



Vital Signs Monitoring and Interpretation for Critically Ill Patients

Vilic, Adnan

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Vilic, A. (2017). *Vital Signs Monitoring and Interpretation for Critically Ill Patients*. Technical University of Denmark, Department of Electrical Engineering.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Technical University of Denmark



VITAL SIGNS MONITORING AND INTERPRETATION FOR CRITICALLY ILL PATIENTS

By Adnan Vilic

PhD Thesis, June 2017

In collaboration with:

Technical University of Denmark, Lyngby

Zealand University Hospital, Roskilde

Bispebjerg University Hospital, Copenhagen

DTU Electrical Engineering
Department of Electrical Engineering



PhD dissertation by: Adnan Vilic

Main supervisor: Associate Professor MSK, MSc.EE, PhD, Helge B. D. Sørensen
Department of Electrical Engineering
Technical University of Denmark, Lyngby

Clinical supervisors: Professor, MD, PhD, Troels Wesenberg Kjær
Department of Neurology
Zealand University Hospital, Roskilde

Clinical Associate Professor, MD, PhD, John Asger Petersen
Intensive Care Unit Z
Bispebjerg University Hospital, Copenhagen

© Adnan Vilic, 2017

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Technical University of Denmark
Department of Electrical Engineering
DK-2800 Kgs. Lyngby Denmark

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Technical University of Denmark.

PREFACE

This dissertation is prepared at the Department of Electrical Engineering, at Technical University of Denmark (DTU). It is a partial requirement for obtaining the degree of Doctor of Philosophy. The research has been carried out in a close cooperation between Technical University of Denmark (DTU), Bispebjerg Hospital and Zealand University Hospital in Roskilde (ZUH). The dissertation is written in a form that can be read independently of the related composed articles, hence textual and graphical parts are occasionally repeated.

The presented research was carried out in the period between March 2013 and June 2017. Within this were two timespans with leave of absence which, when combined, accounted for one year. Besides conducting the research, other activities included supervising bachelor- and master degree students, teaching, participating in conferences, and upgrade and maintenance of laboratory equipment. All applications and approvals for data collection and management were written and obtained during the studies. In terms of publications, the research resulted in two journal papers and three conference papers.

The project was internally funded.

Adnan Vilic
Lyngby, June 2017

ACKNOWLEDGEMENTS

I would first like to thank my supervisors, Associate Professor, MSK, PhD, Helge B. D. Sørensen from Technical University of Denmark (DTU), Clinical Associate Professor, MD, PhD, John Asger Petersen from Bispebjerg University Hospital and Professor Troels W. Kjær from Zealand University Hospital (ZUH) for their continuous optimism, enthusiasm and support from throughout the study. I am very thankful to Chief Physician, PhD, Troels Wienecke from Zealand University Hospital for taking time to share his knowledge and giving me practical insight into treatment of stroke patients despite a very busy schedule.

After a collaborating company ceased development on their product, which this thesis was highly dependent on, I was very fortunate that Professor Troels W. Kjær joined the project and proposed a collaboration with ZUH. It was a warm experience to have a workplace created on a short notice and to be welcomed and included by staff from ZUH during the stroke related projects; special thanks to Department Chairman, Jesper Gyllenborg and Head Medical Secretary Tanja W. Bo. for enabling this.

I am grateful for the many hours spent with Associate Professor Sadasivan Puthusserypady and Associate Professor Thomas Sams on philosophical discussions related to research and life in general. PhD students Darius Adam Rohani, Alexander Neergaard Olesen and Jakob Skadkær Møller always found the time for reviewing and supplying constructive feedback that improved the documentation and research quality. Mehrdad Khodaverdi and Samra Skrijelj furthermore deserve sincerest thanks proofreading parts of the dissertation, and for their support when it was needed.

Sincere thanks go to the many students for whom I was a teaching assistant, but also to the unfamiliar ones who had been recommended to consult me for guidance. Although time demanding, they taught me irreplaceable and valuable methods of communication and introduced me to interesting problems that expanded my knowledge in areas outside my own research topic.

Finally, I would like to thank my family and friends for their understanding and support even though they have been greatly neglected on many occasions. Especially my brother, Kenan Vilic, deserves my deepest appreciation and utmost gratitude for countless hours spent on motivating, helping and guiding me through the academic career. It is through his selfless sacrifices that I learned how one can always acquire new knowledge from helping others.

ABSTRACT

In current clinical practice, vital signs such as heart rate, blood pressure, oxygen saturation level, respiratory rate and temperature are continuously measured for critically ill patients. Monitored by medical devices, each vital sign provides information about basic body functions and allows medical staff to intervene if health deteriorates. It has been documented that most of the alarms provided by the devices do not require actions, and that this occurs mainly because the signals are treated individually without context. The overload in alarms forces medical staff to make priority decisions, and can cause critical scenarios leading to a patient's death be overseen. The focus of this project was investigating clinical applicability of combining vital signs for critically ill patients. Several approaches were developed and tested with increasingly homogeneous patient groups.

The first study presents a data-driven approach to representation of a patient's physiological condition by combining vital signs into Early Warning Scores (EWS). Data were collected for 57 critically ill patients who had each been admitted to the intensive care unit at Bispebjerg Hospital for several days. To evaluate the estimation of physiological condition, text-based electronic health records (EHR) were collected, and time-labeled entries were extracted through algorithms from Natural Language Processing (NLP). The combination of EWS and NLP enabled the development of a system which could present and quantify a physiological condition timeline for patients. Promising results were obtained with EWS as measure, in which patients with $EWS \geq 8.5$ passed away while all patients who were admitted for over 53 hours with $EWS < 6.5$ survived.

The second study focused on ischemic stroke patients at Zealand University Hospital. Since all patients had same cause of admission and similar comorbidities, they were a more homogeneous critical patient group than in the first study. To predict the degree of disability after one day of admission, features based on vital signs and medical history were used in two prediction models. An introduced queue-based multiple linear regression (qMLR) model achieved best results with a root mean square error (RMSE) of $RMSE = 3.11$ on a Scandinavian Stroke Scale (SSS) where degree of disability ranged from 0 - 46. Worse outcomes were observed in patients who had pulse > 80 and a negative correlation between systolic and diastolic blood pressures during the first two hours of admission.

The final study dealt with classification of diabetes mellitus (DM) in ischemic stroke patients, where current findings indicate that one third of patients have unrecognized DM. A support vector machine was trained using vital signs and medical history, and correctly classified whether patients had DM with an accuracy of 87.5%.

The overall conclusion is that vital signs have high potential in applications for critically ill patients. Context-awareness through grouping with existing admission data is a prerequisite, unless vital signs are used to detect a specifically defined pathological events.

RESUME

I nuværende klinisk praksis bliver vitale parametre som hjerterytme, blodtryk, blodets iltmætning, vejrtrækning og temperatur målt kontinuert for kritisk syge patienter. Hver vital parameter giver information om basale kropsfunktioner og måles ved hjælp af medicinsk udstyr, som muliggør, at personale griber ind hvis patients helbred forværres. De fleste af alarmerne som sendes fra udstyret kræver ikke handlinger, og forekomsten skyldes, at parametre betragtes enkeltvis og uden kontekst. De mange alarmer tvinger medicinsk personale til selv at prioritere situationer, hvilket i værste fald kan medføre at kritiske tilstande overses og en patient dør. Projektets fokus har været at undersøge den klinisk anvendelse og kombinationer af vitale parametre i forbindelse med kritisk syge patienter.

Det første studie introducerer en datastyret tilgang til at repræsentere en patients fysiologiske tilstand ved at kombinere vitale parametre i Early Warning Scores (EWS). Data blev indsamlet for 57 kritisk syge patienter som har været indlagt på intensivafdelingen på Bispebjerg Hospital i flere dage. For at vurdere estimatet af fysiologisk tilstand, blev tekstbaserede sundhedsjournaler indsamlet for alle patienter, og tidsmarkerede registreringer blev udtrukket ved brug af algoritmer fra Natural Language Processing (NLP) feltet. Kombinationen af EWS og NLP gjorde det muligt at udvikle et system som kunne vise og vurdere en patients fysiologiske tilstand på en tidslinje. Lovende resultater blev opnået med EWS som enhed, hvor patienter med en endelig EWS ≥ 8.5 døde, mens alle patienter som var indlagt i over 53 timer med EWS < 6.5 overlevede.

Et andet studie fokuserede på patienter med iskæmiskslagtilfælde som var blevet indlagt på Sjællands Universitetshospital i Roskilde. Her havde alle patienter samme indlæggelsesårsag, samt lignende komorbiditeter, som gjorde dem til en mere homogen gruppe af kritisk syge patienter end i det første studie. For at forudsæ graden af handicap efter den første indlæggelsesdag, blev deskriptorer baseret på vitale parametre og medicinsk historik, anvendt i to prædiktionsmodeller. Bedste resultatet med opnået med en introduceret kø-baseret multipel lineær regressionsmodel (qMLR) som opnående en RMSE = 3.11 (kvadratroden af middelværdien af de kvadrerede fejlskøn), på Scandinavian Scale hvori graden af handicap kunne gå fra 0-46. Dårligere udfald blev observeret i patienter med en puls > 80 , og en negativ korrelation mellem systolisk og diastolisk blodtryk i løbet af de første to timer af indlæggelsen.

Det sidste studie beskæftigede sig med klassifikation af diabetes mellitus (DM) hos patienter som var indlagt grundet iskæmiskslagtilfælde, hvor nuværende fund indikerer at en tredjedel af patienter med iskæmiskslagtilfælde har DM uden at vide det. En support vector machine blev trænet, og kunne ved brug af vitale parametre samt medicinsk historik korrekt finde om en patient havde DM i 87.5% af tilfælde.

Den overordnede konklusion er, at vitale parametre har stort potentiale til anvendelse i forbindelse med kritisk syge patienter. Det er en forudsætning at parametrene bruges i kontekst-baseret sammenhæng hvori øvrig indlæggelsesrelevant data bør indgå, medmindre man ønsker at detektere specifikke patologiskdefinerede begivenheder.

ABBREVIATIONS

AMI	Acute myocardial infarction
ANS	Autonomic Nervous System
BMI	Body Mass Index
BP	Blood Pressure
bpm	Beats per minute
DBP	Diastolic Blood Pressure
CIS	Critical Information System
CNS	Central Nervous System
CT	Computed Tomography
DBP	Diastolic Blood Pressure
DM	Diabetes Mellitus (commonly referred to as diabetes)
DSL	Danish Language- and Literature Society
DTU	Technical University of Denmark
EHR	Electronic Health Records
EWS	Early Warning Score(s)
GUI	Graphical User Interface
HbA1c	Glycated hemoglobin (Hemoglobin A1c)
ICU	Intensive Care Unit
LASSO	Least Absolute Shrinkage and Selection Operator
LOS	Length of Stay
MLR	Multiple Linear Regression
mmHg	Millimeter of Mercury
MRI	Magnetic Resonance Imaging
mRS	Modified Ranking Scale
NIH	National Institutes of Health
NIHSS	National Institutes of Health Stroke Scale
NIV	Non-Invasive Ventilation
NLP	Natural Language Processing
PCA	Principal Component Analysis
PCT	Patient Condition Timeline
qMLR	Queue-based Multiple Linear Regression
RMSE	Root Mean Squared Error
RR	Respiratory Rate
rt-PA	Recombinant Tissue Plasminogen Activator
SBP	Systolic Blood Pressure
SITS	Safe Implementation of Thrombolysis in Stroke
SITS-MOST	SITS Monitoring Study
SSS	Scandinavian Stroke Scale
SU	Stroke Unit
SVM	Support Vector Machine
TIA	Transient Ischemic Attack
VIEWES	VitalPAC Early Warning Scor
ZUH	Zealand University Hospital, Roskilde

CONTENTS

Motivation	1
1.1 Project Description	2
1.2 Scientific Contributions	3
1.3. Research Approvals	3
Physiology	5
2.1 Vital signs	5
2.1.1: Heart Rate and Pulse	6
2.1.2: Blood Pressure	6
2.1.3: Temperature	7
2.1.4: Respiratory Rate	7
2.1.5: Oxygen Saturation Level	8
2.2 Stroke	8
2.3 Diabetes Mellitus	11
Technical Aspects	14
3.1 Correlation Coefficient	14
3.2 K-fold Cross-Validation	15
Automatized Physiological Condition Assessment of Critically Ill Patients	18
4.1 Background	18
4.1.1 Research Hypothesis	19
4.1.2 Approach Overview	20
4.2 Data	21
4.3 Methods	22
4.3.1 Early Warning Scores	22
4.3.2 Natural Language Processing	25
4.3.3 Combining EHR and EWS	28
4.4 Results	29
4.4.1 Patient Condition Timeline	29
4.4.2 Relationship between Mortality and EWS	30
4.4.3 Mortality as Function of Other Parameters	31
4.4.4 Summarizing EHR Through Word Frequency Distribution	32
4.4.5 Event Detection Using EWS	34
4.5 Discussion	34
4.6 Conclusions	36
Detection of Parameters Involved in Ischemic Stroke Outcome	39
5.1 Background	39
5.1.1 Research Hypothesis	40
5.1.2 Approach Overview	41
5.2 Data	41
5.2.1 Validation	43
5.3 Methods	45
5.3.1 Choosing Stroke Outcome	45
5.3.2 Predicting Outcome through Regression	47
5.4 Results	51
5.5 Discussion	54
5.6 Conclusion	54
Automatic Identification of Diabetes Mellitus in Ischemic Stroke Patients	57
6.1 Background	57
6.1.1 Research Hypothesis	58
6.1.2 Approach Overview	58
6.2 Data	59
6.2.1 Intervariable Comparison for DM and Non-DM	61
6.3 Methods	63

6.3.1 Features.....	63
6.3.2 Classification through Support Vector Machines	64
6.4 Results	66
6.5 Discussion.....	68
6.6 Conclusion	70
Conclusion.....	72
7.1 Future Perspective	73
References	75
Appendix A – Stroke Outcome Scales	83
Appendix B – Publication in Proceedings of the 38 th Annual International Conference of the IEEE Engineering in Medicine and Biology Society	88
Appendix C – Publication in IEEE Journal of Biomedical and Health Informatics	94
Appendix D – Publication in Proceedings of the 39 th Annual International Conference of the IEEE Engineering in Medicine and Biology Society	103
Appendix E – Submission to Proceedings of the 39 th Annual International Conference of the IEEE Engineering in Medicine and Biology Society	103
Appendix F – Submission to Journal of Translational Engineering in Health and Medicine	112

MOTIVATION

With an increasingly older population and a shrinking workforce due to declining birthrates, there will be a period in which resources need to be streamlined. By the year 2040, the Danish population will have half as much workforce while there will be more than twice as many people older than 65 years [1]. In addition, the prevalence of chronic diseases is increasing due to unhealthy lifestyle. E.g. in the US, 86.9% of people older than sixty-five had at least one chronic disease in 1998. By 2008 prevalence increased to 92.2% [2]. Especially cardiovascular complications such as heart failure, hypertension, and diabetes mellitus (DM) are expected to be more prominent [3]. DM alone is expected to rise by at least 60% depending on country by the year 2030 [4] [5] [6] [7].

Cardiovascular and other medical problems can be revealed through vital sign measurements. The measurements are usually performed unobtrusively without invasive equipment. The most common vital signs are pulse, blood pressure, temperature and respiratory rate. Depending on geographical location, the list is further expanded to include oxygen saturation level, consciousness level and urine output over time [8]. In developed countries, most of these vital signs are measured manually by nurses as part of the routine by nurses, to prevent sudden mortality. Since the process is resource costly, the measurement schedule is currently optimized so that measurements are done only as frequently as necessary depending on how sick the patient is [9] [10]. For bedridden patients, bedside vital sign monitors are available in hospitals since the nineteen-sixties, and the modern screens that display vital signs of a lying patient were deployed in 1977 [11]. Since then, the main advancements in technology are better sensors, interfacing with other hospital equipment, and easier interpretable graphics. Recent advancements in wearable technologies have enabled vital signs monitors to be portable. Once wearable devices become widely available at hospitals, essential surveillance tasks will become automated and medical staff efficiency can potentially improve.

Regardless which type of vital sign monitoring device is used, built-in alarming only notifies staff when a single vital sign reaches an abnormal value. This occurs so often in emergency settings, that only one in twenty audible alarms per bed turns out to be false – with the main contributor being abnormal heart rhythms [12]. This simple classification of abnormality persists because decisions need to be taken rapidly, and it is impossible for medical staff to take all available information about the patient into account. Nearly all decisions are therefore based on single parameter measurements, and less evident patterns and interactions between parameters are often overlooked.

1.1 PROJECT DESCRIPTION

This project investigated the feasibility of combining multiple vital signs and other patient information to evaluate the progress of critically ill patients throughout their hospital admission. Other medical information such as health records, medical history and medication were also included to supply additional information which may improve evaluation. Three retrospective studies were conducted in which automatic methods and algorithms were developed and implemented to objectively evaluate usefulness of vital signs monitoring in critical settings.

The first study involved uniting recorded vital signs with electronic health records (EHR) from the Intensive Care Unit (ICU) at Bispebjerg Hospital. A timeline over each patient's health throughout the admission was modeled by combining vital signs into a single value for any given time. Natural language processing (NLP) methods were implemented for relevant data extraction and analysis from textual EHR. This allowed direct mapping of EHR entries over the health condition timeline, and enabled evaluation of the health condition timeline model. Chapter 4 describes the study in detail.

The population of the first study was highly diverse, with patients having varied reasons for admission and receiving different treatments. The second study was therefore conducted on a more homogeneous population of critically ill patients, namely the ones treated for ischemic stroke at the Stroke Unit (SU) at Zealand University Hospital (ZUH) in Roskilde. All patients from the population were admitted for same reasons and received the same treatment, but were initially different in terms of background and severity of stroke. Sudden health deteriorations rarely occur in SU settings, for which reason the focus of the study was on modelling the outcome as function of a disability score, which was continuously assessed by medical staff. Chapter 5 elaborates on the study.

The third study continues with the ischemic stroke population. The available data is combined to identify which patients are diagnosed with diabetes mellitus (DM). It aimed at demonstrating an application of how the combination of vital signs with other data can contribute to detecting patterns in critically ill patients. Significant differences between DM and non-DM patients could furthermore contribute to improving detection of unrecognized DM, allowing for earlier medication and improved life quality. Chapter 6 explains the study in detail.

Objectives

The primary objective of this thesis was to investigate feasibility of using vital signs monitoring in critical settings to investigate patient outcome in critical settings. Overall objectives are:

- Investigating clinical usability of combining multiple vital signs in critical settings to estimate the health condition of a patient at any given time.
- Determining how vital signs are related to degree of disability in a homogenous patient group such as patients treated for ischemic stroke.
- Examining feasibility of using vital signs and personalized data as identifiers for diabetes mellitus in stroke unit patients.

1.2 SCIENTIFIC CONTRIBUTIONS

The main scientific findings throughout the study have been presented in two journal papers and three conference papers. Some findings are not covered in the written articles, but are instead included in the dissertation. All publications have been attached as appendices.

[accepted] A. Vilic, J.A. Petersen, K. Hoppe, and H. B. D. Sorensen, "Visualizing patient journals by combining vital signs monitoring and natural language processing". *IEEE 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2529-2532), 2016.

[accepted] A. Vilic, K. Hoppe, J.A. Petersen, T.W. Kjær, and H.B.D. Sorensen, "Simplifying EHR Overview of Critically Ill Patients Through Vital Signs Monitoring". *IEEE Journal of Biomedical and Health Informatics*, 2016.

[accepted] A. Vilic, J.A. Petersen, T. Wienecke, T.W. Kjær, and H.B.D. Sorensen, "Queue-Based Modelling and Detection of Parameters Involved in Stroke Outcome". *IEEE 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017.

[accepted] A. Vilic, T. Wienecke, T.W. Kjær, and H.B.D. Sorensen "Automatic Identification of Diabetes Based on Vital Signs and Medical History in Post-Stroke Patients". *IEEE 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017.

[submitted] A. Vilic, J.A. Petersen, H.B.D. Sorensen, T. Wienecke, and T.W. Kjær, T, "Identifying Diabetes Patients in Post-Stroke Settings by Combining Vital signs and Medical History". *IEEE Journal of Biomedical and Health Informatics*, 2017.

1.3. RESEARCH APPROVALS

All data were collected with the approval from national health authorities. The author contributed actively in all steps of acquiring the necessary authorizations. Following references are related to the conducted research.

- Danish Data Protection Agency
 - Journal: 2014-41-2743
 - Journal: 2008-58-0020

- Danish Health and Medicines Authority:
 - Case: 3-3013-558/1.
 - Case: 16-000032 (Application through Region Zealand).

PHYSIOLOGY

With the focus of the dissertation being on vital signs monitoring of critically ill patients, this section starts by introducing vital signs, how they are monitored, and what measurements one can expect. The section also briefly describes stroke and diabetes mellitus which were the two primary diseases dealt with throughout the project.

2.1 VITAL SIGNS

Vital signs are physiological measurements of body functions that can be measured non-invasively. The most common vital signs are heart rate or pulse, blood pressure (BP), temperature and respiratory rate (RR). Depending on the medical settings and geographical location, other vital signs may include consciousness or pain, oxygen saturation level, urine output and glucose level [9] [13] [14] [15] [16] [17] [18].

Most vital signs are regulated by the hypothalamus, which is a small structure within the brain, and is controlling the autonomic nervous system (ANS). ANS controls autonomic organ functions such as heart rate and subconscious respiratory rate [19]. Aside health condition, vital signs are dependent on many factors, including age and physical condition and gender. One example being the heart rate of a newborn child of 0-3 months is 100 – 150 at rest, while it is 60 – 100 for a people older than ten years [20]. In trained athletes, the heart is enlarged and can consequently pump more blood, leading to a heart rate of 30 – 60 [21]. There is also a strong intervariable relationship between vital signs in healthy people. Exercising increases breathing (respiratory rate), because oxygen is required by muscles during cellular respiration. The necessary oxygen is acquired through breathing, and enters the blood stream through tiny sacks in lungs and finally goes to the heart. From there it is transported with blood around the body at a rate that is partly decided by how fast blood is pumped by the heart (pulse). When muscles contract, their metabolic rate increases which causes heat production and consequently a rise in body temperature.

Hospitals located in the capital region of Denmark monitor five objectively measurable vital signs in all wards, along with state of consciousness, which is evaluated subjectively by medical staff: Pulse, BP, Temperature, RR and Oxygen Saturation Level. The vital signs to monitor were chosen based on a study by Prytherch et al. The study observed nearly two hundred thousand patients and defined a mortality prediction model for patients in general, named Vipac Early Warning Scores (ViEWS) [10] [15] – Chapter 4 describes this model in detail.

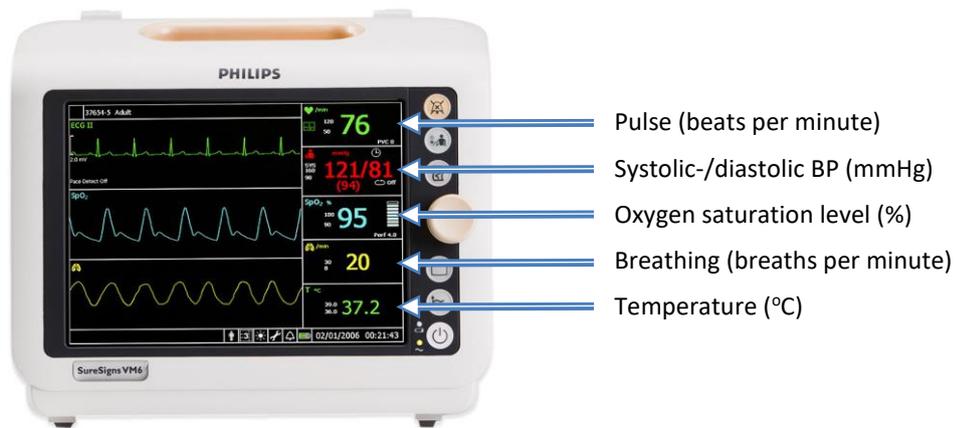


Figure 1: Philips SureSigns VM6 is an example of a commonly used bedside monitor in hospital settings, allowing hospital staff to be alert of abnormal development in vital signs. The type is commonly used after certain surgeries and in emergency departments such as the intensive care unit or stroke unit [22].

Each vital sign subsection includes a table that is aligned to the right of the textual description, and defines what is considered as a normal range of the vital sign in hospital settings. The scores are targeted towards adults and defined by the previously mentioned study [15], where a lower score represents a healthier patient.

2.1.1: HEART RATE AND PULSE

Heart rate is typically measured on the chest in beats per minute (bpm), and signifies at which rate the heart makes contractions and pumps blood throughout the system. Pulse is also measured in bpm, but is measured over large arteries such as the radial artery at the wrist or carotid artery at neck. In addition to heart rhythm it can also indicate the strength of blood flow, because arteries expand at each pulse.

Pulse can reach up to 270 bpm, in which case the occurrence is caused by supraventricular tachycardia due to a rare structural abnormality of the heart [23].

Table 1: Pulse abnormality ranges according to ViEWS [15]

Pulse vs abnormality score	
Pulse	Score
< 41	3
41 – 50	1
51 – 90	0
91 – 110	1
111 – 130	2
> 130	3

2.1.2: BLOOD PRESSURE

When the heart contracts during systole, blood is released into arteries causing them to expand due to a pressure that is measured in millimeter of mercury (mmHg). This is called systolic blood pressure (SBP). Diastolic blood pressure (DBP) measures pressure when blood returns to the heart as it relaxes during diastole. When a person has hypertension, it can be due to several reasons. Arterial stiffness or arteriosclerosis for instance can cause arteries to lose elasticity and a hardening of arterial walls, therefore requiring greater pressure for arteries to expand and blood to get around [24]. DBP is not included in

Table 2: SBP abnormality ranges according to ViEWS [15]

Abnormality score vs SBP	
Systolic BP	Score
< 91	3
91 – 100	2
101 – 110	1
111 – 219	0
> 219	3

the ViEWS model and is therefore not scored, although the relationship between SBP and DBP may be relevant for patient outcome [25].

The gold standard for measuring BP continuously and precise is a cannula that is inserted into an artery, while the non-invasive method is an inflatable cuff that is placed on the upper arm. The cannula is connected to a disposable system that also delivers a saline solution to prevent thrombosis (blood clotting) and thereby occlusion during measurements.

2.1.3: TEMPERATURE

The hypothalamus is also responsible for regulation of body temperature and ensuring homeostasis. It is responsible for causing shivering when the body is cold or sweating in heat. To make the body inhospitable for invading pathogens, it can also elevate the core body temperature to the state of fever.

Temperature is the vital sign with least agreement on standard measurement site. A literature review found that studies use at least eleven different sites for monitoring temperature with one of eight different devices. The device list includes internal probes, catheters and thermometers. The most popular sites were oral, axillary, tympanic and rectal. Oral measurements were found to be most unreliable (error of 0.3°C) and affected by food consumption unless monitored twenty-minutes after consumption. Tympanic measurements using Infrared thermometers have become the standard due to the simplicity and speed of the device even though they are less accurate than rectal measurements (error between 0.1°C – 0.3°C depending on occlusion of ear).

Table 3: Temp. abnormality ranges according to ViEWS [15]

Abnormality score vs Temp.	
Temperature	Score
< 35.1	3
31.2 – 36.0	1
36.1 – 38.0	0
38.1 – 39.0	1
> 39	2

Body temperatures lower than 24°C or greater than 44°C result in either death or serious conditions such as cardiorespiratory collapse and brain damage [18].

2.1.4: RESPIRATORY RATE

Respiratory rate (RR) is measured as the number of breaths that a person takes within a minute. The most common approach to measuring it is by looking at a patient for about fifteen seconds, seeing how often the chest rises, and multiplying by four. Continuous RR monitoring is done with capnography where CO₂ exhales are traced through nasal airways. The approach is somewhat obtrusive, and less invasive alternatives have been tested for decades such as deriving RR based on ECG signals. Results are promising in healthy subjects, but the technology is still immature for real patient admissions where arrhythmia and other cardiovascular complications exist [26] [27] [28].

Table 4: RR abnormality ranges according to ViEWS [15]

Abnormality score vs RR	
RR	Score
< 9	3
9 – 11	1
12 – 20	0
21 – 24	2
> 24	3

A person is assisted by a respirator machine in hospital settings if they are unable to breathe by themselves, so that breathing becomes normal and constant.

2.1.5: OXYGEN SATURATION LEVEL

Oxygen is carried in the blood by being attached to hemoglobin molecules. Each hemoglobin molecule can carry up to four oxygen molecules. Oxygen saturation level is a measure of how much oxygen the blood is carrying as a percentage of the maximum it can carry. It is measured with a pulse oximeter that is most often attached to the index finger. It works through measuring intensity of the color of blood which conveys how much oxygen is carried. There is no specific boundary for which level is fatal as the body can adapt to a slow decrease while a

sudden drop from for example over 90% to 70% can result in death. Circumstances need to be kept in mind when measuring with pulse oximeters though, because inaccurate readings are common depending on the person's skin pigmentation, limb temperature or strength of the pulse [29].

Table 5: O₂ sat. level abnormality ranges according to ViEWS [15]

Abnormality score vs O ₂ sat. %	
O ₂ Satur. (%)	Score
< 92	3
92 – 93	2
94 – 95	1
> 95	0

Oxygen support is given when diseases or other conditions prevent the lungs in receiving sufficient air. This is done either through a nasal cannula, a facemask or tube that is placed into the windpipe after an incision.

2.2 STROKE

Stroke occurs when brain blood flow is compromised. The disturbance in blood supply to the brain frequently causes irreversible damage, making stroke the leading cause of disabilities in adults. Although stroke primarily affects elder people, it is a disease of all age groups depending on factors such as lifestyle, medical history and genetics. There are approximately twelve thousand incidents of stroke in Denmark on an annual basis of which one third are reoccurrences. Of all successfully treated patients, one third have no permanent injuries. According to chief physicians at ZUH, half of the remaining patients have mild disabilities, while the other half are dependent on help from others for basic daily functions such as movement, eating or clothing. Up to 3,000 people die annually in Denmark due to the disease, and there are up to 40,000 annually living with disabilities because of it [30]. Symptoms of stroke may be language difficulties, visual loss, loss of motoric functions and numbness. The symptoms heavily depend on where the stroke occurs and are noticeable in seconds to hours after onset. Untreated, two million neurons die every minute until brain blood flow is restored [31].

There are two main categories of stroke.

- The most common type is ischemic and occurs in at least 80% of all stroke incidents. It happens because blood vessels are clogged due to a formed blood clot (thrombus) that prevents blood supply to tissue (see Figure 2). The cause is often atherosclerosis which is a specific type of the previously mentioned arteriosclerosis, that is related to increase in BP. The lack of blood supply results in tissue death (infarction) in the part of the brain that is no longer receiving necessary blood. A subtype of ischemic stroke is a transient ischemic attack (TIA) that passes by itself within minutes to hours without any treatment, but is important to be aware of, because 40% of TIA patients eventually have an ischemic stroke requiring treatment [32]. Treatment is done with injection of intravenous recombinant tissue plasminogen (rt-PA), also known as alteplase, to dissolve the thrombus. The dosage guideline is based on the patient's weight, and is generally 0.9 mg/kg administered over an hour [33].

It must be administered within a time span of 4.5 hours after onset. Reducing the dosage to 0.6 mg/kg can insignificantly decrease mortality but in exchange for increased disability [34] [35]. A method called “drip and ship” is commonly used if the hospital lacks capacity, or the patient needs surgical removal of the thrombus through thrombectomy. The purpose of this method is to ensure that alteplase is administered timely in a nearby hospital before the patient is moved to the hospital with capacity for treatment [36]. This also applies to ZUH, which is why data was not available for some patients that were initially randomly selected to be included in the study.

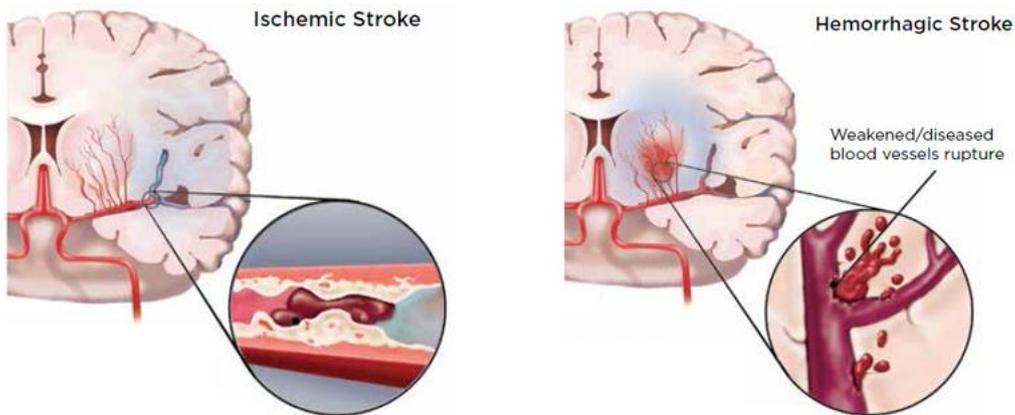


Figure 2: Ischemic stroke and hemorrhagic stroke. Ischemic stroke (left figure) occurs when thrombus prevents supply to brain tissue. Hemorrhagic stroke (right side) is caused by leaking or ruptured blood vessels that cause blood to flow freely in spaces within the brain [37].

- The other type is hemorrhage, and it is caused by blood uncontrollably filling space surrounding the brain due to rupture in arteries or damage to the innermost layers of the skull. Once the blood has left the circulatory system, it clots and brain tissue will be damaged because of increased pressure on the tissue. Although there is no treatment yet for hemorrhagic stroke, it is important to do a brain scan using computed tomography (CT) scan or magnetic resonance imaging (MRI) to know which type of stroke a patient has. Administering alteplase to a hemorrhage patient would prevent ruptured arteries from clotting again, and bleeding would continue.

Figure 3 shows the procedure diagram for stroke treatment at hospitals in Denmark. While the ambulance transports the patient to the hospital, the stroke unit (SU) is informed about who the patient is and which symptoms they exhibit. The SU prepares staff for arrival and reviews available medical history about the patient. After arrival, lab results are collected for important indicators such as blood glucose level, and a neurological examination is performed to estimate the extent of the stroke. Throughout the examination, the patient is also prepared for a CT scan during which the exact location and size of infarction/hemorrhage become visible. Sedatives may need to be administered if the patient is unable to lie still due to for instance shivering. Depending on CT findings, treatment is started according to previously mentioned procedures. In the first twenty-four hours of admission, the patient’s vital signs and neurological functions are closely monitored, and the patient is afterwards transferred to a ward. Three months after discharge, independence of the patient is evaluated to assess the degree of disability. The assessment of disability, progress and neurological functions throughout and after admission are assessed on different scales that are all described in detail in Appendix A.

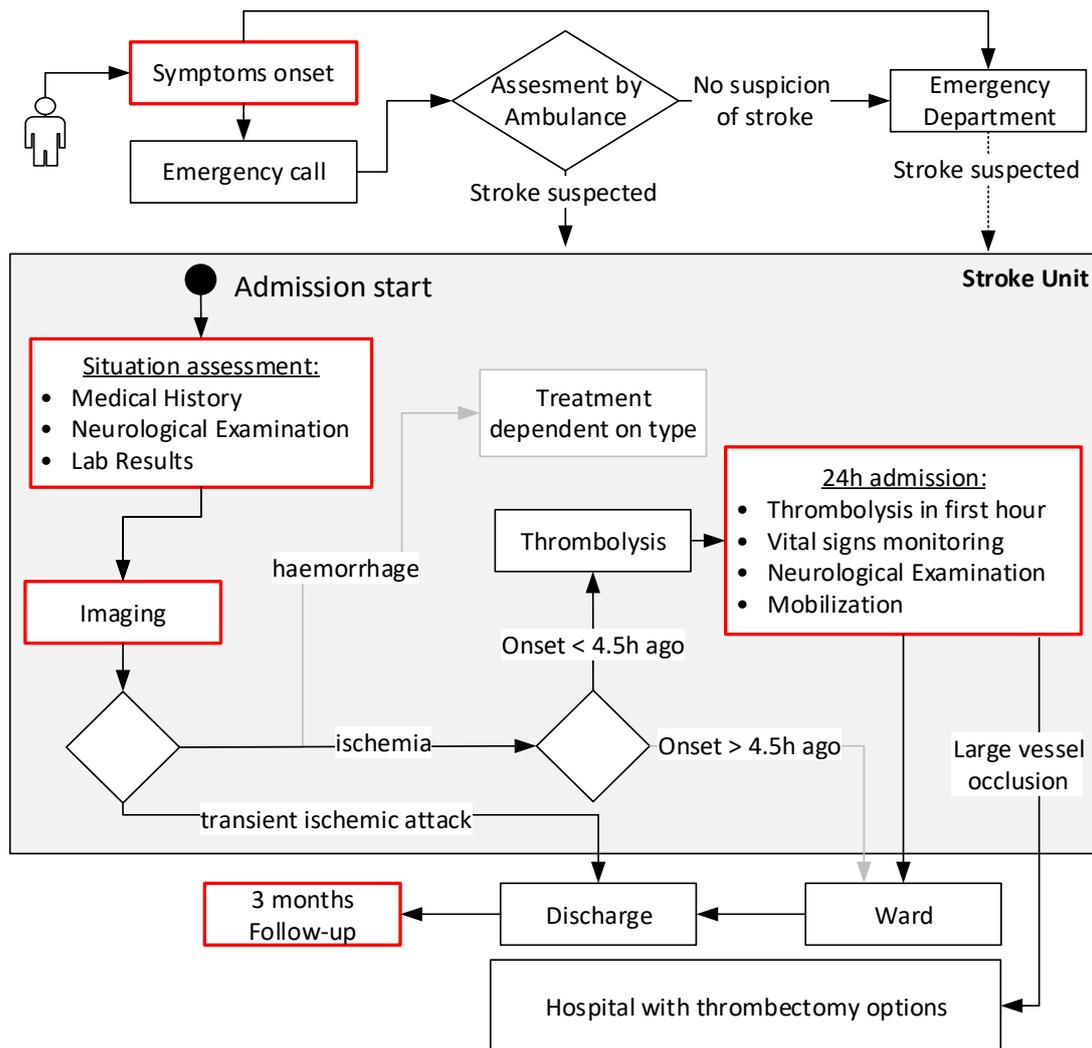


Figure 3: Admission diagram for stroke treatment in Danish hospitals. Going from left to right, the patient or relatives will first emergency dial the health authorities for help, and await an ambulance. The paramedics will do on spot treatment and transport the patient to an emergency department of stroke unit (SU). At the SU, the patient is examined and treatment is started accordingly. The patient is followed up on three months after treatment. Red marks in the diagram indicate quantifiable information that is stored as part of the treatment.

Stroke disrupts the ANS and affects vital signs. Atherosclerosis is the main cause for hypertension in stroke and is caused by the damage to blood vessels that eventually led to the incident. Body temperature in the acute phase also changes. A rise in temperature is often observed following stroke, but the cause is unclear, except for 25-35 % of cases where infection occurs. Normally temperature rises during daytime and declines at the night, as part of a process called circadian rhythm that is controlled by the hypothalamus. Because of affected consciousness and physical inactivity, this rhythm is disrupted and the temperature pattern is changed. The blockage in arteries also deprives tissue of necessary oxygen, and oxygen saturation below 90% has been observed for over 10% of admission time in 20% of patients [38].

2.3 DIABETES MELLITUS

Diabetes mellitus (DM) is a chronic metabolic disease that over time leads to many comorbidities which decrease life quality and increase risk of mortality. The number of diabetics in Denmark almost tripled from 114.000 in 1996 to 320.000 people in 2012. As life expectancy is increasing, so is the prevalence of the disease due to lifestyle factors such as unhealthy diets and lack of exercise, along with better management of comorbidities [3] [4] [5] [6] [7] [39].

In the process of digestion, carbohydrates are transformed into glucose which circulates the blood stream and is absorbed by cells in the body. For cells to absorb glucose efficiently, the pancreas needs to create insulin hormones. In DM, the glucose is not absorbed by cells, it accumulates in the blood stream causing a state of hyperglycemia, that eventually leads to complications of which the most severe are damage of the cardiovascular system, kidneys and nerves. Due to slow progression, the disease may go unnoticed for several years and immediate glucose tests may not reveal it. A potentially more accurate acute estimate of DM status is glycated hemoglobin in the blood (HbA1c), which reflects blood glucose over several prior weeks [40] [41]. There are two different types of DM. In the first type, the pancreas is unable to produce sufficient insulin. In the second type, the produced insulin is ineffective due to less sensitivity of the insulin receptors on the cell surface [42]. The second type, Type 2, accounts for about ninety percent of cases, and is to some extent preventable through a healthy lifestyle. Patients with this type are also more prone to peripheral arterial disease, large-artery atherosclerosis and stroke [40] [43] [44] [45] [46] [47]. People with DM are 2.27 more likely to suffer from ischemic stroke and 1.56 more likely to suffer hemorrhagic stroke [48].

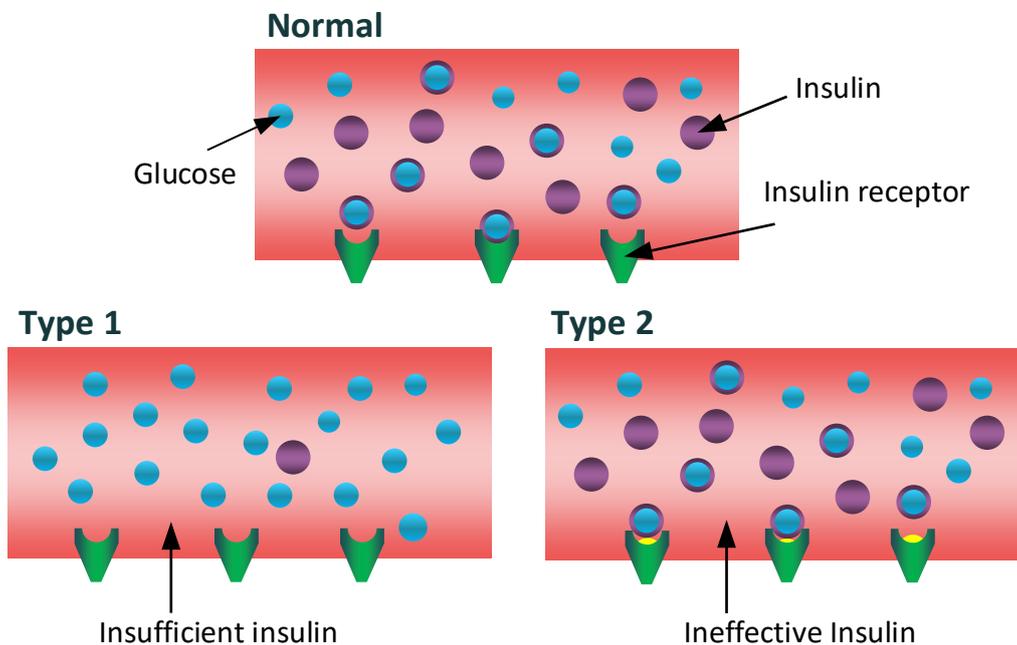


Figure 4: In the normal state (top), pancreas produces sufficient effective insulin for cells to be able to absorb glucose. Diabetes Mellitus Type 1 (left) occurs when the pancreas is unable to produce sufficient insulin to absorb glucose. In Type 2 (right), pancreas produces sufficient insulin, but it is ineffective. When glucose is not absorbed by cells, it accumulates in the blood stream and eventually leads to damage of the cardiovascular system, kidneys and nerves.

