Predicting adverse cardiac events at the emergency department

A deep learning approach

Nyström, Axel

2025

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

Predicting adverse cardiac events at the emergency department

A deep learning approach

AXEL NYSTRÖM
DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY

Predicting adverse cardiac events at the emergency department

Predicting adverse cardiac events at the emergency department

A deep learning approach

by Axel Nyström

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Assoc. Prof. Jakob Lundager Forberg, Prof. Mattias Ohlsson, Prof. Jonas Björk, Prof. Ulf Ekelund
Faculty opponent: Assoc. Prof. Adam Hulman

To be presented, with the permission of the Faculty of Medicine of Lund University, for public criticism at Belfragesalen on Thursday, the 24th of April 2025 at 13:00.

Organization
LUND UNIVERSITY

Department of Laboratory Medicine
Box 124
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2025-04-24

Sponsoring organization

Author(s)
Axel Nyström

Title and subtitle
Predicting adverse cardiac events at the emergency department A deep learning approach

Abstract

The emergency department is a stressful environment, in which physicians are required to make fast and accurate diagnostic assessments amidst an ever increasing flood of clinical information, including a growing body of medical knowledge. Meanwhile, the digitization of medical health records in combination with recent breakthroughs in Artificial Intelligence is ushering in a new era of precision medicine. Deep-learning powered decision support tools represent a promising avenue for improving patient outcomes and reducing work-flow complexity for physicians. The goal of this thesis was to explore ways to apply modern deep-learning algorithms to improve predictions of adverse cardiac events among chest-pain patients at the emergency department.

The first paper investigated the utility of prior electrocardiograms (ECGs) for predicting major adverse cardiac events (MACE). We found that, contrary to clinical recommendations, prior ECGs will not meaningfully contribute to the predictions of MACE. The second paper aimed at quantifying the benefit of transfer learning for models using ECGs to predict acute myocardial infarction (AMI). We found that a simple transfer-learning strategy of pre-training on patient age and sex resulted in a substantial increase in the downstream performance of AMI, while simultaneously enabling the use of larger and more powerful deep-learning model architectures. The third paper explored options for early rule-out of AMI. The results indicated that as many as 16% of chest-pain patients can be safely ruled out based on age and sex alone, suggesting opportunities to further streamline the clinical pathway for low-risk patients. The fourth paper was concerned with identifying and predicting acute coronary occlusion myocardial infarctions (OMI), which are particularly serious and require urgent invasive treatment. We found that 29% of all AMI patients could be classified as OMI, but only 11% of those patients received timely treatment. Our deep-learning ECG model was able to predict the OMI outcome with an AUC of 88% using only the ECG and medical history, and 95.3% when including the initial high-sensitivity cardiac troponin T lab results.

In conclusion, this thesis identified several promising applications and improvements of deep-learning algorithms in predicting adverse cardiac events in chest-pain patients. Much work remains to be done, including prospective clinical trials, but hopefully we are now one step closer to the implementation of a deep-learning powered decision support tool that can help save lives at the emergency department.

Key words
deep learning, machine learning, transfer learning, electrocardiogram, emergency department, chest pain, myocardial infarction, occlusion myocardial infarction, decision support tool

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
1652-8220

ISBN
978-91-8021-691-3 (print)
978-91-8021-691-3 (pdf)

Recipient's notes

Number of pages
181

Price

Security classification

Signature _____ Date 2025-03-13

Predicting adverse cardiac events at the emergency department

A deep learning approach

by Axel Nyström

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the licentiate student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Tipula Maxima, iStock: Rosendo Serrano Valera

Cover illustration back: Tipula Maxima, iStock: sduben

Funding information: The thesis work received funding from the Swedish Research Council, the Swedish Heart-Lung Foundation, and Sweden's Innovation Agency.

Faculty of Medicine, Department of Laboratory Medicine

ISBN: 978-91-8021-691-3 (print)
ISBN: 978-91-8021-691-3 (pdf)

Typeset in LaTeX

Printed in Sweden by Media-Tryck, Lund University, Lund 2025

For my niblings
Bosse and Freja

Contents

List of publications and author contributions

This thesis is based on the following publications, referred to by their Roman numerals. Contributions follow the CRediT (Contributor Roles Taxonomy) system.

