b \in \mathbb{R}$ . The class of each instance can then be found by  $y = \text{sgn}(w^\top \phi(x) + b)$ , where  $\text{sgn}$  refers to the sign function.

The estimation of the boundary  $L$  is performed based on a set of training examples  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) with corresponding class labels  $y_i \in \{-1, 1\}$ , where  $N$  is the number of data points. An optimal boundary is found by maximising the margin that is defined as the smallest distances between  $L$  and any of the training instances. In particular, one is interested in constants  $\mathbf{w}$  and  $b$  that minimise a *loss-function* [12]:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^\top \mathbf{w},$$

and are subject to:

$$y_i(\mathbf{w}^\top \phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, N.$$

By applying the lagrangian to the problem we get

$$\mathcal{L}(\mathbf{w}, b; \alpha) = \frac{1}{2} \|\mathbf{w}\|_2^2 - \left( \sum_{i=1}^N \alpha_i (y_i [\mathbf{w}^\top \phi(x_i) + b] - 1) \right),$$

where  $\alpha_i \geq 0$  are the Lagrangian multipliers for  $i^{\text{th}}$  data point. The resulting classifier in both primal space and dual space are

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^\top \phi(x) + b),$$

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i \phi(x_i)^\top \phi(\mathbf{x}) + b\right).$$

The dot product  $\phi(x_i)^\top \phi(\mathbf{x})$  is computationally expensive, hence, it is replaced with the *kernel* function  $k(x_i, \mathbf{x})$ , this replacement is known as the *kernel* trick. With the *kernel* trick, there is no need to execute the step of feature map as it is implicitly done by the *kernel* function. Hence, the dual space classifier with the *kernel* trick is

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i k(x_i, \mathbf{x}) + b\right).$$

For practical reasons, we suggest to obtain the linear hard margin SVM from the standard SVM formula that tolerate misclassification errors [13]

$$\min_{\mathbf{w}, b; \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^N \xi_i,$$

subject to:

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, 2, \dots, N.$$

where the constant  $C$  denotes the *penalty term* that is used to penalise misclassification through the slack variables  $\xi_i$  in the optimisation process. The linear hard margin SVM can be obtained via penalising the error extremely by giving  $C$  a very high value (e.g.,  $10^{10}$ ). With this trick, we can get a solution with misclassified instances to be investigated through the feature engineering phase.

## 6.4 Feature Engineering

In our previous study [2], we introduced a feature engineering procedure for mortality prediction. In the upcoming subsections, we illustrate both feature extraction and fine-tuning phases.

### 6.4.1 Feature Extraction

Prior to elaborating on feature extraction, it is important to define time windows of feature extraction as well as the prediction horizons during the ICU stay of each patient. The only true label regarding mortality risk is at the moment of discharge, being mortality or recovery (survival). Hence, feature extraction time windows and prediction horizons are defined from the end of the stay, as shown in Figure 6.3, and the discharge reason (mortality/survival) is added

as a label. Moreover, the prediction horizon before that moment of discharge comprised 12 observations (approximately one day). The feature extraction period is defined as 60 observations before the prediction horizon.

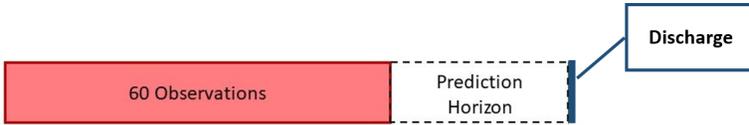


Figure 6.3: A scheme illustrates the last set of observations before discharge: time period for feature extraction (red) to predict mortality/survival at discharge (blue) given a specific prediction horizon (white).

The size of time-windows is determined based on a medical input and several analytical test trials to find the optimal window-size. Moreover, this average period of 3.5 days (60 observations) agrees with the experience of clinical experts in the field. This agreement is based on the fact that there is no standard at the moment that refers to a minimum or maximum of observations to use, in order to make the best decision.

The extracted features comprises three categories, namely , statistical, dynamic, and physiological features.

#### 6.4.1.1 Statistical Features

The first category of features to be extracted is the set of statistical features which represent the basic characteristics of each time-series within segmented, non-overlapping time windows: *minimum, maximum, mean, median, standard deviation, variance, and energy.*

Statistical features are extracted within windows whose sizes are defined by the number of observations and not by a specific time period due to the nonuniform sampling rate (hourly to bi-hourly) as mentioned before. Extraction is based on the raw measurements of the vital signs and their first derivatives as well as the calculated standard early warning scores (EWS) of these measurements based on ZOL hospital standards as shown in Table 6.1. A weak point about statistical features is the static nature of these features as they do not reveal the dynamic behaviour of the time-series. Therefore, an additional category of features is required to be explored, namely dynamic features.

Table 6.1: Early Warning Scores system used by ZOL Hospital based on vital signs' values.

SCORE	3	2	1	0	1	2	3
Temperature ( $^{\circ}C$ )		<35.1	35.1-36.5	36.6-37.5	>37.5		
Heart Rate (BPM)		<40	40-50	51-100	101-110	111-130	>130
Respiration Rate (BPM)		<9		9-14	15-20	21-30	>30
Oxygen Saturation (%)	<91	91-93	94-95	>95			
Systolic Blood Pressure (mmHg)	<70	70-80	81-100	101-180	180-200	>200	

### 6.4.1.2 Dynamic Features

The extracted dynamic features are *Pearson correlation coefficients*, *crossing-the-mean count*, *outlier-occurrence count*, and *outlier indicator*. *Correlation coefficient* is computed between each pair of vital signs within each window. For this feature, it is necessary to be applied to the *z-score* of the vital signs. *Crossing-the-mean count* of a vital sign is determined by counting the number of times that the recorded vital sign crosses its mean value within each window. This feature indicates the abrupt changes in the vital sign from one observation to another. *Outlier-occurrence count* is computed by counting the number of outliers detected within each window. An outlier is detected by the statistical definition: any point outside the range  $\mu \pm 3\sigma$  for a normally distributed variable is an outlier. For this feature, it is not expected to work with the vital sign of oxygen saturation ( $SpO_2$ ) as it is negatively skewed, however, it will be tested as a feature to prove the concept. Finally, the *outlier indicator* is determined by the difference between the mean and the median of the records within each window.

### 6.4.1.3 Physiological Features

Finally, the physiological features comprises the statistical and dynamic features of Pulse pressure (PP) and its first derivative. Pulse pressure is defined by the difference between systolic and diastolic blood pressure. Moreover, the considered physiological features also comprise the features of  $SpO_2$  and RR values count below specific thresholds (77% and 5 BPM respectively). These features are further referred to *low- $SpO_2$  count* and *low-RR count*. Furthermore, another physiological feature that is imported directly from the patients' medical record is their positive and negative diagnosis with cardiovascular diseases (CVD). The methods of extracting the aforementioned features are illustrated in the article [2] in more detail.

### 6.4.2 Feature Fine-Tuning

After defining three different categories of features, it is necessary to fine-tune the proposed features in order to obtain the most efficient combination and representation of them. As was shown in Section 5.4, the error performance can drop after combining features from different categories. One interpretation of this drop is that some features are strictly efficient for a group

of patients and confusing for the rest. In order to limit this effect a fine-tuning step is performed.

The feature fine-tuning phase is based on the selection of vital signs instead of the selection of dimensions which is in contrast with existing automatic and conventional feature-selection techniques. Indeed, the rows of the input matrix of our data correspond with the different subjects in the study and contain the different features calculated on multiple windows (e.g., the statistical feature of mean is extracted from  $m$  vital signs within  $n$  time-windows resulting in  $m \times n$  columns for each subject). Conventional feature-selection techniques select the columns of the matrix that are most representative for the study [14]. However, in this way feature values within a specific time-window can be excluded leading to features that are hard to interpret. For this reason, we propose a backward selection approach where a feature (corresponding to multiple columns in the input matrix) can be excluded from the set of features. Moreover, prior knowledge is used in order to reduce the randomness in the selection process of the features. For instance, we will exclude the statistical and dynamic features of the  $HR$  guided by the prior knowledge that the heart is a main actuator in the control system of a human body that responds to different excitations (e.g., medication), not only critical events [15]. The effect on the performance score of this selection will be discussed in Section 5.4.

The followed steps of the feature fine-tuning (FT) procedure are imported from our previous work [2] and shortly described below:

1. FT1: For  $HR$  extracted features, it is found that excluding both statistical and dynamic features enhances the error performance.
2. FT2: The *correlation coefficients* feature is found more efficient when presented in both real and absolute values.
3. FT3: *Outlier-occurrence count*, is found most efficient when applied to  $SBP$ ,  $MAP$ ,  $RR$  and  $PP$  excluding  $DBP$  and  $SpO_2$ . Moreover, the *outlier-occurrence count* is found more efficient when presented in an aggregate form instead of individually except for the vital sign  $SBP$ .
4. FT4: The *correlation coefficients* feature is providing the best performance when computed only between  $HR$  and  $SBP$ . Together with considering the features *low- $SpO_2$  count* and *low- $RR$  count* the classification performance is improved.
5. FT5: *crossing-the-mean count* is found more efficient when applied only to  $SBP$  and  $RR$  and represented in the aggregate form.
6. FT6: The dynamic feature of *outlier indicator* is more efficient when applied only to  $SBP$  and  $DBP$ .
7. FT7: Ultimately, considering the physiological feature of  $CVD$  enhanced the performance.