Vital signs are also affected in people with DM as they are with stroke when it comes to the cardiovascular system. Diabetic neuropathy adds to the problem by being a common disorder that impairs ANS functions, and can cause vital signs to change slower during for example change in body position [49] [50]. A study found that the oxygen saturation of DM patients need to be monitored through blood gas analysis, as standard pulse oximetry becomes more unreliable the greater HbA1c is [51]. Respiratory rate is also affected because lung functions are reduced, requiring more breaths to supply the body with sufficient oxygen [52].

TECHNICAL ASPECTS

This chapter briefly introduces theory of the two very basic but essential mathematical methods that reoccur throughout the dissertation. This first method, correlation coefficient, is commonly applied to determine relationship or similarity between two signals. The second method, cross-validation, is used to partition data in classification and prediction problems. It increases the likelihood of the final model working on unknown data, and not only the data it was trained on. The theory behind all other models and methods are described only in the relevant chapters where they are applied.

3.1 CORRELATION COEFFICIENT

The similarity or dependence relationship between two signals was estimated through Pearson's correlation coefficient. If two signals are equally long, the correlation coefficient is defined as:

$$r(S_1, S_2) = \frac{\sum(S_1 - \bar{S}_1)(S_2 - \bar{S}_2)}{\sqrt{(\sum S_1 - \bar{S}_1)^2} \sqrt{(\sum S_2 - \bar{S}_2)^2}} \quad (3.1)$$

S_1 and S_2 are the two respective signals, and the overline represents the given signal's mean value. The figure below illustrates four signals where the blue signal is original and the remaining ones are alterations of it in amplitude and shape. If the signal is correlated with itself and thereby identical, the correlation coefficient is $r = 1$. Comparing it with its inverse (orange) results in $r = -1$ regardless of the amplitude. The two remaining signals are fully random (purple), and semi-random in which every second sample follows the shape of the original (yellow), with respective correlation coefficients of $r = 0$ and $r = 0.5$. In the semi-random case, the coefficient will vary depending on which samples are replaced.

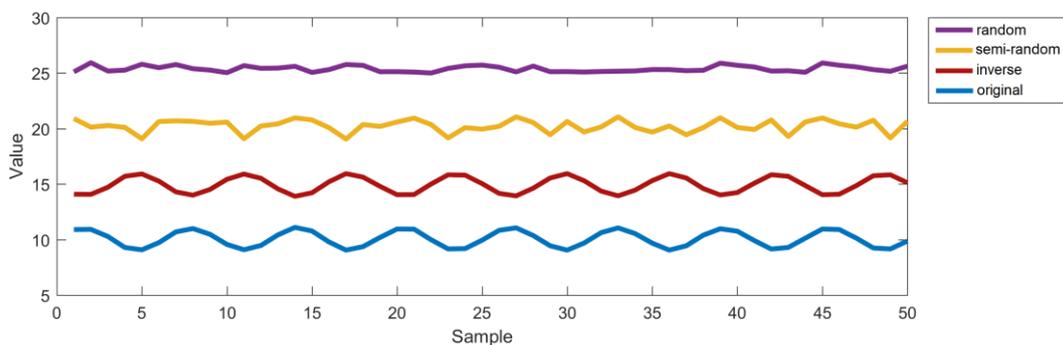


Figure 5: Signal resemblance illustration. Computing the correlation coefficient of the original signal (blue) with itself will result in $r = 1$. Comparing it to the inverse (red), the correlation is $r = -1$.

3.2 K-FOLD CROSS-VALIDATION

Models that predict and classify data are often at risk of being overfitted. The condition occurs when the underlying model is fitted to an extent where it performs with high accuracy on the existing data that it was trained on, but fails on unknown data. Figure 6 illustrates this in a two-dimensional space where samples belong to two different classes and arranged according to two features (p_1 , p_2). Each of the subfigures has a different separation model outlined in green. The left figure illustrates overfitting through a model that separates the data points perfectly during training, whereas the linear model in the right figure has lower accuracy because it ignores outliers. When the models are tested with new data, it is likely that new X-points are misclassified in the overfitted model because they are likely to appear in the void belonging to class 'O' due to outliers during training.

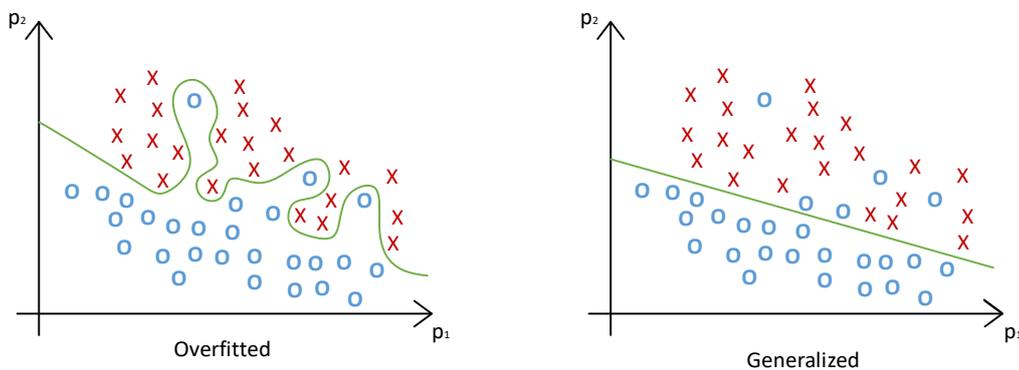


Figure 6: Overfitting (left) shows a polynomial approach where all points are correctly separated. The simpler model (right) has lower accuracy because it does not take outliers into account. If the overfitted model is tested with new data, X-points that resemble the outliers in class O are likely to be misclassified.

One way of avoiding overfitting is through a k-fold cross-validation. It is a data partitioning technique, where the entire data set is subsampled into k-number of folds. The desired fitting model is then trained as many times as there are folds. In each iteration, the subsample is used for validation while the remaining data from the entire set is used for training (see Figure 7). This results in k-number of fitted models where the true accuracy is the average of all models. The advantage of splitting the data in this manner is that every sample from the data set is used exactly once for validation.

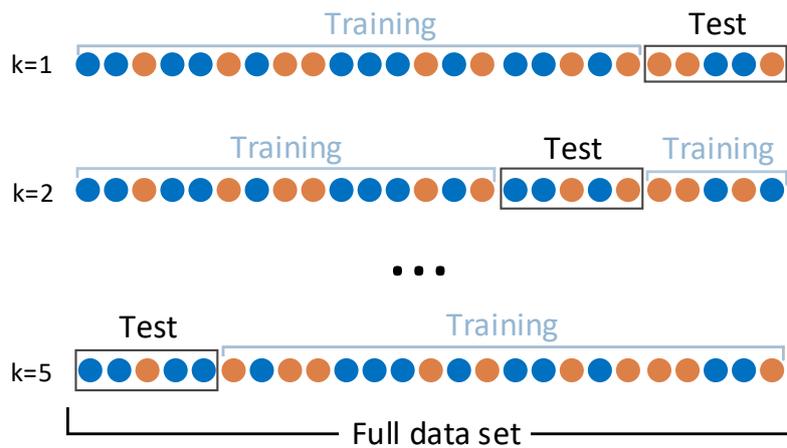


Figure 7: k-fold cross-validation. The dataset is first partitioned into random folds of equal size. A different fold is then used in each iteration as test set to validate the fitting model, until all samples from the dataset have been validated.

The number of folds can be from $k = \{2 \dots n\}$, where n is the number of samples in the dataset. For fitting models in this project, 5-fold cross-validation was applied to avoid overfitting because of a small dataset. In the third study (Chapter 6), the cross-validation was stratified, meaning that the randomness of fold generation was reduced in exchange for ensuring that all classes are well represented in each fold.

AUTOMATIZED PHYSIOLOGICAL CONDITION ASSESSMENT OF CRITICALLY ILL PATIENTS

Objective: Patients in the intensive care unit (ICU) are in severe and life-threatening conditions. They are more staff-demanding, require continuous monitoring of all body functions, and the most advanced equipment available. The patient is discharged once they are no longer in a physiologically critical condition. Condition is evaluated by physicians and the means for quantitative assessment have yet to be developed. A quantitative assessment exists outside the ICU in most wards through an Early Warning Score (EWS) model, which has contributed to more efficient monitoring and reduced mortality [53]. The model has yet to be investigated as for how well it is in represents a patient's condition. This study makes use of the high-quality data that is recorded in ICU settings to evaluate EWS at any given time and thereby create a physiological condition timeline. The EWS timeline can then be validated by being combined with the patient's electronic health record (EHR) through natural language processing (NLP) methods. Two scientific contributions have been published on this topic (Appendices B and C).

4.1 BACKGROUND

Intensive care unit (ICU) patients are constantly monitored with electronic devices, so any parameter changes that are sudden or deviate from healthy ranges will always alarm staff, allowing appropriate actions to be taken swiftly. These data are typically processed and stored in clinical information systems (CIS), which have proven effective for optimization of resources and decision making [54] [55] [56] [57]. The monitoring procedure however leads to hundreds of daily alarms per patient of which most are false and require no actions. Drew et al observed 72 beds at a hospital, of which 32 were at an ICU, over the period of a month to see why important deteriorations are sometimes not acted upon. Each patient generated 187 audible events per day in average of which 89% were false alarms [12]. Frequency or significance of these alarms could be reduced if multiple parameters are considered. This has been proven through telehealth projects where mortality in intensive care units was reduced by streaming data from the hospitals to outside facilities. Experienced employees at remote locations were then responsible for alarming staff whenever a patient deteriorated in several parameters [58] [59].

Automatized algorithms for alarming medical staff initially need to be similar to existing decision-making models that the staff is familiar with. A study aiming to detect patient deterioration employed logistic regression and included 36 variables, but performance or impact could not be verified because

staff did not act on alerts [60]. The first step towards quantitative assessment of a patient's physiological condition therefore needed to be a retrospective tool in which the output is based on existing models [14].

The main challenge in the development of algorithms for estimation of physiological condition is the feedback on performance. Unless patient condition scoring is standard practice, or more staff are employed during a project, it is unlikely that sufficient surplus in mental resources will be available to attend to critically ill patients and to remember scoring.

Since scoring is not part of standard practice in the ICU, the next best thing is to extract text from electronic health records (EHR), which are continuously being updated by medical staff throughout admission. Developed countries have strict regulations and laws to ensure that every institution creates, documents and stores all medically relevant information regarding a patient in form of an EHR [61]. The EHR consists of scans, test results, schemes, medication subscriptions, and a textual part containing details about the medical history, circumstances, events and so on. The textual parts are especially long for critically ill patients who are were admitted for multiple days, because their condition is often unstable and all changes need to be documented.

An average of four pages of pure text are generated per patient every day at the ICU at Bispebjerg Hospital, Denmark. EHR of patients who have been admitted for longer periods can therefore become almost overwhelming. Irrelevant and redundant information will be present in respect to automated condition assessment, but it is also later burdensome for medical staff to follow up on, if they for instances want to know how a patient responded to treatments during admission. A common way of dealing with this is to quantify textual data by applying Natural Language Processing (NLP) to do a text summarization. The overall aim is to reduce the presented information to a minimum by, for example removal of redundancy, or applying statistics to rank sentences [62] [63] [64].

This study approached the problem of quantifying physiological condition through the development of a data-driven tool, where vital sign measurements were combined with textual information from the EHR. The automatized tool displays a visual representation of a patient's physiological condition over time and maps it against time-labelled entries in their EHR. It enabled an alternative approach to navigate in the EHR, and quality checking documentation against assumptions of the patient's condition. Finally, once the model for evaluating a patient's condition is validated, it becomes possible to find EHR entries related to deterioration, causes and resolution.

Immediate changes in parameters are already handled by built-in alarm systems in medical devices. It was therefore of greater interest for the collaborating physicians at Bispebjerg Hospital to inspect progress in physiological changes over hours rather than instantaneously. A motivating factor for them was to detect sepsis which may be overseen by devices because it occurs more slowly; unless it's a septic shock [65]. The tool therefore focuses on progress of a patient's condition over hours and days.

4.1.1 RESEARCH HYPOTHESIS

Increased monitoring has effectively reduced mortality in hospital settings. The effect is either achieved because the models act as an insurance so that patients are not neglected, or because the models can approximately evaluate physiological condition. The objectives were therefore related to

quantitative assessment of physiological condition through Early Warning Scores (EWS) [9] [15] [66]. Our hypotheses were:

1. Early Warning Scores can evaluate development in physiological condition of a patient over time if vital signs are measured continuously. Higher mortality is expected for patients who score high in the model.
2. An overview of admission related topics, such as medication, complications, and cause of admission can be derived from individual electronic health records. This can be achieved by applying text summarization through natural language processing.
3. All physiological condition changes throughout an admission are registered in electronic health records (EHR). The EHR will therefore have data listed in the hour, or hours, when condition changes.

4.1.2 APPROACH OVERVIEW

The hypotheses were tested through an approach that consists of two parallel modules that are merged at the end, as illustrated in Figure 8. The top branch extracted and separated entries from the EHR after locating timestamps using NLP. The branch below was dedicated to estimating the patient's condition over time through processing of vital signs measurements. Lastly, an interactive timeline was created which showed physiological condition of patients along with markers of entries.

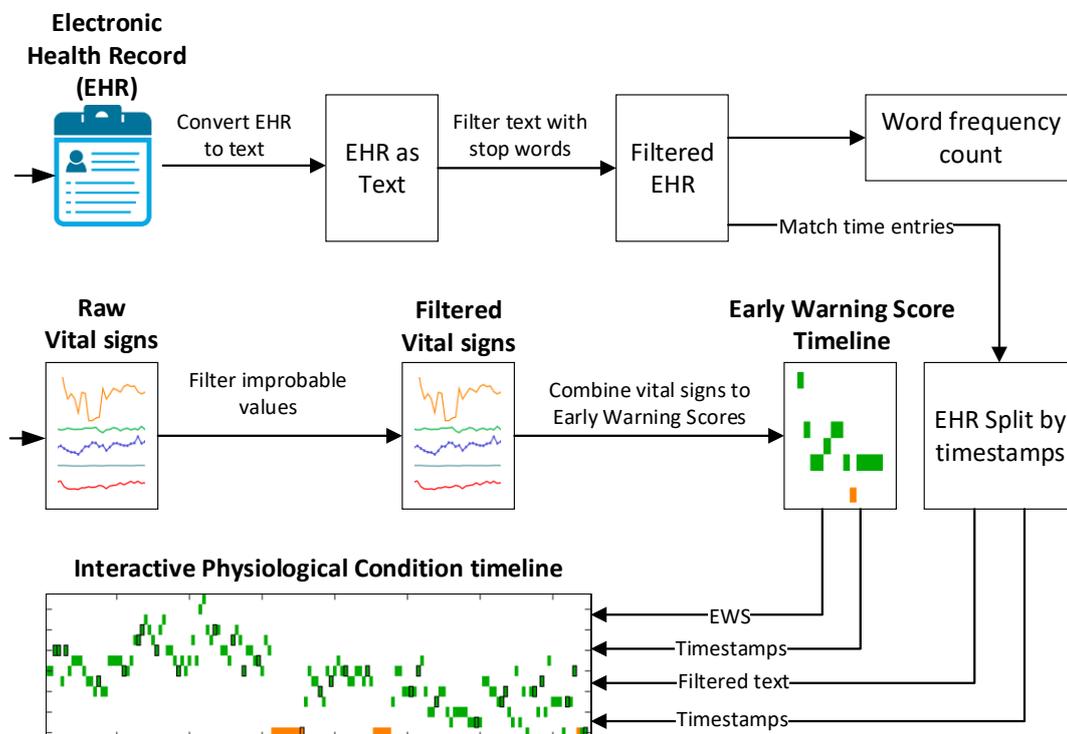


Figure 8: The physiological condition timeline for a patient is created by parallelly processing vital signs and the electronic health records (EHR). The textual EHR parts are converted to a text file, and filtered for names and insignificant words. Meanwhile, vital signs are converted into Early Warning Scores which are then combined with the EHR by matching timelines.

In addition, a word frequency count was done on the EHR part to get an overview of the challenges that were reoccurring throughout the admission.

4.2 DATA

The population consists of ICU patients whose primary admission cause was cardiac arrest, sepsis or respiration insufficiency. These patients have a high risk for complications/events during admission, and were considered as common by the clinical physicians at site. In the period from February 2013 to May 2013, Bispebjerg Hospital had 184 ICU admissions with an average length of stay of 4.5 days. Inclusion criteria were that patients needed to be admitted for at least two days because clinical physicians expected events to be unlikely in short term patients. This reduced patient of interest to 84. Not all patients had recordings of the vital signs described in Chapter 2 so the final population was reduced to 57 patients.

Table 6: Overview of patient info for patients who were admitted to Bispebjerg Hospital's ICU between February 2013 – May 2013, and met the inclusion criteria.

Age	66.70 ± 11.48 years
Length of stay	9.87 ± 8.34 days
Surgery related admission	17 patients
Admitted for other reasons	40 patients
Released to ward	29 patients
Deceased in ICU	18 patients
Transferred to another ICU	2 patients
Unique diagnoses and events	78
Total diagnoses and events	310
Unique treatment types during admission	45
Total treatments, scans and surgeries performed	474

Data management and storage for the ICU was handled by a company, Daintel, who were unreachable regarding exporting data. Medical staff were also unaware of data management, so it was not possible to confirm whether all actions during admission were stored in the same system. Collection of data for this project was performed by inspecting each patient's file and exporting EHR, codes and all raw measurements. The codes included diagnosis types, treatments, scans and surgeries, that were decoded using a database provided by the Health Data Board [67]. Raw measurements included vital signs, drug dosages, equipment settings, and potentially more. Table 6 shows an overview of the collected data about patients, other than vital signs, drug dosages and EHR for the entire admission.

Table 7: Ranges for vital signs that by Bispebjerg Hospital's ICU were considered reliable.

Type (T)	Cut-off ranges	
	Low	High
Pulse (bmp)	-	200
Respiratory Rate (bmp)	-	45
Temperature (°C)	27	45
Systolic Blood Pressure (mmHg)	-	310
Oxygen Saturation Level (%)	70	100
Inspired O ₂	-	-

The measured vital signs were furthermore filtered prior use. Highly improbable measurements according to clinical physicians have been removed. One example is oxygen saturation below 70%, because they were likely caused by electrodes falling off. Table 7 consists of the list of ranges that were allowed for each vital sign. It was not possible to get error rates and inspect what could go wrong with specific devices at site because equipment was not standardized and sometimes interchanged. The boundaries were therefore defined by a chief physician at the ICU staff as the values they regard as reliable.

4.3 METHODS

4.3.1 EARLY WARNING SCORES

Hospitals in developed countries optimize staff resources by using scoring models in wards to assess patient condition and monitoring needs. One such model is Early Warning Scores (EWS), that was introduced by Morgan et al [10] [66]. Each vital sign is assigned a sub-score based on how abnormal the value is compared to a healthy person, and an overall score is then estimated through summation of all sub-scores. Surveillance frequency, form and who is responsible are then adjusted based on overall score. The model is widely applied, but it is not even standardized on national plans [14] [68].

Table 8: EWS Table used in Capital Region of Denmark

Type (T)	Sub score						
	3	2	1	0	1	2	3
Pulse (bmp)	<41		41-50	51-90	91-110	111-130	>130
Respiratory Rate (bmp)	<9		9-11	12-20		21-24	>24
Temperature (°C)	<35.1		35.1-36.0	35.1-36.0	36.1-38.0	38.1-39	>39
Systolic Blood Pressure (mmHg)	<91	91-100	101-110	111-219			>219
Oxygen Saturation Level (%)	>92	93-94	94-95	>95			
Inspired O ₂				Air		Any O ₂	
Central Nervous System				Alert			Voice/Pain Unresponsive

Table 8 presents the full VIEWS table of which parts were presented throughout Chapter 2. Vital signs have a sub-score of 0 in the model if they are normal, and a nearly healthy person would therefore have an EWS of 0. Abnormality increases surveillance times significantly (Table 9), and ranges from follow-ups every 12 hours by nurses to follow-ups every thirty minutes by on-call physicians if a score above 8 is reached. The model is not used in ICUs, because staff is always close to the patient and ready to react. Evaluation of the central nervous system (CNS) is part of the model, but it was omitted in this study, because it was evaluated inconsistently and subjectively.

Table 9: Early Warning Score action table. The ABCDE optimization approach ensures that the patient breathes (Airway and Breathing), Circulation is fine, Disability such as coma or convulsion is avoided, and that patient is inspected for rashes, bleedings and other marks (Exposure) [69].

EWS	Observation interval	Action procedure
0-1	Every 12 hours (+/- 1 hour)	Continue scoring every 12 hours.
2	Every 6 hours (+/- 30 min.)	Caregiver ABCDE* optimizes. Parameter with score of 2 must be reported to nurses. Nurse becomes in charge of ABCDE optimization.
3-5	Every 4 hours	Nurse ABCDE* optimizes Physician evaluates and adjusts observation intervals. Physician is immediately called if any parameter has score of 3.
6	Every 4 hours	Nurse ABCDE* optimizes. On-call physician is summoned. Physician checks on patient and adjusts treatment plan.
7-8	Every hour	Nurse ABCDE* optimizes. On-call physician is summoned. Physician checks on patient and adjusts treatment plan. <i>Consider mobile acute team or ICU admission</i>
9+	Every 30 minutes	Nurse ABCDE* optimizes. On-call physician is summoned. Physician checks on patient and adjusts treatment plan. On-call physician discusses issue with medical specialist, ICU or mobile acute team.

4.3.1.1 Estimating Early Warning Scores

The raw vital sign measurements were down sampled to one value per hour, because progress over time was of interest, and the clinical physicians at site were confident that changes in minutes and seconds would be irrelevant because sudden incidents are detected by staff and built-in equipment alarms. This section demonstrates how EWS were estimated, by going through a real patient example.

Figure 9 shows hourly average vital sign values of patient who was in surgery for peritonitis and was admitted in the ICU due to septic shock – The most unusual development in vital signs throughout the admission happened around fifty hours after admission, where pulse significantly increased and systolic blood pressure began to rise. The sudden rise in pulse was described in the EHR, but the cause was unknown. It was however likely related to the patient receiving noradrenaline from beginning of the admission until about the hour where pulse starts increasing. The assumption is valid because pulse decreased continuously after reintroducing noradrenaline with amiodarone. Likewise, the increase of pulse around ninety hours was reduced with amiodarone.

Next, a quality check was performed to ensure that vital signs had sufficient data points to represent the hour. Respiratory rate, as an example, had sufficient data points if at least two breaths per minute were recorded. Temperature on the other hand changes much slower, so one value per five minutes was sufficient.

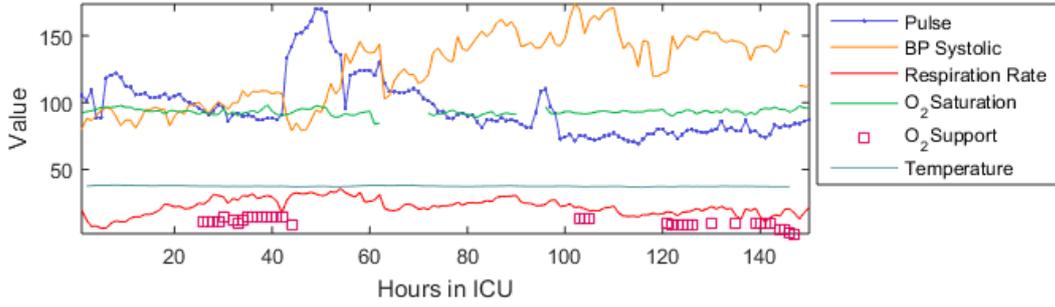


Figure 9: Hourly average vital sign values of a patient. Three of the vital signs were monitored throughout the entire admission. On two occasions, oxygen saturation monitoring was ceased, and there were some periods where slow breathing motivated staff to inspire oxygen.

Each hour of admission was then denoted by one of four classes (1-white, 2- orange, 3-yellow, 4-green) ranging from respectively “non-existent” to “ideal or above”. Figure 10 shows quality check measurements for the same patient. Oxygen saturation was measured with a non-invasive ventilation device (NIV), and was removed in the two periods where measurements are absent. This was done because the equipment was obtrusive, and staff wanted to test if the patient is improving. From the vital signs in Figure 9 it is visible that respiratory rate slowly drops both times after the equipment is removed.

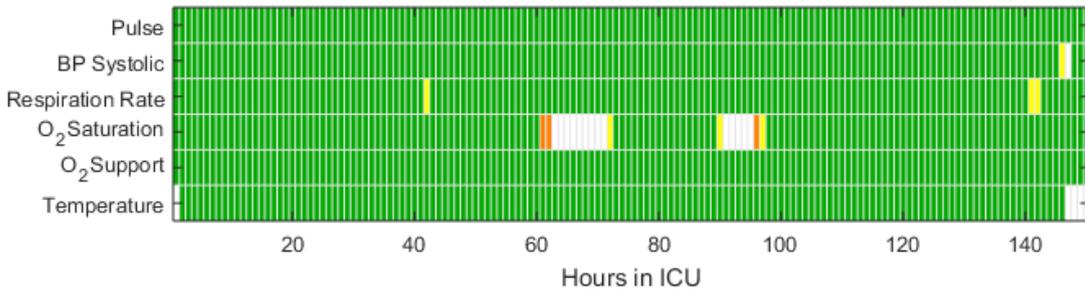


Figure 10: Quality control of recorded vital signs for a patient. Each box represents an hour and the color represents whether sufficient data is available. White indicates that no measurements exist. Orange that there are recordings, but less than half of ideal. Half to ideal number is yellow. Idea number of measurements is green.

With data validated and present, the EWS timeline could be created. $v_h(VS)$ describes the average value of all measurements for a vital sign, VS , throughout an hour, h – each hour starting at h_{start} and ending 60 minutes later at h_{end} . In addition, the number of measurements per hour, $c_{h,VS}$, was stored for the previously mentioned quality control, since sampling rates varied among medical devices. The down sampling is expressed by equation (4.1), where X_{VS} is a collection of all measurements for a given vital sign, and i is the index at a given time.

$$c_{h,VS} = h_{end} - h_{start}$$

$$v_h(VS) = \frac{1}{c_{h,VS}} \sum_{i=h_{start}}^{h_{end}} X_{VS}(i), \quad (4.1)$$

Once the hourly average for a vital sign is calculated, $v_h(VS)$, the value can be directly translated into a vital sign sub-score, SS_{VS} , through lookup in Table 8. I.e. a systolic blood pressure of 95 results in SS_{VS} being 2. The hourly EWS, EWS_h , is then estimated by equation (2) as the sum of SS_{VS} for all vital signs.

$$EWS_h = \sum_{VS=1}^{VS=6} SS_{VS}(v_h(VS)) \quad (4.2)$$

Figure 11 illustrates EWS timeline for the same patient as in previous examples. Every hour is represented by a rectangle. Rectangles are green if there were sufficient measurements for all vital signs, and are otherwise orange with the EWS value zero. The areas with missing data were kept in the graph to emphasize missing surveillance, which in many cases could be explained through tests, examinations and surgeries – in this case, staff tested whether the ventilation device was necessary.

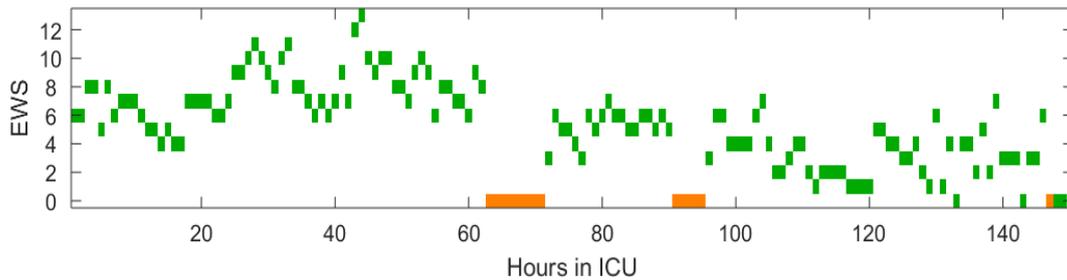


Figure 11: EWS timeline for patient. Based on measurements alone, the patient appeared to slowly but gradually deteriorate until about the 45th hour. This became a turning point and improvement was seen until the 120th hour, after which it again fluctuated.

4.3.2 NATURAL LANGUAGE PROCESSING

Natural texts can also be considered as signals, in which words are perceived as measurement points. Words can be quantified based on how frequently and in which context they appear, and even be assigned a sentimental value representing how positively or negatively they are perceived. Natural Language Processing (NLP) is a scientific field related to dealing with these and more complex tasks when processing natural text. There are even freely available tools and open source projects that can save implementation time [70] [71]. Poor support of Danish language, lack of interfacing with the EWS part, and the necessary conversion of EHR documents to other file formats were drawbacks of these existing tools, and NLP methods were therefore built from scratch for this project.

The following two subsections both deal with NLP. The first section is related to hypothesis 2, of applying text summarization to test create an overview of important admission related topics, such as medication, cause of admission and complications. The second subsection is related to hypothesis 3 of determining whether major physiological changes were registered in EHR around the same time the events occurred.

4.3.2.1 Admission Overview Through Text Filtering

As indicated in the approach overview, in Figure 8, the NLP part was not only developed for extracting and finding event. It also featured the necessary implementation of text summarization and word ranking to provide an overview of what the main concerns were during admission.

The first step in analyzing a text is filtering irrelevant information, commonly referred to as stop words. They may vary on application, but most often include words that are important for the flow and sense of a sentence and are meaningless by themselves. These words are highly frequent in all natural languages, making them easy to detect because they follow Zipf’s Law. The law explains that words in a corpus (natural language text) will appear inversely proportional to their rank if they are sorted by frequency. This means that the most common word in a corpus will appear twice as often as the second most common word, and so on [72]. In descending order, the most common words in English are: “the”, “be”, “and”, “of” and “a” [73]. Knowing that stop words can be found by their frequency, a collection of fifty-six million Danish words was obtained from the Danish Language- and Literature Society (DSL) and subsequently sorted by occurrence [74]. The two-hundred most frequent unique words were selected as stop words. Names were filtered out manually as it was not possible to obtain a database containing these – on the bright side, no eponymously named diseases were removed.

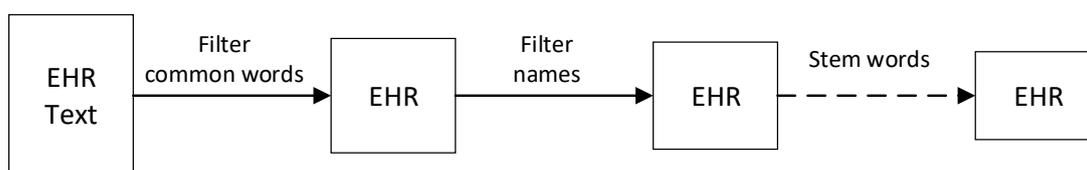


Figure 12: Filtering textual electronic health record (EHR) of a patient. Box size indicates relative size of document after each iteration. Most frequent words in the Danish language and names are filtered from the original text. Stemming would further reduce the size, but the tradeoff between implementation and benefit ruled it out.

The filtering procedure is illustrated in Figure 12. Ideally, it would include the last part of so-called stemming, so that words do not reoccur in different variations. As an example, applying stemming to the three words: “describe”, “described”, and “describing”, would result in a single word “describ” that then occurs three times. Stemming was not implemented as it requires implementation of language rules, and the tradeoff between implementation and gain would have been miniscule.

4.3.2.2 Entry Search in Electronic Health Records

Whenever hospital staff update an EHR, the software that they are using will autogenerate a timestamp indicating current date, time, and location. It safe to assume that abnormal changes are documented after incidents, because the law requires all events to be documented. Therefore, finding when the EHR is updated, results in finding approximately when an event occurs. Figure 13 is an example of an autogenerated timestamp and shows how it can be broken down. The department can change just like time and date, because ICU patient may require surgery or scanning elsewhere – it is also possible that no department is indicated.

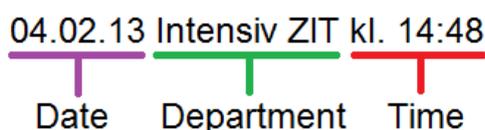


Figure 13: Autogenerated timestamp in ICU settings. Time and date only change in numeric values whereas department may or may not be present and may consist of multiple words.

To find these timestamps and extract the information within, regular expressions were used. A regular expression is a pattern formed as a sequence of characters and symbols. The search pattern is applied to text and responds when a match it found. Two everyday examples are listed below. The first one

is used when one wants to find all files of a specific format, like PDF, on the computer. The asterisk indicates any text fits the criteria if it with “.PDF”. The second example is common whenever online purchases or registrations are made; It verifies that an email address is valid. It specifies that only one “at” sign is allowed and that characters, numbers and a dash can be on either side. After the final period character, the email can only end with two to four characters in the range A-Z.

- File searches: *.pdf
- Email confirmations: \b[A-Z0-9.%+-]+@[A-Z0-9.-]+\.[A-Z]{2,4}\b

The same idea was used to make the regular expression for finding entries in EHR. The following pattern combined into one sequence (excluding text after the hashtag), matches all autogenerated timestamps for the software used at Bispebjerg Hospital (Figure 13):

```

^([0][1-9]|[1-2][2-9]|[3][0-1]) # Day of month
.([0][1-9]|[1][0-2]).13 # Month and Year
[a-zA-ZæøåÆØÅ\s.]* # Department
([0-1][0-9]|[2][0-3]):[0-5][0-9] # Hour of entry

```

A timestamp starts at the beginning of text line with a number between 01-31 (day) followed by a period character, a number from 01-12 (month), another period character, and last two digits of the year (13 for this project). Department name may or may not be present and consists only of letters; the asterisk here indicates department can contain zero to many words. Each timestamp ends with a time ranging from 00:00 to 23:59.

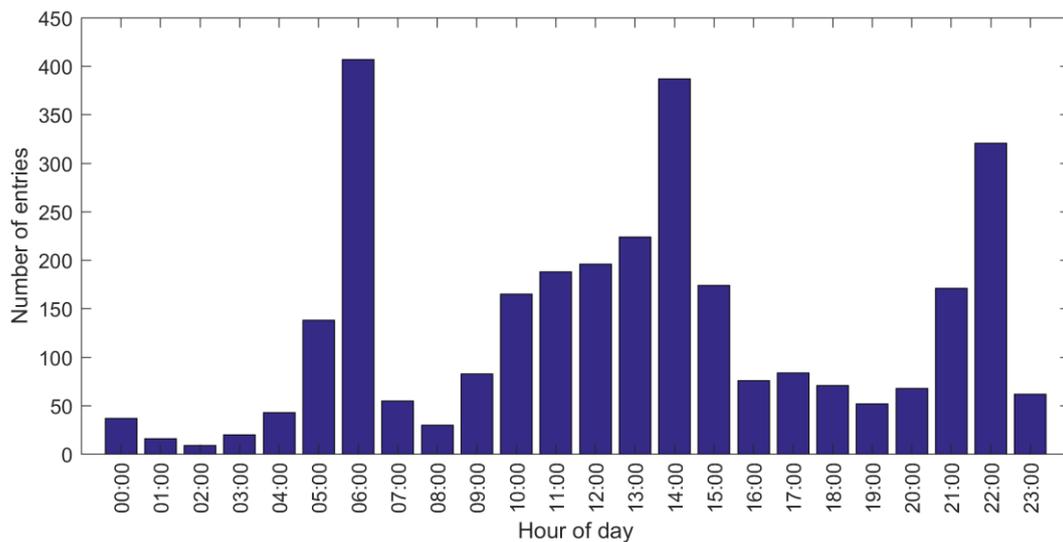


Figure 14: Overall number of EHR entries per hour at Bispebjerg Hospital’s ICU. Most entries occur during shift changes (every eight hours after 06:00). Entries during evenings and night are mostly incident related while planned treatments are scheduled in mornings.

For all patients combined, there were in total 5093 mentions of time, date, or a combination of these. Individually each patient had 116 ± 84 mentions. Autogenerated timestamps accounted for 3077 of these, with 70 ± 51 mentions per patient. There were no misclassifications in detecting autogenerated

timestamps because they differed in format from how staff referred to time. The differences between autogenerated timestamps and manual entries by staff were that, (1) staff omits the year when referring to dates, and (2) manual entries rarely began a line with a date. Figure 14 shows how many timestamps were generated in each hour of the day, which is of interest for understanding how documentation was conducted. The figure shows an overrepresentation in the hour marks at 06:00, 14:00 and 22:00 when shift changes are scheduled and staff write their summary. It is safe to assume that entries throughout the night are related to incidents while morning and partly afternoon are related to planned treatments.

For the patient from the previous examples, the regular expression found 40 EHR entries after the above regular expression was applied. Entries made within 30 minutes of each other in the same hour were merged into the same entry (Figure 15).

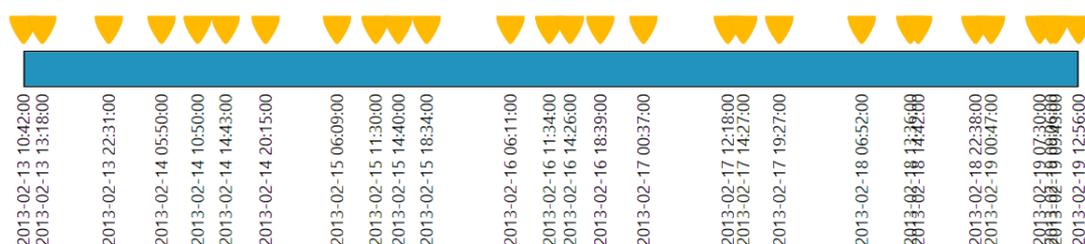


Figure 15: Identified timestamps of all entries in a patient's EHR. The distances between entries show that the EHR was being updated irregularly, and mostly without updates between the hours 20:00 and 06:00.

4.3.3 COMBINING EHR AND EWS

The physiological condition timeline from Figure 11 was then ready to be merged with the EHR based on timestamps illustrated in Figure 15. All extracted entries were merged with the ones registered in the same hour. Next, for every hour on the timeline, a contour border was created around the EWS and the square became interactive, so that it can open the corresponding entry, when selected.

Redundant information would ideally be reduced by extracting all entries and categorizing them based on type (e.g. diagnosis, event, follow-up, procedure, visit, etc.). This was not possible as text often contained spelling errors and was phrased in keyword format with non-standard abbreviations that were only familiar to internal staff members.

The interactive physiological condition timeline for the patient example throughout this chapter is demonstrated in Figure 16. A graphical representation in form of expanding lines has been added to illustrate the selected entry point. The afternoon entry categorically explains the admission situation in very short phrasing even without text filtering. The last two lines of the entry are the important ones, because they address progress, which in this case was continuously slow improvement— this improvement is also visible by the EWS scores.

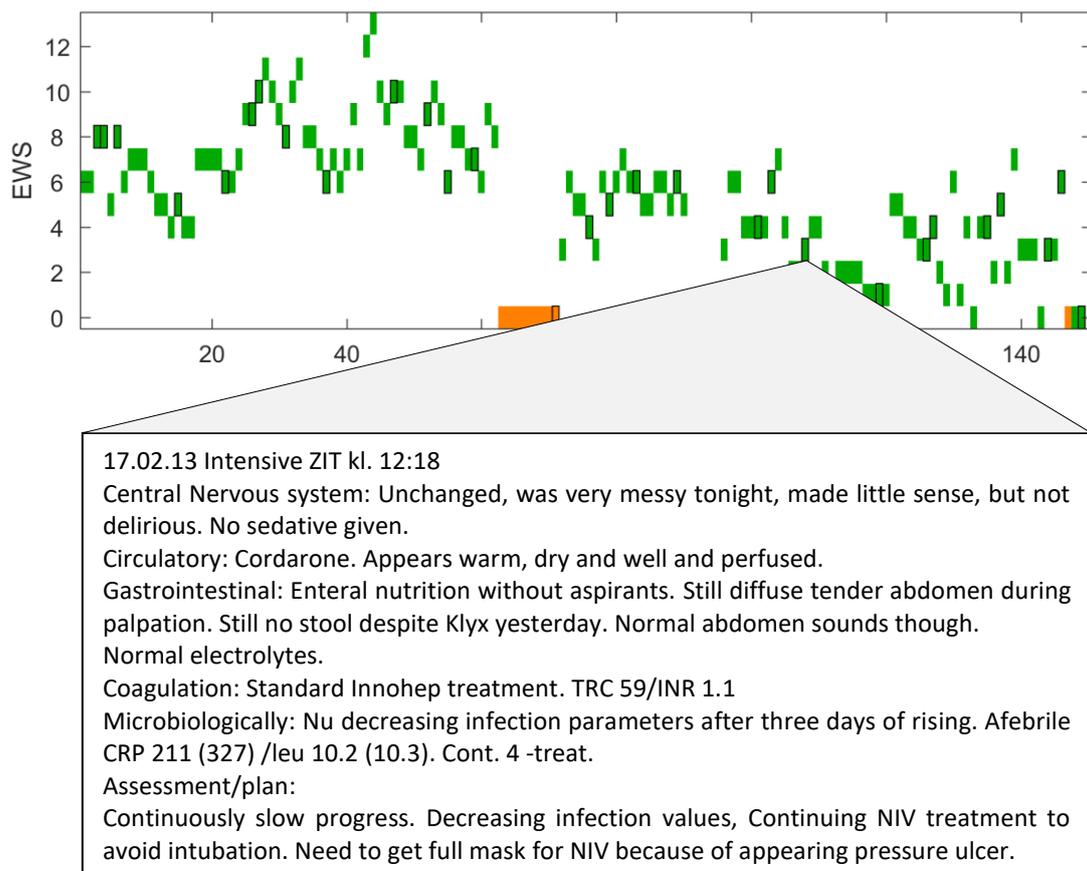


Figure 16: Early Warning Score timeline combined with EHR. If there were entries in the EHR for a given hour in the admission, the EWS value has a black box around it. On selection, the box opens the corresponding EHR entry. For convenience of the reader, the text in the example was translated word by word into English.

4.4 RESULTS

4.4.1 PATIENT CONDITION TIMELINE

Single vital signs values can highlight local changes but they fail to assess the overall condition. Figure 17 shows this, and demonstrates the output of the developed tool for viewing a patient's physiological condition throughout an admission using EWS. The vital signs show that the patient received induced oxygen support from 26 - 42 hours into the admission (top figure), but do not show that during this period, the patient's condition gradually and significantly worsened at 28 hours before improving. A nearby EHR entry revealed that the deterioration was caused by exhausting exercises with a positive expiratory pressure device.

Similarly, it is more difficult from the vital signals to see that the overall health condition is generally improving over time especially after about 100 hours into the admission. This was the entry stating that infection parameters started decreasing for the first time since the admission, and that the patient was improving. The patient was discharged and moved to another department with a stable condition and an EWS of 0. It is however important to note that the patient at discharge was still overhydrated, had an infection, had atrial fibrillation, and was fed through a feeding tube. The fact that all vital signs were normal was therefore heavily due to medication and electrical devices.