I **Prior electrocardiograms not useful for machine learning predictions of major adverse cardiac events in emergency department chest-pain patients**
A. Nyström, P. Olsson de Capretz, A. Björkelund, J. Lundager Forberg, M. Ohlsson, J. Björk, U. Ekelund
Journal of Electrocardiology, 2024, 82, pp. 42-51. License: CC BY 4.0

My contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization.

II **Transfer learning for predicting acute myocardial infarction using electrocardiograms**
A. Nyström, A. Björkelund, M. Ohlsson, J. Björk, U. Ekelund, J. Lundager Forberg
Submitted to *PLOS Digital Health*

My contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization.

III **Stepwise increasing input to machine learning models predicting 30-day AMI or death in emergency department chest-pain patients**
P. Olsson de Capretz, A. Nyström, A. Björkelund, J. Björk, M. Ohlsson, U. Ekelund
Submitted to *Karger Cardiology*

My contributions: Conceptualization, methodology, software, data curation, writing – original draft, writing – review & editing.

IV **Predicting occlusion myocardial infarctions in the emergency department using artificial intelligence**
A. Nyström, A. Björkelund, H. Wagner, U. Ekelund, M. Ohlsson, J. Björk, A. Mokhtari, J. Lundager Forberg
Submitted to *Annals of Emergency Medicine*

My contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization.

Abstract

The emergency department is a stressful environment, in which physicians are required to make fast and accurate diagnostic assessments amidst an ever increasing flood of clinical information, including a growing body of medical knowledge. Meanwhile, the digitization of medical health records in combination with recent breakthroughs in Artificial Intelligence is ushering in a new era of precision medicine. Deep-learning powered decision support tools represent a promising avenue for improving patient outcomes and reducing work-flow complexity for physicians. The goal of this thesis was to explore ways to apply modern deep-learning algorithms to improve predictions of adverse cardiac events among chest-pain patients at the emergency department.

The first paper investigated the utility of prior electrocardiograms (ECGs) for predicting major adverse cardiac events (MACE). We found that, contrary to clinical recommendations, prior ECGs will not meaningfully contribute to the predictions of MACE. The second paper aimed at quantifying the benefit of transfer learning for models using ECGs to predict acute myocardial infarction (AMI). We found that a simple transfer-learning strategy of pre-training on patient age and sex resulted in a substantial increase in the downstream performance of AMI, while simultaneously enabling the use of larger and more powerful deep-learning model architectures. The third paper explored options for early rule-out of AMI. The results indicated that as many as 16% of chest-pain patients can be safely ruled out based on age and sex alone, suggesting opportunities to further streamline the clinical pathway for low-risk patients. The fourth paper was concerned with identifying and predicting acute coronary occlusion myocardial infarctions (OMI), which are particularly serious and require urgent invasive treatment. We found that 29% of all AMI patients could be classified as OMI, but only 11% of those patients received timely treatment. Our deep-learning ECG model was able to predict the OMI outcome with an AUC of 88% using only the ECG and medical history, and 95.3% when including the initial high-sensitivity cardiac troponin T lab results.

In conclusion, this thesis identified several promising applications and improvements of deep-learning algorithms in predicting adverse cardiac events in chest-pain patients. Much work remains to be done, including prospective clinical trials, but hopefully we are now one step closer to the implementation of a deep-learning powered decision support tool that can help save lives at the emergency department.

Populärvetenskaplig sammanfattning på svenska

Hjärtinfarkt är den vanligaste dödsorsaken både i Sverige och resten av världen, och bröstsmärta är en av de vanligaste besöksorsakerna på akuten. Samtidigt är det bara en bråkdel av bröstsmärtepatienterna som faktiskt har en hjärtinfarkt. En snabb och korrekt handläggning av patienter är nödvändig för att de med livshotande och tidskritiska besvär ska få den behandling de behöver, utan att man överbelastar sjukvårdssystemet eller genomför fler kliniska tester än nödvändigt.