## 6.5 Results

This section illustrates the results of mortality prediction for the candidate profiles simulating COVID-19 patients (COVID-19-like). As mentioned earlier, the candidate profiles of ICU patients that simulate COVID-19 patients are those suffering from pulmonary disease, infection, and Pneumonia. There are three subpopulations and the corresponding diseases, number of patients and discharge labels are shown in Table 6.2

Table 6.2: Three subpopulations with different combinations of pulmonary disease, infection, and pneumonia (M:mortality; S: survival).

Subpopulation	Pulmonary	Infection	Pneumonia	N <sup>o</sup> of Patients	M/S
<b>I</b>	x	x		58	28/30
<b>II</b>	x	x	x	100	47/53
<b>III</b>	x		x	85	43/42

For each subpopulation, 2 models are developed. One to get the best possible sensitivity while keeping an acceptable precision. The second is targeting the best precision with acceptable sensitivity. The motivation behind developing those two models is inspired by the use of the predictive model. The first model is important when the main target is to have an alarm regarding the severity of the case. The second model can be useful to support the decision of discharge due to losing the hope of the patient's survival. In the situation of the COVID-19 pandemic, discharge decisions are crucial, especially when the maximum capacity of the ICU is reached.

Applying a similar approach as described in our previous study [2], each profile of the ICU patients requires a specific set of features to predict mortality. Therefore, a forward feature selection procedure is applied for each model of each subpopulation to select the most relevant features that provide the best error performance. As depicted in Table 6.3, a set of features is selected for each model/subpopulation combination. From Table 6.3 it can be seen that the extracted features are mainly from three vital parameters, namely SpO<sub>2</sub>, blood pressure, and respiration rate.

Figure 6.4 and Table 6.4 show the results of both sensitivity and precision of models 1 and 2 for subpopulations I, II, and III. These results are based on extracting features from 60 observations (4 to 5 days).

Based on the first results, medical experts advised to apply these models for less number of observations to approach 2 days for feature extraction as the length of stay at the ICU for COVID-19 patients is varying from two days to more. Hence, another model for each subpopulation is developed based on 30 observations for feature extraction. The developed models here are just based on keeping the balance between sensitivity and precision without prioritising one over the other. The resulting sensitivity, precision and  $F_1$  Score are depicted in Figure 6.5 and Table 6.5.

Table 6.3: The set of extracted features for each model/subpopulation combination.

Sub.	Model	SpO <sub>2</sub>	Blood Pressure	Respiration Rate	Others
I	1	Max. EWS SpO <sub>2</sub> Min. Diff. EWS SpO <sub>2</sub> Var. Diff. SpO <sub>2</sub> Med. Diff. EWS SpO <sub>2</sub>	Median PP Std. Diff. DBP Var. MAP	Std. Diff. EWS RR Std. EWS RR Max. EWS RR	
	2	Max. EWS SpO <sub>2</sub> Min. Diff. EWS SpO <sub>2</sub> Var. Diff. SpO <sub>2</sub> Med. Diff. EWS SpO <sub>2</sub>	Median PP Std. Diff. DBP Var. MAP Median Diff. MAP	Std. Diff. EWS RR Std. EWS RR Max. EWS RR	
II	1	Std. SpO <sub>2</sub> Min. SpO <sub>2</sub>	Median PP Energy PP Median EWS SBP Mean Diff. EWS SBP Median Diff. PP	Var. EWS RR Std. Diff. RR	HR-SBP corr.
	2	Std. SpO <sub>2</sub> Min. SpO <sub>2</sub>	Median PP Energy PP Median EWS SBP Mean Diff. EWS SBP Median Diff. PP Max. PP	Var. EWS RR Std. Diff. RR	HR-SBP corr.
III	1	Max. Diff. SpO <sub>2</sub> Var. SpO <sub>2</sub>	Max. PP Mean Diff. EWS SBP Median PP	Median Diff. RR Var. EWS RR	
	2	Max. Diff. SpO <sub>2</sub> Med. Diff. EWS SpO <sub>2</sub>	Max. PP Mean Diff. EWS SBP Outlier ind. SBP	Median Diff. RR	Outliers Aggregate CVD

Table 6.4: Precision, sensitivity and  $F_1$ -score of models 1 and 2 for subpopulations I, II, and III based on 60 observations for feature extraction.

Subpopulation	Model	Sensitivity %	Precision %	$F_1$ Score
I	1	89.29	89.29	0.8929
	2	85.71	96	0.9046
II	1	89.36	87.50	0.8842
	2	87.23	89.13	0.8817
III	1	93.05	85.11	0.8890
	2	88.37	88.37	0.8837

Table 6.5: Precision, sensitivity and  $F_1$ -score of models for subpopulations I, II, and III based on 30 observations for feature extraction.

Subpopulation	Sensitivity %	Precision %	$F_1$ Score
I	85.71	85.71	0.8571
II	89.36	82.35	0.8570
III	88.37	90.84	0.8959

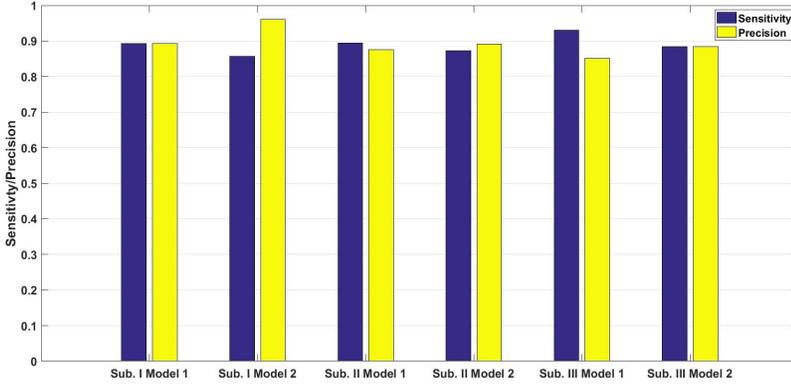


Figure 6.4: Sensitivity and precision of models 1 and 2 for subpopulations I, II, and III based on 60 observations for feature extraction.

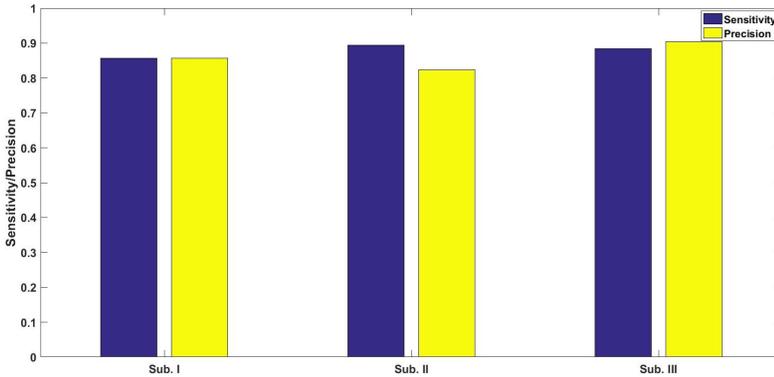


Figure 6.5: Sensitivity and precision of models 1 and 2 for subpopulations I, II, and III based on 30 observations for feature extraction.

## 6.6 Discussion

For COVID-like patients, it is noticed that the number of relevant and effective features is limited (between 7 to 11 features) as depicted in Table 6.3. Moreover, these features are mainly extracted from the vital signs oxygen saturation, blood pressure, and respiration rate. Intuitively, both  $SpO_2$  and RR are directly relevant to the pathology of pulmonary diseases (e.g., COVID-19). However, it is interesting to find the blood pressure, especially pulse pressure, to be relevant as well. This finding is consistent with two clinical findings. The first is about the association between hypertension and thrombosis [16]. The second finding is a recent observation on deceased

COVID-19 patients that they experienced thrombosis mostly caused by pulmonary embolism [17]. Moreover, a recent study [18] observed that patients with hypertension have a two-fold increased risk of mortality due to COVID-19 compared to patients without hypertension.