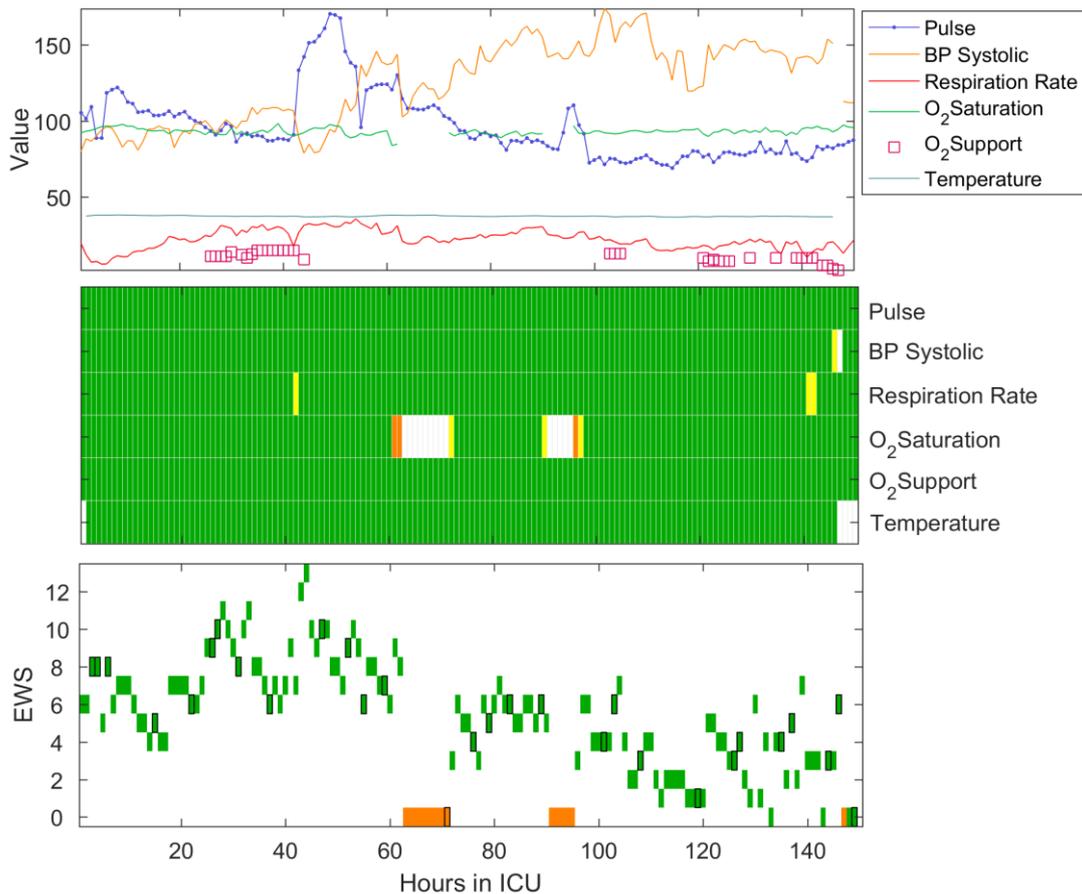


Figure 17: Output of developed tool for displaying a patient's physiological condition throughout admission. Top figure shows vital sign values. Middle figure shows quality of sampling frequency of the vital signs. The bottom figure shows the EWS timeline representation has selectable boxes containing EHR entry content, for the hours where documentation exists.

The EWS timelines for other patients were similar in shape and data, but not useful for patients who were discharged shortly after admission, because their condition tended to remain at a fixed EWS. This also raised questions whether the outcome could be deduced based on the timeline without reading entries. Another fully detailed example was covered in one of our previous studies [75].

4.4.2 RELATIONSHIP BETWEEN MORTALITY AND EWS

Thirty-eight of the ICU patients survived admission. Figure 18 shows the relationships between EWS when each patient was admitted, released and how stable they were – defined by standard deviation. The circle color represents whether they survived, and the size is the length of stay (LOS). EWS on the other hand was a better indicator, especially when the condition was stable. The EWS standard deviation for survivors was lower (1.85 ± 0.37) than for deceased patients (2.19 ± 0.45) which is also documented in other studies [76] [77].

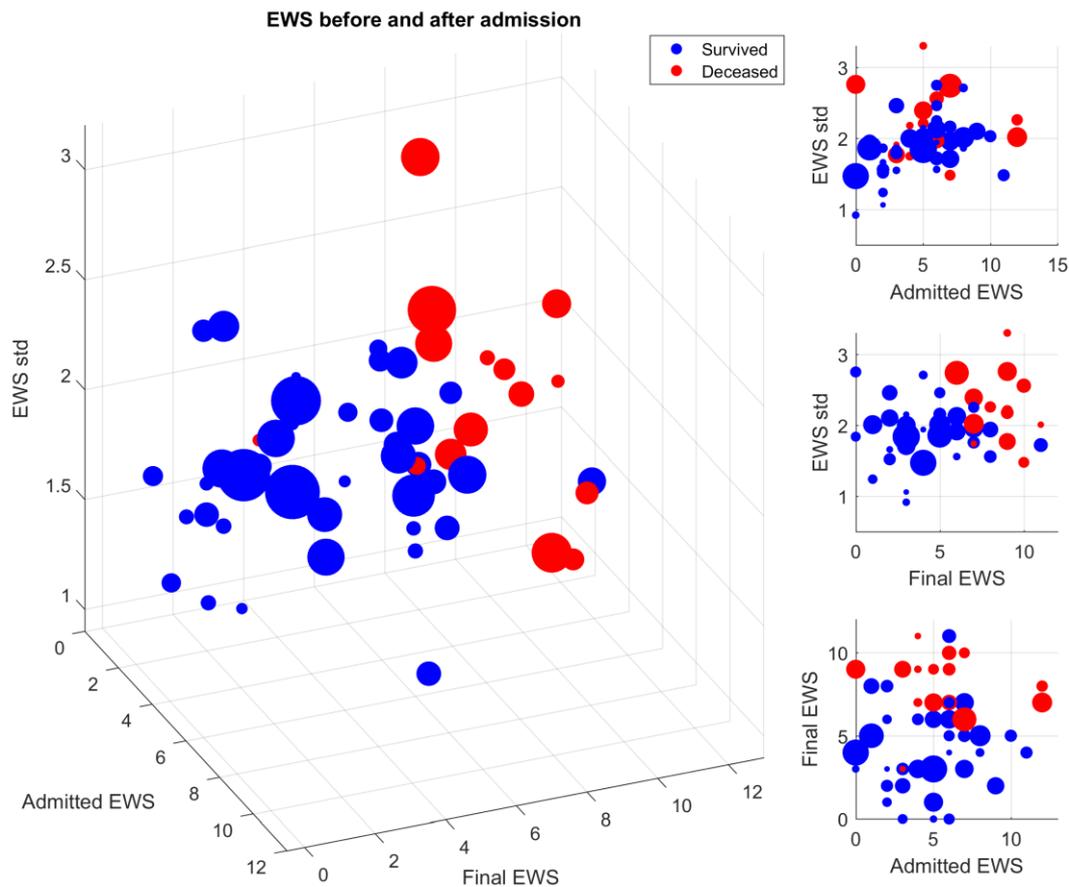


Figure 18: Patient outcome based on EWS and duration of admission. Each patient is represented by a circle, that is colored depending on mortality outcome. The right-side figures are different axis projections of the left larger figure. The best outcome separation is apparent when comparing final EWS with standard deviation.

4.4.3 MORTALITY AS FUNCTION OF OTHER PARAMETERS

Relationships between parameters were further investigated through a computed regression tree for the parameters: admission EWS, discharge EWS, EWS standard deviation, Age, total diagnosis codes, action codes and length of stay.

A regression tree has the same but inverted structure as trees observed in nature. The root is on top, and with every new condition it branches out, until no further branching can be done, and the end nodes become leaves. It was a useful in this case because it recursively partitions data in a non-linear way based on binary conditions. It could for instance be the case that all patients older than eighty with an EWS of over five deceased during admission. The tree was computed with MATLAB's version of the Classification and Regression Trees (CART) algorithm [78]:

1. Examine all binary splits on every predictor in the available data in the node.
2. Select the split predictor that yields maximal separation between classes.
3. Impose the split.
4. Repeat recursively for the two child nodes.

The recursive algorithm can have different stopping criteria based on desired output. The splitting of a node in this project was stopped if any of the two conditions below were met. The second criterium was set to avoid a too deep tree which becomes case specific rather than general.

1. The node consists of only one class and is therefore pure.
2. The node's parent has less than 10 observations.

The regression tree revealed two pure groups. Most patients belonged to the first pure group (1), who were admitted for more than three days but ended the admission with EWS under 7. The second pure group were patients who patient died if their final EWS score was 9 or above. In half of the cases where final EWS was between 7 and 8, the patient died.

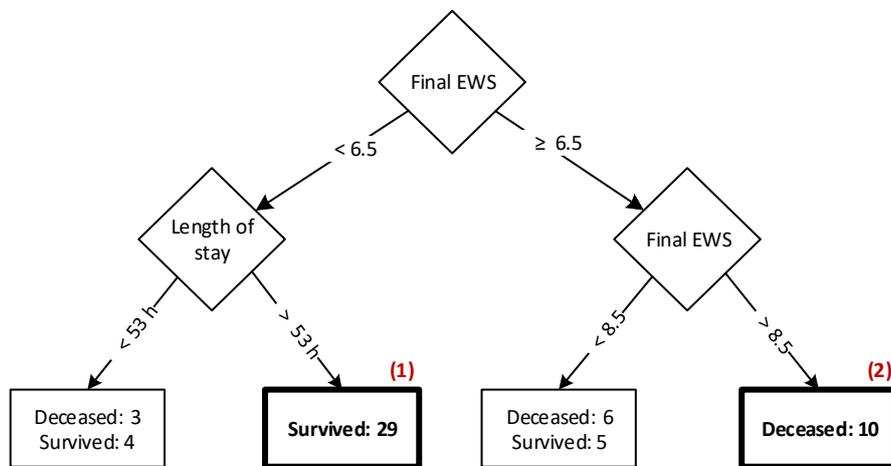


Figure 19: Regression tree for outcome of ICU patients. Based on criteria, it was possible to determine two pure classification sets, which are marked with colored numbers.

These findings slightly contradict our previous study which featured 44 patients (one fifth less than this study) [79]. The regression tree was previously constructed with fewer parameters in which both age and LOS were included, but age was more dominant. Unless EWS was over 7, older patients (over seventy-five years) had lower mortality than younger patients. A contributor was that younger patients had long history of illness and high alcohol consumption while the older patients were healthier prior ICU admission. LOS became a better branching parameter in this study because the 75+ group in this study had three deceased patients in that age group. The mortality for patients with EWS under 7 is ratio-wise still higher for patients under 75 (43% vs 27%).

4.4.4 SUMMARIZING EHR THROUGH WORD FREQUENCY DISTRIBUTION

The first three result-related subsections focused on the first hypothesis and whether EWS can be used for modelling physiological condition. From this subsection on, the focus shifts to assessment of EHR documentation quality and whether EHR entries are reliable source of condition evaluations.

The frequency distribution for words was calculated through individual summations of words that remained after filtering (section 4.3.2.1). Figure 20 shows the thirty most frequent remaining words in the previously introduced patient's EHR. The words in the figure have been translated from Danish for convenience of readers. Several groups of words stand out.

When reading the EHR thoroughly, many different drugs were found to have been administered throughout the admission. After counting words, only five stood out in the EHR: Cordarone, Furix, Klyx, Ultiva and Innohep.

- Cordarone prevents and treats a serious type of irregular fast heartbeats (ventricular tachycardia) that can cause cardiac arrest. It appeared so often because dosage needs to be carefully monitored. In terms of vital signs, this explained the abnormally high pulse.
- Furix forces diuresis (increased urine production) which could be linked to an overhydration problem.
- Klyx is a laxative used for gastrointestinal obstruction.
- Ultiva is a pain reliever, which given the prevalence of the word pain, it was not surprising
- Innohep is an anticoagulant reoccurring often, but even a manual inspection of the EHR did not clarify its purpose in the admission.

Another group of dominant words was related to stoma (stoma, stomas, gut sounds, abdomen, Klyx). This was the actual cause of admission, because an intestinal surgery did not go as planned and Klyx was used together with a temporary stoma to help with a gastrointestinal obstruction.

The last group of words was related to overhydration (overhydrated, diuresis and Furix).

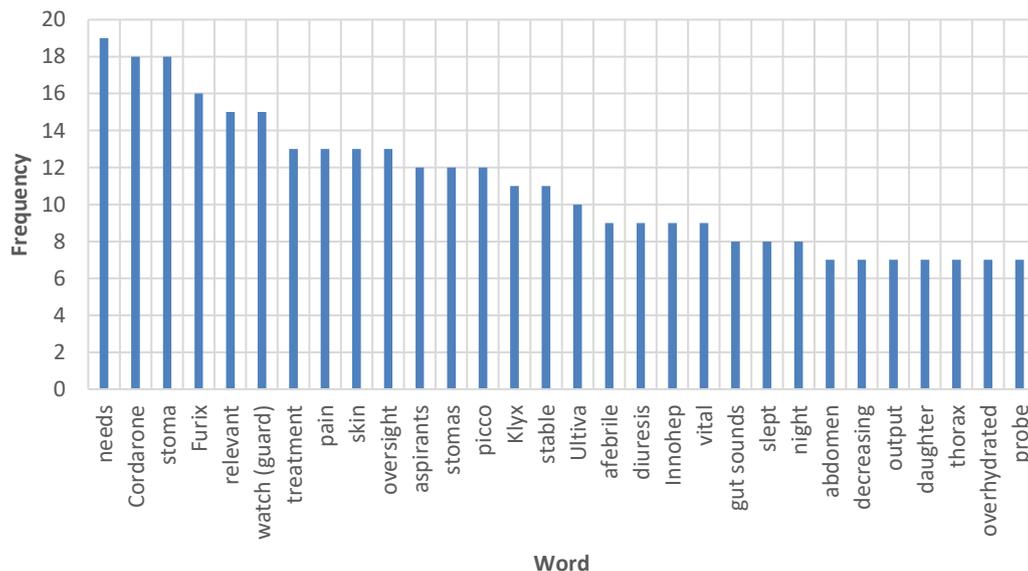


Figure 20: Distribution of the thirty most frequent of a patient's EHR after filtering. Words related to stoma and overhydration reoccur throughout the entire EHR in different variants.

More specific information could be derived from the EHR if neighboring words were extracted based on the context in which they appear. Through implementation of a mechanism known as N-grams, words could be linked together so that word pairs “abdomen” and “pain” would for instance be linked [80]. Similarly, PiCCO is a cardiac output monitor that can have multiple purposes, but it was only mentioned in the EHR as measurement tool for overhydration.

4.4.5 EVENT DETECTION USING EWS

Finally, the last hypothesis is addressed regarding whether physiological changes are registered in the EHR shortly after they occur. This hypothesis was tested by comparing all timestamps for actions and diagnoses timestamps and EHR entry timestamps, and comparing them to the estimated physiological changes using the EWS.

For each patient's EWS timeline, the derivative function was computed, and then compared with respect to diagnoses and treatments. Figure 21 shows the change in EWS in the hour of diagnosis and treatment registrations. Both types follow normal distributions, but the diagnosis distribution is left skewed whereas the treatment is slightly right skewed. Most diagnoses were registered when EWS fell by two points, and the patient's condition presumably improved. This also makes sense as an improvement following treatment would confirm an assumed condition.

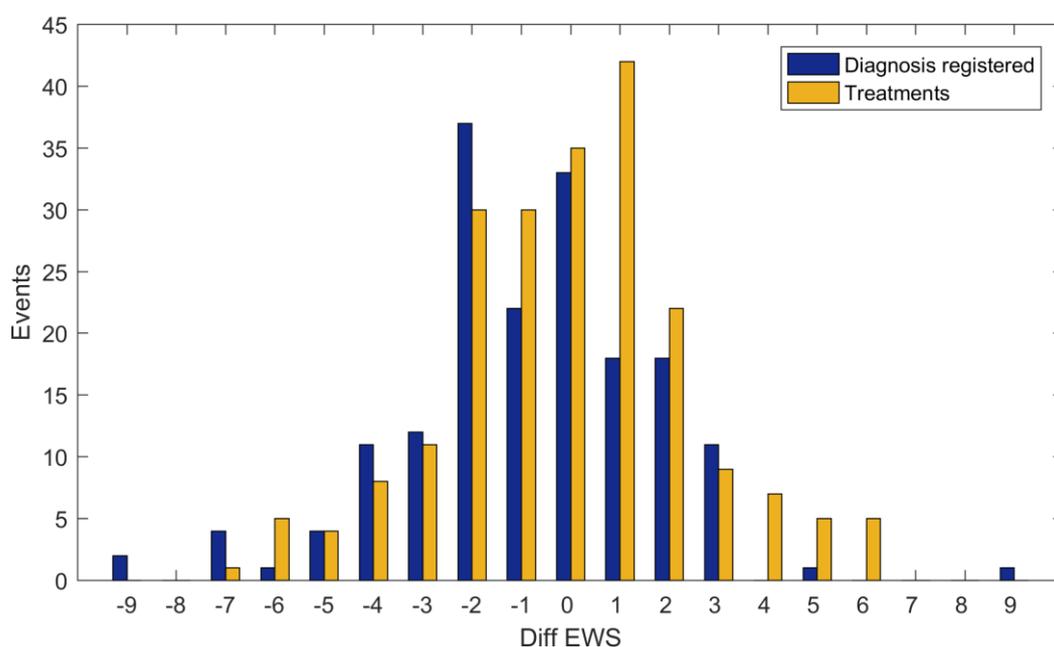


Figure 21: Overview of EWS changes in the hour when a diagnosis or other action is registered within the action codes.

It was expected that the treatment curve would be more right skewed as a deteriorating condition requires action. This was not the case, as all interventions are registered and tasks like catheter switching made up almost half of the treatments during improving EWS (n=34). No entries in the EHR stood out even after filtering shift switches (Figure 14).

4.5 DISCUSSION

The developed tool is in its current state useful for retrospective purposes, and does not feature decision boundaries that can improve treatment during admission. It can be applied for reviewing specific patient admissions or trend analysis of progress for either general patients or across patient groups. It is currently suitable within ICU settings where vital signs are monitored with precise electronic devices that are connected to patients spending most time in bed. Once unobtrusive

wearable monitoring devices become more available in other departments, the modelling will require more sophisticated processing that takes movement and body position into account, since both have influence on vital signs. It could function as an individual application, but implementing it as module in an existing CIS is also an option, since all data from this study was available in a CIS. It would be an improvement of the existing system at Bispebjerg Hospital because the CIS at site only had one review option, namely to look at vital signs individually over short intervals at a time.

Modelling of physiological condition in this study addressed all patients and therefore used the EWS approach. The benefit of using EWS in this study was, that the model is already verified and widely applied by staff. This theoretically also solves the previously mentioned problem of staff being unresponsive towards output [60]. Once unobtrusive wearable devices are deployed in other wards where EWS is already standard procedure, they will enable automatic monitoring of patients. This will reduce manual surveillance and allow staff to focus on more demanding tasks. Missing data due loose electrodes will be a concern since EWS can only be estimated if all values are present. In this study, the issue was addressed by the quality control measure in which EWS were considered objects with a score value and class for how much data was available. Further adjustments for missing data were not considered in this study because the absence of data could mostly be traced to ongoing interventions by staff.

Comparing the EWS timeline with individual EHR revealed that the model could be improved if data about medication, disease history and electronic devices in use were included. A common occurrence within the dataset was that vital signs at times appeared normal, but only because medication or a medical device, such as a respirator, supported the body. The existing EWS model was still a promising starting point, as the regression tree in Figure 19 showed. Looking back at national guidelines from Table 9, an EWS score of six and above is considered critical, and hence the physician takes over the surveillance; Similarly, the automatically generated regression tree for the study population starts by separating patients at an EWS of 6.

Due to the complexity of the content in action codes, it was not possible automatically detect specific events in the EHR and map them against the timeline. It was nevertheless still possible to use NLP to extract full entries and map them against the EWS timeline, allowing manual inspection of what caused physiological changes at any given time. In many cases this was sufficient to find cause and resolution to development around events. The text summarization method through word ranking proved useful because it from EHRs could identify the cause of admission, reoccurring difficulties and administered drugs.

From the hospital's point of view, the selected population had increased risk for complications and deterioration. It is therefore safe to assume that their condition would fluctuate more than for average patients. Deteriorations were therefore more prominent and hence easier to detect. On the other hand, it is still visible through standard deviation in Figure 18, that deceased patients were more unstable than discharged patients. Since the gathered data was from an ICU where surroundings were well controlled, staff may have been a major contributor to fluctuations in their efforts to improve continuously deteriorating patients.

Although the dataset was rather smaller, results reveal that it is possible to make an automatized tool for quality control of past admissions. On individual level, an example of this was demonstrated through Figure 16, where EWS were steadily decreasing and the condition was improving prior the

first staff entry related to certain improvement. On the whole population, sections 4.4.2 and 4.4.2 demonstrated that the deceased and surviving patients could be identified primarily based on EWS. The dataset also confirmed that the ICU was fully in control, but also overwhelmed and unaware by the amount of data available per patient. Thus, it is necessary to develop tools that combine the data in ways that are more intuitive.

There were some limitations throughout the study that need to be addressed in future extensions. Physiological condition is based on vital signs, but it is currently not possible to detect events which are dependent on modalities fulfilling certain circumstances, as it is the case with sepsis [81] [82] [83]. Since the original EWS model was designed to detect significant abnormalities, it may be insensitive for smaller changes – one example being that all SBP values in the range 111-219 are treated equally. Context awareness as to which medication and medical devices the patient is dependent on reveal which vital signs are affected, and make it possible to alter the model. This could be done by penalizing sub-scores further based on type of aid. If aids in the current model only provide a temporary improvement, the developed system serves more as an assessment tool for how well staff is managing the patient, rather than evaluating the patient's actual physiological condition. Since the model is incomplete in this regard, EWS cannot in its current form be used to determine when patients can safely be discharged. While the data for comorbidities and medication were available, they were omitted due to population size. The patients were admitted for similar reasons, but had hundreds of different co-morbidities, diagnoses and treatments (see Table 6). Understanding how patient outcome was affected by the various factors was out of scope for the study.

4.6 CONCLUSIONS

The developed tool demonstrated a novel approach to combining EHR with vital signs measurements to automatically generate an overall overview of a patient's development throughout an admission. A real-life scenario was covered to give the reader an applied example of the patterns revealed with the combination of EHR and vital signs. Another example can be found in our previous study [75]. The EWS model proved viable for detecting changes in physiological condition, making it possible to evaluate whether health is improving, stable or deteriorating.

The first hypothesis was that EWS can evaluate development in physiological condition of a patient over time, and that higher mortality would occur in patients who scored high in the model. The first part of the hypothesis was confirmed through the example up to section 4.4.1 and previous papers on the subject [75] [79]. The second part was addressed in sections 4.4.2 – 4.4.3, where it was determined that that patients who were unstable and whose standard deviation fluctuated more, were at higher risk of mortality. Finally, mortality was greatest among patients with $EWS > 6$.

The second hypothesis concerned whether it was possible to generate an overview of a patient's admission through quantitative measures, to obtain information about the admission. Through natural language processing and text summarization, section 4.4.4 demonstrated how word ranking can outline topics such as medication and complications throughout an admission. Words describing a reoccurring problem in the EHR mainly do so because staff continuously reflects and monitors changes in respect to the problem. Filtering stop words reduced noise in the textual signal, but further improvements can be made through implementation of n-grams and stemming.

The third hypothesis investigated if physiological condition changes throughout an admission were really registered in electronic health records (EHR). Even after filtering out textual entries related to shift summaries, the textual EHR had detailed entries throughout the entire admission, and the entries did not particularly focus on events. The codes related to treatment and diagnoses on the other hand showed that diagnoses are typically registered in the hour following an EWS improvement of 1-2 scores. Treatment was slightly more prevalent when the patient's condition started to decline by 1 EWS.

The study achieved promising results, and demonstrated how vital signs primarily on individual level can be used to model physiological condition of critically ill patients over time. Further investigation of individual vital signs roles requires more focused patient sub-groups where patients have similar backgrounds and co-morbidities. The next chapter and study therefore focus on a more homogenous critically ill patient group, namely ischemic stroke patients.

DETECTION OF PARAMETERS INVOLVED IN ISCHEMIC STROKE OUTCOME

Objective: When ischemic stroke patients are admitted to a Stroke Unit (SU), medical forms are filled out, containing information about the admission itself, along with the patient's medical history, medication etc. Throughout the admission, vital signs are measured and neurological disability is assessed to monitor progress. Although it is known from literature and experience that all registered parameters are of relevance, few studies have looked at quantifying how and which parameters are of greatest importance for stroke outcome. In this study, an automatic algorithm has been designed to determine which parameters are of highest importance through regression models that approximate the degree of disability following the first day of admission. A paper on this study has been accepted for publication and is attached as Appendix D.

5.1 BACKGROUND

The study on intensive care unit (ICU) patients, described in the previous chapter, focuses on health deterioration where patients had high risk of mortality. Health deterioration could be overlooked if multiple parameters were slowly, and simultaneously, deteriorating. Stroke patients are also in the category of critically ill patients, but the disease is reflected by neurological functions, where health deterioration is evaluated through assessment of the degree of disability.

To investigate the role of vital signs in respect to disability outcome in stroke patients, a collaboration was made with the department of neurology at Zealand University Hospital in Roskilde (ZUH). The hospital has been involved in treatment of 792 ischemic stroke patients in the years 2013 – 2015, and treatment follows national guidelines where intravenous recombinant tissue plasminogen (rt-PA) is used to dissolve the ischemia [33]. In severe cases, such as large vessel occlusions, patients were relocated to other hospitals, meaning that patients treated at ZUH are more likely to have better outcomes than at larger hospitals. This makes early warning scores (EWS) unsuitable to analyze the data due to fewer critical cases.

Previous studies have investigated the effects of individual parameters, such as individual vital signs, treatment time and glucose levels [38] [84] [85] [86]. As an example, elevated blood pressure (BP) at admission, is observed in 80% of ischemic stroke cases. Better outcomes in terms of lower mortality are observed for patients who have SBP at 120 – 140 mmHg. The relationship between mortality and SBP is furthermore U-shaped, where the presented ideal range is the nadir and values outside the

range increasingly worsen outcome [87]. The phenomenon is still poorly understood, but lowering BP should therefore be carefully considered. Suggestions to causes range from mental stress related to admission, to the fact that elevation of BP increases the perfusion pressure in the ischemic penumbra [88] [89] [90] [91] [92] [93] [94]. The fact that elevated BP is of advantage for stroke patients also means, that the EWS model needs adjustment before being used as a tool for stroke patients, because elevated BP is not accounted for (see Table 8).

Few papers were identified that investigate stroke outcome as function of multiple parameters. The most wholesome study was conducted throughout the years 2002 – 2006 by the Safe Implementation of Thrombolysis in Stroke MONitoring STUDY (SITS-MOST) on 6136 international patients, and processed through multivariable analysis [95]. It examined parameters involved in different types of outcomes of which functional independence was the only relevant one for this study, because it addressed whether disability was so severe that they needed help with simple daily tasks. Data was analyzed through multivariable logistic regression where outcome was a binary category representing independence and dependence. Parameters were only included if their individual correlation to output had $p \leq 0.25$. The remaining variables were subjected to stepwise exclusion, and in some cases removed, but the exclusion criteria are undefined. Identified parameters involved in outcome were antihypertensive treatment, heart failure, diabetes, diastolic blood pressure (DBP), gender, NIHSS at onset, mRS prior stroke, systolic BP (SBP) and weight. A study by Johnston et al [96] developed a risk model to predict outcome after three months based on infarct volume from imaging, medical history and NIHSS at admission. The study also used logistic regression and determined age, NIHSS at admission, history of diabetes and infarct volume to be of significance for positive outcome, while only infarct volume was related to poor outcome. One study investigated EWS in respect to mortality in stroke patients, where based on twenty-four patients, mortality was lowest for EWS 0-1 (2%) and highest for EWS ≥ 5 (63%) [97].

Although the patients in this study were connected to bedside vital signs monitors, all admission related information including treatment progress were stored in printed forms that were eventually scanned and uploaded onto secured servers. The manual registration of data also meant that the available sampling frequency of vital signs was low.

5.1.1 RESEARCH HYPOTHESIS

The main hypothesis was that due to the homogeneity of the patient population, it is possible to predict the degree of disability after 24 hours based on vital signs monitoring along with neurological assessment. Data about the medical history, admission and treatment initialization were included to determine which parameters contribute most towards this outcome. Our hypotheses were:

1. It is possible to predict the degree of disability for ischemic stroke patients for the first day of admission based on available patient data that is recoded until alteplase dosage is fully administered.
2. The degree of disability after the first day of admission is dependent on development in vital signs during the period.
3. All registered parameters are relevant for ischemic stroke patients, but their impact on degree of disability differs. The most significant parameters in predicting degree of disability after stroke can be determined through automatized algorithms.

5.1.2 APPROACH OVERVIEW

To address the hypotheses above, paper-based medical forms for ischemic stroke patients have been acquired and digitized. Section 5.2 addresses the data acquisition process involving digitization and validation of forms. Section 5.3 addresses the features and models involved in prediction of disability scores. Reflection on the hypotheses are presented in sections 5.5 and 5.6. The approach itself is outlined in Figure 22.

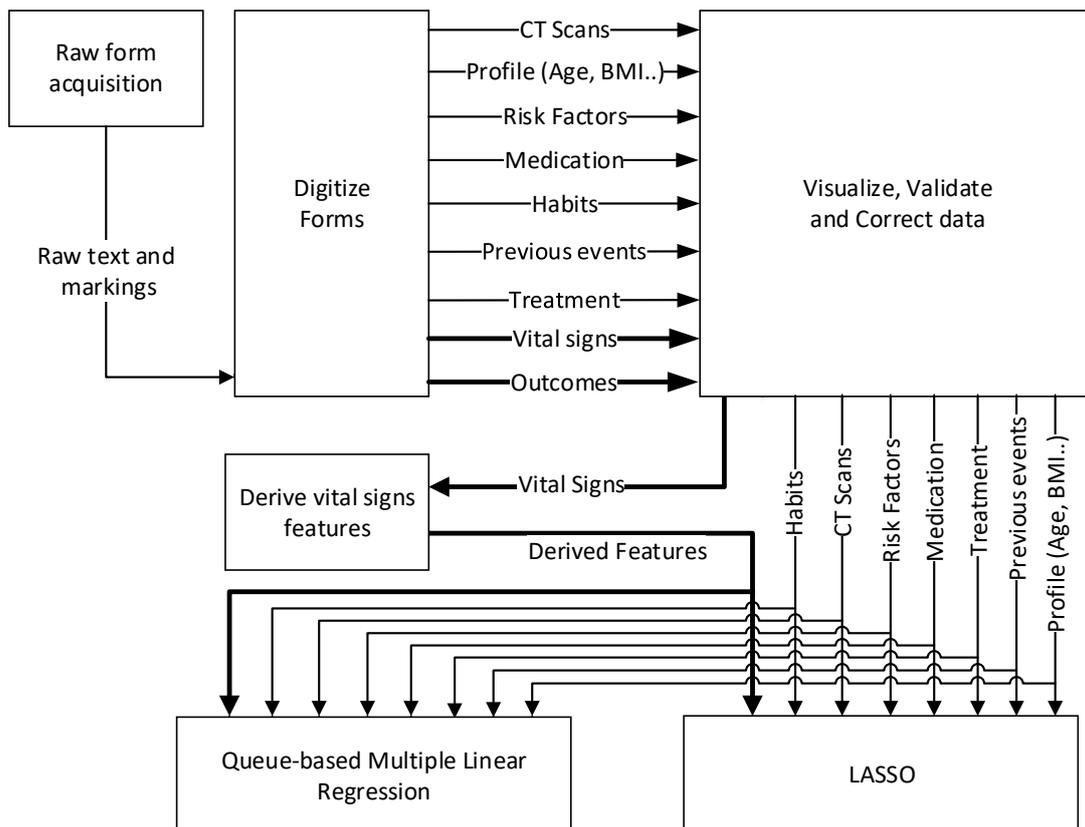


Figure 22: Approach used for prediction of disability scores for ischemic stroke patients for the first day of admission. Prediction was tested with LASSO and a designed variant of multiple linear regression.

5.2 DATA

Data including vital signs, medical history and neurological assessment were obtained from the standard medical forms of sixty-four randomly selected ischemic stroke patients from ZUH. Seven of these were excluded due to missing data, or due to transfers to other hospitals, as part of the drip and ship method (section 2.2) which at ZUH happens within 6 hours after admission. The final population included 57 patients.

The forms were filled out manually by nurses and physicians throughout the admission by following national guidelines for intravenous thrombolysis treatment during ischemic stroke [33]. Throughout the admission, medical history and admission information were registered once, while vital signs and neurological scores were registered more frequently. In the first two hours of treatment, neurological disability is assessed using Scandinavian Stroke Scale (SSS) scores, which examine eight parameters

Table 10 shows all parameters included in the study along with mean, prevalence or distribution where applicable. The SITS-MOST column refers to the population of the study with identical name and serves as a reference [95]. It is important to note that the two groups cannot be directly compared as the SITS-MOST population contains all cases of stroke, including severe ones (such as large vessel occlusion) which are not present in ZUH. The study was furthermore conducted in a time where alteplase was only administered if the age of the patient was below eighty years, and during the first three hours after onset.

All parameters from SITS-MOST are included in this study. Onset to treatment start is in this study split into two separate parameters, because the specific onset time is, unlike arrival to treatment, not always well known. Similarly, the type of antiplatelet is specified. The profile of patients in terms of age, gender and weight shows that the populations are similar. Vascular risk factors show the same patterns except for hyperlipidemia, which is present in 34% of the SITS-MOST population, compared to 78% in our population. Our study had five times more previous stroke incidents, which can be attributed to earlier interventions in risk groups like DM patients [98].

5.2.1 VALIDATION

Data for each patient were visualized, validated and corrected in cases where errors had occurred during digitization. While most errors for numeric values were found through outlier detection and manual inspection, binary and categorical values typically only had one measurement. Validation of these was supplemented through investigation of intervariable relationships, as e.g. if a patient previously had apoplexy, they were expected to be on antiplatelet medication such as Clopidogrel. After combining data for all patients, the correlation coefficient was computed for each combination of two variables, and the results were inspected in collaboration with chief physicians.

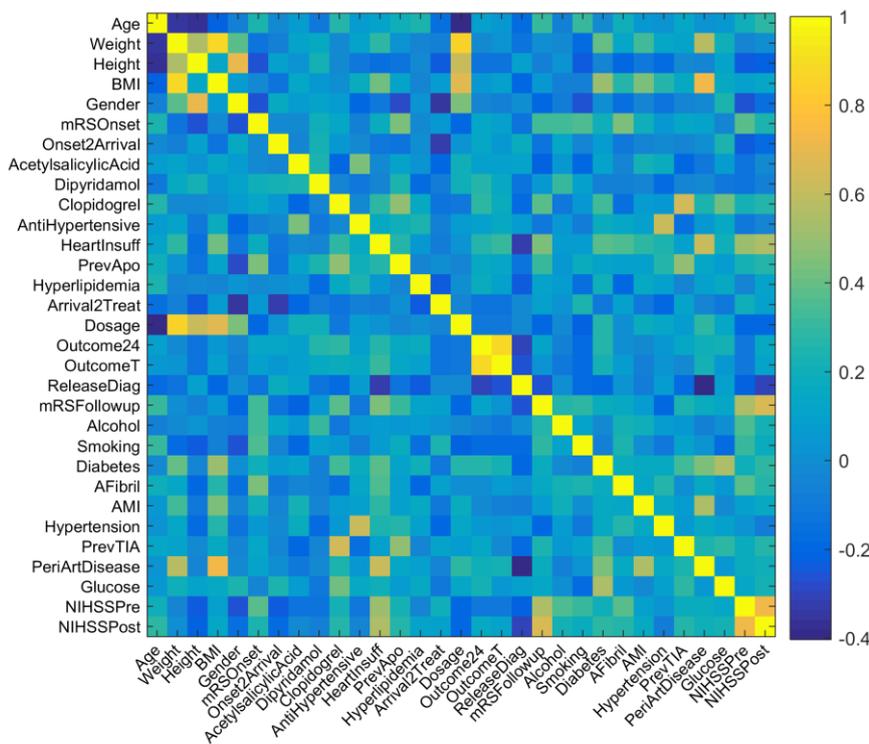


Figure 25: Intervariable relationships shown through correlation coefficients

Figure 25 shows the intervariable relationships after correction of found errors. Unsurprisingly in regards to profile the height, weight, gender and BMI were very strongly correlated. Taller people were in fact thinner which the BMI and height relationship show. Dosage of alteplase was negatively correlated with age, because older patients were lighter. Other examples of obvious correlations are

Table 10: Digitized parameters which are registered as part of national guidelines for treatment of ischemic stroke patients. SITS-MOST column shows all available data from a study with same name [95]. Grey shading for a parameter means that it was recorded

Type	Parameter	Population (n=57)	SITS-MOST (n=6136)	Type
Profile	Age (years)	68.21 ± 11.92	67.80	Numeric
	Weight (kg)	78.16 ± 17.84	75.55	Numeric
	Height (cm)	170.79 ± 9.30		Numeric
	BMI (kg/m ²)	26.65 ± 5.25		Numeric
	Gender (female)	44 %	40 %	Ordinal Binary
Imaging (CT scans)	Local bleeding	-		Categorical
	Remote bleeding	-		Categorical
	Cerebral edema	-		Categorical
Treatment at hospital	mRS onset	0.42 ± 1.02		Ordinal (0-5)
	Onset to arrival (min)	108.12 ± 47.3	140	Numeric
	Arrival to treat. (min)	30.63 ± 14.24		Numeric
	Dosage rt-PA (mg)	69.91 ± 13.40	68	Numeric
	Glucose (mmol/L)	7.04 ± 2.18	6.47	Numeric
Vascular risk factors	Arterial fibrillation	18 %	24 %	Ordinal Binary
	Diabetes	16 %	16 %	Ordinal Binary
	Heart insufficiency	8 %	8%	Ordinal Binary
	Hypertension	61 %	60 %	Ordinal Binary
	Hyperlipidemia	79 %	34 %	Ordinal Binary
	Periph. arterial disease	4 %		Ordinal Binary
Previous events	Acute myocard. infarc.	11 %		Ordinal Binary
	Prev. apoplexy (ever)	50 %	10 %	Ordinal (1-3)
	Transient ische. attack	11 %		Ordinal Binary
Relevant drugs	Acetylsalicylic acid	23 %	30 %	Ordinal Binary
	Clopidogrel	11 %	7%	Ordinal Binary
	Dipyridamole	2 %		Ordinal Binary
	Anti-hypertensive	37 %	47 %	Ordinal Binary
Habits	Alcohol (>recommend)	11 %		Categorical
	Smoking (prev./curr.)	51 %	23 %	Ordinal (1-4)
Outcomes	Bleeding present	2 %		Ordinal Binary
	mRS (after 3 months)	2.32 ± 1.75		Ordinal (0-6)
	Outcome 24h	-		Ordinal (1-5)
	Outcome at discharge	-		Ordinal (1-5)
Neurological assessment (24H monitoring)	Scandinavian Stroke Scale, NIH Stroke Scale			Numeric
Vital signs (24H monitoring)	pulse, temperature, blood pressure, oxygen saturation level			Numeric

visible, such as that patients having a history of hypertension were on anti-hypertensive medication ($r = 0.60, p < 0.01$). Similarly, the antiplatelet Clopidogrel, was in this population used by patients with previous incidents of transient ischemia attack (TIA) ($r = 0.63, p < 0.01$). The time from hospital arrival to treatment was slightly shorter for females ($r = -0.35, p < 0.01$), and for patients who arrived quicker to the hospital ($r = -0.31, p < 0.02$). An explanation for the phenomenon was not found.

5.3 METHODS

This section describes how stroke outcome was defined and compares the different types of outcome measures assessed by the stroke unit at ZUH. It then describes the features that were included and omitted in the study, and ends with a detailed description of an automatic outcome modelling algorithm.

5.3.1 CHOOSING STROKE OUTCOME

Stroke units often monitor and evaluate outcome on multiple different and independent scales for documentation and research purposes. To address the first hypothesis of predicting degree of disability, it is sufficient to work with one outcome scale. All outcome related scales used by ZUH are described and shown in Table 11 and a full description of each scale is attached as Appendix A.

Table 11: Outcome assessment scales at the stroke unit in ZUH. Lower scores represent positive outcome, and lower degree of disability, in all cases except SSS. Detailed description of scales is attached as Appendix A.

Name	Description	Scale/Classes
Modified Ranking Score (mRS)	Severity assessment after three months (to what degree is assistance from others required in daily tasks)	0 – 6
National Institutes of Health Stroke Scale (NIHSS)	Individual and combined assessment of all motoric abilities and consciousness.	0 – 42
Outcome 24h	Assessment of progress from admission to 24 hours later, with 3 representing no change.	0 – 5
Outcome Discharge	Assessment of progress until discharge, with 3 representing no change.	0 – 6
*Scandinavian Stroke Scale (SSS)	Continuous individual assessment of motoring abilities and consciousness during the first 24 hours	////
*SSS does not have a combined score, and is the only scale where ascending scores represent positive outcome		

It could have been interesting to see if mRS is predictable because it was used for outcome in the SITS-MOST study, but since some mRS classes were underrepresented in the dataset, this was not possible. NIHSS and SSS were the first choices for prediction, because “Outcome 24h” and “Outcome Discharge” only quantify intra-patient progress. Studies have already proven that NIHSS and SSS are interchangeable, although it may require a conversion equation [99] [100]. In our findings, even negating NIHSS and scaling in respect to SSS was sufficient to have a high correlation with SSS ($r =$

0.96 at admission and $r = 0.91$ at discharge). Since SSS is also monitored as often as vital signs, it furthermore gave the option of testing an approach like the EWS system.

Before moving on to feature extraction, the progress of SSS over time was examined. A patient's, p , neurological function, N , was extracted at a given time, t . It was then normalized through division by the highest possible score for the given neurological ability. This was done for all patients and average was estimated so that the output is a representation of the combined functionality among all patients (equation 5.1).

$$all(t, N) = \frac{1}{p} \sum_{p=1}^{p=57} \frac{SSS(p, N, t)}{\max(SSS_N)} \quad (5.1)$$

Figure 26 shows this for the first two hours followed by every sixth hour. The closer the value is to one (black), the fewer disabilities patients had. The consciousness level was always high, and most prominent disabilities were related to fine motor skills. The combined column shows the same representation but for the overall SSS score, demonstrating that the greatest change happened in the first two hours. Since the greatest change was observed over all categories combined, it was sufficient to only focus on how the overall SSS is influenced over time.

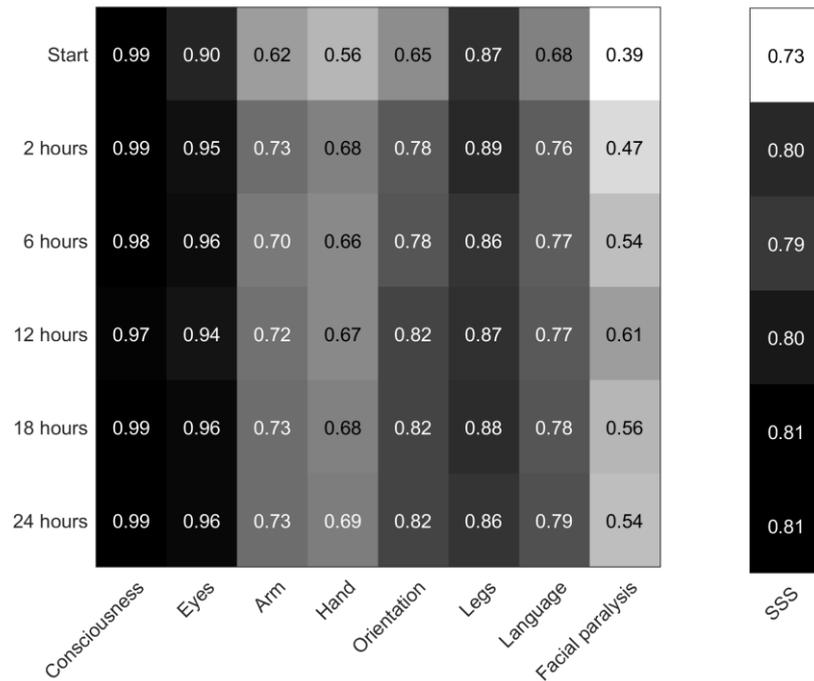


Figure 26: Progress of neurological ability for individual motoric areas of all patients (left), and combined SSS for all patients at specific hours during the first 24h of treatment (right). Higher scores represent lower the disability.