Målet med den här avhandlingen har varit att tillämpa så kallade maskininlärningsalgoritmer på stora mängder patientdata, för att skapa prediktionsmodeller som kan hjälpa sjukvårdspersonal på akuten att fatta bättre och snabbare beslut för patienter med bröstsmärta. Maskininlärning är en form av artificiell intelligens som automatiskt lär sig att identifiera mönster i data genom att träna på exempel. I min forskning har det framför allt handlat om att hitta samband mellan elektrokardiogram (EKG) vid ankomsttillfället och diagnosticerad hjärtinfarkt inom 30 dagar.

EKG är ett viktigt redskap för läkare att identifiera hjärtinfarkt, speciellt för att hitta de mest akuta formerna av hjärtinfarkt som kräver omedelbar invasiv behandling. Analysen av EKG:et är till stor del manuell, och även om det fungerar bra, så finns det forskning som visar att det kan bli ännu bättre. Avhandlingen är baserad på fyra delarbeten, som alla syftar till att utveckla eller förbättra metoder för att predicera hjärtinfarkt på akuten, med hjälp av EKG och annan på akuten lättillgänglig information.

I det första delarbetet undersökte vi tilläggsvärdet av tidigare EKG:er för prediktion av hjärtinfarkt och andra allvarliga utfall. Det är nämligen rekommenderat att jämföra ett nytt EKG med tidigare EKG från samma patient, om sådant finns tillgängligt, eftersom EKG-markörer för hjärtinfarkt ibland är permanenta. Vi fann dock att de tidigare EKG:erna inte ledde till några förbättringar i prediktionerna.

Maskininlärningsmodeller fungerar oftast bättre ju mer data de har tillgång till. Men att skaffa mer data är inte så lätt, särskilt i den medicinska världen. Ett alternativ är att använda så kallad överförd inlärning, där man utnyttjar annan data som finns tillgänglig, men som på olika sätt skiljer sig från den ursprungliga problemdomänen. Detta har varit en mycket framgångsrik strategi i många andra tillämpningar, men har inte tidigare studerats i kontexten av att förutsäga hjärtinfarkter med hjälp av EKG:er. I det andra delarbetet utnyttjade vi närmare en miljon EKG:er från patienter som sökt vård för annat än bröstsmärta. Vi tränade en modell som kunde gissa patienternas ålder och kön med stor träffsäkerhet, och visade att denna sedan kunde lära sig att bli mycket bättre på att predicera hjärtinfarkter än motsvarande modell som inte först fick börja med att träna på ålder och kön.

På akuten är det inte bara viktigt att fatta rätt beslut, utan det spelar också roll hur snabbt man kan fatta rätt beslut. I det tredje projektet undersökte vi olika tidpunkter under ett akutvårdsbesök som ny information normalt sett blir tillgänglig, och byggde maskininlärningsmodeller för att predicera hjärtinfarkt givet den information som fanns just då. Det visade sig bland annat att omkring 15% av alla bröstsmärtepatienter med goda säkerhetsmarginaler kan skickas hem innan man ens hunnit ta några prover över huvud taget. Runt hälften av alla patienter kan skickas hem redan efter ett första blodprov, och därigenom slippa vänta på ett andra blodprov som är det normala förfarandet. Det verkar alltså finnas goda möjligheter till att effektivisera akutvårdsförloppet genom att utesluta de allvarligaste utfallen ännu tidigare.

Den allvarligaste formen av hjärtinfarkt är den som orsakas av en total blockering (ocklusion) av ett kranskärl. Detta är ett akut livshotande tillstånd som kräver omgående invasiv behandling för att mekaniskt lösa upp ocklusionen. I praktiken identifieras denna form av hjärtinfarkt med hjälp av EKG, och många liv har räddats genom att man numera rutinmässigt avläser EKG:et i ambulansen, med möjligheten att åka direkt till hjärtmottagningen. I det fjärde delarbetet har vi retrospektivt identifierat patienter med ocklusionsinfarkter, och hittat att de utgör omkring en tredjedel av alla hjärtinfarkter i Skåne. Endast en bråkdel av dessa fick vård inom rekommenderad tid. Vi byggde en maskininlärningsalgoritm för att identifiera dessa patienter utifrån EKG och labb-värden, och visade att den kunde hitta fler än dubbelt så många fall av ocklusionsinfarkter som om man använde traditionella EKG-mönster.