In their recent study [19], Yadaw et al. developed parsimonious machine learning models for mortality prediction of COVID-19 patients. Although the investigated population in their study was not defined as an ICU population, they observed findings relevant to our study. The main findings in their study are regarding the impact of specific features on the mortality prediction. These features are age, minimum oxygen saturation ( $SpO_2$ ) and type of patient encounter (i.e. inpatient, outpatients and telehealth visits) [19]. Regarding age, we notice that the majority of pulmonary patients (67%) are of age higher than 65 years old, as shown in Figure 6.6. Therefore, most of the pulmonary population in our study is at old age. Moreover, age as a feature is not significantly contributing to our developed models as old age (>65) is present in both mortality and survival examples. The second significant feature in their study, minimum  $SpO_2$  is also present as an important feature in our developed models for subpopulation II, which is considered the closest to the COVID-19 profile. For the third significant feature, namely patient encounter type, our entire studied population is of inpatient encounter type (i.e. hospitalised patients). Which is consistent with their investigated population as over 78% of them in total are of inpatient type, and over 98% of the deceased patients are also of inpatient encounter type.

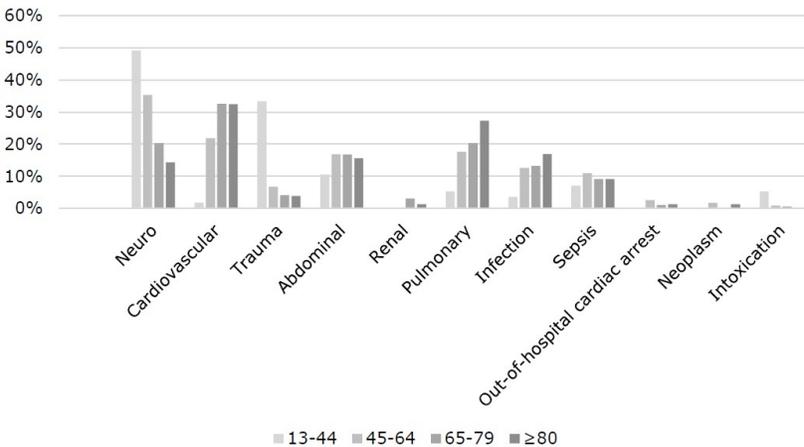


Figure 6.6: Distribution of the patient population and their reason for admission. The population was divided into age categories of 13–44, 45–64, 65–79 and >80 years of age.

In another study, Chansik et al. developed several machine learning model COVID-19 patients without specifying them as ICU patients [20]. Some of the observed significant features are old age, hypertension, and infection route (cluster infection or infection from personal contact). In our study, we observed the importance of blood pressure and its extracted features for our developed models. For subpopulation II, both pulse pressure and systolic blood pressure were

the primary sources of features, as shown in Table 6.3. Moreover, the feature of correlation between heart rate and systolic blood pressure was present. Therefore, our finding regarding the importance of blood pressure is consistent with this study.

In their study [21], Vaid et al., investigated mortality and critical events prediction for COVID-19 patients using machine learning. They found a set of potential predictors for mortality, amongst which is tachypnea. This finding is consistent with our results regarding the impact of respiration rate and its extracted features on the mortality prediction error performance.

Ultimately, it is worth mentioning that our study is based on time-series extracted features, in contrast with the discussed studies [19, 20, 21]. Therefore, we expect that our model would also be useful for real-time prediction, making it potentially useful for monitoring of COVID-19 patients.

## 6.7 Conclusions

Based on the mortality prediction results, we conclude that our developed models are reliable for ICU mortality prediction for COVID-like patients. From a modelling perspective, the developed models are considered simple, since they are based on linear hard margin approach of SVM and a limited number of easily extracted features. Therefore, the parsimony principle is satisfied in addition to partial model interpretability and explainability. From a clinical perspective, the input features of these models are extracted from only three vital signs, namely, oxygen saturation, respiration rate and blood pressure. For future research, we propose to evaluate our models in online mode, which requires a clinical assessment. Furthermore, we suggest applying our models for monitoring COVID-19 patients admitted to ICU.

# Bibliography

- [1] Jason P., et al.. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *The Lancet Respiratory Medicine*, 2020, doi:10.1016/S2213-2600(20)30161-2.
- [2] Amer A. Y., et al.. Feature Engineering for ICU Mortality Prediction Based on Hourly to Bi-Hourly Measurements. *Applied Sciences* 2019, 9(17), 3525.
- [3] Jiang F., Deng L., Zhang L., Cai Y., Cheung C. W., and Xia Z.. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). *Journal of general internal medicine*, 2020, doi:10.1007/s11606-020-05762-w.
- [4] World Economic Prospects Monthly. *Economic Outlook*, 2020, 44, pp. 1–33, doi:10.1111/1468-0319.12473.
- [5] Wang D. et al.. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*. 2020, 323(11), doi:10.1001/jama.2020.1585.
- [6] Klok F. A., et al.. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thrombosis research*, 2020.
- [7] Zhou F. et al.. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet (London, England)*, 2020, 395(10229), pp. 1054–1062, doi:10.1016/S0140-6736(20)30566-3.
- [8] Weiss P. and Murdoch D. R.. Clinical course and mortality risk of severe COVID-19. *Lancet (London, England)*, 2020, 395(10229), pp. 1014–1015, doi:10.1016/S0140-6736(20)30633-4.
- [9] Boser B.E., Guyon I.M., Vapnik V.N.. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
- [10] Cortes C., Vapnik V.. Support-vector networks. *Mach. Learn.* 1995, 20, 273–297.

- [11] Suykens J.A.K., Vandewalle J.. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300
- [12] Suykens J.A.K., Van Gestel T., De Brabanter J., Vandewalle J.. *Least Squares Support Vector Machines*; World Scientific Publishing Co.: Singapore, 2002.
- [13] Abu-Mostafa Y.S., Malik M.-I., Hsuan-Tien L.. *Learning from Data*; AMLBook: New York, NY, USA, 2012.
- [14] Kuhn M., Johnson K.. Feature engineering and selection: A practical approach for predictive models. *Chapters 10 and 11, CRC Press*, **2019**.
- [15] Grodins Fred S. *Control Theory and Biological Systems*, Columbia University Press: New York, NY, USA, 1963.
- [16] Lip G. Y. H.. Hypertension and the prothrombotic state. *Journal of human hypertension*, **2000**, *14(10)*, 687–690.
- [17] Leisman D. E., Clifford S. D., and Matthieu L.. Facing COVID-19 in the ICU: vascular dysfunction, thrombosis, and dysregulated inflammation. *Intensive Care Medicine*, **2020**.
- [18] Chao G. et al.. Association of hypertension and antihypertensive treatment with COVID-19 mortality: a retrospective observational study. *European Heart Journal*, **2020**, 1–9. doi:10.1093/eurheartj/ehaa433
- [19] Yadaw A. S., Li Y. C., Bose S., Iyengar R., Bunyavanich S., Pandey G.. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digital Health* *2(10)*, **2020**, e516–525.
- [20] An C., Lim H., Kim D. W., Chang J. H., Choi Y. J., Kim S. W.. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Scientific reports* *10(1)*, **2020**, 1–11.
- [21] Vaid A. et al.. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *Journal of medical Internet research* *22(11)*, **2020**, e24018.

## Chapter 7

# General Discussion

As mentioned in the introduction, uncertainty in medicine and healthcare has a crucial impact on the clinical decision-making process. Therefore, the main objective of this thesis is to develop performant machine learning algorithms to handle the challenges that promote data-related uncertainty (**technical uncertainty**) (Section 1.1.2). These challenges are data and model-related **challenges** (Section 1.3.3) that are observed in the context of human health applications. Moreover, these challenges were identified: class-imbalance, non-uniform data distribution, ambiguity (confusion), continuously increasing data, model personalisation, black-box nature, online modelling and streaming analytics as illustrated in Figure 1.2b. Furthermore, these challenges can be linked to the elements of technical uncertainty [1]. To recapitulate:

- Starting with probability, which reflects the risk resulting from the indeterminacy of future outcomes. Its essential challenge is concerned with time-based prediction accuracy. This accuracy is directly related to the challenges of **class-imbalance**, **non-uniform data distribution** and **ambiguity**.
- Ambiguity reflects the indecisiveness resulting from imprecise predictions, lack of evidence and conflicting information. Ambiguity's essential challenges are linked to relation-based prediction accuracy, data sufficiency and engineered features' efficiency. The relation-based prediction (e.g. classification) is a time-independent prediction and its accuracy is influenced by **class-imbalance and non-uniform data distribution** and **ambiguity**. Data sufficiency which is required for strong evidence, and it is directly linked to the amount of data and its challenges **increasing data-size** and **online modelling and streaming analytics**. Moreover, sufficient data shall include several sources of data (e.g. multiple subjects for human health applications) which imposes the challenge of **model personalisation**. Engineered features' efficiency is needed to synthesise the different information pieces and resolve the conflict between them, which is linked to model explainability (**black-box nature**).

- Complexity results from the multiplicity of influential factors and variables or the lack of interpretability. Its essential challenges are related to the difficulty of considering many influential factors and model interpretability. Therefore, we proposed machine learning algorithms (e.g. SVM's) to handle the computational problem as they can handle a large number of input variables. On the other hand, model interpretability is linked to the challenge of [black-box nature](#).