It is evident from Figure 26 that there is little change in the overall SSS after two hours. An algorithm that attempts to predict final SSS based on initial SSS would therefore also be highly accurate. More interesting was therefore whether additional features could improve accuracy.

5.3.2 PREDICTING OUTCOME THROUGH REGRESSION

To be able to address all three hypotheses, two regression approaches were tested as prediction models. One was a “least absolute shrinkage and selection operator” (LASSO) regression model which incorporates variable/feature selection, and the other was a custom designed queue-based multiple linear regression model. Alternative methods, such as ridge regression were discarded, because they lack direct dimensionality reduction – which in this study was desirable. Before implementing the regression models, features needed to be defined.

5.3.2.1 Features

All parameters from Table 10 were included as features except for the ones describing outcome types. They were further supplemented with potential features derived from vital signs and SSS for the first two hours after treatment onset, to see if vital signs contribute to SSS and to address the second hypothesis of this study. The potential features were:

Mean (μ)	Standard deviation (σ)	Other:
– SBP	– SBP	– Correlation between SBP and DBP
– DBP	– DBP	– SSS _{DISCHARGE-ADMISSION}
– Pulse	– Pulse	
– Temperature	– Temperature	
– Oxygen Saturation level	– Oxygen Saturation level	
– SSS	– SSS	

The correlation coefficient was included because previous studies have found that there is a positive linear relationship between SBP and DBP in healthy subjects, which can be influenced depending on diseases such as hypertension and DM. The expectation was that correlation coefficient is at first widely distributed, and later stabilizes around $r = 0.74$ for most patients [25]. If BP had been measured three times per hour throughout the entire admission, the change in SBP versus DBP could have been used to extract information about arterial stiffness [25] [101].

The scatter plots in Figure 27 show each patient’s SSS and BP correlation for respectively two first hours and after twenty-four hours. The bold markers represent the center of mass. The patients improve in SSS by an average of 4.08, and in terms of correlation a change is seen from $r = 0.43$ to $r = 0.52$. The histograms furthermore show a uniform distribution after two hours which after twenty-four hours is converging towards 0.6.

5.3.2.2 Multiple Linear Regression

Multiple Linear Regression (MLR) is a fitting or prediction model, where a dependent variable y , or response, can be estimated through a relationship of a set of variables $x_1 \dots x_n$. As an example, one could model the response, heart rate, as a function of age, height, weight and history of hypertension. Mathematically, it is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (5.2)$$

x_i is substituted with the value of the i ’th variable x , and β_i is the i ’th coefficient denoting the variable’s contribution in the relationship model. β_0 is called the intercept and is the response when all predictors are zero.

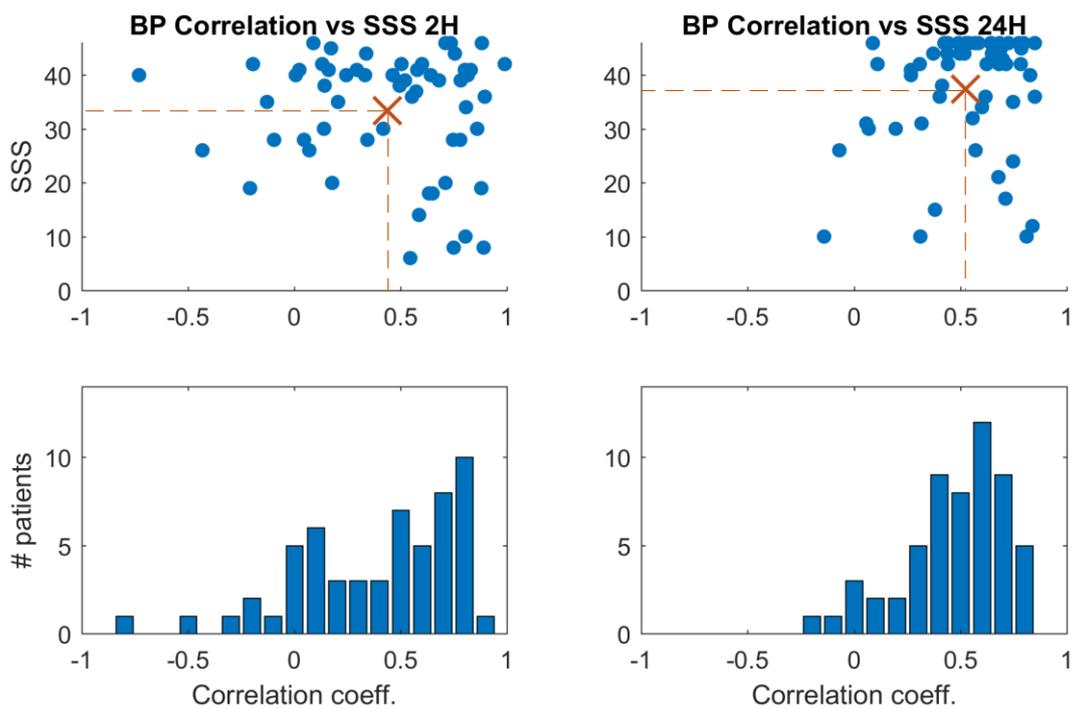


Figure 27: Correlation between DBP and SBP after 2 hours and 24 hours after initialized treatment. It is assumed that BP correlation coefficient increases as patients get healthier [25]. Disability improves from SSS = 33.35 to SSS = 37.44, while BP correlation increases from $r = 0.43$ to $r = 0.52$.

The same model applies when dealing with binary predictors. Ordinal/categorical predictors on the other hand need to be split into as many binary variables as there are categories, minus one. The predictor *previous apoplexy* in the dataset for example, has three categories: *no previous events*, *yes, over 3 months ago* and *yes, recently*. By splitting it into two variables, they can together represent all three states (Table 12).

Table 12: History about previous apoplexy is registered as one of three categories. Two variables represent whether the patient had apoplexy within the past three months, and if both are 0, the patient belongs to the third category of no previous apoplexy events.

	v_1 (3+ months ago)	v_2 (recent)
No previous event	0	0
Yes, 3+ Months ago	1	0
Yes, Recent	0	1

The summation of intercept and each of the variables multiplied by their respective coefficients give an estimate for heart rate. This response will mostly deviate from the true value. Likely causes are that measurements of predictors are inaccurate, or the model incomplete because predictors are missing – such as whether the person is standing or sitting. The overall error by how much each response deviates from actual measurements is referred to as root mean squared error (RMSE). It is computed as the square root of the average sum of squares for predictions subtracted by actual measurements (equation 5.3).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.3)$$

Before creating a regression model, one needs to be aware of the sizes of samples and predictors. With 57 patients and 39 predictors prior splitting ordinal categories, a relationship model would be overfitted, and would represent the given data well, but poorly predict new observations (section 3.2). The minimum number of patients per sample depends on the complexity of the model, how well classes are separated and the application purpose. Using Monte Carlo simulation Knofczynski et al generated twenty-three million samples to create an overview of the number of necessary samples for prediction using MLR. They found that 35 samples were required for five predictors, and after that, additional five samples were required for each additional predictor. The study concluded with statements that sample size should always be found through a statistical power analysis where possible [102].

5.3.2.3 Least Absolute Shrinkage and Selection Operator (LASSO)

The previous studies, which both approached the problem through logistic regression, either included few predictors or removed them by manually inspecting their individual correlation with outcome [95] [96]. Alternatively, the Least Absolute Shrinkage and Selection Operator (LASSO) could have been considered Initially, the same model is created as for MLR in equation (5.1), with all variables present, but some are then eliminated by having their coefficients shrunk to zero.

$$\min_{\beta_0, \beta} \left(\frac{1}{2p} \sum_{p=1}^{p=57} (y_p - \beta_0 - x_p^T \beta)^2 + \lambda \sum_{j=1}^v |\beta_j| \right) \quad (5.4)$$

The operator solves the minimization problem from (5.4), where p is the number of patients. y is the expected outcome, β is coefficient that is being minimized, x is the variable vector, v is the number of variables, and T denotes the transpose operation.

The LASSO approach furthermore requires that all variables are standardized prior use, so that variables are penalized in the same manner. Categorical variables are split before the standardization, which leads to the disadvantage that variables cannot directly be compared anymore [103].

5.3.2.4 Queue-based Multiple Linear Regression

Alternatively, a queue-based regression multiple linear regression (qMLR) model was designed to determine the most significant variables related to SSS at discharge, and at the same time ensure that the model would not become too complex. It shares similarities with traditional forward stepwise regression, but differs by (1) not selecting the most correlated variable with SSS in each iteration, (2) including variables if their contribution is significant to SSS even if only through interaction, and (3) not including all variables in the final model.

A queue is a data-structure where data is accessed like the real-world equivalent. In first-in-first-out queues, data is dequeued (removed from queues) in the order in which they were enqueued (inserted into the queue). Variables were stored in a circular queue where they were dequeued and tested in the MLR model. They were then either kept as part of the model or enqueued. Since categorical

predictors are represented by multiple variables, enqueueing and dequeuing occurred collectively (see Figure 28).



Figure 28: Example of first-in-first-out queue. As multiple variables are needed to represent a categorical predictor, dequeuing and enqueueing can operate with multiple variables at a time.

When predictor is introduced to the qMLR model during an iteration, its significance is tested and its variables are included in the model if $p < 0.20$. This threshold was empirically defined, but should in a future version be systemically determined. If the RMSE is furthermore lower than that of previous iterations, the current fitting model is stored as the best so far. If $p > 0.20$, the predictor is requeued. Since the entire regression model is updated in every iteration, the coefficients and p-values are also updated. Therefore, a correction step is necessary. When the entire queue was tested without any changes to the model, all predictors with $p > 0.20$ are requeued. This results in a continuously improving fitting model which includes predictors with $p < 0.20$. The process is illustrated in Figure 29.

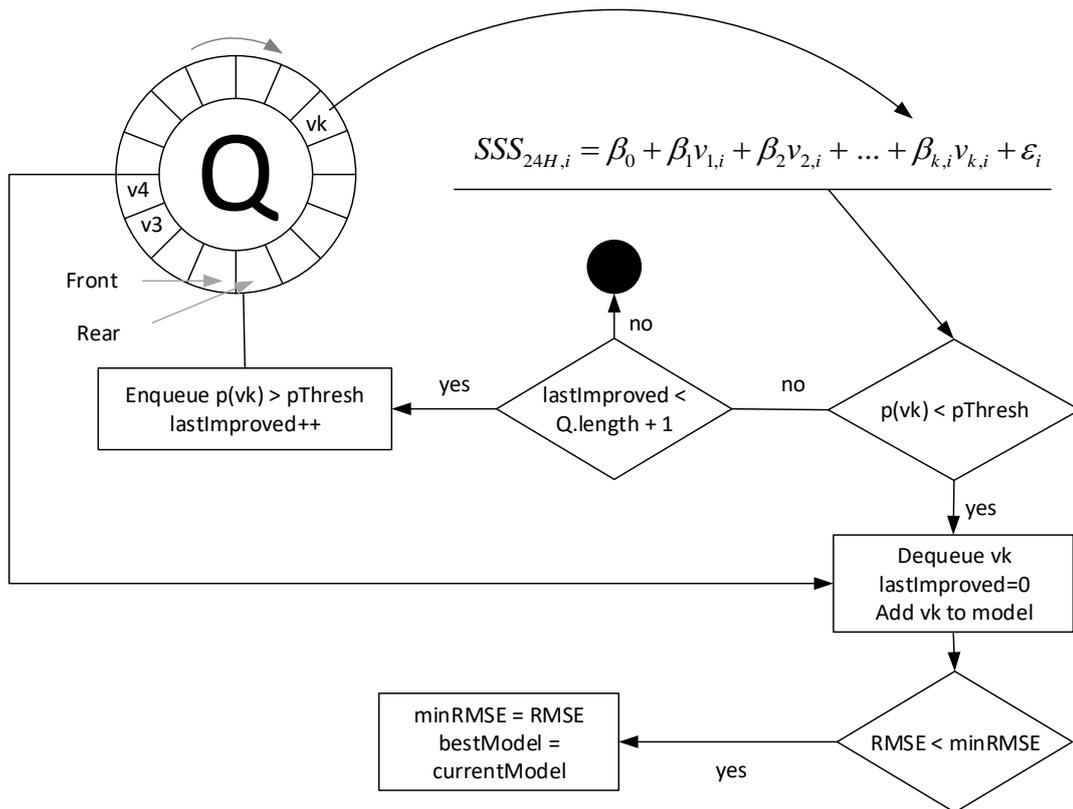


Figure 29: Generation of outcome prediction model by adding or removing predictors iteratively and testing them in a qMLR model. It terminates when remaining possible combinations with the current model no longer improve RMSE.

Before inspecting which predictors were of greatest importance to SSS at discharge, the model's robustness was verified. Depending on the order in which predictors were introduced to the MLR model, the final model may vary. This is because predictors are added and removed one by one, and interactions between variables may be overlooked.

Data was split so that seventy percent of observations were trained on, and thirty percent were left for testing. In Figure 30, both lines show the training error during fitting. The blue line represents fitting when predictors were provided to the queue in random order. The orange line shows the same, but with predictors initially ordered by their significance from the previous model. In both cases, the best model included the same variables, and fitting did not improve after reaching RMSE = 1.18. The reason for RMSE being 2.6 from start was because the first random predictor was onset SSS – which is highly correlated with SSS at discharge ($r = 0.86$). The mean SSS in the first two hours was even more correlated with final SSS ($r = 0.93$). The SSS scores from the first two hours alone were enough to fit final SSS with RMSE = 2.01, but by including other variables, a more accurate fitting was achieved.

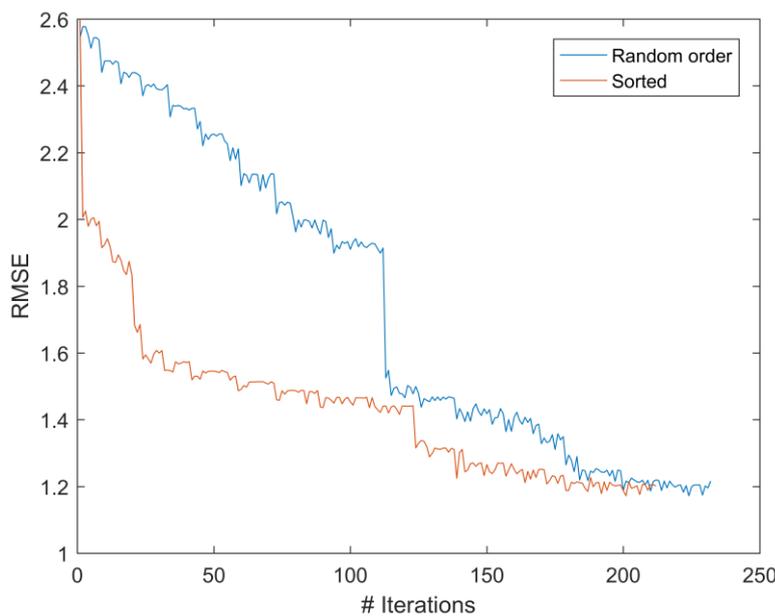


Figure 30: Optimization of qMLR during training for SSS prediction. Blue model was fitted when predictors were added randomly to the model. Orange shows fitting with predictors from previous best-fit model.

5.4 RESULTS

The two approaches were compared with respect to RMSE to see which model approximates SSS at discharge the most; with the least complex model. LASSO had 10 predictors by the end of training whereas qMLR had 11 predictors. On the test set, qMLR performed with RMSE = 3.11, whereas LASSO performed worse and reached RMSE = 5.96. One classification outlier in the qMLR model contributed to an RMSE increase of 0.55.

Figure 31 shows the predicted SSS compared to actual outcome, revealing that LASSO always undervalued SSS values.

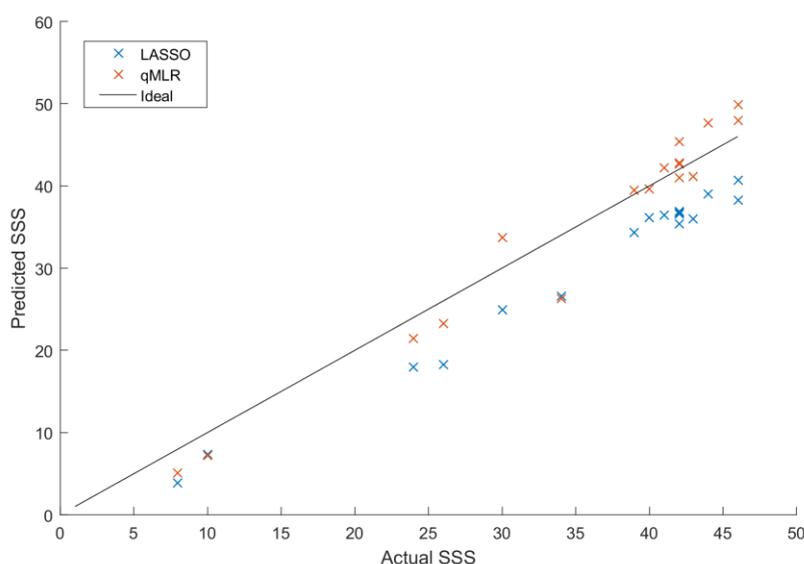


Figure 31: Prediction of SSS at discharge with qMLR and LASSO

Both approaches included four of the same predictors in their final models: μ Pulse, BP correlation, μ SSS, and mRS at onset. The remaining predictors in the LASSO were more focused on vital signs than it was the case with qMLR (see Table 13 and Table 14). An interesting observation was, that in the second to last step of qMLR training, it also included σ SBP ($p = 0.78$), hyperlipidemia ($p = 0.63$), high alcohol intake ($p = 0.34$), and arrival to treatment ($p = 0.31$), which were all present in the final LASSO model. They were all automatically removed because p-values exceeded the threshold.

Looking at the coefficients in the LASSO approach (Table 13), the greatest positive contributions towards a low disability was SSS during the first two hours, and the change in SSS during the period. Other positive contributors were an increased BP correlation and mRS at onset. The positive contribution of mRS was due to underrepresentation of other classes within the predictor. Even though a trend in data was visible where an increase of mRS overall reduced outcome by 10, several patients with mRS = 0 had very poor outcomes. Hyperlipidemia and high alcohol consumption were the strongest predictors for poor outcome. The influence of vital signs was significant enough to contribute to the model for SBP and pulse. Higher than average measurements for the population were unfavorable.

Table 14 shows the predictors that were included in the determined qMLR model. The insignificant predictors, which are still within the predefined threshold, were not removed as they contributed through interaction. The t-statistics show the major contribution of SSS, whereas other parameters are closer to each other. Unlike LASSO, no penalization was involved and parameters did not need to be standardized to assure fairness, which also allows for more direct interpretation. As an example, being male increases SSS almost by one in the dataset, which related to how males improved more than females. They were however also admitted with higher SSS. Males improved from a mean of SSS = 35.96 to 39.12 while females moved from SSS = 30.00 to 33.72.

Although alteplase treatment was expected to follow the guideline of 0.9mg/kg, and not become part of the final model because it should not vary, this was not the case, and in the qMLR it reduces SSS slightly. Closer investigation of administered alteplase revealed that six patients received a smaller

than recommended dosage and had an average outcome of SSS = 40 (38.5 at onset) while six other patients received a higher than recommended dosage and had average outcome SSS = 28.83 (23.83 at onset). The decisions of administering more than 0.9mg/kg is therefore likely due to the critical condition of the patient at admission. The dosage guidelines were still followed with respect to maximum dosage.

Table 13: Best-fit LASSO model for stroke outcome for dataset population, with RMSE = 5.96

Variable	Estimate
(Intercept)	6.718
Hyperlipidemia	-0.694
Arrival to treatment	-0.004
Alcohol_High	-0.540
σ Systolic BP	-0.002
μ Systolic BP	-0.007
μ Pulse	-0.027
μ SSS	0.952
mRS onset	0.001
σ SSS	1.577
BP Correlation	0.076

Table 14: Best-fit qMLR model for stroke outcome for dataset population with RMSE = 3.11

Variable	Estimate	SE	t-Statistic	p-Value
(Intercept)	-1.66	4.66	-0.36	0.72
Gender (Male)	0.94	0.64	-1.45	0.15
Dosage	-0.12	0.04	-3.32	<0.01
μ Pulse	-0.05	0.02	-2.76	<0.01
SSS onset	-0.49	0.07	-6.46	<0.01
μ SSS	1.61	0.09	17.46	<0.01
mRS onset	1.29	0.34	3.82	<0.01
Previous TIA	2.67	1.04	2.55	0.02
Age	-0.05	0.03	-1.21	0.07
Peripheral Artery Disease	-3.88	3.19	-1.21	0.24
BMI	0.40	0.13	3.07	<0.01
BP Correlation	3.76	0.79	4.73	<0.01

Boundaries need to be considered when interpreting such results, as is seen from SSS at onset. The reason why a higher SSS at onset has negative impact on outcome is, because a good recovery is possible even in severe cases, allowing patients to even make recoveries from SSS = 28 to 46. Since SSS = 46 is the upper boundary, patients with a high SSS at onset can numerically not improve by much.

The greatest contributor to poor outcome for the population at first appeared to be peripheral arterial disease, and although it could make sense, it is important to note that only two patients in the population had the diagnosis (see Table 10). From a medical point of view, the most unexpected results were increased BMI and previous TIA being linked to a more positive outcome in the dataset.

5.5 DISCUSSION

The combination of circular queuing and MLR provided a better model for predicting disability outcome stroke patients than LASSO did. Both approaches reduced dimensionality so that they did not overfit, and it was possible to deduce predictors related to outcome. A shortcoming was that the few samples limit how many predictors could be included, and the final models were therefore likely to have a high RMSE in test compared to training. Another limitation of both approaches lies in the possibility of overseeing contributions of variables when combined. Predictors could be left out in the qMLR, if they together significantly contributed to the model but individually, in no combination, have a $p < 20$.

Based on the results, disability after stroke (SSS at discharge) can primarily be derived by SSS during the first two hours. Contributions of other factors are much smaller and possibly overseen if they are well regulated. As an example, the arrival to treatment would be considered insignificant if all patients were treated at onset, because the dataset lacked representation on what happens when the patient is untreated for hours after onset.

From the vital signs, pulse and BP correlation coefficient contributed to the model. Patients who had no symptoms at discharge (SSS = 46) had a BP correlation coefficient of $r = 0.5$, while patients with negative BP correlation ended their first twenty-four hours with average SSS = 33.3.

In terms of treatment, dosage, modified ranking score (mRS) – and in the LASSO case, arrival at hospital to treatment start were the most important. It is interesting that onset to arrival at hospital was discarded in both scenarios, but arrival to treatment was not. A plausible explanation is that, once patients were admitted, treatment was initialized quickly unless difficulties occur. If the patient was restless, they needed a sedative before CT scan could be performed and the stroke type verified.

From a medical point of view, the most unexpected results were increased BMI and previous TIA being linked to a more positive outcome in the dataset. Some classes were underrepresented in the dataset, such as peripheral arterial disease and heart insufficiency, making them unreliable. The next step would therefore be to expand the dataset to include more patients, allowing a more realistic determination of parameters involved in stroke outcome.

5.6 CONCLUSION

It was possible to determine stroke outcome for the first day based on the first two hours, given 10 – 11 predictors that were selected by automatized respectively qMLR and LASSO. Two automatized models were implemented, which both could be improved through additional data. The advantage of qMLR over LASSO for this specific problem was not only the performance but also that it was easier to interpret results because variables were not standardized and penalized. With substantially more data and variables, LASSO may however be necessary because it is computationally faster and does not need to generate a new model in each iteration. The designed qMLR model combined with multiple linear regression shows promising results as an alternative to stepwise regression where significance and interpretability of coefficients are preferred over computational speed.

The first hypothesis was that the degree of disability following a stroke is predictable after the first day of admission, based on available data from the first two hours. A RMSE = 1.18 was achieved during

training, which increased to RMSE = 3.11 at test. Aside from the outlier classification, the remaining estimates missed by the actual disability score by SSS = -0.41 ± 2.60 . The hypothesis is therefore confirmed for ischemic stroke patients who do not require thrombectomy.

The second hypothesis investigates if the degree of disability at discharge will have been affected by vital signs. BP correlation coefficient and pulse played a significant role in determining the outcome. Both predictors were in the final models. A higher pulse was in fact associated with worse outcome. For the entire population, patients with a mean pulse > 80 during the first two hours had at discharge SSS = 30 ± 13.72 , whereas patients with pulse < 80 had SSS = 39 ± 8.23 . Patients whose symptoms were no longer present after the first day of admission had a BP correlation coefficient of $r = 0.5$, whereas an observed negative BP correlation in average resulted in SSS = 33.30.

The aim of the third hypothesis was to examine whether the degree of disability at discharge could be predicted with only the most significant parameters, and that these could be automatically detected. The presented qMLR algorithm narrowed down parameters of importance for stroke outcome from 39 to 11. LASSO even reduced them to 10. Both approaches agreed on four of the parameters. For qMLR, data about SSS from the first two hours alone lowered training RMSE to 2.01 but the remaining variables further improved accuracy and lowered RMSE. All but three of the identified parameters had $p < 0.05$. In clinical settings, LASSO's advantage would have been that it consistently predicted SSS to be lower, so medical staff would be more alert.

Compared to the overall critically ill patient population, concentrating on a homogeneous sub-group simplifies to process of creating outcome prediction models. Same treatment, complications and progress during admission made the prediction fair accurate even for patients who scored poorly on the degree of disability scale. The next chapter and study segment ischemic stroke patients further to investigate the influence of the comorbidity diabetes mellitus.

AUTOMATIC IDENTIFICATION OF DIABETES MELLITUS IN ISCHEMIC STROKE PATIENTS

Objective: Diabetes mellitus (DM) prevalence is increasing because of obesity, increasing life span length and a growing population. It is a risk factor for long-term development of complications because of damage to blood vessels. This leads to cardiovascular diseases and damage to tissue and nerves. Since the vessels are very affected by the disease, DM patients are at least twice as likely of suffering from a stroke. It is estimated that one-third of stroke patients have unrecognized DM. This study investigated whether parameters that were recorded as part of standard stroke treatment can be used to differentiate between recognized DM and non-DM patients. An accurate model that can separate between the two subgroups could contribute to revealing unrecognized DM patients and thereby ensure earlier treatment. Two scientific contributions have been submitted on this topic of which one was accepted, and the other is submitted for review (Appendices E and F).

6.1 BACKGROUND

Diabetes mellitus (DM) and hypoglycemia after stroke are associated with poor outcome weeks and months after ischemic stroke patients are discharged [45] [46] [47] [86] [88]. Studies found that while there was no difference in outcome for DM and non-DM during the first three months after admission, the worse outcome can be documented for DM patients after six months. Additionally, the mortality in DM patients increases one year after stroke in especially patients under 50 years [45] [46]. Kaarisalo et al found that twenty-eight days after admission, the recovery from stroke-related disabilities takes more than twice as long for patients with recognized DM [88].

Recognized DM patients are overrepresented in stroke units, but there are cases where patients have the disease unknowingly. Determining whether a patient has DM is challenging because the endocrine and cardiovascular systems behave abnormally for all ischemic stroke patients during the acute phase. While approximately 16% of all stroke patients have recognized DM (Table 10 and [40]), hyperglycemia, which is commonly associated with DM, is reported in 30-40% of all stroke patients [38] [43] [104]. Although HbA1c is a more accurate estimate in the acute case, it may not be sensitive enough. A group studied sixty-two post-stroke patients for twelve weeks, where HbA1c was compared to oral glucose tests, and found that potentially one third of stroke patients may have unrecognized DM [40]. In the most recent related study, Zahra et al collected data for ischemic stroke patients based on screening DM type 2 guidelines by the American Diabetes Association. They found that 20% of their population (n = 250) had the disease but were undiagnosed. Most common risk

factors were in descending order: Hypertension, smoking, hyperlipidemia atrial fibrillation, and myocardial infarction [105] [106].

No previous studies were identified that look at differences between DM and non-DM patients in stroke settings. DM should ideally be detected before the stroke occurs, but evidence suggests that it is still an overlooked problem even after the stroke. Determining the disease timely would lead to better management of comorbidities and fewer stroke reoccurrences – e.g. through better control of glycemic levels and BP [48]. Since screening for DM in post-stroke settings appears to be neglected, this study focused on an automatic approach to separating the two groups: recognized DM and non-DM stroke patients. Achieving high accuracy through automatic separation of the two groups would provide further insight into which markers to focus on when identifying DM patients, and is the first step towards finding unrecognized DM patients.

6.1.1 RESEARCH HYPOTHESIS

The affected cardiovascular system in DM patients contributes to vital signs behaving differently from non-DM patients. The differences between the two patient groups should, therefore, be evident even before stroke treatment is initialized. The differentiation can be achieved with data that is registered as part of standard stroke treatment, as they include the parameters which previous studies have found to be of importance [106].

The hypotheses were:

1. It is possible to differentiate patients with and without recognized diabetes mellitus (DM) within the first admission day following an ischemic stroke. The data that is registered as part of guidelines for ischemic stroke treatment are sufficient to achieve this.
2. In addition to commonly associated parameters with DM (high glucose level and high BMI in western countries), vital signs associated with the cardiovascular system are decisive factors in distinguishing between recognized DM and non-DM ischemic stroke patients.
3. Parameters registered according to the ischemic stroke treatment guidelines are all relevant for progress and treatment but are not all necessary for identifying patients with recognized DM. Registered parameters that are relevant for differentiating between recognized DM and non-DM patients can be determined through an automatized algorithm.

6.1.2 APPROACH OVERVIEW

The overall approach is illustrated in Figure 32 and followed the same procedure as in the previous stroke related study until the feature selection step. In the validation and correction step, data was furthermore statistically compared between the two patient groups for other potential features and differences. The significant features were determined through bidirectional feature selection and used as input for a two-class support vector machine (SVM) classifier, which was trained to distinguish the groups. The classification was tested with data from the two first hours of admission (until the alteplase is fully administered) and twenty-four hours of admission. The two-hour period was tested to see if differences were easier detectable prior treatment.

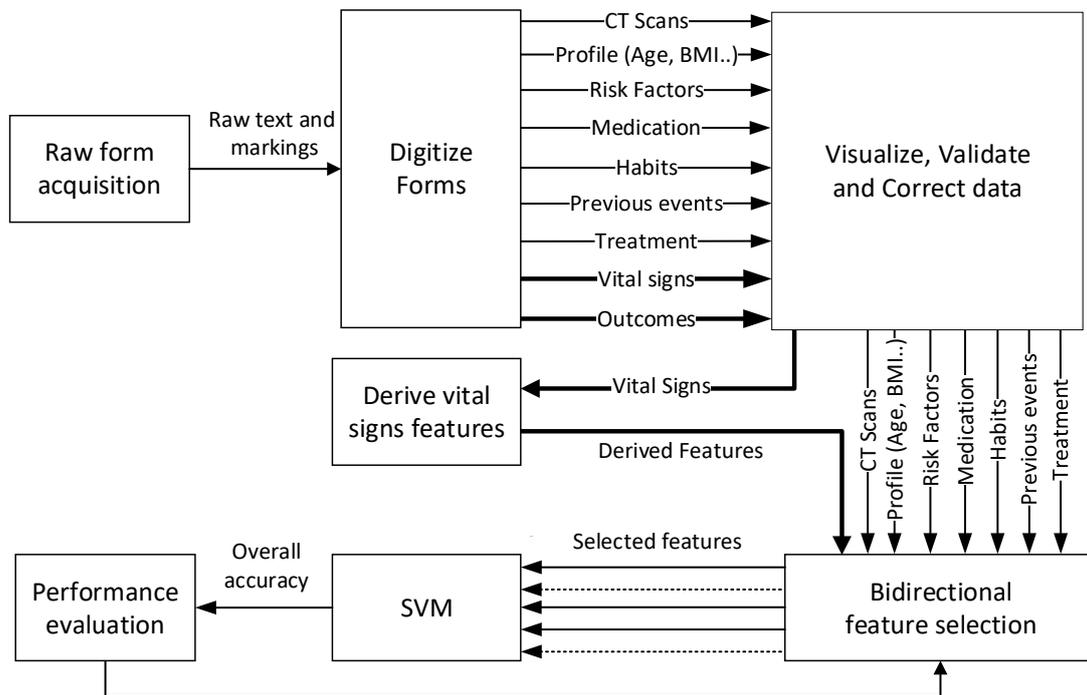


Figure 32: Approach for differentiating between recognized DM and non-DM patients. The dataset was different from the study in chapter 5, but included the same features. Bidirectional feature selection was applied in combination with a two-class support vector machine classifier for class separation.

6.2 DATA

The study started with the same dataset as in Chapter 5, and was increased in population size because only nine patients had recognized DM (Table 10). After increasing the population to seventy-eight randomly selected patients, a bias was introduced to include additional thirteen randomly selected DM patients (Figure 33). Nineteen patients were in total excluded because they were relocated to another hospital within few hours after admission. A non-DM patient was excluded from the study due to the history of gestational diabetes.

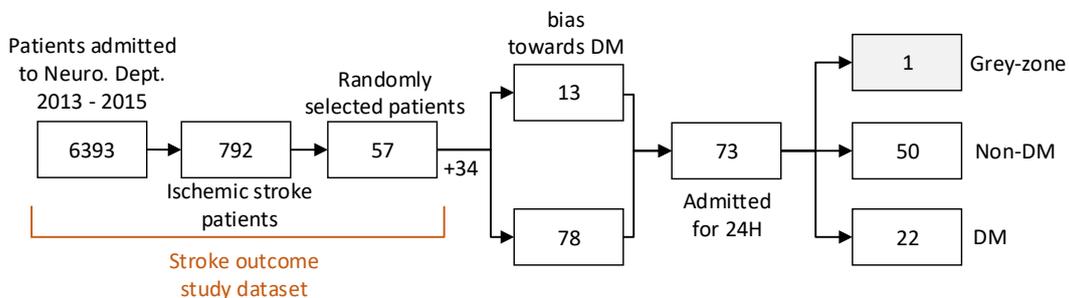


Figure 33: The dataset from stroke outcome study in chapter four was expanded by 34 patients from ZUH. Thirteen of the patients were randomly selected with a bias towards DM to have a reasonable representation of the population. One patient was excluded because of a history of gestational diabetes.

The population for this study consequently consisted of 72 patients. A patient was considered having DM if the information was registered in the national patient database, any previous medical records or the evaluation form during stroke-related admission.

Table 15 shows all parameters included in the study along with mean, prevalence or distribution where applicable. The ischemic stroke patients were separated in the third and fourth columns depending on whether they had recognized DM prior and throughout their admission. Both groups received the same treatment. Most differences between the two groups are in accordance with our findings and expectations about DM. Examples are greater hypertension, BMI, somewhat higher glucose and higher prevalence of cardiovascular diseases. Since the risk of apoplexy increases with DM, more cases of previous apoplexy were expected in the group. Consequently, DM patients have

Table 15: Digitized parameters that were registered as part of national guidelines for treatment of ischemic stroke patients, and then split for DM and Non-DM patients.

Type	Parameter	Diabetes (n=22)	Non-Diabetes (n=50)	Type
Profile	Age (years)	69.27 ± 9.91	68.12 ± 11.92	Numeric
	Weight (kg)	90 ± 18.04	75.88 ± 14.56	Numeric
	Height (cm)	171.09 ± 6.60	171.14 ± 9.62	Numeric
	BMI (kg/m ²)	30.80 ± 6.45	25.72 ± 3.39	Numeric
	Gender (female)	50 %	56 %	Ord. Binary
Imaging (CT scans)	Local bleeding	–	–	Categorical
	Remote bleeding	–	–	Categorical
	Cerebral edema	–	–	Categorical
Treatment at hospital	mRS onset	0.55 ± 0.91	0.32 ± 0.96	Ordinal (0-5)
	Onset to arrival (min)	124.68 ± 60.67	105.68 ± 43.71	Numeric
	Arrival to treat. (min)	33.36 ± 21.15	31.04 ± 14.91	Numeric
	Dosage rt-PA (mg)	78.27 ± 10.04	69.10 ± 13.32	Numeric
	Glucose (mmol/L)	9.04 ± 2.96	6.47 ± 1.37	Numeric
Vascular risk factors	Arterial fibrillation	23 %	14 %	Ord. Binary
	Heart insufficiency	18 %	4 %	Ord. Binary
	Hypertension	86 %	58 %	Ord. Binary
	Hyperlipidemia	86 %	66 %	Ord. Binary
	Periph. arterial disease	18 %	0 %	Ord. Binary
Previous events	Acute myocard. Infrac.	18 %	10 %	Ord. Binary
	Prev. apoplexy (ever)	50 %	24 %	Ordinal (1-3)
	Trans. ischemic attack	23 %	8 %	Ord. Binary
Relevant drugs	Acetylsalicylic acid	41 %	21 %	Ord. Binary
	Clopidogrel	32 %	10 %	Ord. Binary
	Dipyridamole	9 %	2 %	Ord. Binary
	Anti-hypertensive	64 %	34 %	Ord. Binary
Habits	Alcohol (> recommend)	4 %	2 %	Categorical
	Smoking (prev. / curr.)	77 %	46 %	Ordinal (1-4)
Outcomes	Bleeding complications	4 %	0 %	Ord. Binary
	mRS (after 3 months)	2.36 ± 1.94	2.16 ± 1.69	Ordinal (0-6)
	Outcome 24h	–	–	Ordinal (1-5)
	Outcome discharge	–	–	Ordinal (1-5)
Neurological assessment (24H monitoring)	Scandinavian Stroke Scale, NIH Stroke Scale			Numeric
Vital signs (24H monitoring)	Pulse, temperature, blood pressure, oxygen saturation level			Numeric

greater disabilities prior reoccurred stroke – evaluated by modified Ranking Score (mRS). The time from arrival at hospital to treatment indicates that circumstances during treatment were similar, so it is interesting that the time from onset to arrival at hospital is longer for DM patients. A possible explanation is that reduced mobility (higher mRS), and heavy weight contribute to slower boarding of ambulance. Although HbA1c may be a useful predictor when identifying DM, it is not part of national guidelines for stroke treatment yet, and therefore not measured in most cases [33]. It was possible to obtain most recent HbA1c measurements which were taken at different points unrelated to the ischemic stroke admission for all but three patients. The rest had values that were under 42 mmol/mol, meaning that both recognized and unrecognized DM patients were well regulated.

6.2.1 INTERVARIABLE COMPARISON FOR DM AND NON-DM

The data collection and digitization process followed the same procedure as the previous study and are described in detail in section 5.2. From a medical point of view, it was afterwards of interest to see intervariable relationships for DM versus non-DM patients. Figure 34 shows the intervariable relationships for each of the groups through computed and colored correlation coefficients for each pair of parameters; as well as the difference between groups. By splitting the data into two populations, parameter relations were also influenced. As an example, an expectedly high correlation between height and weight was in the mixed population $r = 0.53$, but increased in non-DM (to $r = 0.76$, $p < 0.01$) compared to DM ($r = 0.16$, $p < 0.01$).

Starting with the non-DM patients, a high correlation in the profile parameters was visible (Figure 34, top left) along with dosage which was primarily administered based on weight. As mRS rose, a higher proportion of patients had atrial fibrillation, with everyone with $mRS > 2$ having the condition. Note that only 12% of non-DM patients had $mRS > 1$. Previous stroke incidents (PrevApo) stand out together with Clopidogrel, but do so only because no patients received Clopidogrel without a previous event of stroke, while both patients who had stroke within three months were on the medication. Looking at patients who suffered from stroke more than three months ago, five of twelve were on the medication. There were furthermore correlations between antihypertensive medication and a history of hypertension, and previous stroke events were related to previous TIA events which was also unsurprising. Finally, there were no incidents of peripheral artery disease.

Moving on to the DM group, smaller relationships were present in in profile parameters. Patients with previous cases of stroke (PrevApo) had lower glucose level at admission ($n = 11$, 7.54 ± 1.86 mmol/L) than patients without previous events ($n = 11$, 10.53 ± 3.01 mmol/L). This could indicate that DM patients with a history of stroke were more aware of their medication intake. While there were only four cases of transient ischemic attack, none of the affected DM patients were on anti-hypertensive medication. Similarly, hypertension and glucose were negatively related ($r = -0.53$, $p < 0.01$), because two of totally three DM patients without hypertension had the highest glucose levels of the entire population. Dipyridamole was positively correlated with arrival to treatment ($r = 0.81$, $p < 0.01$), but since only two patients received Dipyridamole it was not possible to reflect on the finding. The relationship between acetylsalicylic acid and Clopidogrel did not stand out for non-DM patients, because the majority received neither of the two. It does however for DM patients ($r = -0.57$, $p < 0.01$), where most took one the medications, but never simultaneously. Further research revealed that the national DM treatment guidelines recommend patients to be on acetylsalicylic acid, unless the patient responds poorly, in which case Clopidogrel is used [107].

Between the two populations, there were four major differences (Figure 34, bottom figure). The most evident difference being absence of peripheral artery disease in non-DM patients. Next heart insufficiency was negatively correlated with weight in non-DM patients ($r = -0.22, p < 0.01$) whereas

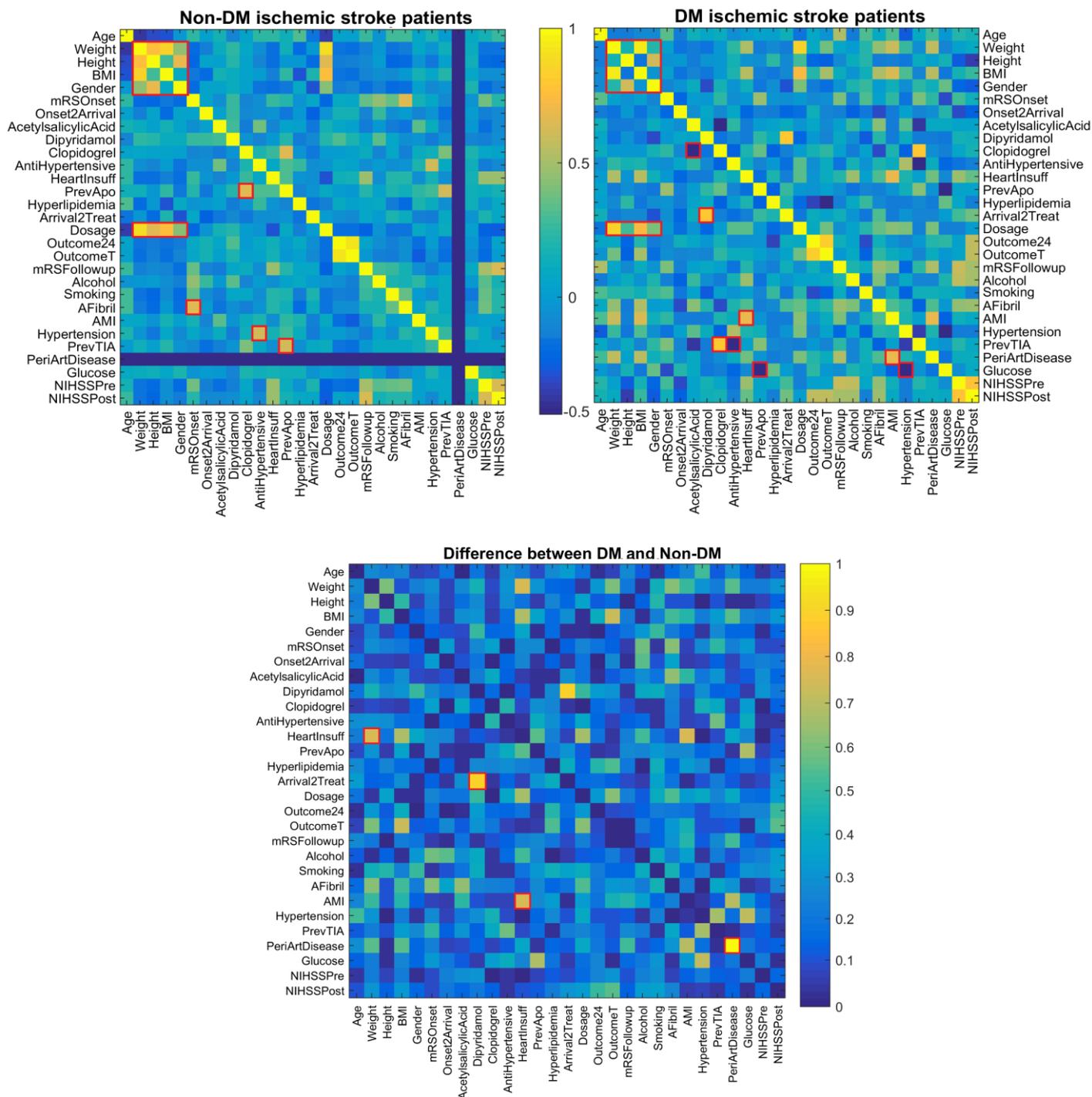


Figure 34: The top figures show Intervariable correlation coefficients for respectively non-DM and DM ischemic stroke patients. Peripheral artery disease was only present in the DM population. The highest and lowest correlating parameters are outline with red. The bottom figure illustrates the largest differences between the two other correlation matrices.

it was positively correlated for DM patients ($r = 0.55$, $p < 0.01$). The correlation was expected to be higher for DM, but positive for both populations; the cause for negative correlation likely being the underrepresentation of heart insufficiency in non-DM population. The arrival to treatment relationship with Dipyridamole was, as previously mentioned, caused by underrepresentation of data. Finally, most DM patients who suffered from heart insufficiency also suffered from acute myocardial infarction (AMI) in the diabetes population, while the same relationship was not observed in non-DM patients. The same association has been documented in literature [108] [109] [110].