Sammanfattningsvis har jag i de olika delprojekten som utgör den här avhandlingen tillämpat moderna maskininlärningsalgoritmer för prediktion av hjärtinfarkter på akuten. Mycket återstår att göra, inte minst i form av kliniska valideringsstudier, men förhoppningsvis kan dessa algoritmer till slut komma att användas i form av beslutstödsystem på akuten, med snabbare och mer tillförlitlig vård som resultat.

Abbreviations

ACO	Acute Coronary Occlusion
AMI	Acute Myocardial Infarction
ANN	Artificial Neural Network
API	Application Programming Interface
AUC	Area Under the receiver operating characteristic Curve
CABG	Coronary Artery Bypass Graft surgery
CAG	Coronary Angiography
CNN	Convolutional Neural Network
DL	Deep Learning
ECG	Electrocardiogram
ED	Emergency Department
ESC	European Society of Cardiology
ESC-TROP	Effectiveness and Safety of a Clinical assessment and 0h/1h Troponin Rule-Out Protocol
FN	False Negatives
FP	False Positives
FPR	False Positive Rate
GPU	Graphical Processing Unit
hs-cTn	High-sensitivity cardiac Troponin
hs-cTnI	High-sensitivity cardiac Troponin I
hs-cTnT	High-sensitivity cardiac Troponin T
ICD10	International Classification of Diseases, 10th revision
MACE	Major Adverse Cardiac Event
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
OMI	acute coronary Occlusion Myocardial Infarction
PCI	Percutaneous Coronary Intervention
POC	Point-of-Care
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
ROC	Receiver Operating Characteristic
SEM	Skåne Emergency Medicine
STEMI	ST-Segment Elevation Myocardial Infarction
TN	True Negatives
TP	True Positives
TPR	True Positive Rate

Chapter 1

Introduction

The research presented in this dissertation is inter-disciplinary: we are applying methods from computer science and engineering to solve problems in emergency medicine. In this chapter, I provide a brief overview of the most relevant parts of each respective discipline. The first three sections cover the symptoms, pathophysiology, and management of patients with chest pain at the emergency department (ED). This is followed by a brief introduction to the most relevant machine learning concepts, and an overview of previous research. The chapter concludes by summarizing the aims of this thesis.

Chapter two covers the data and methods specific to this thesis, and chapter three provides a description and commentary of each of the four papers. Chapter four contains a brief summary, followed by an outlook of the future.

1.1 A strangling feeling in the chest

Slut ömt i ditt sköte min smäktande kropp
Förkväv i ditt famntag min smärta
I maskar lös tanken och känslorna opp
I aska mitt brinnande hjärta
— Erik Johan Stagnelius, Till förruttnelsen

Heart and pain, as observed by the romantic poets, rhymes in the Swedish language. Chest pain is among the most common complaints at the emergency department, accounting for approximately 8% of all visits in southern Sweden [1]. The underlying

causes for chest pain range from benign, such as musculoskeletal disorders, to life-threatening, including heart attacks. Chest pain originating from the heart is called *angina pectoris* – a strangling feeling in the chest – and the most common of the serious outcomes is the acute myocardial infarction (heart attack), which is found in about 4%–6% of all chest-pain visits [2, 3]. The primary focus of this thesis is to develop computer models that can help physicians at the emergency department to identify patients undergoing myocardial infarction.

The heart, like all organs, require oxygen to function. This oxygen is delivered via the blood, through coronary arteries that wrap around the heart. If the bloodflow through the coronary arteries is insufficient, for instance as a result of an occlusion (blockage), parts of the heart become *ischemic* – oxygen starved. If the ischemia continues for too long, the affected cells die, which is called *infarction*. An infarction in the heart is called an acute myocardial infarction (AMI), or more colloquially, a heart attack.