Next we will discuss our sub-objectives in the light of the investigated human health applications. The discussion will be in terms of answering two main research questions:

1. To what extent the proposed localised learning algorithm of kNN-LS-SVM is capable of handling both data and model-related challenges that affect the model performance?
2. To what extent can a feature engineering procedure integrated with a linear hard margin approach of SVM provide interpretability and explainability while providing an acceptable error performance for prediction?

## 7.1 Question N<sup>o</sup>1

In the following subsections, we will discuss to what extent the proposed method of localised learning handled the data-related challenges and then the model related challenges. The discussion will be organised according to the [sub-objectives](#) defined in section 1.4.1.

### 7.1.1 Data-related Challenges

The first sub-objective is aiming to investigate the challenges of class-imbalance, non-uniform data distribution, ambiguity, and continuously increasing data-size. These challenges are observed in many human health and medical applications, especially those relying on continuous monitoring and wearable technology. In this thesis, we investigated three human health applications based on continuous monitoring and wearable technology, namely human activity recognition (HAR), thermal comfort prediction, and vital signs prediction. Below, we discuss how these challenges are handled in each of these applications and the experienced limitations. In chapter 2, we investigated the aforementioned data-related challenges in the context of HAR. For this purpose, we simulated the challenges of class-imbalance, non-uniform data distribution, and ambiguity using [synthetic data](#) (Section 2.4). The datasets depicted in Figures 2.3a, 2.3b, 2.3c, and 2.5a simulate these challenges in two dimensional space for binary classification problems.

#### 7.1.1.1 Class-imbalance and Non-uniform Data Distribution

The class-imbalance is obtained in synthetic data by controlling the ratio between the two classes; also, the non-uniform distribution is obtained by controlling the heterogeneity of the data points within each class. The proposed localised learning algorithm of kNN-LS-SVM was applied to

these synthetic datasets, and it outperformed both global (LS-SVM) and local (kNN) algorithms especially when the challenges level increases (e.g., extreme class-imbalance of 1.25% of imbalance percentage). After showing the relative efficiency of kNN-LS-SVM over both LS-SVM and kNN algorithms, we applied kNN-LS-SVM to four [real-world datasets](#) (Section 2.5). In these HAR datasets, the challenge of class-imbalance was present in three of them, namely WISDM v1.1, Daphnet FoG, and WISDM v2.0. For the two versions of WISDM dataset, they are multiclass problems with extreme imbalance percentage of 4.4% (standing) and 1.9% (stairs) for v1.1 and v2.0 respectively. For the Daphnet FoG dataset, the binary classification problem has an imbalance percentage of 10%. As shown in Tables 2.1, 2.2, 2.3, and 2.4, the proposed algorithm of kNN-LS-SVM showed a comparative to better performance over the other global and local algorithms except for some cases when compared to kNN. For instance, kNN outperformed kNN-LS-SVM and the other benchmark algorithms for the Daphnet FoG dataset. This is possible due to the fact that for each local model, both the number of neighbours  $k$  and the distance metric were optimised which increased the computational and temporal complexities as shown in Table 2.5. In contrast, kNN-LS-SVM local models have fixed  $k$ -number and distance metric for the same dataset.

In chapter 3, we investigated the possibility to develop an online personalised thermal comfort predictive model based on easy-to-measure variables applying kNN-LS-SVM algorithm. Similar to HAR, thermal comfort prediction faces the challenges of class-imbalance and non-uniform data distribution. The challenge of class-imbalance is present in the collected data, as shown in Figure 3.7a. It is obvious that thermal comfort states Hot and Cold (i.e. 3 and -3) are with extreme class-imbalance. In Table 3.5, the resulting confusion matrix of LS-SVM is depicted, and it shows the poor performance of the classifier especially with the extremely minor state of Cold as zero true positives were recognised. In contrast, as shown in Table 7.1, the resulting confusion matrix of kNN-LS-SVM shows a higher performance than that of LS-SVM by either removing the Else option (undefined class) or the higher true positive percentage of Cold state (23.81%). Moreover, for the Hot state, the true positives increased from 40% to 78.57%. Therefore, the localised learning approach proved to outperform the global one for this application as well. This outperformance can be understood in the light of the observed superiority of localised learning algorithms over global ones for imbalanced small-size datasets as shown in the investigation of [synthetic data](#) (Section 2.4) in Chapter 2. However, for more reliable performance, classes were reduced from seven to three by merging them in three different configurations as shown in Figures 3.7 b, c, and d.

Table 7.1: The normalized confusion matrix of Model II.

		Actual Label (j)						
		Cold %	Cool %	Slightly Cool %	Neutral %	Slightly Warm %	Warm %	Hot %
Predicted Label (i)	Cold	23.81	5.88	0	0	0	0	0
	Cool	23.81	70.59	12.90	0	0	0	0
	Slightly Cool	38.10	17.65	48.39	7.69	3.70	0	0
	Neutral	14.28	5.88	35.48	73.08	50.0	18.87	0
	Slightly Warm	0	0	3.22	19.23	42.59	35.85	0
	Warm	0	0	0	0	3.07	37.74	21.43
	Hot	0	0	0	0	0	7.55	78.57

### 7.1.1.2 Ambiguity (Confusion)

Ambiguity was introduced in the [synthetic data](#) (Section 2.4) by either overlapping the distributions or the adjacent nonlinear patterns of the two classes. Ambiguity is imposed to the synthetic data by changing the variance of the positive class to be higher and lower than that of the negative class as shown in Figures 2.3b and 2.3c. For the second synthetic dataset, the ambiguity is present by the nonlinear patterns of the two adjacent classes, as shown in Figure 2.5a. For both synthetic datasets, kNN-LS-SVM kept outperforming the global classifier LS-SVM after imposing ambiguity as shown in Figures 2.4b, 2.4c, and 2.5b.

For HAR real-world datasets 2.5.1, the challenge of ambiguity is experienced as well. For instance, in WISDM v1.1, the activities of walking up and downstairs are confusing especially for global algorithms (e.g., LS-SVM). Moreover, ambiguity is observed between the activities of *sitting* and *lying down* for WISDM v2.0 dataset. Therefore, all algorithms' error performance dropped for these two activities compared to the other activities as shown in Table 2.3. However, kNN-LS-SVM outperformed the other algorithms as shown in Figure 2.8.

For thermal comfort prediction (Chapter 4), it is observable that due to the subjective nature of thermal comfort annotation, confusion was present. As shown in Figure 7.1, there are overlaps between the thermal comfort perceptions regarding the three levels of temperature (5, 24, and 37°C) between subjects. This ambiguity is evident in the confusion matrix of LS-SVM (Table 3.5), since the global classifier was confused between the different levels of thermal comfort. For instance, the neutral state which is at 24°C only recognised correctly by 60.6% and the rest was confused with the adjacent states and further states as well. On the other hand, kNN-LS-SVM enhanced this percentage to be 73%, and the rest were confused only with the adjacent states (Slightly Cool and Slightly Warm) as shown in Table 7.1. Hence, the localised learning approach could overcome the ambiguity problem as expected better than the global algorithm as it could handle the local properties of the test instances such as the overlap between different subjects' votes. Moreover, this confusion will be discussed again later in this section under model personalisation challenge.

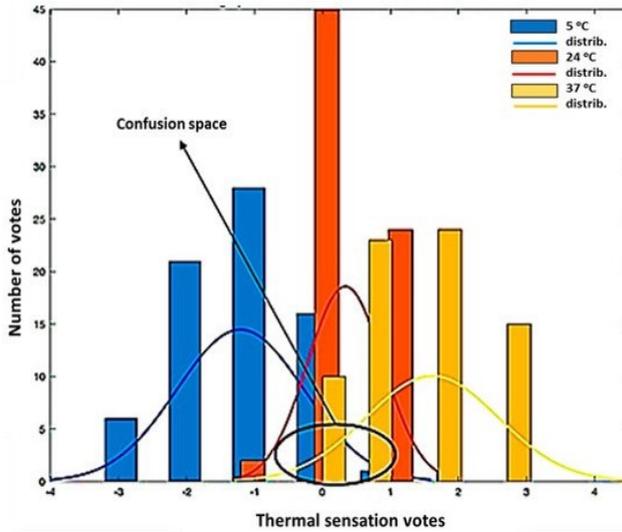


Figure 7.1: Distribution of the 25 participants' votes over their thermal sensation at the three environment temperatures of 24, 5, and 37° C. (from [4])

### 7.1.1.3 Continuously Increasing Data

This challenge is present in the first three applications (Chapters 2, 3, and 4) as variables were monitored with a minimum rate of 1 Hz (e.g., acceleration, heart rate, heat flux). Two analytical problems result from the continuous data-size increment: increasing computational complexity and varying data distribution in the input space. These problems are correlated since the data-size influences the computational complexity once retraining or updating the model is required to handle the data distribution change. This change can result from collecting data from different and new subjects (sources), promoting inter-subject variability. Moreover, it may result from intra-subject variability as the subject's status can be time-dependent (e.g. vital signs variation due to health status). We proposed an approach (kNN-LS-SVM) to control the computational complexity by training local models with a relatively small number of data points. Hence, the computational complexity of the predictive model is (partially) isolated from the increasing data-size. However, our approach depends on kNN which is influenced by the data-size, which will be discussed in the limitations section. A reflection on the computational complexity for the three applications is discussed below.