6.3 METHODS

This section describes the classification process used for differentiating between recognized DM and non-DM patients that were admitted at ZUH following an ischemic stroke. It describes how features were selected, and the considerations regarding classifier design.

6.3.1 FEATURES

The same features were used in this study as described in section 5.3.2 including the derived ones. The correlation between SBP and DBP was particularly interesting because it is expected to behave differently for DM patients, but it has not been tested in stroke settings [25]. The BP correlation along with Scandinavian Stroke Scale (SSS) are shown in Figure 35 after two hours of admission and again after twenty-four hours. As in chapter 5, the expectation was that was a stabilization around $r = 0.74$ for non-DM patients. The bold markers represent the center of mass for each group. DM patients.

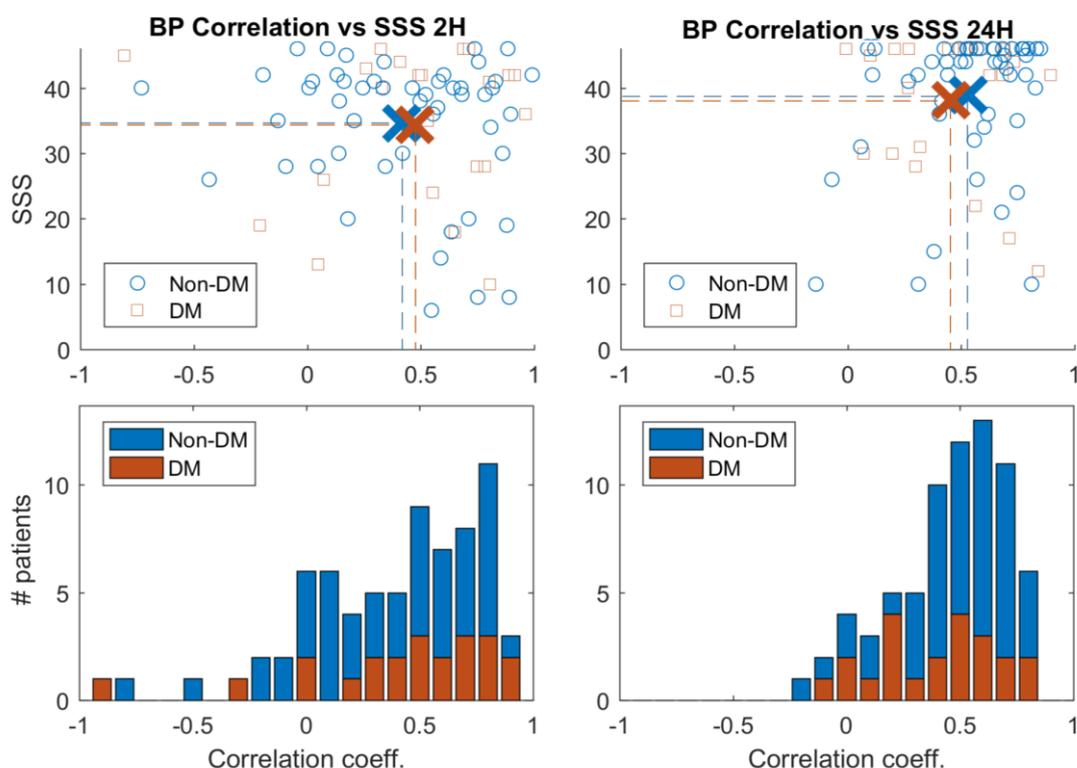


Figure 35: Correlation between diastolic and systolic blood pressure, two and twenty-four hours after initialized treatment, with center of mass outlined. Center of mass for non-DM moves from {0.42, 34.58} to {0.53, 38.92}, whereas DM move from {0.47, 34.36} to {0.45, 38.14}.

improved in SSS by an average of 3.77 while non-DM improved by 4.34. In terms of correlation, a decrease was seen in DM patients from $r = 0.47$ to $r = 0.45$, while it increased in non-DM from $r = 0.42$ to $r = 0.53$. The histograms show the correlation distribution with a convergence towards 0.6 for non-DM over time, and slightly positive, yet widely distributed, correlation for DM patients. All correlations were significant ($p < 0.01$).

After increasing the number of included patients from previous study, it was still not sufficient to deal with the low patient to parameter ratio. Classifier would be overfitted and have poor generalization if all parameters were to be included as features. To reduce dimensionality, principal component analysis (PCA) was considered. The way it works is by placing data in an n -dimensional space of which the center is the mean value of each parameter. It then finds the direction with the greatest variance and projects all data onto the component. This is repeated to create other components under the condition that each variance computation must be orthogonal compared to its predecessors [111]. PCA was discarded, as it is not suited for nominal and binary parameters as it operates on variance which only makes sense to compute for continuous data.

6.3.2 CLASSIFICATION THROUGH SUPPORT VECTOR MACHINES

There are two types of automatic classification algorithms: supervised and unsupervised. The first category is trained on features, and is given information about which class given features represent, so that a pattern can be detected. Unsupervised classifiers on the other hand attempts to detect classes themselves, and may therefore detect fewer or more classes than exist. With sufficient samples per feature, an unsupervised classifier would be preferred for distinguishing between the patient groups because it is likely that some patients have unrecognized DM [40] [106]. Since supervised classifiers recognize provided patterns, they are in risk of being taught to classify a patient as non-DM who may have unrecognized DM. Still, supervised classification was chosen for this study because the low patient to parameter ratio would cause unsupervised classifiers to generate many classes based on few observations per class. This problem is dealt with in supervised learning by having a fixed set of output classes to which only a subset of features is provided.

The supervised classifier chosen to distinguish the two classes was a support vector machine (SVM) (Figure 36) [112] [113] [114] [115]. It automatically detects decision boundaries based on provided features, but also introduces a “slack” parameter to account for noise and outliers – a likely scenario as some non-DM patients were suspected to have DM. Classes are separated through a hyperplane (black line in figure) which is created between them so that they are furthest possible apart. The support vectors are then the data points closest to the hyperplane and are located on the margin (orange lines) which is an area perpendicular with the hyperplane. They are the points that resemble the opposing class the most. After training, data points that lie within the margin belong to neither class unless a soft margin is introduced.

The separating hyperplane and the margin area are defined in (equation 6.1), where \mathbf{w} and \mathbf{w}_0 respectively indicate its direction and position, and \mathbf{x} is the input sample. All input vectors at are on either side of the margin ($\geq +1$ or ≤ -1) will be assigned to their corresponding class.

$$\begin{aligned} \text{hyperplane: } g(\mathbf{x}) &= \mathbf{w}_T \mathbf{x} + \mathbf{w}_0 = 0 \\ \text{margin:} & \quad \mathbf{w}_T \mathbf{x} + \mathbf{w}_0 = \pm 1 \end{aligned} \tag{6.1}$$

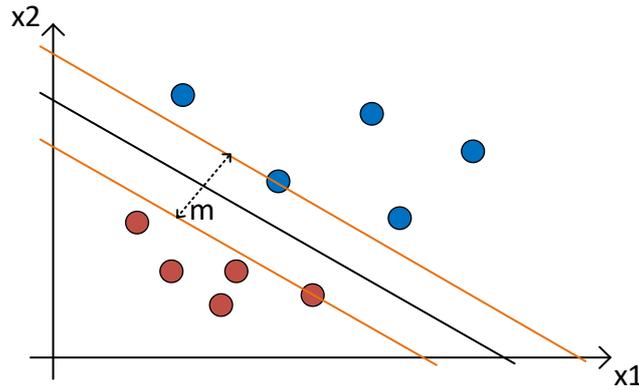


Figure 36: Support Vector Machine example in two dimensions, separated by a line which becomes an n-dimensional hyperplane when n parameters are present. The area between orange lines is the margin and data points closest to the margin are the support vectors.

While this works well in training and generation of the model, there is a risk that yet unknown samples to the classifier end up in the margin. It is for those cases necessary to have a soft margin, so that every sample gets assigned to a class, but with a penalty that increases with distance from boundary. If a sample is on the correct side of the hyperplane and outside the margin, its “slack variable” is zero, $\xi = 0$. If it is on the correct side of the hyperplane and within the margin is $0 < \xi \leq 1$, whereas samples with $\xi > 1$ indicates misclassification.

The optimal hyperplane for distinguishing the two classes can then be found through solving the minimization problem in 6.2. C is a box constraint that controls the penalization so that misclassified samples are punished harder, and n is the sample number.

$$\min_{\xi_n \in \mathbb{R}^+} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n \right\} \text{ subject to } y_n (\mathbf{w}_T \mathbf{x}_n + w_0) \geq 1 - \xi_n \quad (6.2)$$

Features may not always be linearly separable, but SVM are inherently linear. The problem can be dealt with by mapping the features into a higher dimension and then applying the linear classifier [116]. By Applying the so-called kernel trick, achieves this while remaining in the linear feature space. This is done by applying a kernel function which differs based on application. Several kernels were tested for separating DM from non-DM patients, and the best results were achieved with a quadratic polynomial ($q = 2$) SVM, with its corresponding kernel function (6.3). \mathbf{x} and \mathbf{z} are input vectors.

$$K(\mathbf{x}, \mathbf{z}) = (\gamma \mathbf{x}_T \mathbf{z} + 1)^q \quad q > 0 \quad (6.3)$$

The SVM kernel scale, γ , was found through a heuristic subsampling procedure, which was empirically determined by built-in MATLAB functions.

Features were selected by using bidirectional feature selection, where forward selection first included features that increased classification accuracy of the SVM. Features were then individually tested eliminated to see whether accuracy changed after their removal. From the parameters available in Table 15, the ones with no or few occurrences in either group were removed (bleeding related to

dosage, bleeding findings in CT/MR, cause of death), as well as irrelevant/redundant (discharge diagnosis, Outcome 24h, Outcome discharge, NIHSS).

The data was split for training and testing using 5-fold cross validation, with at least four DM patients presented in each fold. Because of ratio differences in parameters, the data was standardized prior to training through estimation of z-scores.

Two alternative classifiers were considered for implementation. The first being logistic regression which could have been approached similarly as in the previous chapter, and the second being feed-forward backpropagation neural network. Logistic regression could provide approximately same results if the SVM had a linear kernel, and was discarded because SVM can solve more complex problem when using other kernels. The primary disadvantage of neural networks was the training time which is longer than for SVM, and would become worse in this problem because the features of the final model were unknown. Each time that features would change, the network structure would need to be changed several times, because it may contain too many or too few so-called hidden neurons.

6.4 RESULTS

The best performance was obtained when combining medical history with vital signs monitoring of the initial two hours of admission. For both two and twenty-four hours, the same features resulted in best performance (Table 16).

Table 16: Features determined to give best separation between DM and non-DM patients based on both the first two and twenty-four hours of admission

Profile + History	Admission	Vital signs
BMI	From onset to arrival time	SBP Mean
Previously had apoplexy	From arrival to treatment time	SBP Standard Deviation
Previously had TIA	Dosage	DBP Mean
History of hypertension	Glucose at arrival	DBP Standard Deviation

Figure 37 shows confusion matrices that summarize which classes the SVM classifier determined that patients belonged to, compared their actual classes. The matrices are shown with respectively two hours and twenty-four hours of vital signs data. Most misclassifications happened with DM patients under both conditions. This was expected because the recognized DM class was underrepresented, and there were few samples from the DM group that could assist as support vectors. It is also possible that some non-DM patients had unrecognized DM and were so alike well-regulated DM patients that they were inseparable [40].

The top performance, with an accuracy of 87.5%, was obtained with two hours of data. Increasing measurements of vital signs from two hours to twenty-four hours caused the accuracy to drop to 80.6%. A plausible cause supports the assumption that patient groups have more differences at stroke onset, e.g. DM patient's autonomic nervous system is more affected than it is for non-DM [44]. Both groups become more stable as time passes, thereby making them more similar and harder to distinguish. It is important to keep in mind that even though the performance is high, over two third

of the population belong to the same class. If a classifier had found all patients to be non-DM for this dataset, the accuracy would consequently be 69.4%.

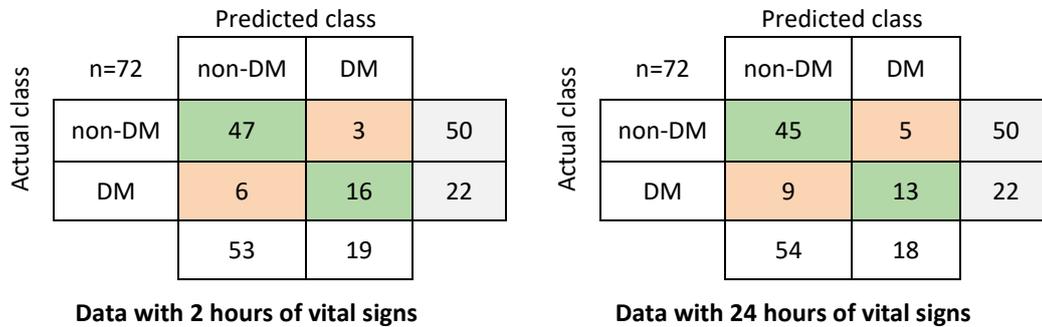


Figure 37: Confusion matrices for SVM given medical history and respectively two hours and twenty-four hours of admission data. The classifier achieved an accuracy of 87.5% with two hours of data (left side), whereas accuracy dropped to 80.6% with 24 hours (right side).

The optimal classifier was further examined to investigate causes for misclassifications, and the possibility of dealing with them. If the classifier was uncertain which of the two classes a sample belongs to, it may be improved through thresholding or adjusting the soft margin in the SVM. Figure 38 shows scoring levels of the SVM classifier for each patient. Each patient is represented by a bar that is colored accordingly to the patient's class. The scoring level was afterwards negated for misclassified cases. As an example, the seventeenth patient was registered as not having DM, but the classifier determined the patient to belong to the recognized DM population with classification score level = 0.75. The high classification score level in the misclassification cases revealed that either more samples or more/better features were required to improve classification. Misclassified non-DM patients had a history of hypertension and received a higher than average dosage of rt-PA; one received the maximum allowed dosage, and another received 6% higher dosage than recommended for their weight, but still within the guidelines. Two of the three patients had glucose levels greater than one standard deviation above average. The last patient had glucose level slightly above average, but additionally also a history of TIA and apoplexy. The misclassified DM patients had no features that stood out.

The twelve standardized features used in classification are shown in Figure 39. Line colors again represent the correct classes, misclassifications are represented through dashed lines. As in Table 15, increased mean and standard deviation in BMI and glucose for DM patients is visible. It also shows that misclassified non-DM patients had a history of hypertension whereas only one misclassification was unrelated to hypertension.

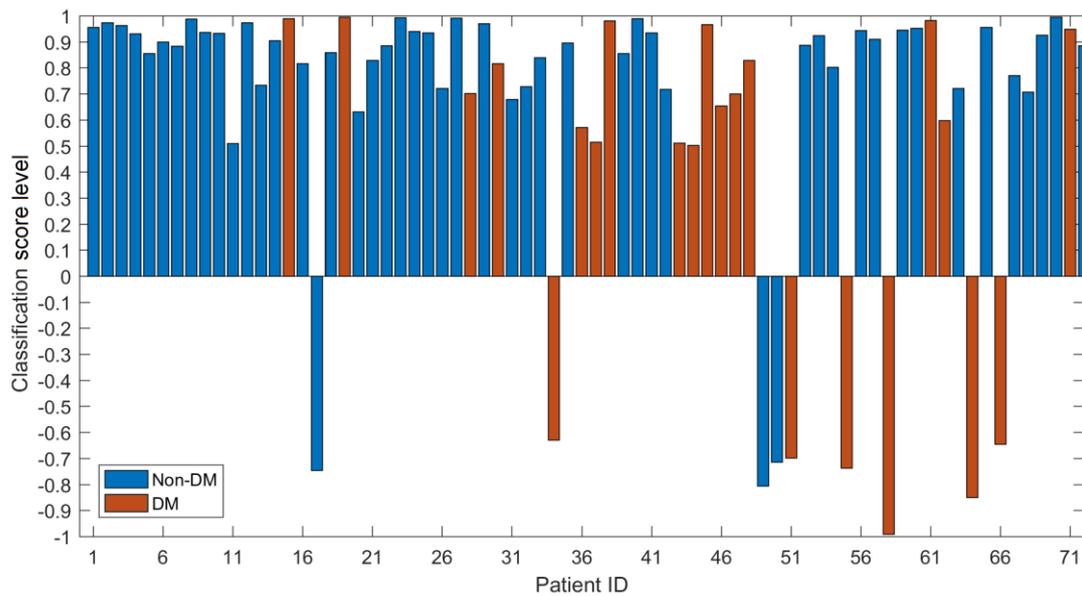


Figure 38: Classification score levels for Support Vector Machine classifier. Each patient is colored based on their expected class. The score level is negated for misclassifications.

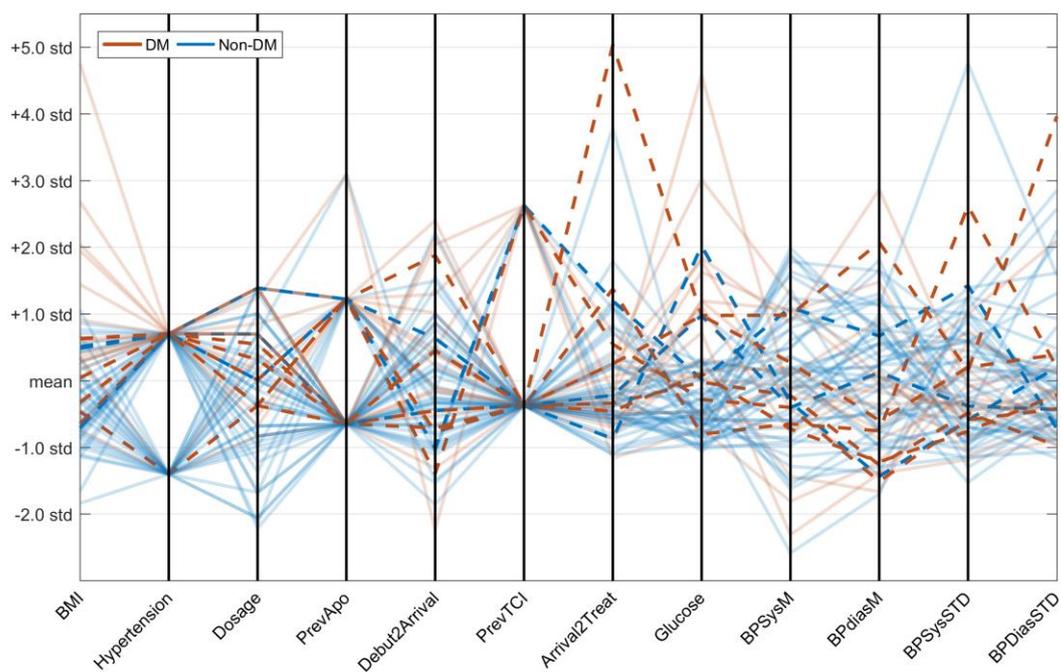


Figure 39: Concurrent standardized feature comparison. Line colors indicate the correct class, and are bold and dashed in case of misclassifications.

6.5 DISCUSSION

SVM identified, in combination with bidirectional feature selection, twelve features in the dataset for detecting whether ischemic post-stroke patients had recognized DM. Even though only four vital sign measurements per hour were registered during the first two hours of admission, BP measurements were among the most contributing features. It was also unsurprising that dosage was selected as one

of the primary features after BMI, as administration guidelines for rt-PA take weight into account. The intervariable dependency is not seen as problematic in the model, because dosage varies when considering other factors, such as age and degree of disability. From the most dominant features determined by Zahra et al, only a history of hypertension was in this study identified as decisive. All parameters mentioned in the previous research were included in this study. The previous research did not explicitly mention all included parameters, and it is therefore unknown if they also included the remaining eleven parameters that were included in the final model [106].

The SVM classifier demonstrated that the two patient groups were separable, by correctly identifying whether a patient was diagnosed with DM in 87.5% cases. Misclassifications were likely primarily caused by the differences in population sizes and presence of unrecognized DM. No unrecognized cases of DM were found while reading the patient's textual EHR related to the admission. Most entries in the EHR did however suggest that there was suspicion of DM, and this was further investigated in the department of neurology, but finally left for the general practitioner to follow up on.

Vital signs measurements were initially investigated for the first twenty-four hours of admission. As patients improve over time during admission (see Figure 35), their vital signs become more similar and they become harder to distinguish. One exception was pulse, where the groups were admitted with 81.72 ± 1.43 bpm, and were 3-5 bpm apart until the twelfth hour, after which DM patients remain around 75.38 bpm but non-DM slowly drop for three more hours and then stay at 67.61 bpm. Pulse could therefore still serve as a feature after twelve hours of admission, but it is first necessary to understand why the phenomenon occurred. A decrease in BP was observed for both patient groups from admission start until the fifth hour of admission after which it steadily increased for DM patients whereas it continued to decrease for non-DM patients until the twelfth hour. After reaching minima, a steady increase was observed for both groups. It would have been interesting to investigate how BP changes after ischemic stroke for the two patient groups since DM patients have increased risk of impaired cerebral autoregulation. This would however require measurements to be monitored at a resolution in seconds instead of every fifteen minutes [117]. The accuracy drop from that occurs when using twenty-four hours of data instead of two-hours, and the changes in development of vital signs over time illustrate the necessity of monitoring regularly so that small changes are traced.

As in previous studies on this subject, a limitation was the lack of HbA1c as a potential feature, which is expected higher for patients with unrecognized DM [40] [106]. Although some physicians perform the measurement, it is currently not a guideline requirement to do so. It was possible to retrieve the most recent HbA1c sample measurements for most patients, but they were in almost all cases taken at times which no longer covered the period of ischemic stroke admission. All values were lower than 42 mmol/mol which is lower than the national boundary of 48 mmol/mol for a DM diagnosis [39], and indication that even DM patients were well controlled.

To further improve the current approach, vital signs should be monitored at a higher resolution to detect smaller changes, and Hb1Ac readings at time of admission, or within weeks after stroke need to be available. Follow-ups are necessary to identify and correct for unrecognized DM.

6.6 CONCLUSION

Promising results have been achieved for automatically identifying DM in early hours after stroke, which, when combined with information from follow-ups after discharge, could detect patients with unrecognized DM.

The first hypothesis was that it is possible to differentiate between DM and non-DM patients within the first day of admission following an ischemic stroke, given already registered data. This was confirmed through the developed classifier, which successfully identified which group patients belong to with an accuracy of 87.5%. The approach was only based on data that were already available and are registered as part of standard treatment.

The second hypothesis assumes that vital signs are decisive for distinguishing between DM and non-DM patients – especially the ones related to the cardiovascular system (pulse, BP and oxygen saturation level). The hypothesis was partially confirmed but requires data at a higher resolution to be conclusive. Oxygen saturation varies very little and would be addressed by staff if it falls below 90, which are reasons for the parameter not being present in the final model. Through bidirectional feature selection, BP parameters were automatically selected as being some of the greatest contributors. Significant changes in pulse were observed between the two patient groups twelve hours after admission, but the best model was determined by only using vital signs for the first two hours. While the classifier accuracy is likely to improve if pulse was included after twelve hours, the lack of physiological explanation for the phenomenon or a very large dataset are required before one can justify mixing data from different periods of admission.

The third hypothesis was that not all registered parameters are necessary to differentiate the two patient groups, and that the relevant parameters could be automatically determined. This was confirmed through the approach of combining bidirectional feature selection together with a SVM. The feature selection narrowed down the number of necessary parameters to separate the patient groups to twelve.

The first two studies demonstrated how vital signs in critically ill patients can be used to create models that explain current and future health based on progress. This study demonstrated that they can also be used to differentiate between different patient groups and thereby be used to detect missed diagnoses. In all three studies, vital signs were valuable for evaluation, prediction and classification, but they also demonstrated the necessity of context-awareness through additional supplemented data.

CONCLUSION

The aim of the project was investigating feasibility and applicability of using vital signs for critically ill patients. This was conducted by starting with a broad population of all high-risk patients, and then moving on to the subgroup of ischemic stroke patients. Finally, ischemic stroke patients were segmented into DM and non-DM patients with the prospective applicability of reducing readmissions. Three overall objectives, that are also stated in the first chapter, were investigated.

- Investigating clinical usability of combining multiple vital signs in critical settings to estimate the health condition of a patient at any given time.

Previous studies have investigated EWS as a model for outcome prediction in terms of mortality, but have been unable to address how well it approximates current health status. The first study in this project resulted in the development of a prototype system that deals with the problem by combining EWS and EHR. The output is an interactable timeline of the patient's health condition, that enabled for direct comparison between how staff perceive condition against model estimates. The research thus formed the basis for further improvement and development of models for estimation of current health condition.

- Determining how vital signs are related to degree of disability in a homogenous patient group such as patients treated for ischemic stroke.

Ischemic stroke is mainly treated with a two-hour administration of alteplase. Progress is evaluated through degree of disability in neurological functions, and vital signs are monitored as precaution for sudden decline in health. In this research, the degree of disability after twenty-four hours was predicted with an error of RMSE = 3.11, by using only two hours of vital signs measurements and medical history data. BP and pulse were the most important vital signs for outcome. The correlation coefficient between SBP and DBP was $r = 0.5$ for patients with no disabilities at discharge (SSS = 46), while patients with a negative correlation had an average SSS = 33.30. An average pulse > 80 resulted in a 23% worse outcome.

- Examining feasibility of using vital signs and personalized data as identifiers for diabetes mellitus in stroke unit patients.

An impaired ANS in DM patients suggests that vital signs may change differently in events such as ischemic stroke, making it an opportunity to detect presence of yet unrecognized DM. The third study in fact showed that recognized DM and non-DM patients are more diverse during the first two hours

of admission than later. An implemented SVM identified with an accuracy of 87.5% whether a patient had recognized DM. Four of the twelve contributing features derived from SBP and DBP.

This research has overall provided novel approaches for interpretation and applicability of vital signs data for critically ill patients. It has demonstrated that interpretation of vital signs greatly benefits from synergy with admission related data to solve problems related to evaluation, prediction and classification.

7.1 FUTURE PERSPECTIVE

Our findings show the importance of context-awareness when dealing with vital signs monitoring through all three studies. Specific pathological events are detectable through uni- and multimodal combination of vital signs, but for overall condition assessment, it is necessary to include historical and current data about the patient. Fortunately, with increasingly more data being digitized, this will be possible in near future.

The following future work can improve the applicability of developed models:

1. Medication, medical history and data about current circumstances (e.g. connected medical devices) need to be included in models that estimate physiological condition based on multiple vital signs. This includes time of day, so that the temperature related circadian rhythm can be modelled and changes to the ANS be detected [38].
2. Vast untapped knowledge exists in EHR. Through word frequency counting, much could be derived about patient's admission, complications and progress. Implementing sophisticated NLP algorithms such as semantic relatedness, n-grams and stemming will automatize the process of interpreting content.
3. The second and third studies provided very promising results but had low observation to parameter ratio. The number of patients should be increased and parameter significance should be evaluated while all parameters are present in the model. The data already exist in paper form but need to be digitized.
4. Most alarms in controlled ICU settings where the patient is lying in the bed most of the time do not require actions to be taken. Once vital signs monitoring is deployed outside the ICU, algorithms need to ensure that correct feedback (e.g. disconnected electrodes) is provided to staff to avoid even greater overloads in alarm rate.
5. Vital signs need to be monitored continuously to allow for advanced time-series algorithms to be applied. Even with up to four measurements per hour, it was possible to see that BP correlation was more positive for non-DM patients that pulse stopped improving earlier in DM patients.

REFERENCES

- [1] »Future of an Ageing Population,« Government Office for Science, 2014.
- [2] W. W. Hung, J. S. Ross, K. S. Boockvar og A. L. Siu, »Recent trends in chronic disease, impairment and disability among older adults in the United States,« *BMC Geriatrics*, årg. 11, nr. 47, 2011.
- [3] T. M. Dall, P. D. Gallo, R. Chakrabarti, T. West, A. P. Semilla og M. V. Storm, »An Aging Population And Growing Disease Burden Will Require A Large And Specialized Health Care Workforce By 2025,« *Health Affairs*, årg. 32, nr. 11, pp. 2013-2020, 2013.
- [4] J. P. Boyle, T. J. Thompson, E. W. Gregg, L. E. Barker og D. F. Williamson, »Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence,« *Population Health Metrics*, årg. 8, nr. 12, 2010.
- [5] E. S. Huang, A. Basu, M. O'Grady og J. C. Capretta, »Projecting the Future Diabetes Population Size and Related Costs for the U.S.,« *Diabetes Care*, årg. 32, pp. 2225-2229, 2009.
- [6] R. Waldeyer, R. Brinks, W. Rathmann, G. Giani og A. Icks, »Projection of the burden of Type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040,« *Diabetic Medicine*, årg. 30, nr. 8, pp. 999-1008, 2013.
- [7] M. Javanbakht, A. Mashayekhi, H. R. Baradaran, A. A. Haghdoost og A. Afshin, »Projection of Diabetes Population Size and Associated Economic Burden through 2030 in Iran: Evidence from Micro-Simulation Markov Model and Bayesian Meta-Analysis,« *PLOS One*, årg. 10, nr. 7, 2015.
- [8] M. Elliott og A. Coventry, »Critical care: The eight vital signs of patient monitoring,« *Critical care*, årg. 21, nr. 10, pp. 621-625, 2012.
- [9] RJM Morgan, F. Williams og M. Wright, »An early warning scoring system for detecting developing critical illness,« *Clinical Intensive Care*, årg. 8, nr. 100, 1997.
- [10] R. Morgan og M. Wright, »In defense of early warning scores,« *British Journal of Anaesthesia*, årg. 99, nr. 5, pp. 747-748, 2007.
- [11] H. Shubin og M. H. Weil, »Efficient Monitoring with a Digital Computer of Cardiovascular Function in Seriously Ill Patients,« *Annals of Internal Medicine*, årg. 65, nr. 3, pp. 453-460, 1966.
- [12] B. J. Drew, P. Harris, J. K. Zègre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding og X. Hu, »Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients,« *PloS one*, årg. 9, nr. 10, p. e110274, 2014.
- [13] W. R. Mower, G. Myers, E. L. Nicklin, K. T. Kearin, L. J. Baraff og C. Sachs, »Pulse Oximetry as a Fifth Vital Sign in Emergency Geriatric Assessment,« *Academic Emergency Medicine*, årg. 5, nr. 9, pp. 858-965, 1998.
- [14] U. Kyriacos, J. Jelsma og S. Jordan, »Monitoring vital signs using early warning scoring systems: a review of the literature,« *Journal of Nursing Management*, årg. 19, pp. 311-330, 2011.
- [15] D. R. Prytherch, G. B. Smith, P. E. Schmidt og P. I. Featherstone, »ViEWS - Towards a national early warning score for detecting adult inpatient deterioration,« *Resuscitation*, årg. 81, pp. 932-937, 2010.

- [16] J. Smith og R. Roberts, *Vital Signs for Nurses*, Wiley, 2015.
- [17] R. Sparks, B. Celler, C. Okugami, R. Jayasena og M. Varnfield, »Telehealth Monitoring of Patients in the Community,« *Journal of Intelligent Systems*, årg. 25, nr. 1, pp. 37-53, 2016.
- [18] T. J. B. Institute, »Vital signs,« *Best Practice: Evidence Based Practice Information Sheets for Health Professionals*, årg. 3, nr. 3, 1999.
- [19] W. Jänig, »Autonomic Nervous System,« i *Human Physiology*, Springer, 1989, pp. 333-370.
- [20] C. Leon, R. A. Samson, S. M. Schexnayder og M. F. Hazinski, *Pediatric Advanced Life Support: Provider Manual*, American Heart Association, 2012.
- [21] A. D'Souza, A. Bucchi, A. B. Johnsen, S. J. R. Logantha, O. Monfredi, J. Yanni, S. Prehar, G. Hart, E. Cartwright, U. Wisloff, H. Dobryznski, D. DiFrancesco, G. M. Morris og M. R. Boyett, »Exercise training reduces resting heart rate via downregulation of the funny channel HCN4,« *Nature Communications*, årg. 5, p. 3775, 2014.
- [22] »SureSigns VM6 Bedside Patient Monitor,« Phillips, [Online]. Available: <http://www.philips.co.in/healthcare/product/HC863065/suresigns-vm6-bedside-patient-monitor>.
- [23] R. A. Colucci, M. J. Silver og J. Shubrook, »Common Types of Supraventricular Tachycardia: Diagnosis and Management,« *American Family Physician*, årg. 82, nr. 8, pp. 942-952, 2010.
- [24] D. K. Arnett, G. W. Evans og W. A. Riley, »Arterial Stiffness: A New Cardiovascular Risk Factor?,« *American Journal of Epidemiology*, årg. 140, nr. 8, pp. 669-682, 1994.
- [25] B. Gavish, I. Z. Ben-Dov og M. Bursztyn, »Linear relationship between systolic and diastolic blood pressure monitored over 24h: assessment and correlates,« *Journal of Hypertension*, årg. 26, nr. 2, pp. 199-209, 2008.
- [26] M. Chana, D. Estève, J.-Y. Fourniols, C. Escriba og E. Campoa, »Smart wearable systems: Current status and future challenges,« *Artificial Intelligence in Medicine*, årg. 56, pp. 137-156, 2012.
- [27] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick og S. Tan, »Respiration rate monitoring methods: A review,« *Pediatric Pulmonology*, årg. 46, nr. 6, pp. 523-529, 2011.
- [28] M. Folke, L. Cernerud, M. Ekström og B. Hök, »Critical review of non-invasive respiratory monitoring in medical care,« *Medical and Biological Engineering and Computing*, årg. 41, nr. 4, pp. 377-383, 2003.
- [29] W. M. Schramm, A. Bartunek og H. Gilly, »Effect of local limb temperature on pulse oximetry and the plethysmographic pulse wave.,« *Journal of Clinical Monitoring and Computing*, årg. 14, nr. 1, pp. 17-22, 1997.
- [30] C. R. Kruuse, »Apopleksi -blodprop eller blødning i hjernen,« Patienthåndbogen, 26 08 2014. [Online]. Available: <https://www.sundhed.dk/borger/patienthaandbogen/hjerte-og-blodkar/sygdomme/apopleksi/apopleksi-blodprop-eller-bloedning-i-hjernen/>.
- [31] J. L. Saver, G. C. Fonarow, E. E. Smith, M. J. Reeves, M. V. Grau-Sepulveda, W. Pan, D. M. Olson, A. F. Hernandez, E. D. Peterson og L. H. Schwamm, »Time to Treatment With Intravenous Tissue Plasminogen Activator and Outcome From Acute Ischemic Stroke,« *Journal of American Medical Association*, årg. 309, nr. 23, pp. 2480-2488, 23 309 2013.
- [32] »What is TIA,« National Stroke Association, [Online]. Available: <http://www.stroke.org/understand-stroke/what-stroke/what-tia>.
- [33] »Nationale retningslinier for intravenøs trombolyselbehandling ved akut iskæmisk apopleksi,« Dansk Selskab for Apopleksi, 2016, v2.
- [34] C. S. Anderson, T. Robinson, R. I. Lindley, H. Arima, P. M. Lavados, T.-H. Lee, J. P. Broderick, X. Chen, G. Chen, V. K. Sharma, J. S. Kim, N. H. Thang, Y. Cao, M. W. Parsons og C. Levi, »Low-Dose versus Standard-Dose Intravenous Alteplase in Acute Ischemic Stroke,« *The New England Journal of Medicine*, pp. 2313-2323, 24 374 2016.

- [35] K. R. Lees, E. Bluhmki, R. v. Kummer, T. G. Brott, D. Toni, J. C. Grotta, G. W. Albers og M. Kaste, »Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials,« *Lancet*, pp. 1695-1703, 9727 375 2010.
- [36] K. N. Sheth, E. E. Smith, M. V. Grau-Sepulveda, D. Kleindorfer, G. C. Fonarow og L. H. Schwamm, »Drip and Ship Thrombolytic Therapy for Acute Ischemic Stroke: Use, Temporal Trends, and Outcomes,« *Stroke*, årg. 46, nr. 3, pp. 732-739, 2015.
- [37] »What is Stroke,« Heart and Stroke Foundation of Canada, 2015. [Online]. Available: <http://www.heartandstroke.ca/stroke/what-is-stroke/types-of-stroke>.
- [38] A. A. Wong og S. J. Read, »Early changes in physiological variables after stroke,« *Annals of Indian Academy of Neurology*, pp. 207-220, 4 11 2008.
- [39] T. Almdal, H. C. Kjeldsen, A. Sanbæk, H. Vestergaard og C. Sachs, »Diabetes - type 2,« Lægehåndbogen, 12 05 2015. [Online]. Available: <https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/endokrinologi/tilstande-og-sygdomme/diabetes-mellitus/diabetes-type-2/>.
- [40] C. S. Gray, J. F. Scott, J. M. French, K. G. M. M. Alberti og J. E. O'Connell, »Prevalence and prediction of unrecognised diabetes mellitus and impaired glucose tolerance following acute stroke,« *Age and Ageing*, pp. 71-77, 33 2004.
- [41] D. J. Wexler, D. M. Nathan, R. W. Grant, S. Regan, A. L. V. Leuvan og E. Cagliero, »Prevalence of Elevated Hemoglobin A1c among Patients Admitted to the Hospital without a Diagnosis of Diabetes,« *The Journal of Clinical Endocrinology & Metabolism*, årg. 93, nr. 11, pp. 4238-4244, 2008.
- [42] »Diabetes Mellitus,« Patienthåndbogen, [Online]. Available: <https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/endokrinologi/tilstande-og-sygdomme/diabetes-mellitus/>.
- [43] S. E. Capes, D. Hunt, K. Malmberg, P. Pathak og H. C. Gerstein, »Stress Hyperglycemia and Prognosis of Stroke in Nondiabetic and Diabetic Patients,« *Stroke*, pp. 2426-2432, 32 2001.
- [44] R. Chen, B. Ovbiagele og W. Feng, »Diabetes and Stroke: Epidemiology, Pathophysiology, Pharmaceuticals and Outcomes,« *American Journal of The Medical Sciences*, årg. 351, nr. 4, pp. 380-386, 2016.
- [45] Q. Jia, X. Zhao, C. Wang, Y. Wang, Y. Yan, H. Li, L. Zhong, L. Liu, H. Zheng, Y. Zhou and Y. Wang, "Diabetes and Poor Outcomes Within 6 Months After Acute Ischemic Stroke," *The China National Stroke Registry*, pp. 2758-2768, 2011.
- [46] M. J. Luitse, G. J. Biessels, G. E. Rutten og L. J. Kappelle, » Diabetes, hyperglycaemia, and acute ischaemic stroke,« *The Lancet Neurology*, pp. 261-271, 3 11 2012.
- [47] D. Nikneshan, R. Raptis, J. Pongmoragot, L. Zhou, S. C. Johnston og G. Saposnik, »Predicting Clinical Outcomes and Response to Thrombolysis in Acute Stroke Patients With Diabetes,« *Diabetes Care*, pp. 2041-2047, 39 12 216.
- [48] Collaboration, The Emerging Risk Factors, »Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies,« *Lancet*, årg. 375, pp. 2215-2222, 2010.
- [49] D. Robertson, Primer on the Autonomic Nervous System 2nd Edition, Elsevier, 2004.
- [50] J. Fessel og D. Robertson, »Orthostatic hypertension: when pressor reflexes overcompensate,« *Nature Clinical Practice Nephrology*, årg. 2, nr. 8, pp. 424-431, 2006.
- [51] L. J. Pu, Y. Shen, L. Lu, R. Y. Zhang, Q. Zhang og W. F. Shen, »Increased blood glycohemoglobin A1c levels lead to overestimation of arterial oxygen saturation by pulse oximetry in patients with type 2 diabetes,« *Cardiovascular Diabetology*, årg. 11, nr. 110, 2012.

- [52] M. D. Goldman, »Lung Dysfunction in Diabetes,« *Diabetes Care*, årg. 40, nr. 5, pp. 1915-1918, 2003.
- [53] N. Alam, E. Hobbelenk, A. Tienhoven, P. Ven, E. Jansma og P. Nanayakkar, »The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review,« *Resuscitation*, årg. 85, nr. 5, pp. 587-594, 2014.
- [54] F. S. M. A. a. P. K. A. Ehteshami, »Intensive Care Information System Impacts,« *Acta Informan Medica - Journal of Academy of Medical Sciences of Bosnia and Herzegovina*, årg. 21, nr. 3, pp. 185-191, 2013.
- [55] R. Bosman, E. Rood, H. Straaten, J. Spoel, J. Wester og D. Zandstra, »Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery,« *Intensive Care Medicine*, årg. 29, nr. 1, pp. 83-90, 2003.
- [56] E. Levesque, E. Hoti, D. Azoulay, P. Ichai, D. Samuel og F. Saliba, »The implementation of an Intensive Care Information System allows shortening the ICU length of stay,« *Journal of Clinical Monitoring and Computing*, årg. 29, nr. 2, pp. 263-269, 2015.
- [57] M. C. D. J. Fraenkel og P. Daley, »Quality benefits of an intensive care clinical information system,« *Critical Care Medicine*, årg. 31, nr. 1, pp. 120-125, 2003.
- [58] B. S. Nassar, M. S. Vaughan-Sarrazin, L. Jiang, H. S. Reisinger, M. Robert Bonello og P. Cram, »Impact of an Intensive Care Unit Telemedicine Program on Patient Outcomes in an Integrated Health Care System,« *Journal of American*, årg. 174, nr. 7, pp. 1160-1167, 2014.
- [59] T. Buchman, »Building a Triborough Bridge for Critical Care Medicine: Connecting Patients, Data and Real-Time Analytics (slides),« The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2016. [Online]. Available: <https://embc.embs.org/2016/tim-buchman/#abstract>.
- [60] T. C. Bailey, Y. Chen, Y. Mao, C. Lu og G. Hackmann, »A Trial of a Real-Time Alert for Clinical Deterioration in Patients,« *Journal of Hospital Medicine*, årg. 8, nr. 5, pp. 236-242, 2013.
- [61] A.-M. Vangsted og A. L. Petri, »Bekendtgørelse om autoriserede sundhedspersoners patientjournaler,« Styrelsen for Patientsikkerhed, 2016. [Online]. Available: <https://www.retsinformation.dk/Forms/R0710.aspx?id=183578>.
- [62] P. B. Jensen, L. J. Jensen og S. Brunak, »Mining electronic health records: towards better research applications and clinical care,« *Nature Reviews Genetics*, årg. 13, pp. 395-405, 2012.
- [63] R. Mishral, J. Bian, M. Fisman, C. R. Weir, S. Jonnalagadda og J. Mostafa, »Text Summarization in the Biomedical Domain: A Systematic Review of Recent Research,« *Journal of biomedical informatics*, pp. 457-467, 2014.
- [64] R. Pivovarov og N. Elhadad, »Automated methods for the summarization of electronic health records,« *Journal of the American Medical Informatics Association*, årg. 22, pp. 938-947, 2015.
- [65] Surviving Sepsis Campaign Guideline Committee, »Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock,« *Critical Care Medicine*, årg. 43, nr. 3, pp. 304-377, 2016.
- [66] R. Morgan, F. Williams og M. Wright, »An early warning scoring system for detecting developing critical illness,« *Clinical Intensive Care*, årg. 100, nr. 8, 1997.
- [67] »medinfo.dk,« Sundhedsdatastyrelsen, [Online]. Available: <http://www.medinfo.dk/>.
- [68] A. R. Corfield, F. Lees, I. Zealley, G. Houston, S. Dickie, K. Ward og C. McGuffie, »Utility of a single early warning score in patients with sepsis in the emergency department,« *Emergency Medicine Journal*, årg. 31, nr. 6, pp. 482-487, 2014.
- [69] T. Thim, N. H. V. Krarup, E. L. Grove, C. V. Rohde og B. Løfgren, »Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach,« *International Journal of General Medicine*, årg. 5, pp. 117-121, 2012.