There are different possible causes for myocardial infarction, but the most common is the rupture of a plaque inside a coronary artery, formed by the buildup of cholesterol over the course of several decades. The ruptured plaque can lead to the formation of a thrombus (blood clot) which can get stuck and limit or completely block the flow of blood. Other causes of infarction include a mismatch between oxygen supply and demand, for instance due to fever, increased heart-rate or respiratory failure, and injuries related to coronary revascularization procedures [4].

1.2 Myocardial infarction

And then it happened - the attack, the trance,
Or one of my old fits. There sat by chance
A doctor in the front row. At his feet
Patly I fell. My heart had stopped to beat,
It seems, and several moments passed before
It heaved and went on trudging to a more
Conclusive destination. Give me now
Your full attention.
— Vladimir Nabokov, Pale Fire

1.2.1 Diagnosis

In practice, myocardial infarction is diagnosed as a clinical syndrome: it's a collection of symptoms and measurements that indicate that a myocardial infarction has taken

place [5]. There have been several different definitions in the past 75 or so years, starting with the world health organization in the 1950s [6], all the way up to the present. As technology and biomarkers have improved, the definitions have been refined and unified. The latest de-facto version is the fourth universal definition of myocardial infarction, published in 2018 in collaboration with the American heart association, the European society of cardiology, the American college of cardiology, and the world heart federation [4].

The current definition for diagnosis of AMI require the presence of elevated troponin levels (above the 99th percentile upper reference limit), accompanied by a rise or fall in troponin levels, together with either ischemic electrocardiogram changes, imaging evidence of infarction, or symptoms of acute myocardial ischemia [4]. Symptoms of myocardial ischemia include discomfort in the chest, upper extremity, neck or jaw, or the upper central region of the abdomen (epigastrium), as well as "ischemic equivalents" like shortness of breath (dyspnea) or fatigue [4, 7]. The discomfort is often diffuse, non-localized, and not affected by movement of the region.

1.2.2 Biomarkers

Cardiac troponin I and T are two types of proteins that are part of the building block within the myocardial cell that is responsible for contraction, and they are found almost exclusively in the heart [8]. When heart cells die, the troponin levels in the blood become elevated within 2-3 hours, with a peak after 24 hours, and may remain elevated for several weeks [9]. Although other biomarkers for myocardial necrosis exist, cardiac troponin I and T are preferred for the evaluation of myocardial injury [10, 11]. Highly sensitive and specific laboratory assays are able to detect troponin concentrations lower than what is present in 99% of healthy individuals [12]. Analysis of high-sensitivity cardiac troponin (hs-cTn) in the central laboratory typically takes 30–60 minutes; however, high-sensitivity assays are now available, if not yet widely used, for point-of-care testing within minutes in the ED [13].

The concentration of troponins in the bloodstream following myocardial injury is dynamic, and depend on kidney function, blood flow, and the extent and timing of heart damage, among other things [4]. The typical concentrations follows the trend illustrated in Fig. 1.1. A secondary lab-value is required to properly distinguish myocardial infarction from chronic myocardial injury, in which troponin levels are elevated but stable [4].

ESC guidelines recommend an initial hs-cTn measurement upon ED arrival, followed by a second measurement one hour later, unless the initial test didn't detect any troponins at all [7]. Although this protocol is widely implemented, adhering to the strict