For HAR application (Chapter 2), a local model is trained, validated and tested for each test sample (unseen sample). Moreover, the recorded elapsed time measurements in Table 2.5 are those of online modelling. For kNN-LS-SVM, the recorded elapsed time is that of building the local model in addition to applying the model to the test-point of interest. The recorded elapsed times show that kNN-LS-SVM outperforms all other benchmark methods from a temporal performance perspective. It is worth mentioning that for global models (i.e. LS-SVM and Stack-AE)

and PSVM, the elapsed time is the test time only. On the other hand, for kNN-LS-SVM, kNN and KNN-SVM, the elapsed time is the sum of training, validation, and test times.

For thermal comfort prediction (Chapter 3), as we have shown in the previous section of HAR, the time performance of applying the kNN-LS-SVM to the HAR datasets was outperforming the other techniques. The elapsed time for kNN-LS-SVM online modelling was at maximum (0.0126 seconds), which is less than the sampling period of one second for thermal comfort prediction.

In chapter 4, we investigated vital signs prediction of hospitalised patients. The recording rate of the monitored vital signs (i.e. heart rate, respiration rate, oxygen saturation, systolic, diastolic, and mean arterial blood pressure) is 1 Hz. With this relatively high rate, the data-size is continuously increasing. In this application, we applied kNN-LS-SVM for time-series prediction with a prediction horizon up to 3 hours that is updated every 10 minutes. As mentioned in the section of [results](#) (Section 4.4), the required number of training points for the different profiles of patients is fixed (i.e.  $k=25$  for cardiology and post-operative;  $k=15$  for dialysis patients) regardless the data-size.

Conclusively, by proposing kNN-LS-SVM, we aim to (partially) isolate the predictive model's computational complexity from the continuous increase of the data-size.

### 7.1.2 Model-related Challenges

In this section, we discuss the second and third sub-objectives regarding online modelling, streaming analytics, and model personalisation. These challenges were investigated in chapters 2, 3, and 4. In these chapters, kNN-LS-SVM algorithms for classification and regression are proposed to tackle the challenges as will be discussed below.

#### 7.1.2.1 Online Modelling and Streaming Analytics

The analysed data in the three applications of HAR, thermal comfort, and vital-signs prediction are collected by wearable sensors. These sensors are used to measure the variables of interest continuously (i.e. movement acceleration, ECG, PPG, heat-flux, body and skin temperature). From these variables, several vital-signs are estimated, such as physical activity, heart rate, respiration rate, oxygen saturation, blood pressure, and metabolic rate.

For HAR (Chapter 2), the activities are recognised based on either 4-second or 10-second sliding non-overlapping time windows. Therefore, a high rate (elapsed processing time  $<$  time-window size) activity recognition algorithm is required, which can be obtained by either offline or online learning algorithms, as shown in Table 2.5. However, in a practical approach of HAR application, data will be acquired from multiple subjects to be stored and categorised in a cloud. These different subjects may provide different behaviours for the same activities, which changes the data distribution in the input-space. Therefore, updating HAR models that adapt to the new data distribution may enhance the model's error performance. Moreover, if the option to

add new activities to the HAR system by personal annotation is valid, then the HAR model has to be updated to recognise these new activities as new data points are added to the input space with new labels. Therefore, we used the localised approach of kNN-LS-SVM to provide such online adaptation to any updates to the datasets. The main challenge that faces online modelling, especially to provide real-time predictions is the model's time performance. As shown in Table 2.5, kNN-LS-SVM provides the best time performance compared to the benchmarking algorithms for simulated online modelling (leave-one-sample-out).

For thermal comfort prediction application 3, continuously collecting data from multiple subjects is present. Referring to the proposed framework (Figure 3.8), online modelling and streaming analytics are required. More specifically, data is continuously streaming from wearable sensors to the smartphone to be integrated into the mobile application. Furthermore, the framework requires personal annotation, especially when a new subject uses the system. Therefore, the model has to be updated, considering the new personal data. Hence, streaming analytics is required to handle this in real-time. Using kNN-LS-SVM for this application provided the option to update the model for each new test point; hence, it can support online modelling. Moreover, this update allows considering new streaming data points to train the model which supports streaming analytics. Therefore, the relatively low and limited computational complexity and the acceptable error performance of kNN-LS-SVM makes it a potential candidate to provide both online modelling and streaming analytics for this application.

For vital-signs prediction (Chapter 4), online modelling and streaming analytics are beneficial. Especially that, for time-series prediction of such a dynamic system, it is required to consider the real-time dynamics of vital-signs. As shown in section 4.3.2, features are extracted from 70 minutes period to predict the statistical values of the different vital signs within specific prediction horizons (i.e. +1, +2, and +3 hours). This prediction is designed to be updated every 10 minutes. Moreover, data is acquired from different subjects which affects the data distribution in the input-space similar to the previous applications. Therefore, online modelling and streaming analytics in this application can be beneficial, given the dynamic nature of the biological system of humans in addition to the diversity of data sources. Furthermore, for human health monitoring, it is needed to provide the analytics on time for an efficient clinical decision-making process.

Conclusively, by proposing kNN-LS-SVM, we aim to keep a balance between updating the model and controlling the computational complexity to develop the model given the streaming data. From the obtained results in the first three applications, kNN-LS-SVM provided better combined error and time performance compared to the benchmark local and global methods.

### 7.1.2.2 Model Personalisation

Model personalisation was experienced in both thermal comfort and vital-signs prediction applications (refchapter2, refchapter3). In both applications, the localised learning algorithm of kNN-LS-SVM is proposed given the characteristics of the algorithm. The algorithm that is used introduced another aspect to model personalisation. Hitherto, most of the studies that aim at

developing a personal model rely only on personal data. However, our proposed method relies on the most similar data regardless of whose data is used. This approach's advantage is that a new subject can be analysed efficiently, even if the obtained data from that subject is not enough for training a model. Furthermore, the proposed method prioritises the similarity over personal data. More specifically, data from a different person in the same condition (e.g. fever) can be more informative than personal data in different condition (e.g. normal temperature). For thermal comfort prediction, the obtained personal data of a specific subject can be in a particular situation (healthy). However, the new test point can be in another situation (having a fever). For this purpose, it will be more accurate to train the model with similar condition data from another subject than training the model with personal data. For vital-signs prediction, the same concept is valid as another person with the same condition can be more similar than the same person in a different condition. However, this approach does not ignore the personal data as the nearest data-points can be from the same person or different persons.

Overall, we can conclude that kNN-LS-SVM provided a similar or better performance than the benchmark methods from either error or time performance perspective when the aforementioned challenges are present in the investigated datasets.

## 7.2 Question N<sup>o</sup>2

To what extent can a feature engineering procedure integrated with a linear hard margin approach of SVM provide interpretability and explainability while providing an acceptable error performance for prediction?

### 7.2.1 Model Interpretability and Explainability

In Chapters 5 and 6, the challenges are different and data characteristics as well. Observation rates are relatively low (observation/1-2 hours), the events are crucial (mortality), and the patients are bedridden (ICU patients). In these studies, we focused on features that can inform mortality prediction for ICU patients.

In chapter 5, another analytical perspective is introduced regarding the application of ICU mortality prediction. This study focused on keeping the balance between the computational power of machine learning and the system interpretability for mortality prediction. As shown in chapter 5, the problem of mortality prediction is studied, focusing either on the prediction power (black-box) or the clinical interpretability (white-box). However, our study considers a grey-box approach to tackle this problem as the main objective is to engineer simple explainable features that can provide an acceptable prediction performance. The used machine learning algorithm is a linear hard margin approach of SVM because of its specific characteristics, namely, linearly separating classes and maximising the margin between them with minimal error tolerance.

In this study, we investigated the mortality prediction problem in light of the parsimony principle. This principle tends to choose the simplest solution among equally performant solutions.

This simplicity can be observed in the small number of model parameters and linearity over non-linearity of the model. In this study, we believe that a simple model (linear hard margin SVM) with meaningful features can achieve a reliable error performance. In her article [5], Rudin C. addresses a similar conceptual approach: *"Stop explaining black-box machine learning models for high stakes decisions and use interpretable models instead"*. In her article, she believes that the necessary trade-off between accuracy and interpretability is a myth. In other words, investigating problems with structured data and meaningful features would provide an acceptable performance by either simple or complex models [5]. In our study, we had the opportunity to engineer features that can provide linear separation between the two classes that can inform mortality prediction. Moreover, we aimed to prove a concept that by fitting a set of features to a wide diversity of patients' profiles, these features can be subdivided into subsets of features that perform well for a corresponding profile.