- [70] S. Bird, E. Loper og E. Klein, *Natural Language Processing with Python*, O'Reilly Media Inc, 2009.
- [71] Christopher D Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard og D. McClosky, »The Stanford CoreNLP Natural Language Processing Toolkit,« *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60, 2014.
- [72] D. M. Powers, »Applications and Explanations of Zipf's Law,« *Proceedings of the joint conferences on new methods in language processing and computational natural language learning*, pp. 151-160, 1998.
- [73] M. Davies, »Word Frequency Data,« *Corpus of Contemporary American English*, 2015. [Online]. Available: <http://www.wordfrequency.info/free.asp?s=y>.
- [74] Det Danske Sprog- og Litteraturselskab, »KorpusDK,« Det Danske Sprog- og Litteraturselskab, [Online]. Available: <http://ordnet.dk/korpusdk>.
- [75] A. Vilic, K. Hoppe, J. Petersen og H. B. Sorensen, »Visualizing patient journals by combining vital signs monitoring and natural language processing,« *IEEE 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, nr. 16395001, 2013.
- [76] A. Wang, C. Fang, S. Chen, S. Tsai og W. Kao, »Periarrest Modified Early Warning Score (MEWS) predicts the outcome of in-hospital cardiac arrest,« *Journal of Formosan Medical Association*, årg. 115, pp. 76-82, 2016.
- [77] M. Albur, F. Hamilton og A. P. MacGowan, »Early warning score: a dynamic marker of severity and prognosis in patients with Gram-negative bacteraemia and sepsis,« *Annals of Clinical Microbiology and Antimicrobials*, pp. 15-23, 2016.
- [78] L. Breiman, J. Friedman, R. Olshen og C. Stone, *Classification and Regression Trees*, Chapman & Hall, 1993.
- [79] A. Vilic, K. Hoppe, J. Petersen, T. Kjaer og H. B. Sorensen, »Simplifying EHR Overview of Critically Ill Patients Through Vital Signs Monitoring,« *IEEE Journal of Biomedical and Health Informatics*, nr. 99, 2016.
- [80] K. Ganesan, »What are N-Grams?,« *Text Mining, Analytics & More*, 23 11 2014. [Online]. Available: <http://text-analytics101.rxnlp.com/2014/11/what-are-n-grams.html>.
- [81] A. M. Drewry, T. C. Bailey, B. M. Fuller og R. S. Hotchkiss, »Body temperature patterns as a predictor of hospital-acquired sepsis in afebrile,« *Critical Care*, årg. 17, nr. R200, 2013.
- [82] C. A. Buchan, A. Bravi og A. J. E. Seely, »Variability Analysis and the Diagnosis, Management, and Treatment of Sepsis,« *Current Infectious Disease Reports*, årg. 14, nr. 5, pp. 512-521, 2012.
- [83] M. Singer, C. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. Bernard, J. Chiche, C. Coopersmith, R. Hotchkiss, M. M. Levy, J. Marshall, G. Martin, S. M. Opal, G. Rubenfeld, T. Poll og J. Vincent, »M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. Poll, J.-L. Vincent, a,« *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*, årg. 315, nr. 8, pp. 801-810, 2016.
- [84] J. Reith, H. S. Jorgensen, P. M. Pedersen, H. Nakayama og H. O. Raaschou, »Body temperature in acute stroke: relation to stroke severity, infarct size, mortality, and outcome,« *Lancet*, pp. 422-425, 8899 347 1996.
- [85] F. D. S. S. T. R. L. O. R. T. A. C. Carmine Marini, »Contribution of Atrial Fibrillation to Incidence and Outcome of Ischemic Stroke,« *Stroke*, årg. 36, nr. 6, pp. 115-119, 2005.

- [86] A. Y. Poppe, S. R. Majumdar, T. Jeerakathil, W. Ghali, A. M. Buchan og M. D. Hill, »Admission Hyperglycemia Predicts a Worse Outcome in Stroke Patients Treated With Intravenous Thrombolysis,« *Diabetes Care*, årg. 34, nr. 4, pp. 617-622, 2009.
- [87] K. N. Vemmos, G. Tsivgoulis, K. Spengos, N. Zakopoulos, A. Synetos, E. Manios, P. Konstantopoulou og M. Mavrikakis, »U-shaped relationship between mortality and admission blood pressure in patients with acute stroke,« *Journal of Internal Medicine*, årg. 255, nr. 2, pp. 257-265, 2004.
- [88] M. M. Kaarisalo, I. Raiha, J. Sivenius, P. Immonen-Raiha, A. Lehtonen, C. Sarti, M. Mahonen, J. Torppa, J. Tuomilehto og V. Salomaa, »Diabetes worsens the outcome of acute ischemic stroke,« *Diabetes Research and Clinical Practice*, årg. 69, nr. 3, pp. 293-298, 2005.
- [89] J. Castillo, R. Leira, M. M. García, M. B. Joaquín Serena og A. Dávalos, »Blood Pressure Decrease During the Acute Phase of Ischemic Stroke Is Associated With Brain Injury and Poor Stroke Outcome,« *Stroke*, pp. 520-527, 2 35 2004.
- [90] H. P. Mattle, L. Kappeler, M. Arnold, U. Fischer, K. Nedeltchev, L. Remonda, S. M. Jakob og G. Schroth, »Blood Pressure and Vessel Recanalization in the First Hours After Ischemic Stroke,« *Stroke*, årg. 36, nr. 2, pp. 264-268, 2005.
- [91] H. Christensen, P. Meden, K. Overgaard og G. Boysen, »The course of blood pressure in acute stroke is related to the severity of the neurological deficits,« *Acta Neurologica Scandinavica*, pp. 142-147, 106 2002.
- [92] M. E. S. Prakash, »What Causes the Acute Blood Pressure Elevation After Stroke?,« *Stroke*, p. 2066, 10 36 2005.
- [93] P. J. Lindsberg, »Editorial Comment—High Blood Pressure After Acute Cerebrovascular Occlusion,« *Stroke*, pp. 268-269, 2 36 2005.
- [94] A. I. Quershi, »Acute Hypertensive Response in Patients With Stroke,« *Pathophysiology and Management*, pp. 176-187, 2 18 2008.
- [95] N. A. N. E. F. A. E. B. A. D. T. E. G. A. F. M. G. W. H. M. G. H. M. K. M. K. V. L. K. R. L. T. M. R. O. Nils Wahlgren, »Multivariable Analysis of Outcome Predictors and Adjustment of Main Outcome Results to Baseline Data Profile in Randomized Controlled Trials,« *Stroke*, årg. 39, pp. 3316-3322, 2008.
- [96] K. C. Johnston, A. F. Connors, D. P. Wagner, W. A. Knaus, X.-Q. Wang og E. C. Haley, »A Predictive Risk Model for Outcomes of Ischemic Stroke,« *Stroke*, årg. 31, nr. 2, pp. 448-455, 2000.
- [97] J. Liljehult og T. Christensen, »Early warning score predicts acute mortality in stroke patients,« *Acta Neurologica Scandinavica*, årg. 133, nr. 4, pp. 261-267, 2016.
- [98] American Heart Association Stroke Council, »Factors Influencing the Decline in Stroke Mortality,« *Stroke*, årg. 45, nr. 1, pp. 315-353, 2014.
- [99] L. J. Gray, M. Ali, P. D. Lyden og P. M. Bath, »Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in Acute Stroke,« *Journal of Stroke and Cerebrovascular Diseases*, årg. 18, nr. 6, pp. 466-468, 2009.
- [100] T. Askim, J. Bernhardt, L. Churilov og B. Indredavik, »The Scandinavian Stroke Scale is Equally As Good As the National Institutes of Health Stroke Scale in Identifying 3-Month Outcome,« *Journal of Rehabilitation Medicine*, årg. 48, nr. 10, pp. 909-912, 10 48 2016.
- [101] Y. Li, J.-G. Wang, E. Dolan, P.-J. Gao, H.-F. Guo, T. Nawrot, A. V. Stanton, D.-L. Zhu, E. O'Brien og J. A. Staessen, "Ambulatory Arterial Stiffness Index Derived From 24-Hour Ambulatory Blood Pressure Monitoring," *Hypertension*, pp. 359-364, 47 2006.
- [102] G. T. Knofczynski, »Sample Sizes When Using Multiple Linear Regression for Prediction,« *Educational and Psychological Measurement*, årg. 68, nr. 3, pp. 431-442, 2008.
- [103] R. Tibshirani, »The LASSO Method for Variable Selection in the COX Model,« *Statistics in Medicine*, årg. 16, pp. 385-395, 1997.

- [104] N. D. Kruyt, G. J. Biessels, J. H. DeVries og Y. B. Roos, »Hyperglycemia in acute ischemic stroke: pathophysiology and clinical management,« *Nature Reviews*, pp. 145-155, 3 2010.
- [105] American diabetes association, »Screening for type 2 diabetes,« *Diabetes Care*, årg. 28, pp. S37-S42, 2005.
- [106] F. Zahra, S. S. Kidwai, S. A. Siddiqi og R. M. Khan, »Frequency of Newly Diagnosed Diabetes Mellitus in Acute Ischaemic Stroke Patients,« *Journal of College of Physicians and Surgeons Pakistan*, årg. 22, nr. 4, pp. 226-229, 2012.
- [107] K. L. Christensen og M. H. Olsen, »Arteriel hypertension,« *Dansk Cardiologisk Selskab*, 06 2016. [Online]. Available: <http://nbv.cardio.dk/hypertension>.
- [108] P. S. PH, J. M. JE og MILIS, »The effect of diabetes mellitus on prognosis and serial left ventricular function after acute myocardial infarction: contribution of both coronary disease and diastolic left ventricular dysfunction to the adverse prognosis. The MILIS Study Group,« *Journal of the American College of Cardiology*, årg. 14, nr. 1, pp. 49-57, 1989.
- [109] D. Singer, A. Moulton og D. Nathan, »Diabetic myocardial infarction. Interaction of diabetes with other preinfarction risk factors,« *Diabetes*, årg. 38, nr. 3, pp. 350-357, 1989.
- [110] L. Ryden, »Diabetes mellitus and congestive heart failure,« *European Heart Journal*, årg. 20, nr. 11, pp. 789-795, 1999.
- [111] S. I. Lindsay, »A tutorial on Principal Components Analysis,« 2002.
- [112] A. Kowalczyk, »SVM Tutorial,« 10 2014. [Online]. Available: www.svm-tutorial.com.
- [113] N. Cristianini og J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 2000.
- [114] I. G. Bernhard E. Boser og V. N. Vapnik, »A training algorithm for optimal margin classifiers,« *COLT '92 Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152.
- [115] C. M. Bishop, »Sparse Kernel Machines,« i *Pattern Recognition and Machine Learning*, Springer Science, 2009, pp. 326-339.
- [116] S. Theodoridis og K. Koutroumbas, »Support Vector Machines: The Non-Linear Case,« i *Pattern recognition*, 2009, pp. 198-201.
- [117] Y.-S. Kim, R. Immink, W. Stok, J. Karemaker og J. v. Niels H. Secher, »Dynamic cerebral autoregulatory capacity is affected early in Type 2 diabetes,« *Clinical Science*, pp. 255-262, 8 115 2008.

APPENDIX **A**

APPENDIX A – ISCHEMIC STROKE OUTCOME SCALES

ISCHEMIC STROKE OUTCOME SCALES

There are several outcome measure scales for stroke patients. The below listed ones are all assessed as part of the Danish national guidelines for ischemic stroke treatment [33]. They evaluate the same neurological functionalities, but do so through various internationally comparable scales.

Modified Ranking Score

The first assessment when an ischemic stroke patient is admitted, is a modified ranking score (mRS). The scale evaluates functional disability as dependence on assistance from others for daily tasks (Table 1). It is examined twice as part of each admission. The first time is when the patient is admitted, but addresses the patient's independence prior the stroke. A follow-up is done three months later to address the long-term effects of the treatment.

Table 1: Modified Ranking Score (mRS) scale for assessment of functional disability and the patient's dependence on assistance

At admission	3 months after admission	Description
		Unknown
0	0	No symptoms
1	1	No visible functional disability
2	2	Some functional disability, requires no assistance
3	3	Moderate functional disability, requires assistance with daily tasks
4	4	Moderate/severe functional disability, requires assistance with daily tasks including hygiene
5	5	Severe function disability, bedridden
	6	Deceased

Global Outcome

Global outcome is evaluated twenty-four hours after treatment and when the patient is discharged from the hospital (or on the seventh day). It has no scores but is a subjective follow-up evaluation since last follow-up. It has six categories: major improvement, improvement, unchanged, worse, significantly worse, deceased.

National Institutes of Health Stroke Scale

The National Institutes of Health Stroke Scale (NIHSS) quantifies the degree of disability for eleven neurological functions. Each of these is scored based on individual scales for the given function (Table 2), and then summarized into an overall score. The higher the scores are, the worse the disability – meaning that the patient has no disabilities if the overall score is 0.

Table 2: NIHSS Individual scoring parameters. Arms and legs are scored once for each extremity whereas the others are only scored once. The scoring is done at the beginning of admission and after twenty-four hours.

<p>1a Consciousness 0 = Wide awake, reacting normally 1 = Not awake, responding to minor stimuli 2 = Not awake, can only be woken up through heavy/repetitive stimuli 3 = Coma</p>	<p>1b Questions (date and age) 0 = Answers correctly on both 1 = One answer is correct 2 = Neither are correct</p>
<p>1c Orders (Opening/closing eyes and clenching/opening fist) 0 = Completes both correctly 1 = Completes one correctly 2 = Neither is correctly completed</p>	<p>2 Gaze(only horizontal movements) 0 = Normal movement 1 = Partial palsy 2 = Total palsy</p>
<p>3 Vision 0 = Normal field of view 1 = Partial hemianopsia 2 = Toal hemianopsia 3 = Bilateral blindness</p>	<p>4 Facial paralysis (Ask patient to show teeth, move eyebrows, squeeze eyes) 0 = Normal symmetric movements 1 = Mild palsy 2 = Partial palsy 3 = Complete palsy</p>
<p>5 Motoric function in arm (scored once for each arm) 0 = No drift 1 = Drift, but not all the way down 2 = Some resistance to gravity 3 = Drops all the way 4 = No movement 5 = Cannot be tested</p>	<p>6 Motoric function in leg (scored once for each leg) 0 = No drift 1 = Drift, but not all the way down 2 = Some resistance to gravity 3 = Drops all the way 4 = No movement 5 = Cannot be tested</p>
<p>7 Ataxia extremities (in limbs) 0 = no ataxia 1 = Ataxia in one extremity 2 = Ataxia in both extremities</p>	<p>8 Sensory 0 = Normal sensitivity when pricked 1 = Mild-Moderate sensitivity 2 = No sensitivity</p>
<p>9 Language (patient describes images, objects and reads sentences) 0 = No aphasia 1 = Mild/moderate aphasia 2 = Severe aphasia 3 = Global aphasia</p>	<p>10 Dysarthria (Ask patient to read words) 0 = Normal speech 1 = Mild/Moderate Dysarthria 2 = Not understandable 3 = Unable to score because of physical barrier</p>
<p>11 Extinction and inattention 0 = No abnormality 1 = Inattention and extinction 2 = Hemi-inattention (for more than one modality)</p>	

Scandinavian Stroke Scale

Scandinavian stroke scale (SSS) is similar to NIHSS as it also addresses neurological abilities, but addresses nine parameters in exchange for more frequent evaluation. In the first two hours of treatment, SSS are registered every fifteen minutes. Afterwards, frequency is reduced to once every two hours. The scoring differs from NIHSS in that scores get higher the healthier a patient is – making 0 the worst score.

Table 17: Scandinavian stroke scale (SSS) sub-scores for each category. The higher the sub-score, the better the outcome

<p>1 Consciousness 6 = Fully conscious 4 = Somnolent, can be woken 2 = Reacts on speech 0 = Unconscious</p>	<p>2 Gaze 4 = No palsy 2 = Palsy 0 = Conjugate eye palsy</p>
<p>3 Upper extremities (arm) 6 = Lifting arm with normal force 4 = Lifting arm with reduced force 2 = Movable but cannot lift arm 0 = No movement</p>	<p>4 Hand 6 = Normal force 4 = Reduced force 2 = Movable but cannot reach palm with fingertips 0 = No movement</p>
<p>5 Under extremities (leg) 6 = Lifting leg with normal force 4 = Lifting leg with reduced force 2 = Movable but with flexion in knee 0 = No movement</p>	<p>6 Orientation 6 = Correct month, place and birthday 4 = Correct for Two of the above 2 = Correct for one of the above 0 = Disoriented (none of the above)</p>
<p>7 Language 10 = No aphasia 6 = Anomie 3 = Mainly yes/no answers 0 = Only yes/no without language</p>	<p>8 Facial palsy 2 = No/doubtful palsy 0 = Palsy</p>
<p>9 Walking 12 = Walks 5 meters without aid 9 = Walks with aid 6 = Walks with assisting person 3 = Sitting without support 0 = Bedridden</p>	

APPENDIX B

APPENDIX B – PUBLICATION IN PROCEEDINGS OF THE 38TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY

Visualizing Patient Journals by Combining Vital Signs Monitoring and Natural Language Processing

Authors:

Adnan Vilic, John Asger Petersen, Karsten Hoppe, and Helge B. D. Sorensen

Status: Published

Visualizing Patient Journals by Combining Vital Signs Monitoring and Natural Language Processing

Adnan Vilic¹, John Asger Petersen², Karsten Hoppe³ and Helge B. D. Sorensen¹

Abstract— This paper presents a data-driven approach to graphically presenting text-based patient journals while still maintaining all textual information. The system first creates a timeline representation of a patients’ physiological condition during an admission, which is assessed by electronically monitoring vital signs and then combining these into Early Warning Scores (EWS). Hereafter, techniques from Natural Language Processing (NLP) are applied on the existing patient journal to extract all entries. Finally, the two methods are combined into an interactive timeline featuring the ability to see drastic changes in the patients’ health, and thereby enabling staff to see where in the journal critical events have taken place.

I. INTRODUCTION

Majority of hospitals keep text-based patient records for each patient, and throughout the admission, the records are updated by medical staff. Patient journals contain information about the cause for admission, medical events, activities, family visits and drug consumption. All this information in a single document can become overwhelming for doctors who later need to verify whether the patient was treated appropriately. Particularly in cases where the patient has been admitted for a longer period. At the intensive care unit (ICU) at Bispebjerg Hospital in Denmark, each day of admission corresponds to approximately four pages of text per patient. Although difficult to computationally process, it is still possible to extract valuable information from journals using methods such as Natural Language Processing [1].

Due to the patient’s critical condition at ICU, patients are typically connected to electronic health monitoring devices. Electronic devices already alarm staff members when a patients’ health deteriorates, and have visual displays enabling staff to see the current condition of the patient. The devices can also be connected to servers, allowing data storage and analysis later on. Most public Danish ICU’s use Critical Information System (CIS), developed by Daintel, for storage and presentation of patient records throughout an admission.

To make the overwhelming information in patient journals more manageable, an attempt is made of combining the stored patient data from Bispebjerg Hospitals’ databases together with their corresponding electronic patient journals. The purpose of the application is to be able to display a visual representation of the patient’s physiological condition over time and map that against time-labeled entries in the patient’s journal. This gives staff a different approach to investigating journals. Instead of reading thoroughly through the entire

document, staff can look at the time where changes happen in the patient’s condition and then investigate the patient journal locations prior- and post-event to see what causes the changes.

II. METHODS AND MATERIALS

A) Patients

Data from ten selected patients from the ICU at Bispebjerg Hospital were inspected. All patients were admitted in the period February 2013 to May 2013. Their average age was in 64.88 ± 11.08 years. The admission cause for included patients was respiration insufficiency, cardiac arrest or sepsis. Average admission time for these patients was 11.2 ± 8.44 days. The inclusion criteria were that they have a journal, and vital sign parameters were recorded for:

- Pulse
- Respiration Rate
- Systolic Blood pressure
- Oxygen Saturation level
- Oxygen inspired (if any)
- Temperature

B) System Design

The system consists of two separate independent parts, which are synchronized once preprocessing is completed, as illustrated by the tree in Fig 1. The left branch represents all handling of electronic vital signs measurements. Vital signs are combined to give an estimation of the patient’s condition over time. The right branch focuses on text handling. An electronic patient journal is segmented in time by using timestamps for when entries are made into the journal.

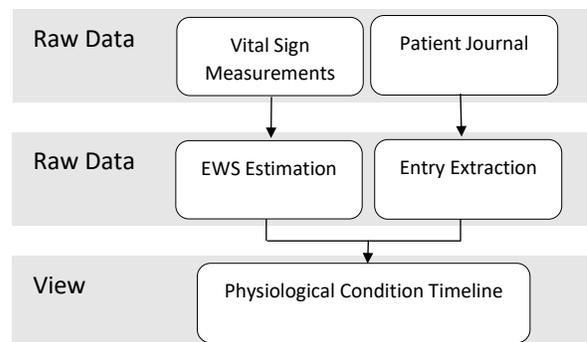


Figure 1 System overview

¹ Technical University of Denmark, Department of Electrical Engineering, Building 349, Oersteds Plads, 2800 Kgs. Lyngby, Denmark.
² Bispebjerg Hospital, Intensive Care Unit, Z, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark
³ DELTA, Microelectronics, Venlighedsvej 4, 2970 Hørsholm, Denmark

TABLE I. EWS table used in Capital Region of Denmark.

Type (T)	Sub score						
	3	2	1	0	1	2	3
Pulse (bmp)	<41		41-50	51-90	91-110	111-130	>130
Respiration Rate (bmp)	<9		9-11	12-20		21-24	>24
Temperature (°C)	<35.1		35.1-36.0	35.1-36.0	36.1-38.0	38.1-39	>39
Systolic Blood Pressure (mmHg)	<91	91-100	101-110	111-219			>219
Oxygen Saturation Level (%)	>92	93-94	94-95	>95			
Inspired O ₂				Air		Any O ₂	
Central Nervous System				Alert			Voice Pain Unresponsive

The last step is synchronizing timestamps and creating the timeline over the physiological condition along with markers of journal entries.

C) *Evaluating Physiological Condition using Vital Signs*

It has become common practice in many developed countries for hospitals to evaluate the patients’ physiological condition using Early Warning Score (EWS). The Early Warning Score system was first introduced by Morgan RMJ et al [2], with the purpose of ensuring that patients who exhibit signs of critical illnesses are monitored frequently [3].

An EWS is estimated by combining measurement readings from different vital signs and, giving sub scores to each individual reading. Finally, all sub scores are combined into a single value. Even though the EWS system is widely used in many hospitals, there is no standard for it. Bispebjerg Hospital is part of the capital region of Denmark, where all hospitals use a slightly modified version of the EWS system ViEWS, proposed by Prytherch et al [4]. The modified ViEWS is presented in Table I.

EWS systems are generally targeted towards wards where patients are not under constant monitoring as a tool to help staff in decide how much attention the patient needs [1,5-6]. The higher the score, the more attention is dedicated to the patient. More specifically, it determines how frequently the patient is visited and by whom. E.g. if all vital signs are normal, the patient will have an EWS of 0, which translates into the patient being checked on every twelve hours. If their vital signs are very abnormal, and they have an EWS of over 8, the patient will be observed every thirty minutes.

At ICUs, electronic devices are constantly observing and monitoring the patient and it does therefore not make sense to use EWS specifically as a tool for assessing how frequently a patient should be observed. It can however still be useful measure by giving an indication of what the patients’ overall health condition is over time.

Before analyzing the recorded measurements, data needs to be preprocessed. First, noisy measurements, that can be caused by electrodes falling off or being wrongly attached, are removed by only including physiologically likely values. This is done by for example removing measurements of systolic blood pressure above 250 mmHg. Similarly, oxygen saturation level measurements below 70% are removed because medical staff renders these readings are unreliable.

The existing alarm systems built into electronic medical devices are sufficient for detecting sudden physiological changes. Therefore, the focus here is to see how the patient’s condition in general develops over several hours or days. The

first step is to downsample so that for each hour, H , an average value of all measurements, $v_H(T)$, describes a vital sign, T . In addition, as sampling rates are different for each device, the amount of samples made throughout an hour, c , are also registered for quality control. It is defined by equation (1), where X is the collection of all measurements during an admission.

$$v_H(T) = \frac{1}{n} \sum_{n=H_{start}}^{n=H_{end}} X_T(i), \quad c_{H,T} = i \quad (1)$$

Following, the hourly EWS can be estimated as a summation of subscores, SS , retrieved by looking up in Table 1 for the average value of each vital sign type, T .

$$EWS_H = \sum_{T=1}^{T=6} SS_T(v_H(T)) \quad (2)$$

It has been decided to omit the central nervous system (CNS) as vital sign, since it is not registered digitally and is evaluated by staff, resulting in a poor and inconsistent registration frequency. Since the model for EWS assumes that all measurements are present to estimate a single score, EWS for one hour is not calculated unless there is at least one value per ten minutes throughout the hour.

Fig 2 shows the timeline for a patient who has been admitted for eight days. It shows how all vital signs develop over time. Looking at the individual parameters, the first thing noticed are spikes in heartrate, which can be explained by the fact that the patient receives adrenaline treatments which have short durations.

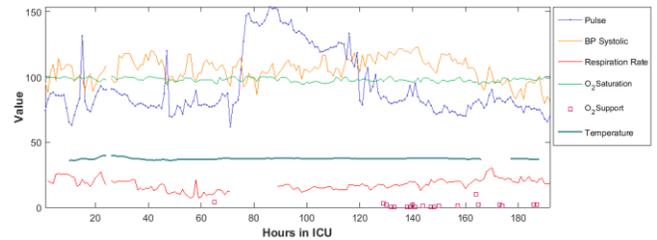


Figure 2 Vital sign measurements for a patient

We see that the patient did not receive any inspired oxygen throughout the first 125 days of admission, and was able to breathe by on their own. If a medical ventilator was helping,

the breathing would have stable and resemble a flatline. The temperature is difficult to see in the figure, but is stable around 37.3°C. The remaining parameters vary significantly throughout the stay and will be revisited later.

Before looking at the EWS estimates, a simple quality check is performed, based on the previously mentioned hourly counts. For every hour, there have to be sufficiently many measurements considering the type of vital sign. E.g. for respiration rate, it is sufficient if in average two breaths per minute are recorded, while temperature changes slower and requires fewer measurements. Four quality classes (1-4) are defined, ranging from non-existent to ideal measuring (1- white, 2- orange, 3- yellow, 4- green). Fig 3 shows the quality of recorded information for the same patient as in Fig 2.

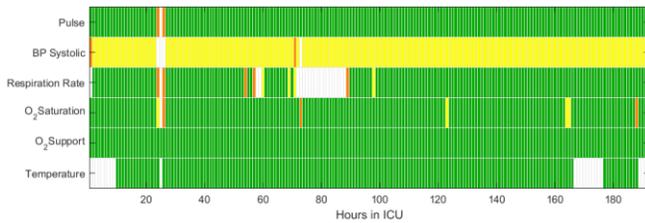


Figure 3 Early Warning Score timeline for a patient. Color represents quality of signal (Green = good, yellow = ok, orange = poor, white = very poor/non-existent)

It shows missing data about twenty hours after admission which is less noticeable in Fig 2, and reveals inconsistencies in recording frequency for especially respiration rate.

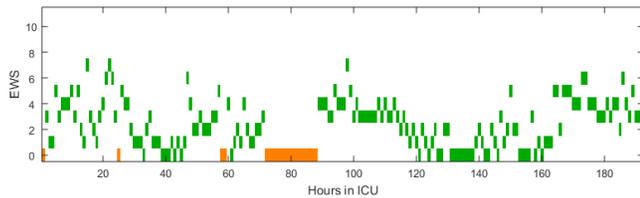


Figure 4 Early Warning Score timeline for a patient. Insufficient data results in orange labeling.

Finally, EWS can be estimated for each hour as illustrated in the figure above. The x-axis is the time spent under surveillance, while the y-axis is the patients' EWS. Each rectangle represents one hour. If enough measurements for each vital sign throughout an hour exist, the rectangle is the actual EWS value and colored green. Otherwise, it is orange with value 0. The points without sufficient measurements are still included, and not left blank, because it emphasizes that something was abnormal at this point – Either that electrodes fell off or that the patient was disconnected from monitoring for some reason.

D) Natural Language Processing in Patient Journals

The next step is linking the patient journal to the EWS timeline. Initially, the idea was to use language models such as n-grams to detect events in patient journals. Since patient journals however are updated by numerous medical employees who phrase things in very different ways, this was not an option. Especially since text was often phrased in keyword format, acronyms and many non-standard abbreviations are used which are only known to some employees.

Since it is not possible to detect when and which events occurred based on measurements, the next best thing is to find when new entries are made into the patient journal and then

manually inspect these entries for events. The assumption here is that shortly after any major change, staff will document what occurred, although there will be a time offset as staff has to attend to the patient before reporting. To extract entries, a regular expression is used. A regular expression is a sequence of characters and symbols which form a search pattern. This pattern can then be used to traverse through text and find content that matches this pattern. The regular expression below follows the date pattern of our patient journals where a line begins with a date, followed by department and hour of entry:

```
^([0][1-9]|[1-2][2-9]|[3][0-1]) // Day of month
.([0][1-9]|[1][0-2]). // Month
1[2-3] // Year of interest
[a-zA-ZæøåÆØÅ\s.]* // Department
([0-1][0-9]|[2][0-3]):[0-5][0-9] // Hour of entry
```

Fig 5 shows an example where the above regular expression is applied on entries of the first three days, to extract information about how frequently reports are made in the patient journal. It shows that reports are done irregularly, but it is still possible that entries are made within an hour of each other. The lack of reporting can also be an indicator of a stable condition. A stable condition will be reflected in timeline for the given period, or through reading the journal.

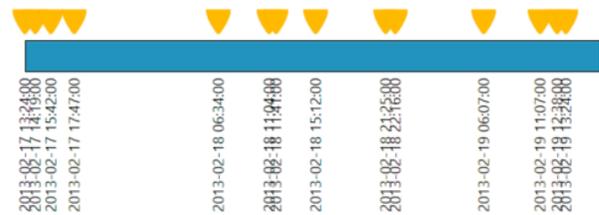


Figure 5 Extracted entry times from a patient journal.

Each entry from the journal is either extracted alone or combined with other entries for the same hour and mapped into the patient's EWS timeline. If there is a journal entry for an hour, the corresponding square gets a border and becomes interactive so that a click results in opening the entry.

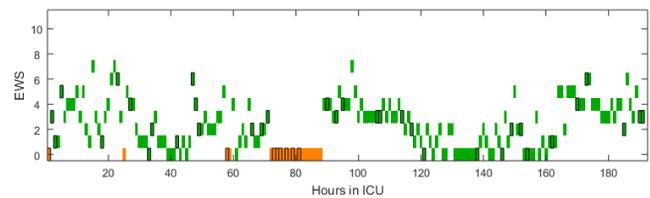


Figure 6 EWS timeline combined with patient journal.

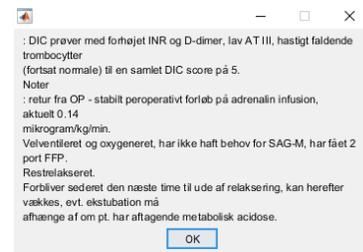


Figure 7 Journal entry 27 hours after admission

III. RESULTS

All figures from the previous section have focused on a single patient. Going through this patient's admission

chronically based on the figures, there are several interesting timespans to investigate. The first is an improvement beginning after missing data around 25 hours after admission. Next, there is sudden deterioration after 45 hours followed by lack of measurements from 70 to 90 hours. Finally, there is a continuous deterioration after 165 hours.

The journal entry for the first timespan (Fig 7), reveals the condition gradually worsened and after a surgery the condition stabilized as reported 27 hours after admission. The deterioration around 45 hours is unexplained in the journal, but an investigation of individual measurements throughout the hour showed significant changes in pulse, oxygen saturation and blood pressure at different times during which the patient was sleeping. In the 70 to 90 hour period, the patient was moved into isolation where they might have been connected to a respiratory device not connected to the used database. Although the last span shows serious deterioration after a stable healthy period, it does not entirely reflect the patients condition. In the final phase, the patient journal states that there is progress and the patient will be discharged soon. The data does not reflect this due to increase in breathing and hypotension caused by medication.

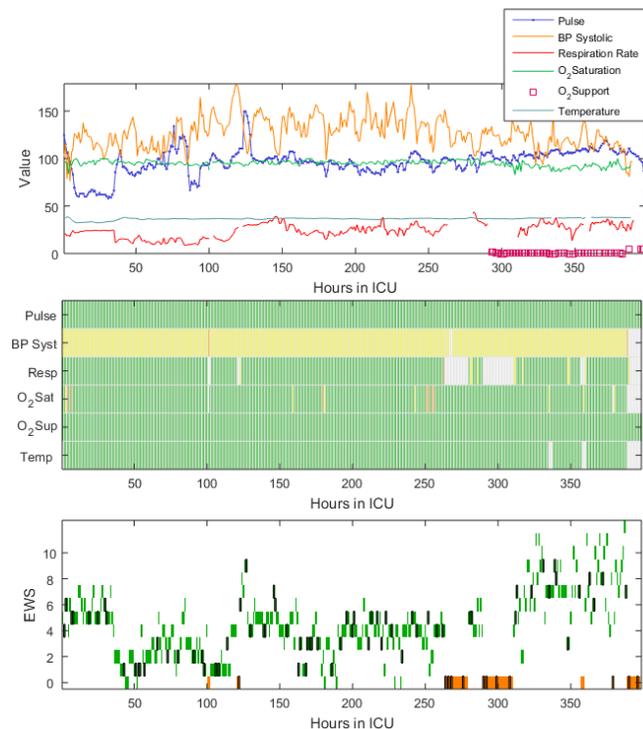


Figure 8 Example of a 16-day admission

Fig 8 shows another example with a much longer admission. Both the vital signs overview chart (Fig 8 top) and the EWS timeline (Fig 8 bottom) clearly show that the condition is overall swinging until around 120 hours where condition worsens and the patient becomes unresponsive to all contact. The patient regains consciousness several hours but status remains critical. Near the end of the admission where the EWS reach values above 9, it is decided to cease treatment and the patient passes away.

IV. DISCUSSION

Reading through the remaining patient journals, the EWS timelines reveal most critical changes, but do not always reflect the patient's condition. According to the EWS timeline, in one case, a patient seemed to be improving and suddenly appeared to worsen severely. According to the patient journal however, the patient gradually got worse and had to be cooled down to avoid brain damage. The sudden decline in EWS was thus caused through necessary treatment.

Although it is difficult to use Natural Language Processing (NLP) for extracting textual information from patient journals due to the nature in which they were written, still valuable information can be extracted as shown with Fig 5-7. While not included in this paper, simple methods like word frequency counts helped understand details about the admission such as which symptoms the staff paid most attention to and which drugs were mainly administered.

The next step of development will include diagnosis information for when and which diagnosis/condition begins and ends to see if vital signs have significant changes related to the diagnosis. This will also reveal whether and for which diagnosis a higher sampling is required.

V. CONCLUSION

The study shows that data alone is insufficient for assessing the patient's condition in ICU settings, especially if the patient is moving or medical staff is interfering. Ideally this model should be tested with a larger patient population outside the ICU who are in a less critical condition and need less care.

The system provides a novel way of visualizing the patient journal, showing when events of interest occurred during the admission.

REFERENCES

- [1] P. B. Jensen, L. J. Jensen og S. Brunak, »Mining electronic health records: towards better research applications and clinical care,« *Nature Reviews Genetics*, årg. 13, pp. 395-405, 2012.
- [2] RJM Morgan, F. Williams og M. Wright, »An early warning scoring system for detecting developing critical illness,« *Clinical Intensive Care*, årg. 8, nr. 100, 1997.
- [3] R. Morgan og M. Wright, »In defense of early warning scores,« *British Journal of Anaesthesia*, årg. 99, nr. 5, pp. 747-748, 2007.
- [4] David R. Prytherch, G. B. Smith, P. E. Schmidt og P. I. Featherstone, »ViEWS - Towards a national early warning score for detecting adult inpatient deterioration,« *Resuscitation*, årg. 81, pp. 932-937, 2010.
- [5] A. R. Corfield, F. Lees, I. Zealley, G. Houston, S. Dickie, K. Ward og C. McGuffie, »Utility of a single early warning score in patients with sepsis in the emergency department,« *Emergency Medicine Journal*, årg. 31, nr. 6, pp. 482-487, 2014.
- [6] U. Kyriacos, J. Jelsma og S. Jordan, »Monitoring vital signs using early warning scoring systems: a review of the literature,« *Journal of Nursing Management*, årg. 19, pp. 311-330, 2011.

APPENDIX C

APPENDIX C – PUBLICATION IN IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS

Simplifying EHR Overview of Critically Ill Patients Through Vital Signs Monitoring

Authors:

Adnan Vilic, Karsten Hoppe, John Asger Petersen, Troels W. Kjær, and Helge B. D. Sorensen

Status: Published

Simplifying EHR Overview of Critically Ill Patients Through Vital Signs Monitoring

Adnan Vilic*¹, Karsten Hoppe², John Asger Petersen³, Troels W. Kjaer⁴, and Helge B. D. Sorensen¹

Abstract— This paper presents a novel data-driven approach to graphical presentation of text-based electronic health records (EHR) while maintaining all textual information. We have developed the Patient Condition Timeline (PCT) tool, which creates a timeline representation of a patients' physiological condition during admission. PCT is based on electrical monitoring of vital signs and then combining these into Early Warning Scores (EWS). Hereafter, techniques from Natural Language Processing (NLP) are applied on existing EHR to extract all entries. Finally, the two methods are combined into an interactive timeline featuring the ability to see drastic changes in the patients' health, and thereby enabling staff to see where in the EHR critical events have taken place.

Index Terms— Electronic Medical Records, Natural Language Processing, Medical Information Systems.

I. INTRODUCTION

DEVELOPED countries have strict regulations that ensure that every institution must create, document and store all medically relevant information regarding a patient in form of an electronic health record (EHR). Aside from scans, tests and schemes, EHR typically consist of a textual part which in details describes the patient's history, circumstances, events, status and planned procedures. It is especially long for critically ill patients who are admitted for several days, because their condition may vary and all changes need to be documented. On average four pages of pure text are generated per patient per day at the intensive care unit (ICU) at Bispebjerg Hospital, Denmark. Therefore, EHR of patients who have been admitted for longer periods become almost overwhelming if medical staff later need to follow up on, for example, how the patient responded to different treatments during admission.

A common way of dealing with this is to quantify textual data

by applying Natural Language Processing (NLP) to do a text summarization. The overall aim is to reduce the presented information to a minimum by, for example removal of redundancy, or applying statistics to rank sentences [1-3].

Unlike at other departments, all patients within ICU's are furthermore monitored continuously with electronic devices, so that any health deterioration is immediately detected, allowing appropriate actions to be taken swiftly. This data is typically and processed with aid of clinical information systems (CIS), which have proven effective for optimization of resources and decision making [4-8]. To the best of the author's knowledge, no previous study has however combined these recordings with content from the patient's EHR to review previous admissions.

This study approaches the problem of overwhelming data through the development of a data-driven tool, where vital signs monitoring is combined with textual information from the EHR. The automatized Patient Condition Timeline (PCT) tool displays a visual representation of a patient's physiological condition over time and maps it against time-labelled entries in their EHR. It enables a different approach to investigate the EHR, and gives the possibility to quality check documentation against assumptions of the patient's condition. Finally, assuming the model for evaluating a patient's condition is accurate, it becomes possible to find EHR entries related to deterioration, causes and resolution.

The PCT is in its current state useful for retrospective purposes, and does not feature decision boundaries that can improve treatment during the admission. It can instead be applied for reviewing a specific patient's admission or trend analysis of progress for either general patients or across patient groups. Although not part of this study, a potential use is the comparison of intra-patient progress for patients who are readmitted into the ICU after e.g. cardiac surgery.

This study is part of a larger initiative which is investigating the feasibility of continuous vital signs monitoring to increase staff efficiency and decrease mortality within hospitals. It is inspired by recent advances in non-obtrusive wearable devices [9-10], and the fact that the general population is aging but birthrates have declined, leading to an upcoming workforce shortage.

A. Modelling Physiological Condition using Vital Signs

Hospitals in developed countries optimize resources by using scoring models to assess condition and needs of individual

¹ Technical University of Denmark, Department of Electrical Engineering, Oersteds Plads, 2800 Kgs. Lyngby, Denmark.

² DELTA, Microelectronics, Venlighedsvej 4, 2970 Hoersholm, Denmark.

³ Bispebjerg Hospital, Intensive Care Unit, Z, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark.

⁴ Zealand University Hospital, Department of Neurology, Sygehusvej 10, 4000 Roskilde, Denmark.

TABLE I: EWS Table used in Capital Region of Denmark

Type (T)	Sub score						
	3	2	1	0	1	2	3
Pulse (bpm)	<41		41-50	51-90	91-110	111-130	>130
Respiration Rate (bpm)	<9		9-11	12-20		21-24	>24
Temperature (°C)	<35.1		35.1-36.0	35.1-36.0	36.1-38.0	38.1-39	>39
Systolic Blood Pressure (mmHg)	<91	91-100	101-110	111-219			>219
Oxygen Saturation Level (%)	>92	93-94	94-95	>95			
Inspired O ₂				Air		Any O ₂	
Central Nervous System				Alert			Voice Pain Unresponsive

patients. One such model uses Early Warning Scores (EWS) introduced by Morgan RMJ et al. [11-12]. It evaluates abnormality of vital signs that can be measured noninvasively. Each vital sign is assigned a sub-score based on how abnormal the measurement is, and an overall score is then estimated through summation of all sub-scores. Surveillance frequency and type is then adjusted based on the overall score. The model is widely applied, but it is yet to be standardized even on national plans [13-14].