Recalling the fourth **sub-objective** (Section 1.4.1), model interpretability and explainability are met and partially tackled in this study. All extracted features are simple and can be explained easily to medical staff. Moreover, clinical insight is obtained from the extracted features, especially those regarding the pulse pressure and its association to mortality. The analysed dataset includes patients that were hospitalised for different reasons and admitted to the ICU for various reasons (e.g., cardiology, neurology, abdominal) as shown in Figure 5.1. Hence, the engineered features are expected to be suitable for several profiles of ICU patients. Moreover, we expected in the conclusion of this study that each profile or combined profiles of the studied patients will require a subset of the resulting features which is approved later in chapter six.

An essential phase in the process of feature engineering in this study is feature fine-tuning. This phase mainly focuses on synthesising a set of the extracted features to enhance the mortality prediction. This feature fine-tuning comprises a human-in-the-loop feature selection method instead of an automatic one. Several advanced model-based and model-free feature selection techniques are developed to tackle similar problems. However, we used the model-based human-in-the-loop approach to integrate the physiological insight into the feature selection process. This physiological insight was considered in the first step of feature fine-tuning by excluding the HR-extracted statistical and dynamic features. This exclusion was motivated by the fact that HR changes occur as a response to several physiological dynamics. Moreover, the physiological insight appeared in including the pulse pressure features. Ultimately, this physiological insight was approved by medical doctors.

Moreover, features are extracted from consecutive time windows of the measured time-series. Therefore, the same feature extracted from the same variable may have multiple dimensions corresponding to the number of time-windows as illustrated in the **discussion** (Section 5.5). For instance, extracting the feature of the mean value of HR from four time-windows provides a feature vector of 4 dimensions. Therefore, for clinical reasons, we based the selection on the feature and the variable not on dimension basis, and this is the main reason to avoid automatic selection techniques.

### 7.2.2 Error Performance and Model Interpretation

In order to prove that a partially interpretable predictive model integrated with explainable and meaningful features can provide a reliable error performance, we applied the proposed method in chapter 5 to a specific profile of ICU patients, namely pulmonary patients with COVID-like symptoms. By selecting a subset of features (Table 6.3) from the previously engineered features in chapter 5, we obtained the mortality prediction performance as depicted in Tables 6.4 and 6.5. These results are obtained from developed predictive models for the three subpopulations shown in Table 6.2. It is worth mentioning that the engineered features in chapter 5 are defined from a smaller dataset (447 patients) compared to the investigated dataset in chapter 6 (801 patients). By focusing on the results of subpopulation II as it is the more inclusive COVID-like population symptom-wise, the obtained results are promising with  $F_1$ -scores of 0.88 and 0.85 for five-day and two-day historical data, respectively. Moreover, the selected features (Table 6.3) for the three subpopulations reflect the relevance of three vital-signs, namely,  $SpO_2$ , respiration rate, and blood pressure. Both  $SpO_2$  and respiration rate are trivially relative to a pulmonary problem. However, adding blood pressure to these vital-signs was not evident till recent studies shown the link between hypertension and mortality risk of COVID-19 patients [6].

As mentioned earlier, we focused on the precision and the recall of mortality prediction. However, in the context of COVID-19 pandemic, it is worth investigating the recovery prediction for the same models as well. Therefore, in Table 7.2, we depict sensitivity, precision, and  $F_1$ -score for the developed models in chapter 6 illustrated in Table 6.4. From Table 7.2, it is evident that the developed models provide acceptable performance for recovery prediction as well. Therefore, based on such reliable models for both mortality and recovery predictions, they can be considered as decision support tools. Such tools can support the decisions of urgent medical intervention or discharge based on the outcome of the predictive models (i.e. mortality or recovery). Ultimately, once these models are validated on COVID-19 data and the whole stay at ICU, they can be used for the pandemic of COVID-19 to predict the future status of mortality or recovery. Such predictions are of paramount importance for ICU's in hospitals where these units experienced under-capacity during the first wave of the pandemic. Therefore, the precise early decision of discharge can save beds to patients with a high need to be admitted.

Conclusively, we can answer the second research question positively. More specifically, the approach of explainable feature engineering procedure integrated to linear hard margin SVM helps to extract explainable and straightforward features. Moreover, the classifier provided a linear hyperplane for linearly separable classes in the input space. Furthermore, the error performance of the proposed classifier with the selected feature was acceptable.

## 7.3 Limitations

In this section, we discuss the experienced limitations in the investigated applications and the proposed methods.

Table 7.2: Sensitivity and precision and  $F_1$ -score of recovery/survival for models 1 and 2 of subpopulations: Pulmonary+Infection (I), Pulmonary+Infection+Pneumonia (II), and Pulmonary+Pneumonia (III) based on 60 observations for feature extraction.

Subpopulation	Model	Sensitivity %	Precision %	$F_1$ Score
Pulmonary+Infection (I)	1	90	90	0.90
	2	96.67	87.88	0.9201
Pulmonary+Infection+Pneumonia (II)	1	88.86	90.38	0.8961
	2	90.57	88.89	0.8972
Pulmonary+Pneumonia (III)	1	92.1	83.34	0.8750
	2	88.1	88.1	0.8810

### 7.3.1 Extra Hyperparameter

For the first three applications (i.e. HAR, thermal comfort and vital-signs prediction), the proposed method was kNN-LS-SVM. One limitation with kNN-LS-SVM is that it introduces a new hyperparameter ( $k$ ) to be optimised. In our algorithm, we optimised  $k$  with other hyperparameters ( $\gamma$  and  $\sigma$ ) during the validation phase. In their study, Karevan et al. [7] introduced the cosine-based similarity method ( $S_i(x) = \frac{x^\top x_i}{\|x\| \times \|x_i\|} + 1$ ) as a similarity measurement function  $S_i(x)$  to weight all training data points  $x$  based on their similarity to the test point  $x_i$ . This approach does not require a new hyperparameter to be optimised. However, when data is nonuniformly distributed the number of similar data points can vary such that, when the number of similar data points is high, the computational complexity of this approach increases. Moreover, fixed or predictable temporal and computational complexities are crucial for online modelling and streaming analytics to assure a reliable real-time processing. Therefore, our proposed approach can be more suitable for these applications compromising the added complexity of the new hyperparameter (i.e.  $k$ ). Future research has to decide how the number  $k$  can be inferred from the distribution of the data points in the input space. This investigation is to determine how many data points can inform the characteristics of different regions in the feature space. However, we claim that using kNN-LS-SVM approach is still suitable for streaming and online analytics recalling the time performance of kNN-LS-SVM, as shown in Table 2.5.

### 7.3.2 Data-size

Another limitation is regarding data-size, it is worth mentioning that data-sizes in all three chapters (2, 3, and 4) are small to middle-sized datasets (100 to 15,000 data points). Data-size can affect the computational complexity of the kNN search algorithm; however, several search algorithms are developed to reduce the search computational complexity [8]. For instance, Wang X. proposed in his study [9] an enhanced kNN search algorithm k-means for kNN (kmkNN) that performed faster than the conventional kNN brute-force algorithm by 30-folds and 70-folds for datasets with  $10^6$  and  $5 \times 10^5$  samples respectively. The computational complexity of the kmkNN algorithm is  $\mathcal{O}(d\sqrt{n} \log dn)$  which is less than that of brute-force kNN algorithm whose compu-

tational complexity is  $O(dn)$  where  $n$  is the number of samples and  $d$  is the number of dimensions. This issue is still an interesting future research direction.

For the thermal comfort application, data shall be extended quantitatively and qualitatively. Data were collected only from 25 healthy young subjects (age: 21–36), which is supposed to be increased in number and age spectrum to provide more diversity. The same limitation regarding data-size was present in vital-signs prediction study, especially for the profiles of post-surgical and dialysis patients with only 10 and 7 patients respectively.

### 7.3.3 Missing a Gold Standard

Regarding missing a gold standard, another limitation was experienced in both clinical studies of vital-signs prediction and ICU mortality prediction. For vital-signs prediction, although the used monitoring device (Somnotouch NIBP) is clinically approved, the estimated vital-signs (e.g. HR) were noisy. Therefore, it would be more efficient to have a gold standard to validate the denoised signals based on synchronous reference measurements. For mortality prediction study, to test our predictive model for real-time mortality risk prediction we need a real-time indicator to mortality risk. Especially that, along the stay, the label is always survival; therefore, another outcome is needed to validate our model. This outcome can be a dynamic version of do-not-resuscitate (DNR) [10] code to be updated during the stay at the ICU based on the physician's assessment of the patient's status. A DNR code status would indicate that the patient would not want cardiopulmonary resuscitation (CPR) performed and would be allowed to die naturally only if their heart stops beating and/or they stop breathing. Hence, the frequently updated version of the DNR code can be considered an indirect indicator of the severity of the case during the stay that reflects the mortality risk.

### 7.3.4 Lack of Data

Ultimately, the main experienced limitation in the study of COVID-like mortality prediction is missing the real COVID-19 data to validate our mortality prediction models of COVID-like patients.

## 7.4 Valorization

### 7.4.1 Human activity recognition

Recently, physical activity was considered a vital sign that reflects an informative health status aspect of subjects, especially those with chronic conditions [11]. Therefore, an integrated approach of HAR and vital-sign monitoring can result in a more complete picture of a subjects health status. Physical activity (PA) assessment is important because of the observed influence of PA intervention on reducing both cardiovascular diseases and cardiometabolic risk factors [12, 13, 14, 15, 16].

Therefore, accurate PA assessment tools are needed to provide adequate feedback to subjects under observation or to their care-givers. Physical activity vital sign tools are methods by which clinicians assess the physical activity (PA) of subjects to obtain indicators of their general physical condition. Moreover, Nelson et al. [17] foresee having future strategies to address or prevent cardiometabolic disease, especially for pediatric patients given the correlation between PA and other factors (e.g., age, sex, and blood pressure).

There are several versions of these tools that measure PA, (e.g., exercise vital sign (EVS), physical activity vital sign (PAVS), and speedy nutrition and physical activity assessment (SNAP)). EVS and PAVS are widely used in the USA health care systems to pay attention to patients with low PA levels; besides these PA indicators are inserted into the patients' electronic medical record (EMR) section of vital signs. These tools are based on questionnaires that are carried out manually by certified nurses while collecting traditional vital signs. These questions are mainly about the physical activities that were performed, their intensity, and their rate within a specific period (e.g., on average, how many minutes per day every week).

A foreseen valorization option is integrating novel wearable technologies (e.g., smartwatches and patches) with an accurate HAR algorithm. Wearable technology is for continuously monitoring physical activities, and the HAR system is a replacement of the manual questionnaires. Therefore, integrating HAR into physical activity vital sign tools can provide autonomy to these tools.

Several requirements are needed to implement our proposed HAR algorithm for continuous monitoring of PA and integrating its outcome into patients' EMR at hospitals. First of these requirements is the used wearable device(s) and their specifications from the clinical perspective. Fortunately, many wearable devices for activity recognition in the market are suitable for this purpose. The second requirement is the need to labelled data for daily-life activities and postures of interest, which can be obtained from publically available datasets. We initiate the training pool for our model by this labelled data till the new data is collected. The first phase of streaming data from patients shall be labelled, contributing to the training pool. Therefore, a labelling procedure has to be followed by the patient or the nursing staff using the device itself, which is an option to be considered in the device specifications. Regarding the required infrastructure for this system, a possible solution is to collect and process the streaming data by the hospital's local server. A next step if several regional hospitals are interested in the system is to store and process the streaming data besides running the model on a cloud that links the different hospitals' servers. The final step is to integrate the model's outcome into the patients' EMR at the hospital.

#### 7.4.2 For Thermal Comfort Prediction

Conceptually, the proposed method in 3 closes the loop of a control system that controls the environmental temperature using an online personalised classification model. Moreover, this application can be useful for hospitalised patients for automatic environment temperature control, considering the physiological and psychological impact of heat on patients.

For instance, in their study, Wagner et al. [18] investigated the effect of thermal comfort on preoperative patients. It is found that thermal comfort and discomfort of patients before and during the surgery has a proportional effect on patients' well-being and their feeling of anxiety [19, 20, 21]. It is worth mentioning that a higher level of anxiety is linked to tachycardia, hypertension, arrhythmias, and increasing levels of pain [22, 23, 24]. Furthermore, the hormonal response of stress is believed to affect the recovery process [24, 25]. Moreover, thermal comfort affects the therapeutic intervention for thermal management during surgery [19, 26, 27]. Unfortunately, the followed procedure to provide the required thermal comfort depends on the manual assessment of the nursing staff and their manual intervention. In addition, this assessment is subjective and patient dependent given the inter-patient variability of the thermal perception and their physiological and psychological status.

A possible valorization option of our proposed method is to assess the patients' thermal comfort automatically prior to and during their surgeries considering their personal perception of temperature. Moreover, our developed thermal comfort prediction model only requires the traditionally monitored vital signs prior to or during the surgery (e.g., heart rate, temperature) in addition to the heat flux which can be estimated efficiently. This autonomous assessment can provide feedback to the nursing staff to intervene manually or to an active warming technology depending on the available resources at the hospital.

Similar to HAR algorithm, the thermal comfort prediction algorithm can be integrated into the hospital's monitoring system for surgery patients. This integration needs some requirements to be satisfied. Firstly is needed to determine the monitoring devices to collect the required vital signs meeting the clinical standards. Furthermore, labelled data are required to initiate the training set. For this application, it is essential to collect data in advance of this profile (i.e. surgery patients) and label it based on personal perception. The collected data can be stored on the local hospital server. The algorithm can run on the same server and provide its outcome to nursing staff or directly to the controller of active warming technology. On a later stage, data from different hospitals can be collected and stored on a cloud to enrich the training pool, but the algorithm can still run locally on hospitals' servers.

### 7.4.3 For Vital-Signs and Physical Activity

In all three applications (Chapters 2-4), we used one common algorithm, which is kNN-LS-SVM. As discussed earlier in these chapters, this algorithm is a simple approach of localised learning that can provide fixed/predictable computational complexity regardless of the data-size increment. By fixed/predictable complexity, we refer to the fixed number ( $k$ ) of the training datapoints to build a new model for each test point.

As shown earlier, the first three applications are dealing with healthy subjects and hospitalised patients at general wards. One main feature that is in common between these applications is the usage of wearable sensors which is shaping the future of healthcare. As using the wearable technology is affording the possibility to achieve continuous and real-time monitoring of

many important human-health related parameters. This possibility to collect this critical data automatically and with minimum human intervention can reduce the load on the medical staff, especially in critical times such as COVID-19 pandemic. Moreover, using wearable technology in the presence of efficient networking infrastructure, patients can be monitored at home, which can limit the under-capacity of hospitals. Integrating wearable technologies with smart algorithms similar to our developed ones can also assist the medical staff in assessing and foreseeing the critical events on time. For at-home monitoring, this requires connectivity to the hospital server through a cloud, which can be achieved by cellular networks or WiFi networks using smartphone applications as an interface. Such an approach is already used with smartwatches connectivity to smartphones and further to a cloud in which each subject's data is stored and processed. In case of lost network connectivity, this would require a local memory integrated into the device to store data for some time.

Regarding the first three applications, namely human activity recognition, thermal comfort, and vital signs prediction, we suggest them as the basis of a framework that merges three applications. This framework from a technology perspective is based on wearable sensors to measure all variables of interest (i.e. physical activity, heart rate, blood pressure, oxygen saturation, respiration rate, temperature, heat flux, and metabolic rate). Fortunately, all of these variables are either measurable or can be estimated by already available wearable sensors. From a data perspective, we expect that the collected data will be in a streaming form because of the high-rate measurement and monitoring. This framework optimally can be designed based on cloud architecture that can sort and store the collected data and make it accessible for analysis. From an analytical perspective, this framework can rely on our proposed localised learning approach, given its efficiency to handle streaming and online analytics. By developing such a framework, it will be possible to continuously and autonomously monitor the health status of patients at either hospitals or homes. With this option of autonomous (tele)monitoring, we can lighten the workload of clinical staff and the admission load to hospitals. Therefore, it will be an efficient approach to integrate this framework with EMR systems at hospitals. For example, in a situation similar to that of COVID-19 pandemic, such a framework can provide important logistic and medical support to healthcare providers since all patients are monitored and assessed continuously with minimal human intervention. The outcome of this framework is expected to be a real-time monitoring and time-series predictions of the measured vital signs. Regarding personal annotation of activities and thermal comfort levels, this can be achieved using the smartphone application. The annotation does not need to be continuous as discrete annotation sessions are needed to adapt the HAR and thermal comfort prediction models for personal use.

#### 7.4.4 WearIT4health

Finally, the nearest valorisation option is linked to [WearIT4health](#) system. This system comprises a wearable vital signs' monitoring device that is integrated into hospitals' EMR systems. The wearable device is still under development and clinical approval, and it is intended to measure



Figure 7.2: The wearable device ©WearIT4health to monitor hospitalised patients' vital signs in real time.

the vital signs of heart rate, blood pressure, oxygen saturation, temperature and respiration rate. Based on the obtained results of the developed algorithms in Chapter 5, we expect to integrate our algorithms into WearIT4health system. Especially that, the used wearable device in [EAGLE](#) study measures the same vital signs. Moreover, the running project [WearIT4COVID](#) intends to continue developing the WearIT4health system focusing on COVID-19 hospitalised patients. In this project we are going to adapt our developed [EWS computation](#) and [prediction](#) models to COVID-19 patients.