Bispebjerg Hospital, and the rest of the capital region of Denmark use a modified version of the model called ViEWS, which was introduced by Prytherch et al [15], and is presented in Table I. Although present in the model, evaluation of the central nervous system (CNS) is omitted as a vital sign in this study, because it is registered inconsistently at varying time interval and evaluated subjectively by staff

Assuming all vital signs are normal, each vital sign will have a sub-score of 0, resulting in EWS being 0, and nurses will check on the patient at least every twelve hours. Abnormality increases surveillance times significantly, and ranges from follow-ups every 12 hours by nurses to follow-ups every thirty minutes by on-call physicians if a score above 8 is reached, as illustrated in Fig 1. The on-call physician takes over surveillance once a score of 6 is reached.

ICUs do not have models for scoring physiological conditions of patients, because staff is always present close to the patient and ready to react. In the context of following up previous admissions, ViEWS may however still be a useful measure by giving an indication of what the patients' overall health condition was throughout the admission.

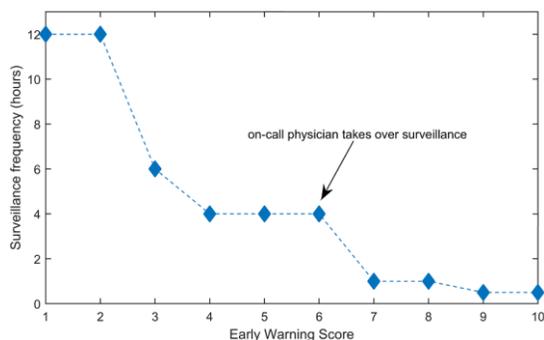


Fig. 1. Surveillance based on Early Warning scores

II. PATIENTS

Data was extracted from the hospital's database system for

fifty randomly selected patients admitted to the ICU with primary cause of admission being respiration insufficiency, cardiac arrest or sepsis. These causes for admission were taken into consideration, because the patient groups they represent are frequently admitted into the ICU, and have a higher risk of complications and deterioration during admission. All patients were admitted and discharged in spring 2013. For patients to be included in the study, the duration of admission had to be at least one day, EHR needed to be available, and the following vital sign parameters had to be recorded:

- Pulse
- Respiration Rate
- Systolic Blood pressure
- Oxygen Saturation level
- Oxygen inspired (optional)
- Temperature

Forty-four subjects fulfilled the inclusion criteria, while the remaining ones did not have recordings for two or more vital sign parameters. The admission time for included patients was 10 ± 8 days. Their average age was in 65.68 ± 10.65 years.

III. METHODS

A. System Overview

The PCT is composed of two independent modules that are merged at the very end, as illustrated in Fig 2. The left branch is dedicated to estimating the patient's condition over time through processing of vital signs measurements. The right branch extracts and separates entries from the EHR after locating timestamps using NLP.

Lastly, a timeline is displayed which shows the patient's physiological condition along with markers of entries.

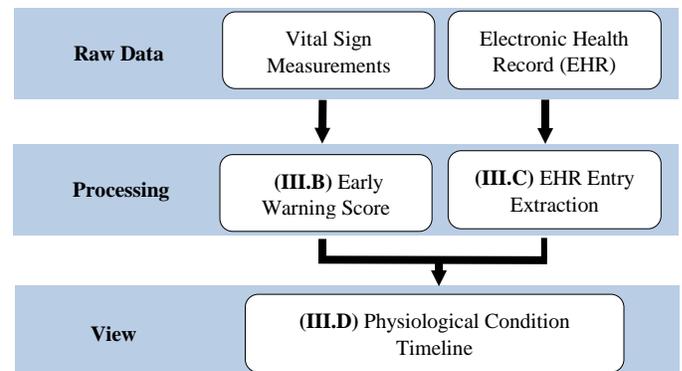


Fig. 2. Patient Condition Timeline (PCT) overview with section numbers under which methods are described in detail.

B. Early Warning Score Estimation

Medical monitoring devices have built in alarm systems that will detect immediate physiological changes. The PCT tool therefore focuses on progress of a patient’s condition over hours and days.

The first step is collecting data for vital signs listed in section 2, and pre-processing them. Unreliable and highly improbable measurements, such as oxygen saturation below 70% are removed, because they are likely caused by electrodes falling off and medical staff renders them as unreliable. Table II consists of a list of ranges that are allowed for each vital sign. Since the hospital uses different devices, which are sometimes also switched, it was not possible to get boundaries based on equipment specifications. Instead, the boundaries were defined by ICU staff as the values they regard as reliable. Data is then further downsampled into hours because we are interested in trend over time.

TABLE II: Included ranges for vital signs

Type (T)	Cut-off ranges	
	Low	High
Pulse (bpm)	-	200
Respiration Rate (bpm)	-	45
Temperature (°C)	27	45
Systolic Blood Pressure (mmHg)	-	310
Oxygen Saturation Level (%)	70	100
Inspired O ₂	-	-

Fig 3 shows hourly average vital sign measurements for a patient who was in surgery for peritonitis and was admitted in the ICU due to septic shock – The most unusual development in vital signs throughout the admission being around fifty hours after admission, where pulse significantly increases and systolic blood pressure begins to rise. The sudden rise in pulse is described in the EHR, but the cause for it is not clear. It is likely related to the patient receiving noradrenaline from beginning of the admission until about the hour where pulse starts increasing. Pulse successfully decreased continuously after reintroducing noradrenaline with amiodarone. Likewise, the increase of pulse around ninety hours is reduced with amiodarone.

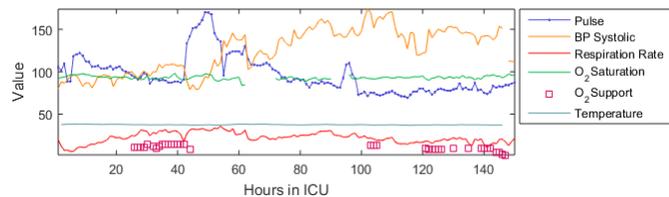


Fig. 3. Hourly average vital signs measurements for a patient

Next, a quality check is performed to ensure that every vital sign has sufficient data points. E.g. for respiration rate, there are sufficiently many data points if in average two breaths per minute (bpm) are recorded, while temperature changes slower, so it is sufficient with a reading every five minutes. As a quality

measure, each hour of admission is then denoted by one of four classes (1-white, 2- orange, 3-yellow, 4-green) ranging from respectively “non-existent” to “ideal or above”.

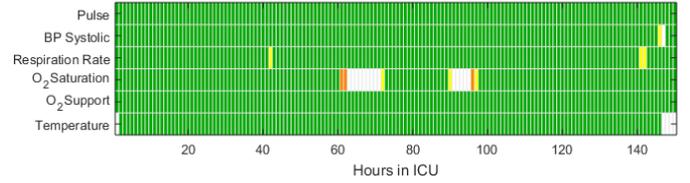


Fig.4. Quality control for a patient (1-white=non/existent, 2- orange=poor,3-yellow=ok, 4-green=ideal or above)

Fig 4 shows quality control of vital signs measurements for the same patient. The two periods with lack of monitoring of oxygen saturation is due to the parameter being measured with a non-invasive ventilation (NIV) device. In both cases, NIV was temporarily removed because it was obtrusive, and staff wanted to test if the patient is improving. With data validated and present, the EWS timeline can now be created.

$v_h(VS)$ describes the average value of all measurements for a vital sign, VS, throughout an hour, h – each hour starting at h_{start} and ending sixty minutes later at h_{end} . In addition, the amount of samples made throughout an hour, $c_{h,VS}$, are noted for the previously mentioned quality control, as sampling varies among medical devices used. The downsampling is expressed by equation (1), where X_{VS} is a collection of all measurements for a given vital sign during an admission, and i is the index at a given time.

$$c_{h,VS} = h_{end} - h_{start} \quad (1)$$

$$v_h(VS) = \frac{1}{c_{h,VS}} \sum_{i=h_{start}}^{h_{end}} X_{VS}(i),$$

Once the hourly average for a vital sign has been estimated, $v_h(VS)$, the value is directly translated into a vital sign sub-score, SS_{VS} , through lookup in Table 1. I.e. a systolic blood pressure of 95 results in SS_{VS} being 2. The hourly EWS, EWS_h , is then estimated by equation (2) as the sum of SS_{VS} for all vital signs.

$$EWS_h = \sum_{VS=1}^{VS=6} SS_{VS}(v_h(VS)) \quad (2)$$

Fig 5 illustrates EWS timeline for the same patient as in previous examples. The x-axis is the time since admission, and the y-axis is the patients’ EWS. Each hour is represented by a rectangle. Green rectangles indicate that there are sufficient measurements for all vital signs; and are otherwise orange with the EWS value zero. The areas with missing data are kept in the graph to emphasize abnormality of missing surveillance, which in many cases was due to examinations, surgeries and tests.

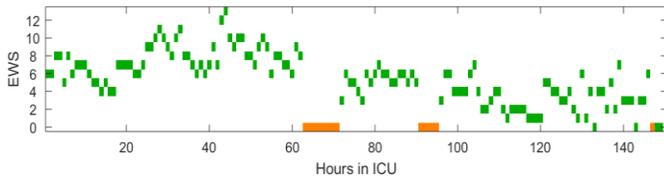


Fig. 5. EWS timeline for a patient

C. Entry Extraction from Electronic Health Records

Based on the study population, vital sign measurements alone are not enough to explain which and when events occur as there can be many reasons for abnormalities in vital signs. The next best thing is to extract text from EHR, which is continuously being updated by medical staff throughout admission whenever changes occur.

With each change, the internal software automatically generates a timestamp (Fig 6). The updating policy makes it safe to assume that abnormal changes in vital signs are documented shortly after their occurrence. Consequently, the EWS timeline can be tied to the patient's EHR.

04.02.13 Intensiv ZIT kl. 14:48
 Date Department Time

Fig. 6. Automatically generated timestamp example. Date and Time can only change in numbers while department may not be present, but if it is, it can consist of one or several words

All entries are extracted using a regular expression that matches only timestamps assigned by internal hospital software. A regular expression is a search pattern that is formed from a sequence of characters and symbols. The pattern traverses through the desired text and responds only to content that matches it. The following expression combined into one sequence of characters and symbols (excluding text after the hashtag), matches all generated timestamps for the software used at Bispebjerg Hospital (Fig 7):

```
^([0][1-9]|[1-2][2-9]|[3][0-1]) # Day of month
.([0][1-9]|[1][0-2]) # Month
1[2-3] # Year of interest
[a-zA-ZæøåÆØÅ\s.]* # Department
([0-1][0-9]|[2][0-3]):[0-5][0-9] # Hour of entry
```

The pattern requires a text line to start with a number between 01-31 (day) followed by a period, then a number from 01-12 (month), followed by another period, and finally either a 12 or 13 (year, which is only denoted by two numbers). Next, there may or may not be a department name(s) listed which are unknown in both length and content, although they do not contain numbers. The timestamp always ends with a time ranging from 00:00 to 23:59. Applying above regular expression on the same patient's EHR, confirms that entries are made irregularly, as it is seen from Fig 7.

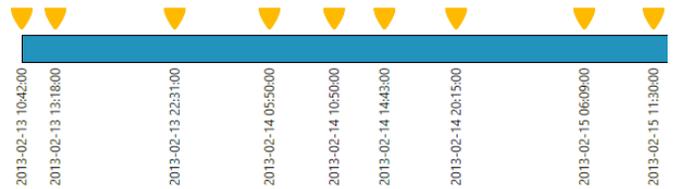


Fig. 7: First part of timeline over extracted entry timestamps for a patient EHR

Each patient's EHR contained 115.75 ± 84.64 ($n=5093$) mentions of date, time or a combination of these. From these, 69.93 ± 50.72 ($n=3077$) were autogenerated timestamps of interest. Although it is common for staff to write dates and time inside the EHR, the regular expression correctly extracted all, and only, the autogenerated timestamps. Comparing manual time entries with the autogenerated ones, there were two primary causes for this performance.

- Staff always omits the year when referring to dates.
- No manual entries began a single line with a date and ended it with a timestamp.

D. Combining EHR Entries with Estimated Physiological Condition

The final step is making an overlay of EHR entries over the EWS timeline (Fig 8). After extracting all entries, the ones from same hour are merged. Next, for every hour on the timeline, a border is created around the EWS and the square becomes interactive, so that it opens the corresponding entry, when pressed on.

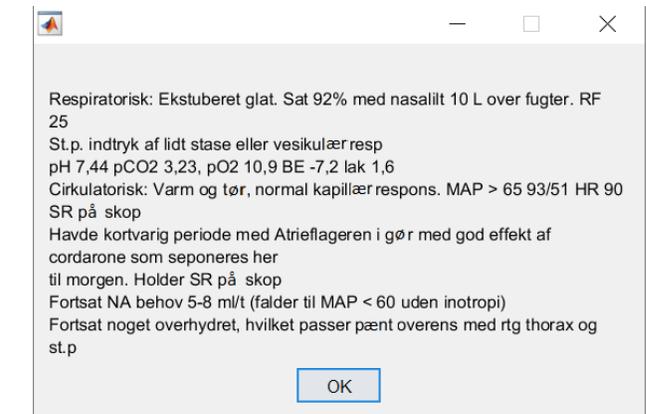
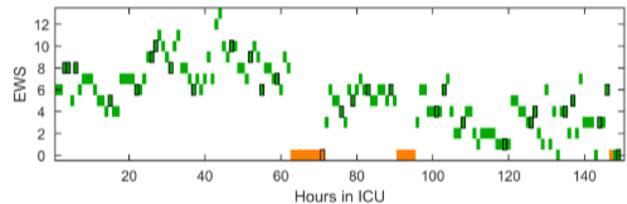


Fig. 8. EWS timeline with entry overlays (top) with example of EHR opening an entry (bottom)

Ideally, one would extract all entries and categorize them based on type (e.g. diagnosis, event, follow-up, procedure, visit, etc.), and thereby reduce redundant information. This was not possible with the current dataset as text often contained spelling errors and was phrased in keyword format with non-standard abbreviations that were only familiar to internal staff members.

IV. RESULTS

The raw vital signs highlight local changes in a single parameter but fail to assess the overall condition, as it can be seen in Fig 9. It is clear from the raw signals that the patient receives induced oxygen support from 26 to 42 hours into the admission, but they do not show that during this period, the patient’s condition gradually and significantly worsened at 28 hours before slowly improving. A nearby entry reveals that the deterioration was caused by exhausting exercises with a positive expiratory pressure (PEP) device.

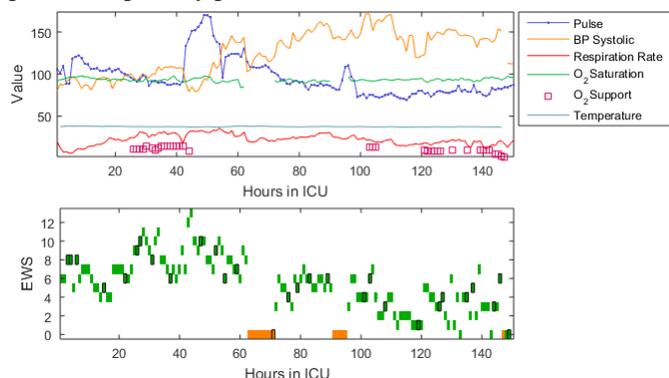


Fig. 9. Raw vital signs (top) vs EWS (bottom) for the same patient

Similarly, it is more difficult from the raw signals to see that the overall health condition is generally improving over time especially after about 100 hours into the admission. At this point an entry states that infection parameters have started decreasing for the first time since the admission, and that the patient appears to be improving. The patient is discharged and moved to another department with a stable condition and an EWS of 0. It is however important to note that the patient is overhydrated, still shows measurable signs of infection, has atrial fibrillation and is fed through a feeding tube. The fact that all vital signs are normal is therefore heavily due to medication and electrical devices.

The EWS timelines for other patients are similar in shape and data, but not useful for patients who are discharged shortly after being admitted, because their condition tends to be stable around the same EWS. This also raises questions whether the outcome can be deduced based on the timeline without reading entries. From the dataset, 29 patients survived and the remaining 15 died.

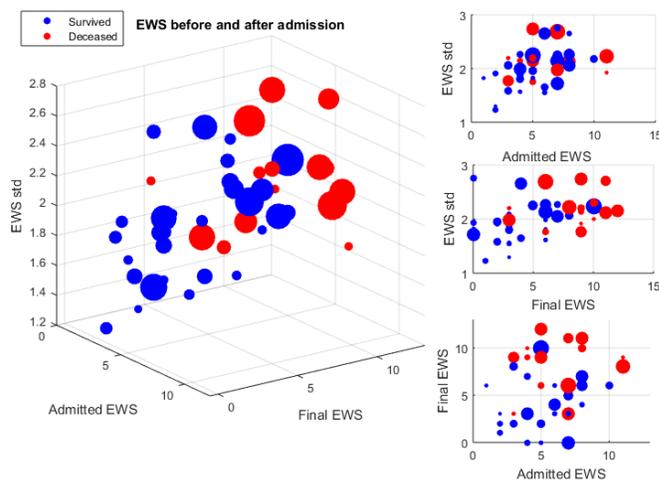


Fig.10. Patient outcome based on EWS and duration of admission

Fig 10 shows the relationships between EWS when the patient is admitted (x-axis), released (y-axis) and stability (standard deviation) during admission (z-axis). Length of stay is represented by size of circle.

The length of admission is insignificant to outcome, which is in accordance to our related literature findings [16]. EWS on the other hand appears to be a better indicator, especially when the condition is stable. The standard deviation of EWS for survivors was lower (1.96 ± 0.37) than for deceased patients (2.19 ± 0.31) which is also documented in other studies [17-18].

To investigate relationships between parameters further, a regression tree is computed for the parameters {EWS at admission, EWS at discharge, standard deviation in EWS, Age}. Length of admission was originally included, but is discarded because it quickly causes overfitting.

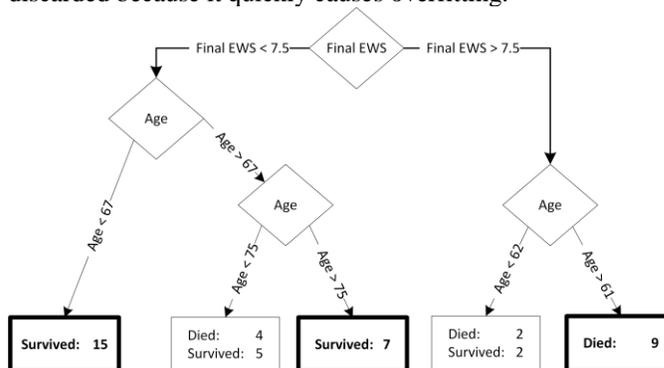


Fig. 11: Regression tree model for outcomes of ICU patients in study

The regression tree reveals three pure groups. If the person at discharge had a final EWS above 7 and is older than 61, the person died. If the person on the other hand had a final score of 7 or lower and were younger than 66, they survived.

The last branching are patients with final EWS lower than 8 but older than 68. Interestingly, these patients are split into two age groups where the older group has higher survival rate – The reason for this is unknown, but one assumption is that the younger of the two groups appears to have a long history of

illnesses and high alcohol consumption while the older patients are healthier prior to ICU admission.

V. DISCUSSION

In its present state, the PCT is a retrospective review tool that graphically enables staff to see changes in the patient's condition. It can work as an independent tool, but could also be implemented as module in an existing CIS, since CIS already contain much of the data mentioned in this study. In the used setup, the existing CIS contained all data, but the only review option is to look at vital signs individually over short intervals at a time. The model is currently suitable in the ICU where vital signs are monitored with precise electronic devices that are connected to patients spending most time in bed. Once unobtrusive wearable monitoring devices become more available in other departments, the modelling will require more sophisticated processing that takes movement and body position into account, since both have influence on vital signs.

Modelling of physiological condition in this study addresses all patients and therefore uses the EWS approach. Other models should be applied when detecting events and dealing with specific groups where it makes sense to combine multiple modalities as it is the case with sepsis [19-21]. Since the EWS model is designed to detect significant abnormalities, it may be too insensitive for smaller changes and pattern behavior. The main reason for choosing EWS in this study is, that the model is already widely applied, and with help of wearable technologies, the long-term aim is to monitor patients outside the ICU. This enables automatic monitoring of patients, reduces manual surveillance by medical staff and allows staff to focus on more demanding tasks. Missing data is serious concern since EWS can only be estimated if all values are present. This is currently addressed through the quality control measure. Further adjustments were not considered because the absence of data could often be traced to ongoing interventions by staff. It is furthermore expected that within few years, unobtrusive wearable devices, that can measure most vital signs, will be available for continuous monitoring – leading to either all or no data being available simultaneously depending on connection.

Comparing the EWS timeline with individual EHR reveals that the PCT model can be improved if data about medication, disease history and electronic devices in use are included. A common occurrence within the dataset has been that sometimes a vital sign appears normal, but only does so because a medication or a medical device, such as a respirator, supports the body. Since the model is incomplete in this regard, EWS cannot currently be used to determine when patients can safely be discharged. Nevertheless, the existing EWS model is a promising starting point, as the regression tree in Fig 11 shows. Looking back at national guidelines from Fig 1, an EWS score of 6 and above is considered critical, and hence the physician takes over the surveillance; Similarly, the automatically generated regression tree for our population starts by separating patients at an EWS of 7.

Due to the complexity of the content in EHR entries, it was not possible automatically detect events in the EHR and map them against the timeline. It was nevertheless still possible to use NLP to extract full entries and map them against the EWS

timeline, allowing manual inspection of what caused physiological changes at any given time. In many cases this was sufficient to find cause and resolution to development around events.

Although the dataset is rather small, patterns are already visible, indicating that it is possible to make an automatized tool for quality control of past admissions. It also confirmed that the ICU is fully in control of what goes on within, but is overwhelmed and unaware by the amount of data available per patient. Thus, it is necessary to develop tools that combine the data in a way that is easily understandable to most medical staff.

The selected population has, from the hospital's point of view, increased risk for complications and deterioration. It is therefore safe to assume that their condition fluctuates more than for average patients. This affects observations because deterioration is more prominent and therefore easier to detect. On the other hand, it is still visible through standard deviation and Fig 10, that deceased patients were more unstable than discharged patients. Since the data is gathered from an ICU where surroundings are well controlled, staff may be a major contributor to fluctuations in their efforts to improve continuously deteriorating patients.

The next steps in development, prior to inclusion of decision boundaries are event registrations by staff, and modelling the effects of medication and comorbidities. While the data for comorbidities and medication are available in this study, they are omitted due to population size. The patients were admitted for similar reasons, but their initial co-morbidities, and the medication they receive, are widely distributed. Therefore, few received the same treatments, and when they did, various other factors also affected the outcome. Unfortunately, with a population of 44 patients who are admitted from different departments, there is little overlap, making it impossible to draw unbiased conclusions.

VI. CONCLUSION

The PCT tool demonstrates a novel approach to combining EHR with vital signs measurements to automatically generate an overall overview of a patient's development throughout an admission. The EWS model proved viable for detecting changes in physiological condition, making it possible to evaluate whether health is improving, stable or deteriorating.

The overall analysis of mortality among patients in the dataset revealed that EWS of 8 and above is the most critical. It also showed the next focus of development should be on including medical history to improve the model for patients who have had longer history of illnesses.

The next step will be improving the model for physiological condition by including medication, comorbidities, and disease history. This can ideally be done through a clinical study in which experienced staff registers events and manually scores the patient's condition in fixed intervals, allowing for a better validation of the model.

ACKNOWLEDGMENT

The authors would like to thank PhD student Darius Rohani for input, proof reading and enthusiasm towards the study from

the very beginning. We also appreciate Associate Professor Sadasivan Puthusserypady and postdoc Markus Waser for their valuable support and suggestions throughout the study.

APPROVALS

All data were collected with the approval from the Danish Data Protection Agency (journal number: 2014-41-2743) and the Danish Health Authority (case: 3-3013-558/1).

REFERENCES

- [1] R. Mishra, J. Bian, M. Fiszman, C. R. Weir, Siddhartha Jonnalagadda, J. Mostafa and G. D. F. , "Text Summarization in the Biomedical Domain: A Systematic Review of Recent Research," *Journal of biomedical informatics*, pp. 457-467, 2014.
- [2] R. Pivovarov and N. Elhadad, "Automated methods for the summarization of electronic health records," *Journal of the American Medical Informatics Association*, vol. 22, pp. 938-947, 2015.
- [3] P. B. Jensen, L. J. Jensen and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics* , vol. 13, pp. 395-405, 2012.
- [4] A. Ehteshami, F. Sadoughi, M. Ahmadi and P. Kashefi, "Intensive Care Information System Impacts," *Acta Informatica Medica - Journal of Academy of Medical Sciences of Bosnia and Herzegovina*, vol. 21, no. 3, pp. 185-191, 2013.
- [5] R. Bosman, E. Rood, H.O.-v. Straaten, J. d. Spoel, J. Wester and Durk Zandstra, "Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery," vol. 29, no. 1, pp. 83-90, 2003.
- [6] E. Levesque, E. Hoti, D. Azoulay, P. Ichai, D. Samuel and F. Saliba, "The implementation of an Intensive Care Information System allows shortening the ICU length of stay," *Journal of Clinical Monitoring and Computing*, vol. 29, no. 2, pp. 263-269, 2015.
- [7] D. J. Fraenkel, M. Cowie and P. Daley, "Quality benefits of an intensive care clinical information system," *Critical Care Medicine*, vol. 31, no. 1, pp. 120-125, 2003.
- [8] E. Levesque, E. Hoti, S. d. L. Serna, H. Habouchi, P. Ichai, F. Saliba, D. Samuel and D. Azoulay, "The positive financial impact of using an Intensive Care Information System in a tertiary Intensive Care Unit.," *International Journal of Medical Informatics*, vol. 82, no. 3, pp. 177-184, 2012
- [9] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba and E. Campo, "Smart wearable systems: Current status and future challenges," *Artificial Intelligence in Medicine*, vol. 56, no. 3, pp. 137-156, 2012.
- [10] P. J. Soh, G. A. Vandenbosch, M. Mercuri and D. M.-P. Schreurs, "Wearable Wireless Health Monitoring: Current Developments, Challenges, and Future Trends," *IEEE Microwave Magazine*, vol. 16, no. 4, pp. 55-70, 2015.
- [11] R. Morgan and M. Wright, "In defense of early warning scores," *British Journal of Anesthesia*, vol. 99, no. 5, pp. 747-748, 2007.
- [12] R. Morgan, F. Williams and M. Wright, "An early warning scoring system for detecting developing critical illness," *Clinical Intensive Care*, vol. 8, no. 100, 1997.
- [13] A. Corfield, F. Lees, I. Zealley, G. Houston, S. Dickie, K. Ward and C. McGuffie, "Utility of a single early warning score in patients with sepsis in the emergency department.," *Emergency Medicine Journal*, vol. 31, no. 6, pp. 482-487, 2014.
- [14] U. Kyriacos, J. Jelsma and S. Jordan, "Monitoring vital signs using early warning scoring systems: a review of the literature," *Journal of Nursing Management*, vol. 3, no. 19, pp. 311-330, 2011.
- [15] David R. Prytherch, G. B. Smith, P. E. Schmidt and P. I. Featherstone, "ViEWS - Towards a national early warning score for detecting adult inpatient deterioration," *Resuscitation*, vol. 81, pp. 932-937, 2010.
- [16] T. A. Williams, K. M. Ho, G. J. Dobb, J. C. Finn, M. Knuiman and S. A. R. Webb, "Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients," *British Journal of Anesthesia*, vol. 104, no. 4, pp. 459-464, 2010.
- [17] M. Albur, F. Hamilton and A. P. MacGowan, "Early warning score: a dynamic marker of severity and prognosis in patients with Gram-negative bacteraemia and sepsis," *Annals of Clinical Microbiology and Antimicrobials*, 15:23, 2016.
- [18] A.-Y. Wang, C.-C. Fang, S.-C. Chen, S.-H. Tsai and W.-F. Kao, "Periarrest Modified Early Warning Score (MEWS) predicts the outcome of in-hospital cardiac arrest," *Journal of Formosan Medical Association*, vol. 115, pp. 76-82, 2016.
- [19] A. M. Drewry, T. C. Bailey, B. M. Fuller and R. S. Hotchkiss, "Body temperature patterns as a predictor of hospital-acquired sepsis in afebrile," *Critical Care*, vol. 17, no. R200, 2013.
- [20] C. A. Buchan, A. Bravi and A. J. E. Seely, "Variability Analysis and the Diagnosis, Management, and Treatment of Sepsis", *Current Infectious Disease Reports*, vol. 14, no. 5, pp. 512-521, 2012
- [21] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. Poll, J.-L. Vincent, and D. C. Angus, "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)", *Journal of the American Medical Association*, vol. 315, no. 8 pp. 801-810, 2016

APPENDIX D

APPENDIX D – PUBLICATION IN PROCEEDINGS OF THE 39TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY

Queue-Based Modelling and Detection of Parameters Involved in Stroke Outcome

Authors:

Adnan Vilic, John Asger Petersen, Troels Wienecke, Troels W. Kjær, and Helge B. D. Sorensen

Status: Published

Queue-Based Modelling and Detection of Parameters Involved in Stroke Outcome

Adnan Vilic^{*1}, John Asger Petersen², Troels Wienecke³, Troels W. Kjaer³ and Helge B. D. Sorensen¹

Abstract— We designed a queue-based model, and investigated which parameters are of importance when predicting stroke outcome. Medical record forms have been collected for 57 ischemic stroke patients, including medical history and vital sign measurement along with neurological scores for the first twenty-four hours of admission. The importance of each parameter is identified using multiple regression combined with a circular queue to iteratively fit outcome. Out of 39 parameters, the model isolated 14 which combined could estimate outcome with a root mean square error of 1.69 on the Scandinavian Stroke Scale, where outcome for patients were 36.75 ± 10.99 . The queue-based model integrating multiple linear regression shows promising results for automatic selection of significant medically relevant parameters.

I. INTRODUCTION

Stroke occurs when brain blood flow is compromised. This happens either through bleeding (hemorrhage) or when a formed blood clot prevents blood supply to tissue (ischemia). At least 80% of incidents are ischemic, and treatable with intravenous recombinant tissue plasminogen (rt-PA), also known as alteplase, to dissolve blood clots. The dosage guideline is based on the patient’s weight, and is generally 0.9 mg/kg administered over an hour. Reducing the dosage to 0.6 mg/kg can insignificantly decrease mortality but in exchange for an increase in disability [1]. Untreated, two million neurons die every minute until brain blood flow is restored [2]. Symptoms of ischemic stroke may be language difficulties, visual loss, loss of motoric functions and numbness. The symptoms heavily depend on where the ischemia is formed, and are noticeable in seconds to hours after onset.

Previous studies have investigated the contributions of individual parameters to outcome in ischemic stroke patients, examining e.g. treatment time, vital signs, glucose levels [3-5]. An already incorporated finding is, that treatment with rt-PA 4.5 hours after onset increases mortality [6]. Once this window of opportunity runs out, hospitals advise physicians not to administer rt-PA.

On admission, and the next 24 hours, neurological functionality is assessed using Scandinavian Stroke Scale (SSS) scores, which examine eight parameters explaining the patient’s progress. Combined, the score can be between 0 – 46, with 46 representing no visible disabilities being present. In this study, we apply multiple linear regression combined with circular

queue data structures to quantify which of already monitored parameters contribute to stroke outcome after 24 hours.

II. METHODS AND MATERIALS

A. Patients

In the years 2013 – 2015, Zealand University Hospital treated 792 ischemic stroke patients with alteplase. Data was collected for sixty-four randomly selected patients, of which seven were excluded because of missing data. The final population is therefore 57 patients. All data, listed in Table 1, were recorded manually by nurses and physicians as part of standard hospital treatment routine, following national guidelines for intravenous thrombolysis treatment during stroke. All variables are only registered once, except for vital signs and assessment of neurological scores. Blood pressure, pulse and SSS are registered every fifteen minutes for the first two hours of admission. Frequency reduces to once every thirty minutes until the eighth hour of admission, and then once per hour for the remaining sixteen hours. Every other time measurements are registered, temperature and oxygen saturation are also included; except for the hours between second to fourth, fourth to sixth and sixth to eighth.

B. Approach overview

Hand-written medical forms have been acquired for each patient and digitized. The digitized features, along with derived features are then used as input in a circular queue. Features are iteratively tested in a multiple regression model to fit patient’s stroke outcome after 24 hours.

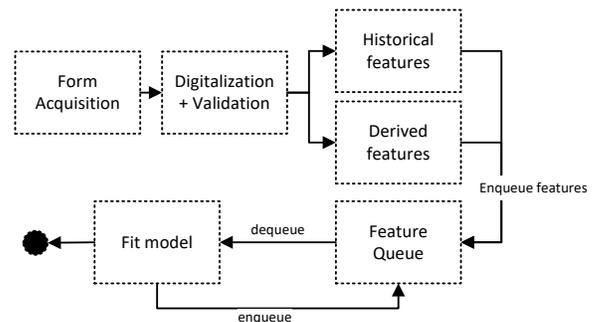


Fig 1: Approach overview

TABLE I: Available data from neurological, vital signs, and medical history forms

Type	Parameter	Population (n=57)	Type
Profile	Age (years)	68.21 ± 11.92	Numeric
	BMI (kg/m ²)	26.65 ± 5.25	Numeric
	Gender (Female)	M=32 F=25	Ordinal Binary
Treatment at hospital	Modified Ranking Score (mRS) onset	0.42 ± 1.02	Ordinal (0-5)
	Onset to arrival (min)	108.12 ± 47.34	Numeric
	Arrival to treatment (min)	30.63 ± 14.24	Numeric
	Dosage rt-PA (mg)	69.91 ± 13.40	Numeric
	Glucose (mmol/l)	7.04 ± 2.18	Numeric
Vascular risk factors	Arterial fibrillation	18 %	Ordinal Binary
	Diabetes	16 %	Ordinal Binary
	Heart insufficiency	8 %	Ordinal Binary
	Hypertension	61 %	Ordinal Binary
	Hyperlipidemia	79 %	Ordinal Binary
	Peripheral arterial disease	4 %	Ordinal Binary
Previous events	Acute myocardial infraction (AMI)	11 %	Ordinal Binary
	Previous apoplexy	50 %	Ordinal (1-3)
	Transient ischemic attack (TIA)	11 %	Ordinal Binary
Relevant drugs	Acetylsalicylic acid	23 %	Ordinal Binary
	Clopidogrel	11 %	Ordinal Binary
	Dipyridamole	2 %	Ordinal Binary
	Anti-hypertensive	37 %	Ordinal Binary
Habits	Alcohol (exceeding national recommendation)	11 %	Categorical
	Smoking (prev. or currently)	51 %	Ordinal (1-4)
Outcomes	Bleeding complications	2 %	Ordinal Binary
	Modified Ranking Score (mRS) after 3 months	2.32 ± 1.75	Ordinal (0-6)
Neurological assessment (24H monitoring)	Scandinavian Stroke Scale, NIH Stroke Scale admission		Numeric
Vital signs (24H monitoring)	pulse, temperature, blood pressure, oxygen saturation level		Numeric

C. Data digitalization and validation

Data is then validated to ensure that it was not corrupted during the digitization process. Parameters that were only recorded once, were further examined to see intervariable relationships, as illustrated in Fig 2. It illustrates a correlation matrix between variables through colors ranging from dark blue to yellow, representing correlation coefficients from -1 to 1. Obvious correlations are visible, such as patients having a history of hypertension are on anti-hypertensive medication ($r = 0.60, p < 0.01$). Similarly, clopidogrel, which is an antiplatelet used to prevent stroke, was in this population used by patients with previous incidents of transient ischemia attack (TIA) ($r = 0.63, p < 0.01$). The time from hospital arrival to treatment was slightly shorter for females ($r = -0.35, p < 0.01$), and for patients who arrived quicker to the hospital ($r = -0.31, p < 0.02$). The shorter pre-hospitalization time indicates fewer complications in terms of compliance, leading to quicker treatment after arrival at the hospital.

D. Derived Features

Potential features are also derived from SSS and vital signs measurements for systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, temperature and oxygen saturation level. For each parameter, mean and standard deviation are estimated during the first two hours of admission. Also, the correlation between SBP and DBP is examined as a higher correlation may be associated with better outcome [7]. So, in addition to the features from Table I, following other

features are also examined:

- Mean and Standard Deviation: DBP, SBP, Oxygen saturation, Pulse, SSS, Temperature.
- Correlation coefficient between DBP and SBP.

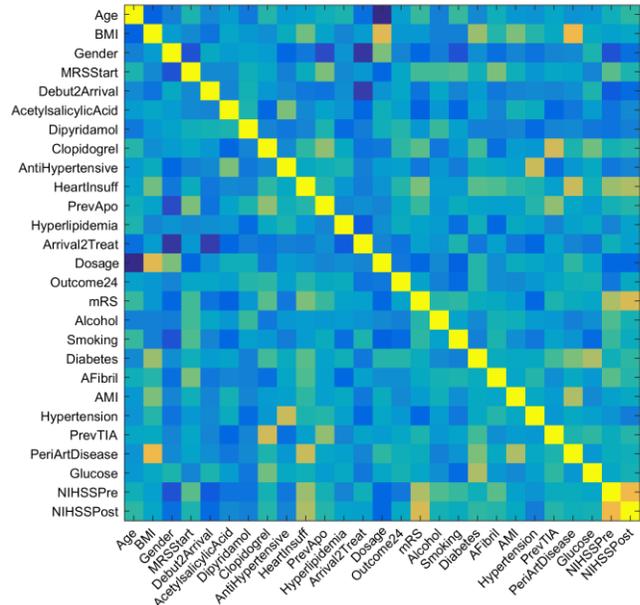


Fig 2: Intervariable relationships through correlation

E. Feature selection and classification

Queues are data structures where data is accessed like their real-world equivalents. In first-in-first-out queues, data is being processed in the order in which they arrived. The queue becomes circular when elements are placed at the back into the queue after they were taken out. Queues are in this study used for storing features that are to be tested in the fitting model. In each iteration, a feature is dequeued (removed from the queue), tested in a fitting model, and either kept as part of the model or enqueued (put back at the end of the queue).

The fitting model is a multiple linear regression model with mixed nominal and continuous variables, and SSS after 24 hours as the dependent variable. When a new feature is introduced during an iteration, its confidence is tested and the feature is included in the model if its $p < 0.20$. If the root mean square error (RMSE) is furthermore lower than that of previous iterations, the current fitting model is stored as the best so far. If the $p > 0.20$, the feature is enqueued again. Since the entire regression model is updated in every iteration, the estimate and confidence will also be updated. Therefore, a correction step is necessary. When the entire queue is tested without any changes to the model, all features with confidences of $p > 0.20$ are enqueued again. This results in a continuously improving fitting model which includes features with confidences up to $p < 0.20$. The process is illustrated in Fig 3.

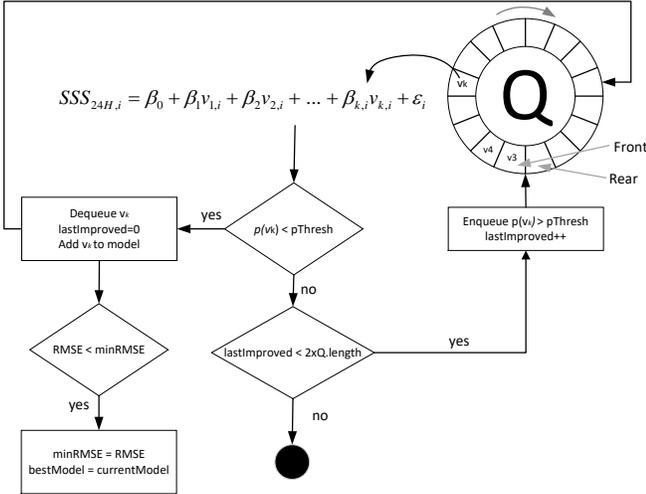


Fig 3: Scandinavian Stroke Scale outcome fitting model where features are iteratively tested in a multiple regression model with a circular queue providing features

III. RESULTS

Before investigating which parameters are of greatest importance to stroke outcome, the system's robustness is verified. Depending on the order in which variables are introduced to the regression model, the final model is likely to vary. This is because variables are added and removed one by one, and it is possible that interaction between variables are overlooked. In Fig 4, the orange line outlines fitting progress when variables are provided in random order. To verify the fitting, it is tested again (blue line), but with variables in the queue being provided in the order of their confidence from the previous model.

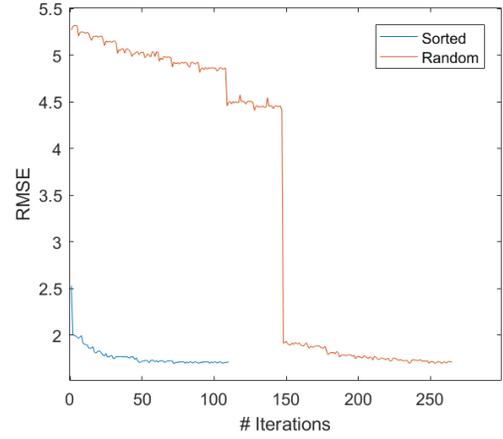


Fig 4: Optimizing fitting model for stroke outcome. Orange model is fitted with variables added randomly to the model. Blue model shows fitting with variables from previous best-fit model.

In both cases, the best model included the same variables and fitting did not improve after reaching $RMSE = 1.69$. The reason for RMSE being 5.23 already from start in the random model is because the first random variable was onset SSS – which is highly correlated with SSS at discharge ($r = 0.86$). The mean SSS in the first two hours is even more correlated with outcome ($r=0.93$), which is why the ordered model starts at $RMSE = 2.6$. The SSS readings from the first two hours alone are enough to fit outcome with $RMSE = 2.01$, but by including other variables, a more accurate fitting is achieved.

TABLE II: Best-fit multiple linear regression model for stroke outcome for dataset population with $RMSE 1.69$

Variable	Estimate	SE	t-Statistic	p-Value
(Intercept)	6.90	3.95	1.75	0.09
SSS_onset	-0.43	0.07	-6.03	<0.01
Hyperlipidemia	-0.94	0.61	-1.53	0.13
Arrival2Treatment	-0.04	0.01	-2.13	0.04
Alcohol_High	-1.30	0.93	-1.40	0.17
SBP_std	-0.05	0.04	-1.33	0.19
Dosage	-0.05	0.03	-1.60	0.12
Pulse_mean	-0.05	0.02	2.52	0.01
SSS_mean	1.42	0.08	17.23	<0.01
mRS_onset	0.50	0.27	1.86	0.07
PrevTIA	1.70	0.80	2.13	0.04
Age	-0.04	0.03	-1.43	0.16
PeripArterDisease	-4.94	2.08	-2.37	0.02
BMI	0.20	0.10	2.04	0.04
BP Correlation	1.35	0.79	1.72	0.09

Table II shows the variables that are included in the determined model. Variables with significant confidences ($p < 0.05$) are highlighted bold. The insignificant ones are not removed as they contribute through interaction. The t-statistics clearly show the contribution differences between SSS measures and the remaining variables which are quite close to each other. The greatest contributor to poor outcome for the population was peripheral arterial disease, and although it could make sense, it is important to note that only two patients in the population had the diagnosis (see Table I).

TABLE 18: Best-fit multiple linear regression model for stroke outcome for dataset population (without SSS). RMSE=6.30

Variable	Estimate	SE	t-Statistic	p-Value
(Intercept)	63.44	8.24	7.70	<0.01
Pulse_mean	-0.19	0.06	-3.00	<0.01
Arrival2Treatment	-0.16	0.07	-2.42	0.02
BMI	-0.40	0.36	-1.11	0.19
mRS_onset	-1.63	0.98	-1.67	0.10
Dosage	0.28	0.11	2.53	0.02
BP Correlation	-4.32	2.85	-1.52	0.13
Hyperlipidemia	-8.52	2.36	3.61	<0.01
Alcohol_High	-8.41	3.29	-2.56	0.01
SBP_std	-0.47	0.18	-2.68	0.01
HeartInsuff.	-14.86	3.96	-3.75	<0.01
Hypertension	3.53	1.99	1.77	0.08
Pulse_std	-0.54	0.23	-2.31	0.02

Table III demonstrates the fitting model when data about SSS data is excluded. The intra-variable estimate ratio increases, but RMSE becomes 6.30. Most variables are the same as in the best-fit model from Table II.

IV. DISCUSSION

Through a combination of circular queuing and multiple linear regression, it was possible to reduce dimensionality and deduce important factors of outcome for stroke patients. The main limitation of the approach lies in the possibility of overseeing contributions of variables when combined. Variables can be left out if they together significantly contribute to the model but individually, in no other combination have a confidence of $p < 20$.

Stroke outcome can primarily be derived by SSS which evaluates the severity of the stroke and its progress. Contributions of other factors are much smaller and possibly overseen if they are well regulated. As an example, the arrival to treatment would be considered insignificant if all patients were treated at onset, because the dataset lacks representation on what happens when the patient has been left untreated for hours after onset.

From the vascular risk factors, most influential parameters were hyperlipidemia, heart insufficiency, hypertension and peripheral arterial disease. While hyperlipidemia and hypertension were well represented in the dataset, only two patients suffered from peripheral arterial disease and four patients had heart insufficiency.

In terms of treatment, dosage, modified ranking score (mRS) and arrival at hospital to treatment start were the most important. It is interesting that onset to arrival at hospital was discarded in all scenarios, but arrival to treatment was not. A plausible explanation is that, once patients are admitted, treatment is initialized quickly unless difficulties occur. If the patient is restless, the patient needs a sedative before CT scan can be performed and the stroke type verified.

From the vital signs, pulse and SBP contributed to the model along with BP correlation. In the best-fit model that includes SSS, BP correlation positively contributes to outcome, while it has the opposite effect in the model without SSS. In both cases, the contribution is insignificant and serves as an interaction variable.

The results presented in tables II and III demonstrate that individual estimate contributions of variables can be linked to the model but not directly to SSS. Although high alcohol consumption reduces outcome by -8.41, it only reduces outcome by that much in a model that exactly includes all variables from table III. From a medical point of view, the most unexpected results were increased BMI and previous TIA being linked to a more positive outcome in the dataset.

V. CONCLUSION

The presented queue-based modelling algorithm narrowed down variables of importance for stroke outcome from 39 to 14 variables. Data about SSS from the first two hours alone lowered RMSE to 2.01 but the remaining variables further improved accuracy and lower RMSE. Although half of the identified variables had a significance of $p < 0.05$, the less significant variables contributed to lowering RMSE from 2.01 to 1.69. Some classes were underrepresented in the dataset, such as peripheral arterial disease and heart insufficiency, making them unreliable. The next step is therefore to expand the dataset to include more patients, allowing a more realistic determination of parameters involved in stroke outcome.

The designed queue-based model combined with multiple linear regression shows promising results as a general approach, with wide applicability, for automatic selection of significant parameters.

REFERENCES

- [1] C. S. Anderson, T. Robinson, R. I. Lindley, H. Arima, P. M. Lavados, T.-H. Lee, J. P. Broderick, X. Chen, G. Chen, V. K. Sharma, J. S. Kim, N. H. Thang, Y. Cao, M. W. Parsons and C. Levi, "Low-Dose versus Standard-Dose Intravenous Alteplase in Acute Ischemic Stroke," *The New England Journal of Medicine*, vol. 24 no. 374, pp. 2313-2323, 2016.
- [2] J. L. Saver, G. C. Fonarow, E. E. Smith, M. J. Reeves, M. V. Grau-Sepulveda, W. Pan, D. M. Olson, A. F. Hernandez, E. D. Peterson and L. H. Schwamm, "Time to Treatment With Intravenous Tissue Plasminogen Activator and Outcome From Acute Ischemic Stroke," *Journal of American Medical Association*, vol. 23 no. 309, pp. 2480-2488, 2013.
- [3] J. Reith, H. S. Jorgensen, P. M. Pedersen, H. Nakayama and H. O. Raaschou, "Body temperature in acute stroke: relation to stroke severity, infarct size, mortality, and outcome," *Lancet*, pp. 422-425, 1996.
- [4] A. Y. Poppe, S. R. Majumdar, T. Jeerakathil, W. Ghali, A. M. Buchan and M. D. Hill, "Admission Hyperglycemia Predicts a Worse Outcome in Stroke Patients Treated With Intravenous Thrombolysis," *Diabetes Care*, vol. 34, no. 4, pp. 617-622, 2009.
- [5] A. A. Wong and S. J. Read, "Early changes in physiological variables after stroke," *Annals of Indian Academy of Neurology*, vol. 4 no. 11, pp. 207-220, 2008.
- [6] K. R. Lees, E. Bluhmki, R. v. Kummer, T. G. Brott, D. Toni, J. C. Grotta, G. W. Albers and M. Kaste, "Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials," *Lancet*, vol. 375 no. 9727, pp. 1695-1703, 2010.
- [7] B. Gavish, I. Z. Ben-Dov and M. Bursztyn, "Linear relationship between systolic and diastolic blood pressure monitored over 24h: assessment and correlates," *Journal of Hypertension*, vol. 26, no. 2, pp. 199-209, 2008.

APPENDIX E

APPENDIX E – SUBMISSION TO PROCEEDINGS OF THE 39TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY

Automatic Identification of Diabetes Based on Vital Signs and Medical History in Post-Stroke Patients

Authors:

Adnan Vilic, Troels Wienecke, Troels W. Kjær, and Helge B. D. Sorensen

Status: Accepted

Automatic Identification of Diabetes Based on Vital Signs and Medical History in Post-Stroke Patients

Adnan Vilic*¹, Troels Wienecke², and Troels W. Kjaer² and Helge B. D. Sorensen¹

Abstract— This study investigates which parameters most accurately identify diabetes mellitus patients from data of the first 24 hours of admission to the stroke unit. Data consists of medical records after admission, and forms in which vital signs and admission progress are registered. It includes 72 patients of which 22 had a pre-admission diagnosis of DM. Through bidirectional feature selection, 12 features were isolated and used in a quadratic Support Vector Machine classifier. The classifier successfully identified DM diagnosed patients with an accuracy of 87.5%.

I. INTRODUCTION

Stroke is one of the most prominent causes of death in developed countries. It is estimated that around 85% of cases are ischemic, meaning that the blood supply to the brain is blocked by clots closing the arteries. Diabetes mellitus (DM) after stroke is associated with poor outcome weeks and months after patients are discharged, but determining whether a patient has DM is a challenge since the endocrine and cardiovascular system behave abnormally. A study by Gray et al found that even glycated hemoglobin in the blood (HbA1c), which reflects blood glucose over several prior weeks, is not sensitive enough, and that potentially one third of stroke patients may have unrecognized DM [1].

This study investigates an automatic data-driven approach to determining the prominent factors in separating DM from non-DM stroke patients through use of the data that is already available in stroke units. Achieving a high accuracy in automatic separation the two groups is the first step towards identification of unrecognized DM.

II. METHODS

Data were obtained for 72 patients, of which 22 were considered having DM because of registered information in the national patient database, previous medical records or the evaluation form during admission. It included profiling (age weight etc.), treatment (stroke onset time, time from arrival to treatment, dosage etc.), vascular risk factors, previous stroke events, relevant prescribed drugs, neurological assessment, image findings and statistical features derived from vital signs – such as mean and standard deviation for systolic/diastolic blood pressure, pulse, temperature and oxygen saturation level.

Significant features are determined by using bidirectional feature selection, where forward selection includes features that increase classification accuracy of a support vector machine (SVM) classifier, that is trained to distinguish DM from non-DM post-stroke patients. The classifier is tested with data for the entire twenty-four hours, and for only the first two hours. The SVM was quadratic with a kernel scale of 2.77 in the final

setup. Training was done with 5-fold cross validation, with at least 4 DM patients presented in each fold. Data is standardized prior to training through estimation of z-scores, because of ratio differences in parameters.

III. RESULTS

Best performance is obtained when using vital signs for the first two hours of admission. The final model consists of Twelve parameters: BMI, time from onset to hospital arrival, time from arrival to treatment, previous apoplexy or transient ischemic attack, dosage, hypertension, glucose, mean and standard deviation for systolic and diastolic blood pressure.

		Predicted class		
		non-DM	DM	
Actual	n=72 non-DM	47	3	50
	DM	6	16	22
		53	19	

Fig 1: Confusion matrix for Support Vector Machine classification, given medical history and two hours of admission data

Fig 1 shows the confusion matrix for the SVM classifier when above features are used with up to two hours of vital signs data. Overall obtained accuracy is 87.5%, which is decent, but it is important to keep in mind that slightly over two third of the population belong to the same class.

IV. DISCUSSION

Using signal processing techniques, we can isolate features of importance when dealing with diseases such as DM, allowing us to optimize and improve recording of data. In this study, bidirectional feature selection found that even four blood pressure readings per hour can have significant impact on separating DM from non-DM patients in post-stroke settings. When increasing vital signs measurements from two hours to twenty-four hours, the accuracy drops to 80.6% because the patient groups have more differences once stroke occurs e.g. one group's autonomic nervous system is more affected than the other's. As time passes however patients become stable and similar, making them harder to distinguish.

REFERENCES

- [1] C. S. Gray, J. F. Scott, J. M. French, K. G. M. M. Alberti and J. E. O'Connell, "Prevalence and prediction of unrecognized diabetes mellitus and impaired glucose tolerance following acute stroke," *Age and Ageing*, vol. 33, pp. 71-77, 2004.

APPENDIX F

APPENDIX F – SUBMISSION TO JOURNAL OF TRANSLATIONAL ENGINEERING IN HEALTH AND MEDICINE

Identifying Diabetes Patients in Post-Stroke Settings by Combining Vital signs and Medical History

Authors:

Adnan Vilic, John Asger Petersen, Helge B. D. Sorensen, Troels Wienecke, and Troels W. Kjær

Status: Submitted

Identifying Diabetes Patients in Post-Stroke Settings by Combining Vital signs and Medical History

Adnan Vilic^{*1}, John Asger Petersen², Helge B. D. Sorensen¹, Troels Wienecke³, and Troels W. Kjaer³

Abstract— One third of stroke patients suffer from potentially unrecognized diabetes mellitus (DM) which is associated with increased morbidity and mortality if left untreated. This study investigates data from the first 24 hours of admission to the stroke unit, to identify the parameters most accurately identifying DM patients. Medical records of the first 24 hours after admission for acute ischemic stroke were collected and digitized. Totally, 72 patients of which 22 had a pre-admission diagnosis of DM were included. Through bidirectional feature selection, 12 features were isolated and used in a quadratic Support Vector Machine classifier. The classifier successfully identified DM diagnosed patients with an accuracy of 87.5%. Our findings indicate that the differences in patient groups are greater in the first hours of admission and that the development of blood pressure is an important contributor.

Index Terms— Biomedical monitoring, Diabetes, Diseases, Hypertension, Classification algorithms, Machine learning, Support vector machines

I. INTRODUCTION

STROKE is one of the most prominent causes of death in developed countries. The disturbance in blood supply to the brain frequently causes irreversible damage, making stroke the leading cause of disabilities in adults. Although stroke primarily affects elder people, it is a disease of all age groups depending on many factors such as lifestyle, medical history and genetics. It is estimated that 10-15% of stroke incidents are due to hemorrhage, while around 85% are ischemic, meaning that the blood supply to the brain is blocked by clots closing the arteries. When a stroke occurs, the symptoms are typically visible for observers in form of loss in motoric or language functions, or subjective to patients as visual disturbance or numbness in parts of the body. Depending on the location and severity of the stroke, signs and symptoms can be noticed in seconds to hours after onset.

The standard treatment of ischemic stroke uses intravenous recombinant tissue plasminogen (rt-PA), also known as alteplase, to dissolve blood clots. The dosage is generally 0.9 mg/kg with 10% as bolus, and the rest is administered continuously over an hour. Anderson et al. demonstrated that reducing the dosage to 0.6 mg/kg insignificantly decreases mortality but causes an increase in disability [1].

Previous studies show that diabetes mellitus (DM) and hypoglycemia after stroke are associated with poor outcome weeks and months after patients are discharged [2-6]. A review

by Luitse et al [5] has found that mortality in DM is not increased compared to patients without DM (non-DM) in the first three months but is increased for DM after one year, especially for patients younger than 50 years; Poor outcome could even be documented after six months in DM patients [2]. When it comes to disabilities Kaarisalo et al determined that recovery after stroke was slower for 51% of DM patients in a population of 1103 after a period of twenty-eight days [6].

There is an overrepresentation of patients with a diagnosis of DM in stroke units, but there are also cases where patients unknowingly have the disease. In the acute phase determining whether a patient has DM is a challenge since the endocrine and cardiovascular system behave abnormally. As an example, hyperglycemia is reported in 30-40% of stroke patients, of which most were non-DM [7-9]. A potentially more accurate acute estimate of DM status is glycated hemoglobin in the blood (HbA1c), which reflects blood glucose over several prior weeks. Measurement of HbA1c is however not sensitive enough. Gray and coworkers studied sixty-two post-stroke patients for twelve weeks, where HbA1c was compared to oral glucose tests, and found that potentially one third of stroke patients may have unrecognized DM [10].

This study investigates an automatic data-driven approach to determining the prominent factors in separating DM from non-DM stroke patients through use of the data that is already available in stroke units. Achieving a high accuracy in automatic separation the two groups is the first step towards identification of unrecognized DM patients.

II. PATIENTS

A total of 792 ischemic stroke patients were treated with alteplase at the department of neurology at Zealand University Hospital in the period from 2013 – 2015. From these, data was collected for seventy-eight randomly selected patients and later expanded with additional thirteen randomly selected DM patients (See Fig 1). Nine non-DM patients were excluded because they were relocated to another hospital within few hours after admission. Additional nine patients, of which one had DM, were excluded due to missing data. A non-DM patient was excluded from the study due to history of gestational diabetes. The final population therefore consists of 72 patients.

A patient was considered having DM if the information was registered in the national patient database, any previous medical records or the evaluation form during stroke-related admission.

The collected and processed data were all recorded manually

¹ Technical University of Denmark, Department of Electrical Engineering, Oersteds Plads, 2800 Kgs. Lyngby, Denmark.

² Bispebjerg Hospital, Intensive Care Unit, Z, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark.

³ Zealand University Hospital, Department of Neurology, Sygehusvej 10, 4000 Roskilde, Denmark.

by nurses and physicians as part of the standard hospital treatment routine, which follows national guidelines for intravenous thrombolysis treatment during acute ischemic stroke. Table I contains a list of all information gathered throughout the study.

All measures are registered once per patient, except for the “24 hours monitoring” measurements which follow a different pattern depending on the variable. Blood pressure, pulse and Scandinavian Stroke Scale (SSS) are registered every fifteen minutes for the first two hours of admission. The frequency is then reduced to once every thirty minutes until the eight hour of admission, then once per hour for the remaining sixteen hours. Every other time that measurements are registered, temperature and oxygen saturation are also included; except for the hours between second to fourth, fourth to sixth and sixth to eighth.

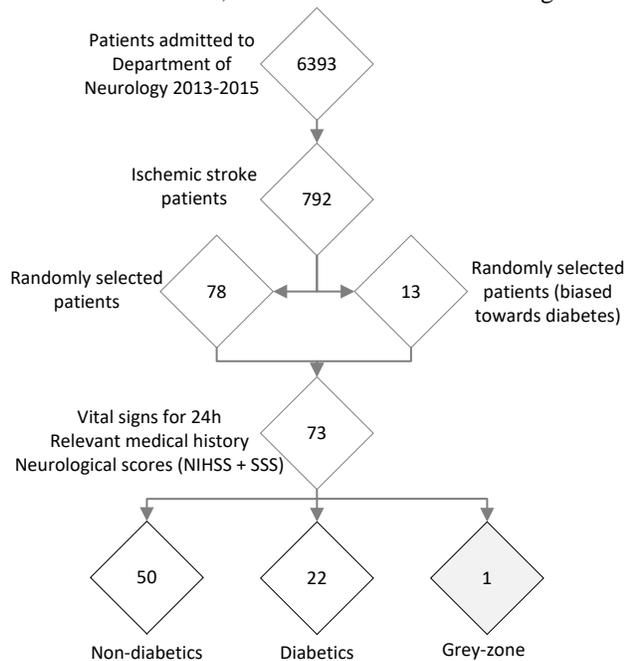


Fig 1: Patient selection overview

III. METHODS

A. Approach overview

The overall approach is illustrated in Fig 2, and starts with acquiring of hand-written evaluation forms for each patient. If neurological scores and vital signs measurement for twenty-four hours are available, along with the medical history, they are digitalized, validated and corrected in case of obvious mistakes. The significant features are determined and used as input for a support vector machine (SVM) classifier trained to distinguish between DM and non-DM post-stroke patients. SVM classification is tested with data for the entire twenty-four hours, and for the first two hours. The two-hour period is tested partly to see if differences can be detected early in the admission, but also to see how the groups react to treatment during the first hour during – and shortly after – bolus is administered.

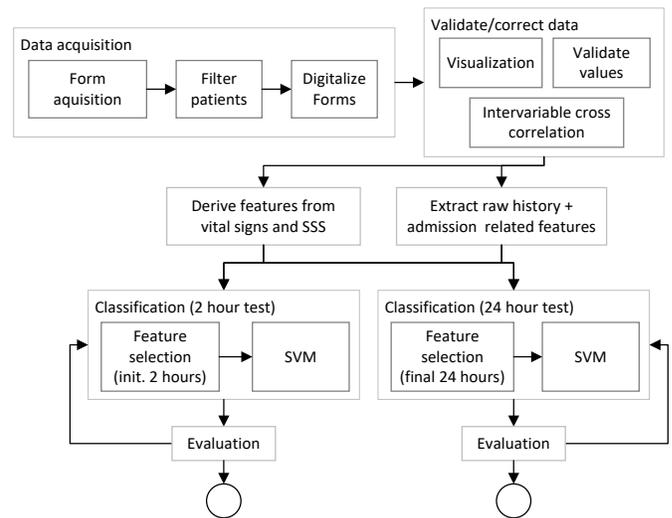


Fig 2: Overall approach of study to identify potentially diabetic patients

B. Data digitalization and validation

The digitalization process from paper to digital data was split into three parts with respect to frequency of recording. Values that were only recorded once, were put into the same comma-separated values (CSV) file, Scandinavian Stroke Scale (SSS) and vital signs into one, and finally the National Institutes of Health Stroke Scale (NIHSS) into its own file because the values that make up NIHSS are the only ones measured once at admission and once after 24 hours. Fig 3 illustrates this, where all papers for a patient make up three CSV files, which are in the end merged into a MATLAB file.

Data are then visualized and validated to ensure that all have been correctly digitized, and aligned appropriately. Since some parameters are only recorded once, the intervariable relationships are also examined and illustrated in Fig 4.

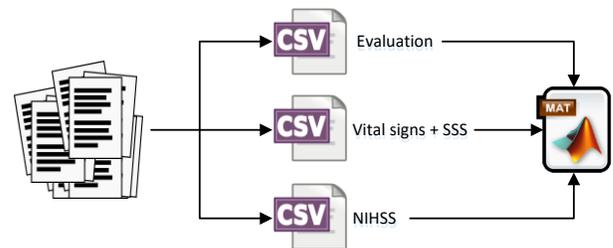


Fig 3: digitalization process from paper to MATLAB files

Data are then visualized and validated to ensure that all have been correctly digitized, and aligned appropriately. Since some parameters are only recorded once, the intervariable relationships are also examined and illustrated in Fig 4. This figure displays correlations between variables color coded from dark blue to yellow, representing correlation coefficients from

TABLE 19: Available data from neurological, vital signs, and medical history forms

Type	Parameter	Diabetes (n=22)	Non-Diabetes (n=50)	Type
Profile	Age (years)	69.27 ± 9.91	68.12 ± 11.92	Numeric
	BMI (kg/m ²)	30.80 ± 6.45	25.72 ± 3.39	Numeric
	Gender (Female)	50 %	56 %	Ordinal Binary
Treatment at hospital	mRS prior onset	0.55 ± 0.91	0.32 ± 0.96	Ordinal (0-5)
	Onset to arrival (min)	124.68 ± 60.67	105.68 ± 43.71	Numeric
	Arrival to treatment (min)	33.36 ± 21.15	31.04 ± 14.91	Numeric
	Dosage rt-PA (mg)	78.27 ± 10.04	69.10 ± 13.32	Numeric
	Glucose (mmol/l)	9.04 ± 2.96	6.47 ± 1.37	Numeric
Vascular risk factors	Arterial fibrillation	23 %	14 %	Ordinal Binary
	Heart insufficiency	18 %	4 %	Ordinal Binary
	Hypertension	86 %	58 %	Ordinal Binary
	Hyperlipidemia	86%	66 %	Ordinal Binary
	Peripheral arterial disease	18 %	0 %	Ordinal Binary
Previous events	Acute myocardial infraction	18 %	10 %	Ordinal Binary
	Apoplexy	50 %	24 %	Ordinal (1-3)
	Transient ischemic attack (TIA)	23 %	8 %	Ordinal Binary
Relevant drugs	Acetylsalicylic acid	41 %	21 %	Ordinal Binary
	Clopidogrel	32 %	10 %	Ordinal Binary
	Dipyridamole	9 %	2 %	Ordinal Binary
	Anti-hypertensive	64 %	34 %	Ordinal Binary
Habits	Alcohol intake (exceeding national recommendation)	4%	2%	Categorical
	Smoking (prev. or currently)	77 %	46 %	Ordinal (1-4)
Outcomes	Outcome 24h	-	-	Ordinal (1-5)
	Outcome discharge	-	-	Ordinal (1-5)
	Bleeding complications	4 %	0 %	Ordinal Binary
	Discharge diagnosis	-	-	Categorical
	mRS after 3 months	2.36 ± 1.94	2.16 ± 1.69	Ordinal (0-6)
	Cause of death (if applicable)	-	-	Categorical
Image findings	CT/MR bleedings	0 %	4 %	Ordinal Binary
Neurological assessment (24 hours monitoring)	Scandinavian Stroke Scale, NIH Stroke Scale admission			Numeric
Vital signs (24 hours monitoring)	pulse, temperature, diastolic blood pressure, systolic blood pressure, oxygen saturation level			Numeric

-1 to 1. As an example, we see an almost perfect match ($r = 0.89, p < 0.01$) between the variable Outcome24 (Outcome 24 hours after admission) and OutcomeT (Outcome at discharge) because most patients are discharged from the unit after 24 hours. Similarly, a higher than average positive correlation is seen between alteplase dosage and BMI ($r = 0.72, p < 0.01$), which is expected because the guidelines for administering alteplase uses body mass to calculate dosage. The reason that dosage is negatively correlated with age ($r = -0.41, p < 0.01$) is because the older patients weighed less, as it is also visible from the relationship between age and BMI. More surprisingly was the fact that the time to get from arrival at hospital to treatment start was longer for females (gender = 0). We have no explanation for this phenomenon but are considering it prospectively.

C. Derived Features

Historical information is supplemented with potential features derived from vital signs after admission to investigate whether this data helps separating DM from non-DM patients. The additional features are mean and standard deviation of SSS and all vital signs (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, temperature and oxygen saturation level). standard deviation of SSS and all vital signs (systolic blood pressure (SBP), diastolic blood pressure (DBP),

pulse, temperature and oxygen saturation level). Two additional features are examined: Correlation between SBP and DBP, and change in SSS from admission to 24 hours later.

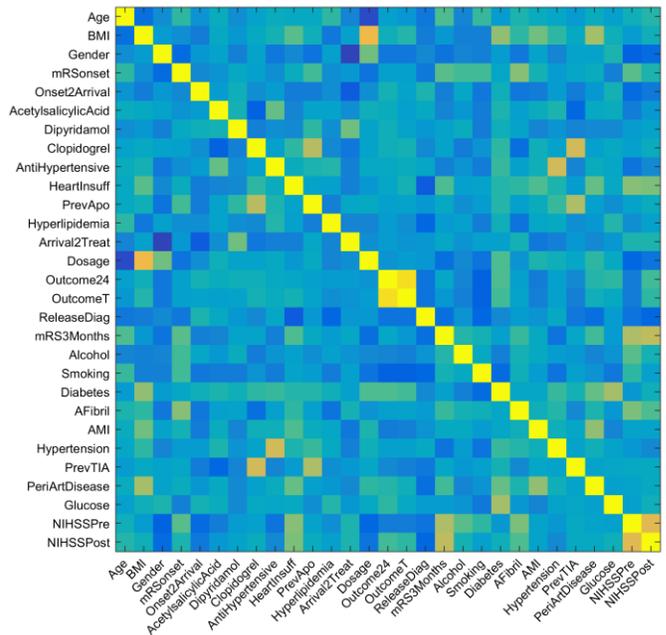


Fig 4: Intervariable relationships through correlation. The correlation coefficient is included since previous studies

have found that there is a positive linear relationship between SBP and DBP which is influenced if the patient has DM [11]. If blood pressure had been measured e.g. three times every hour throughout the entire admission, the change in SBP versus DBP could have been used to extract information about arterial stiffness [12-13]. The expectation was that correlation coefficient is at first widely distributed, and later stabilizes around $r = 0.74$ for most patients [11]. The change is uncertain for DM patients.

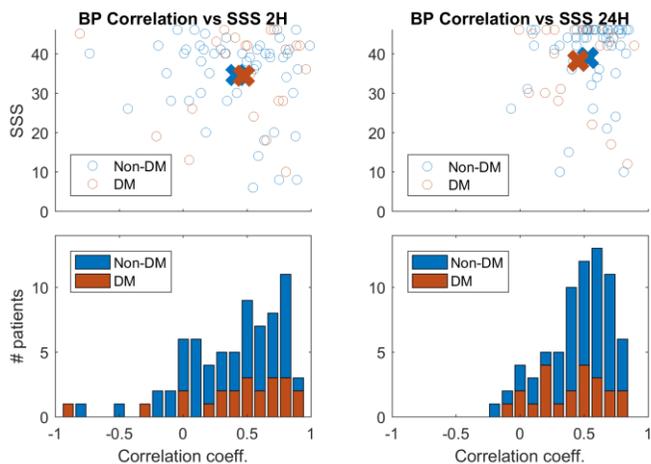


Fig 5: Correlation between diastolic and systolic blood pressure, two and twenty-four hours after initialized treatment.

The two scatter plots in Fig 5 show each patient’s SSS and blood pressure correlation for respectively two first hours and after twenty-four hours. The bold markers represent the center of mass for each group. DM patients improve in SSS by a mean of 3.77 while non-DM improve by 4.34. In terms of correlation however, a decrease is seen in DM patients from $r=0.47$ to $r=0.46$, and in non-DM it increased from $r=0.41$ to $r=0.54$. The histograms show the correlation distribution with a convergence towards 0.6 for non-DM over time, and slightly positive, yet widely distributed, correlation for DM patients.

To summarize, in addition to the features from Table I, following other features are examined:

- Mean: DBP, Oxygen saturation level, Pulse, SBP, SSS, Temperature
- Standard deviation: DBP, Oxygen saturation level, Pulse, SBP, SSS, Temperature
- Correlation coefficient between DBP and SBP.

D. Feature selection and classification

Feature selection poses several challenges. On top of the fact that one third of the patients may have unrecognized DM, the dataset has a low patient to variable ratio. This results in overfitting and poor generalization if all variables are included as features. To reduce dimensionality, Principal Component Analysis (PCA) was considered but discarded because of categorical variables being present and likely of importance; examples being history of hypertension or previous TIA which

are of higher risks in DM patients.

Features were instead selected by using bidirectional feature selection, where forward selection is first used to include features that increase classification accuracy of the SVM, and then features are individually tested eliminated to see if accuracy is changes after their removal. From the variables, available in Table I, some were dismissed prior to this process because they were either rare occurrences in the dataset (bleeding related to dosage, bleeding findings in CT/MR, cause of death) or irrelevant/redundant (discharge diagnosis, Outcome 24h, Outcome discharge, NIHSS). NIHSS was discarded as it is interchangeable with, and addresses many of the same parameters as, SSS [14-15] – but NIHSS was registered only at admission and discharge while SSS was monitored continuously. In terms of vital signs, blood pressure correlation and SSS features, two SVM classifiers were trained. The first, where all readings for the twenty-four hours are included, and the other only for readings from the first two hours (see Fig 2).

The SVM was quadratic and MATLAB’s ‘auto’ setting was used to determine the kernel scale of 2.77 in the final setup. The data was split for training and testing using 5-fold cross validation, with at least 4 DM patients presented in each fold. Because of ratio differences in variables, the data is standardized prior to training through estimation of z-scores.

IV. RESULTS

The best performance was obtained when combining medical history with vital signs monitoring of the initial two hours of admission. For both two and twenty-four hours, the same features resulted in best performance:

- BMI
- From onset to arrival time
- From arrival to treatment time
- Previously had apoplexy
- Dosage
- History of hypertension
- Previously had TIA
- Glucose at arrival
- SBP Mean
- SBP Standard Deviation
- DBP Mean
- DBP Standard Deviation

Fig 6 shows the confusion matrix for the SVM classifier when above features are used with up to two hours of vital signs data. Most misclassifications are with DM patients, which is also to be expected. Ratio-wise the uneven grouping causes the non-DM to be favored due to population size. Another reason can be that, since some patients have unrecognized DM, the milder cases of recognized DM may be similar, and thus end up grouped together. In the same way, wrongly classified non-DM patients could potentially have unrecognized DM. Nevertheless, the overall obtained accuracy is 87.5%, which is

decent, but it is important to keep in mind that slightly over two third of the population belong to the same class.

		Predicted class		
		non-DM	DM	
Actual class	n=72			
	non-DM	47	3	50
DM		6	16	22
		53	19	

Fig 6: Confusion matrix for SVM given medical history and two hours of admission data

When increasing the vital signs from two hours to twenty-four hours, the accuracy drops to 80.6% (see Fig 7). This is likely due to the patient groups having more differences once stroke occurs e.g. one group’s autonomic nervous system is more affected than the other’s. The more time passes however, the more stable and similar both groups become, making them harder to distinguish.

		Predicted class		
		non-DM	DM	
Actual class	n=72			
	non-DM	45	5	50
DM		9	13	22
		54	18	

Fig 7: Confusion matrix for SVM given medical history and twenty-four hours of admission data

Continuing with the better classifier, the confidence levels are examined in further detail. If the confidence is low in the case of misclassifications, the sensitivity of an eventually implemented system can be improved through thresholding, so that only classifications with high confidence are accepted.

Fig 8 shows confidence levels for the SVM classifier. The colors indicate corresponding correct class, and the sign is negated in case of misclassification. I.e. patient seventeen does not have recognized DM but SVM determined with a confidence of $p=0.25$, that the person had DM. Due to the high classification confidences, false classifications cannot be avoided unless additional features are added.

The common traits for misclassified non-DM patients were, that all had hypertension, and received a higher than average dosage of rt-PA. Two of the three patients had glucose levels greater than one standard deviation above average, while the last patient had above average glucose level but a history of TIA as well as apoplexy. No single features stood out in terms of misclassified DM patients.

Fig 9 shows all twelve features after standardization. All blue and brown lines represent non-DM and DM patients respectively, and the dashed lines represent misclassifications. As in Table 1, we see increased mean and standard deviation in

BMI and glucose for DM patients. The figure also shows that all misclassified non-DM patients had hypertension while there was only one misclassification was unrelated to hypertension.

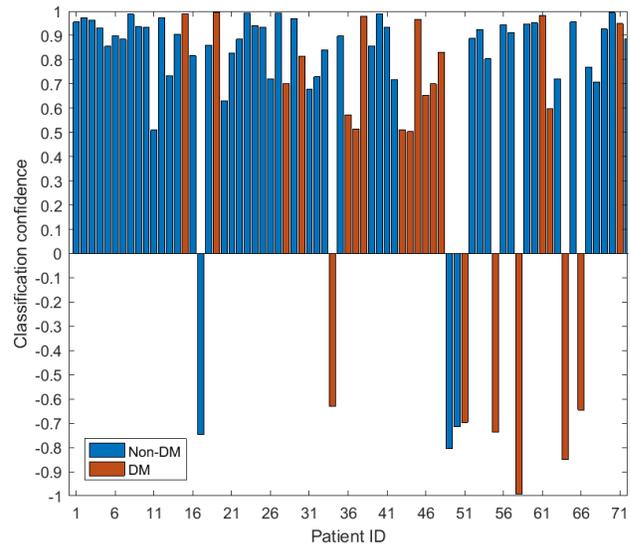


Fig 8: Classification confidence for SVM. Confidence is negated for misclassifications

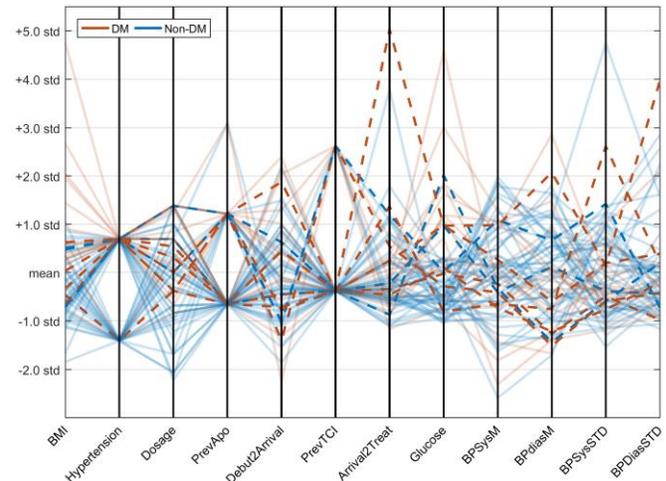


Fig 9: Concurrent feature comparison. Line colors indicate the correct class, and are bold and dashed in case of misclassifications.

V. DISCUSSION

Using signal processing techniques, we can isolate which features are of importance when dealing with diseases such as DM, allowing us to optimize and improve recording of data. In this study, bidirectional feature selection found that, when detecting DM, even four blood pressure readings per hour can have significant impact on separating DM from non-DM patients in post-stroke settings.

At first, vital signs were only investigated for the entire twenty-four hours. Most patients improve greatly throughout the admission however, meaning that their vital signs become more similar over time, as seen from Fig 5. An attempt was also made to combine data from two hours of vital signs monitoring

with the cross-correlation coefficient of blood pressure after twenty-four hours, because the coefficient is more likely to be over $r=0.5$ for non-DM patients. The feature was still automatically eliminated during generation of the SVM model. Elevated blood pressure is seen in about 80% of patients after acute ischemic stroke, and is to some degree preferable, as too high or too low pressure is documented to be associated with poor outcome. Many suggestions have been provided to what causes the elevation, ranging from mental stress due to sudden admission, to the fact that elevation of blood pressure increases the perfusion pressure in the ischemic penumbra [9,16-21]. Like in our findings, Carstensen et al recognized that change in blood pressure could already after four hours be linked to outcome [17]. Since DM patients have increased risk of impaired cerebral autoregulation, it would be interesting to investigate how blood pressure changes after stroke for DM versus non-DM patient. To do so, vital signs should be monitored at a higher resolution, preferably in seconds, rather than every fifteenth minute [22].

In terms of feature selection, an important limitation of this study is the lack of HbA1c as a potential feature which is expected to be higher for patients with unrecognized DM. Since it is a historical measure and does not represent the current state of the stroke patient, it is not standard procedure to test at admission – although some physicians test it. The most recent HbA1c measurements were obtained for all but three patients from the dataset, but many measurements were unrelated to the stroke admission. All values were under the standard boundary <42 mmol/mol, meaning that even DM patients were well controlled. It was also not surprising that dosage was selected as a primary feature after BMI, as guidelines for administration of rt-PA take weight into account. The intervariable dependency is not seen as problematic though, because dosage varies when other factors, such as age and present disability degree, are considered.

The SVM classifier correctly found if a patient is diagnosed with DM in 87.5% cases, proving that there are significant differences allowing the two groups to be separated. Nevertheless, one needs to keep in mind that due to population size, random guessing would result in an accuracy of 69.4%. The population difference is likely the primary cause for misclassifications although unrecognized DM may have a role in it [10]. It was not possible to quantify how many patients in the dataset had undiagnosed DM. Entries in the medical record during admission suggested in many patients that there was a suspicion of DM. This was further investigated in the department of neurology but left for the general practitioner to follow up on.

VI. CONCLUSION

The developed classifier successfully identified which patients have diagnosed DM with an accuracy of 87.5%. The approach is based on data that is already available and registered in many developed countries. Although vital sign recordings were available for twenty-four hours, best performance was achieved with data from the first two hours of admission.

To further improve this approach, vital signs need to be monitored at higher resolution and Hb1Ac readings at time of admission, or within weeks after stroke need to be available. Finally, follow-ups are recommended to identify and correct for unrecognized cases of DM.

Promising results have been achieved for automatically identifying DM in early hours after stroke, which, when combined with information from follow-ups after discharge, could potentially help detecting patients with unrecognized DM.

ACKNOWLEDGEMENT

The authors would like to thank Jakob Skadkær Møller for his aid with access to data, data storage and interest in discussions throughout the project. We also appreciate PhD students Darius Adam Rohani, Ivan Chrilles Zibrandtsen and associate professor Sadasivan Puthusserypady for feedback related to feature and classifier decisions. Finally, special thanks go to Samra Skrijelj for proofreading.

APPROVALS

All data were collected with the approval from the Danish Data Protection Agency (journal number: 2008-58-0020 2014-41-2743) and Region Zealand (journal number. 16-000032).

REFERENCES

- [1] C. S. Anderson, T. Robinson, R. I. Lindley, H. Arima, P. M. Lavados, T.-H. Lee, J. P. Broderick, X. Chen, G. Chen, V. K. Sharma, J. S. Kim, N. H. Thang, Y. Cao, M. W. Parsons and C. Levi, "Low-Dose versus Standard-Dose Intravenous Alteplase in Acute Ischemic Stroke," *The New England Journal of Medicine*, vol. 374, no. 24, pp. 2313-2323, 2016.
- [2] Q. Jia, X. Zhao, C. Wang, Y. Wang, Y. Yan, H. Li, L. Zhong, L. Liu, H. Zheng, Y. Zhou and Y. Wang, "Diabetes and Poor Outcomes Within 6 Months After Acute Ischemic Stroke," *Stroke*, vol. 42, pp. 2758-2768, 2011.
- [3] A. Y. Poppe, S. R. Majumdar, T. Jeerakathil, W. Ghali, A. M. Buchan and M. D. Hill, "Admission Hyperglycemia Predicts a Worse Outcome in Stroke Patients Treated With Intravenous Thrombolysis," *Diabetes Care*, vol. 34, no. 4, pp. 617-622, 2009.
- [4] D. Nikneshan, R. Raptis, J. Pongmoragot, L. Zhou, S. C. Johnston and G. Saposnik, "Predicting Clinical Outcomes and Response to Thrombolysis in Acute Stroke Patients With Diabetes," *Diabetes Care*, vol. 36, no. 7, pp. 2041-2047, 2013.
- [5] M. J. Luitse, G. J. Biessels, G. E. Rutten and L. J. Kappelle, "Diabetes, hyperglycaemia, and acute ischaemic stroke," *The Lancet Neurology*, vol. 11, no. 3, pp. 261-271, 2012.
- [6] M. M. Kaarisalo, I. Raiha, J. Sivenius, P. Immonen-Raiha, A. Lehtonen, C. Sarti, M. Mahonen, J. Torppa, J. Tuomilehto and V. Salomaa, "Diabetes worsens the outcome of acute ischemic stroke," *Diabetes Research and Clinical Practice*, vol. 69, pp. 293-298, 2005.
- [7] S. E. Capes, D. Hunt, K. Malmberg, P. Pathak and H. C. Gerstein, "Stress Hyperglycemia and Prognosis of Stroke in Nondiabetic and Diabetic Patients," *Stroke*, vol. 32, pp. 2426-2432, 2001.
- [8] N. D. Kruyt, G. J. Biessels, J. H. DeVries and Y. B. Roos, "Hyperglycemia in acute ischemic stroke: pathophysiology and clinical management," *Nature Reviews*, vol. 3, pp. 145-155, 2010.
- [9] A. A. Wong and S. J. Read, "Early changes in physiological variables after stroke," *Annals of Indian Academy of Neurology*, vol. 4, no. 11, pp. 207-220, 2008.

- [10] C. S. Gray, J. F. Scott, J. M. French, K. G. M. M. Alberti and J. E. O'Connell, "Prevalence and prediction of unrecognised diabetes mellitus and impaired glucose tolerance following acute stroke," *Age and Ageing*, vol. 33, pp. 71-77, 2004.
- [11] B. Gavisha, I. Z. Ben-Dov and M. Bursztyn, "Linear relationship between systolic and diastolic blood pressure monitored over 24 h: assessment and correlates," *Journal of Hypertension*, vol. 2, no. 26, pp. 199-209, 2008.
- [12] Y. Li, J.-G. Wang, E. Dolan, P.-J. Gao, H.-F. Guo, T. Nawrot, A. V. Stanton, D.-L. Zhu, E. O'Brien and J. A. Staessen, "Ambulatory Arterial Stiffness Index Derived From 24-Hour Ambulatory Blood Pressure Monitoring," *Hypertension*, vol. 47, pp. 359-364, 2006.
- [13] G. Schillaci and G. Pucci, "The dynamic relationship between systolic and diastolic blood pressure: yet another marker of vascular aging?," *Hypertension Research*, vol. 33, pp. 659-661, 2010.
- [14] Laura J Gray, M. Ali, P. D. Lyden and P. M. Bath, "Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in Acute Stroke," *Journal of Stroke and Cerebrovascular Diseases*, vol. 6, no. 18, p. 466-468, 2009.
- [15] T. Askim, . Bernhardt, . Churilov and . B. Indredavik, "The Scandinavian Stroke Scale is Equally As Good As the National Institutes of Health Stroke Scale in Identifying 3-Month Outcome," *Journal of Rehabilitation Medicine*, vol. 10, no 48, pp. 909-912, 2016.
- [16] J. Castillo, R. Leira, M. M. García, M. B. Joaquín Serena and A. Dávalos, "Blood Pressure Decrease During the Acute Phase of Ischemic Stroke Is Associated With Brain Injury and Poor Stroke Outcome," *Stroke*, vol. 2, no. 35, pp. 520-527, 2004.
- [17] H. Christensen, P. Meden, K. Overgaard and G. Boysen, "The course of blood pressure in acute stroke is related to the severity of the neurological deficits," *Acta Neurologica Scandinavica*, vol 106, pp. 142-147, 2002.
- [18] M. E. S. Prakash, "What Causes the Acute Blood Pressure Elevation After Stroke?," *Stroke*, vol. 10, no. 36, pp. 2066, 2005.
- [19] H. P. Mattle, L. Kappeler, M. Arnold, U. Fischer, K. Nedeltchev, L. Remonda, S. M. Jakob and G. Schroth, "Blood Pressure and Vessel Recanalization in the First Hours After Ischemic Stroke," *Stroke*, vol. 2, no. 36, pp. 264-268, 2005.
- [20] P. J. Lindsberg, "Editorial Comment—High Blood Pressure After Acute Cerebrovascular Occlusion," *Stroke*, vol. 2, no. 36, pp. 268-269, 2005.
- [21] A. I. Quershi, "Acute Hypertensive Response in Patients With Stroke," *Pathophysiology and Management*, vol. 2, no. 18, pp. 176-187, 2008.
- [22] Y.-S. Kim, R. Immink, W. Stok, J. Karemaker and J. v. Niels H. Secher, "Dynamic cerebral autoregulatory capacity is affected early in Type 2 diabetes," *Clinical Science*, vol. 8, no. 115, pp. 255-262, 2008.