# Bibliography

- [1] Beresford EB. Uncertainty and the shaping of medical decisions. *Hastings Cent Rep.*, 21(4), 1991, 6–11.
- [2] Kangmoon K., and Lee Y-M.. Understanding uncertainty in medicine: concepts and implications in medical education. *Korean journal of medical education*, 30(3), 2018, 181–188.
- [3] Han P.K., Klein W.M. and Arora N.K.. Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*, 31(6), 2011, 828–838.
- [4] Youssef A., Caballero N., Aerts J.-M.. Dynamic Model-Based Monitoring of Human Thermal Comfort for Real-Time and Adaptive Control Applications. *Biomedical Journal of Scientific and Technical Research, Biomedical Research Network*, 19(4), 2019, 14526–14532.
- [5] Rudin C.. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 2019, 206–15.
- [6] Chao G. et al.. Association of hypertension and antihypertensive treatment with COVID-19 mortality: a retrospective observational study. *European Heart Journal*, 2020, 1–9. doi:10.1093/eurheartj/ehaa433
- [7] Karevan Z., Feng Y., Suykens J.A.K.. Moving Least Squares Support Vector Machines for weather temperature prediction. *In ESANN*, 2017.
- [8] Deng Z., Zhu X., Cheng D., Zong M., Zhang S.. Efficient kNN classification algorithm for big data. *Neurocomputing*, 195, 2016, 143–148.
- [9] Wang X.. A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality. *International Joint Conference on Neural Networks, IEEE*, 2011, 1293–1299.
- [10] Becker C., Manzelli A., Marti A., Cam H., Beck K., Vincent A., Keller A., Bassetti S., Rikli D., Schaefer R., Tislar K.. Association of medical futility with do-not-resuscitate (DNR) code status in hospitalised patients. *Journal of medical ethics*, 2021.

- [11] Golightly Y.M., Allen K.D., Ambrose K.R., Stiller J.L., Evenson K.R., Voisin C., Hootman J.M. and Callahan L.F., 2017. Peer Reviewed: Physical Activity as a Vital Sign: A Systematic Review. *Preventing Chronic Disease*, 14(123), 2017.
- [12] Biswas A., Oh P. I., Faulkner G. E., et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. *Ann Intern Med.*, 162, 2015, 123–132
- [13] Sofi F., Capalbo A., Cesari F., Abbate R., Gensini G. F.. Physical activity during leisure time and primary prevention of coronary heart disease: an updated meta-analysis of cohort studies. *Eur J Cardiovasc Prev Rehabil*, 15,2008, 247–257.
- [14] Stovitz S. D.. Contributions of fitness and physical activity to reducing mortality. *Clin J Sport Med.*, 22, 2012, 380–381.
- [15] Fagard R. H., Cornelissen V. A.. Effect of exercise on blood pressure in hypertensive patients. *Eur J Cardiovasc Prev Rehabil*, 14,2007, 12–17.
- [16] Kelley G. A., Kelley K. S., Tran Z. V.. Exercise, lipids, and lipoproteins in older adults: a meta-analysis. *Prev Cardiol.*, 8, 2005, 206–214.
- [17] Nelson V. R., Masocol R. V., Asif I. M.. Associations between the physical activity vital sign and cardiometabolic risk factors in high-risk youth and adolescents. *Sports Health*, 12(1), 2020, 23–28.
- [18] Wagner D., Byrne M., Kolcaba K.. Effects of comfort warming on preoperative patients. *AORN journal*, 84(3), 2006, 427–448.
- [19] Fossum S., Hays J., Henson M. M.. A comparison study on the effects of prewarming patients in the outpatient surgery setting. *Journal of PeriAnesthesia Nursing*, 16 2001, , 187–194.
- [20] Kolcaba K., Wilson L.. Comfort care: A framework for perianesthesia nursing. *Journal of PeriAnesthesia Nursing*, 17, 2002, 102–114.
- [21] Wagner V. D.. Impact of perioperative temperature management on patient safety. *SSM*, 9, 2003, 38–43.
- [22] Caumo W. et al.. Risk factors for preoperative anxiety in adults. *Acta Anaesthesiologica Scandinavica*, 45, 2001, 298–307.
- [23] Boker A., Brownell L., Donen N.. The Amsterdam preoperative anxiety and information scale provides a simple and reliable measure of preoperative anxiety. *Canadian Journal of Anesthesia*, 49, 2002, 792–798.

- [24] Munafò M., Stevenson J.. Anxiety and surgical recovery. Reinterpreting the literature. *Journal of Psychosomatic Research*, 51, 2001, 589–596.
- [25] Grieve R. J.. Day surgery preoperative anxiety reduction and coping strategies. *British Journal of Nursing*, 11, 2002, 670–678.
- [26] Sessler D. I. et al.. Optimal duration and temperature of pre-warming. *Anesthesiology*, 82, 1995, 674–681.
- [27] Vanni S. M. et al.. Preoperative combined with intraoperative skin-surface warming avoids hypothermia caused by general anesthesia and surgery. *Journal of Clinical Anesthesia*, 115, 2003, 119–125.



## Chapter 8

# Conclusions and Future Work

Throughout this thesis, we investigated the fundamental problem of data-related technical uncertainty in medicine by addressing its different aspects. By understanding the various aspects, our hypothesis is to restrain this problem by providing reliable decision support systems. In this thesis, we developed analytical methods to comprise the analytical component of the decision support system. These analytical methods include machine learning algorithms and feature engineering methodology. The developed machine learning algorithms are mainly localised learning approaches for either classification or time-series prediction. By evaluating these algorithms, we conclude that the localised learning algorithm of kNN-LS-SVM provides high comparative performance. This recognised performance is computationally, temporally and accurately proven. Moreover, this approach can be suitable for streaming analytics and online modelling, which we foresee as an efficient combination with wearable technologies. Furthermore, the algorithm of kNN-LS-SVM is easily implemented, which is beneficial from practical perspective.

An important concept that is proven in this thesis, more specifically, in the mortality prediction application is that a simple model and easily extracted features can predict a critical event such as mortality with acceptable efficiency. Moreover, this simplicity can provide partially interpretable and explainable predictive models which limits the black-box effect. Furthermore, the followed methodology of feature engineering helped to obtain a clinical insight from the analysed data. Ultimately, fitting a set of features to a widely diverse dataset (e.g., multiple profiles) can provide subsets of features that provide efficient performance for each profile.

For future work, we foresee possible applications and some research points to be investigated by future research. Regarding possible applications, we propose two potential applications:

- The first is the integrative framework that merges the applications of HAR, thermal comfort and vital-sign prediction. By such a framework, it will be possible to monitor continuously activity and vital-signs of subjects using wearable sensors which can lighten the workload on medical staff and admission load to hospitals. Moreover, this framework with proper infrastructure can be extended to at-home monitoring which is helpful in such a

situation like COVID-19 pandemic.

- The second application is a platform for mortality risk prediction at ICU's based on the monitored vital-signs of admitted patients. Such a platform shall be integrated into the EMR system at hospitals to access the required data in addition to submitting and displaying the mortality risk in real-time.

Regarding the research points to be furtherly investigated, five of them are stated below:

- Firstly we propose concerning the algorithm of kNN-LS-SVM to develop a methodology that reduces the complexity of optimising k-number. Such a method can be obtained by having an insight into the data density in different regions in the input space in advance by which the k-number can be determined to avoid a hyperparameter tuning phase for k-number.
- A second research point is to consider all analysed vital-signs and variables (i.e. physical activity, heart rate, respiration rate, oxygen saturation, blood pressure, heat flux, metabolic rate, body and skin temperature) in one predictive model hypothesising that these variables are influencing each other. For this purpose, it is required to monitor these variables simultaneously by reliable sensors.
- The third research approach is to re-evaluate the standard early warning score using a continuous calculation of the EWS using wearable technology to be compared to the clinical gold standard determined EWS score. The purpose of this research is to investigate the temporal precision of the standard EWS. In other words, whether the standard rate of EWS calculation (2-3 times per day) reflects a precise picture of the health status dynamics or not. In case this rate is not sufficient, a high-rate of EWS using wearable technology can be a proper replacement.
- The fourth research point is regarding the mortality prediction problem as we propose to assess the mortality risk prediction during the stay by a clinical input or a reference biomarker that informs mortality risk. Afterwards, we can evaluate the real-time performance of the mortality prediction model. Furthermore, such a predictive model can be used in combination with standard ICU scoring systems (e.g., APACHE II) to provide real-time assessment of severity-of-disease at ICU.
- Furthermore, predictive models of vital-signs prediction will be adjusted and validated for COVID-19 patients. Therefore, these predictive models would provide an early warning system dedicated to COVID-19 patients. This research plan is to be carried out by our research groups as a key activity in the Interreg project of WearIT<sub>4</sub>COVID.