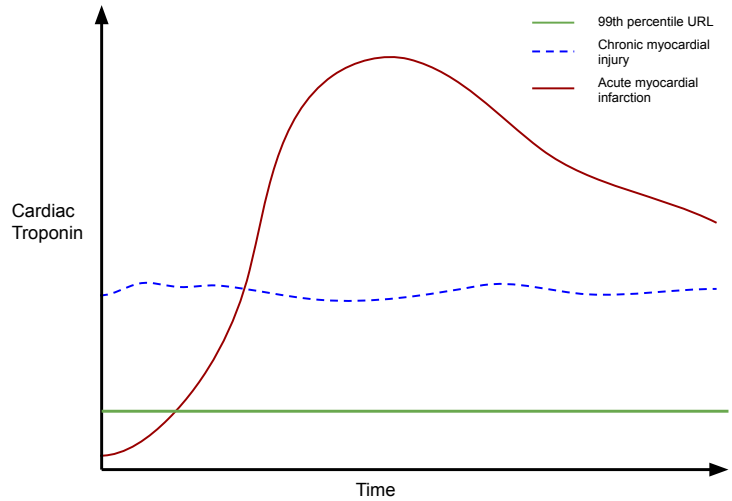

Figure 1.1: Idealization of how the concentration of cardiac troponin in the bloodstream changes over time, following a myocardial infarction. URL = Upper reference limit.

timing can be challenging in a busy ED. In real-world practice, the repeated troponin measurement often occurs within 1–3 hours [14].

1.2.3 The electrocardiogram

The contraction and relaxation of the heart muscle is regulated by electric signals. These signals can be measured as changes in electric potential by placing electrodes on different parts of the body. Such a measurement is called an electrocardiogram (ECG), and is highly informative of the health the cardiovascular system. There are many variations of ECG machines, but the clinical standard has become to place 10 electrodes on specific parts of the chest, arms and legs, as illustrated in Fig. 1.2. The potential difference between the leads can then be combined to form 12 parallel time-series, known as leads, typically plotted on grid paper, as in Fig. 1.3. Normally this 12-lead ECG is measured at a rate of 500 Hz, for ten seconds, while the patient is in a resting position. Each lead can be conceptualized as capturing the activity of the heart from a certain angle, and if there's an infarction, it may or may not be visible depending on what area of the heart is affected [15].

A normal heart beat presents on the ECG as a waveform with a characteristic shape, idealized in Fig. 1.4. The peaks and valleys are traditionally labeled P, Q, R, S, and T. A large part of analyzing the ECG is concerned with the *morphology* of the heart beat: different shapes, patterns, and deviations, which can be expressed[1] as functions

[1]Sometimes only with considerable effort.

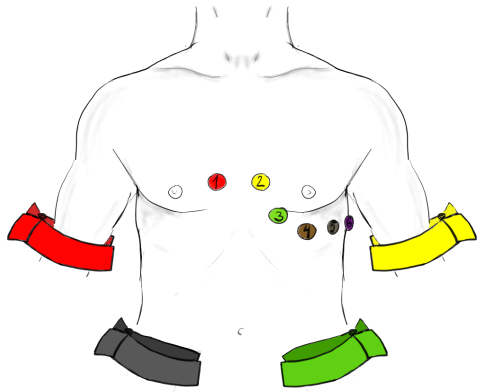

Figure 1.2: Electrode placement of a standard 12-lead ECG. Image by Baklazan99, Wikimedia Commons.

of the PQRST points. Over the years, a wealth of knowledge has been accrued that associate ECG patterns with clinical outcomes and conditions. Identifying the most consequential of these remains an important part of diagnosing AMI [4].

1.2.4 The spectrum of myocardial ischemia

Infarctions come in all sizes, and the underlying ischemia, as we have seen, can be caused by different things. One of the most dangerous types of infarctions are those caused by a complete reduction of blood-flow: acute coronary occlusion (ACO). In these cases, it is recommended that the patient undergo invasive treatment as soon as possible to clear the occlusion [4, 7]. The primary method by which ACO is detected in clinical practice is with the help of the ECG. The defining pattern is the elevation of the ST-segment in at least two adjacent leads,[2] and the resulting diagnosis is known as ST-segment elevation myocardial infarction (STEMI). Early detection systems to prioritize and speed up the management of patients with STEMI has massively improved the outcome of what, for much of human history, was essentially a death sentence. For these reasons, guideline documents, including the fourth universal definition of myocardial infarction, classify infarctions as either STEMI or non-STEMI infarctions (NSTEMI), with correspondingly different treatment options [4, 7].

[2]There are additional and somewhat complicated fine prints and contraindications, but the ST-segment elevation is the main pattern.

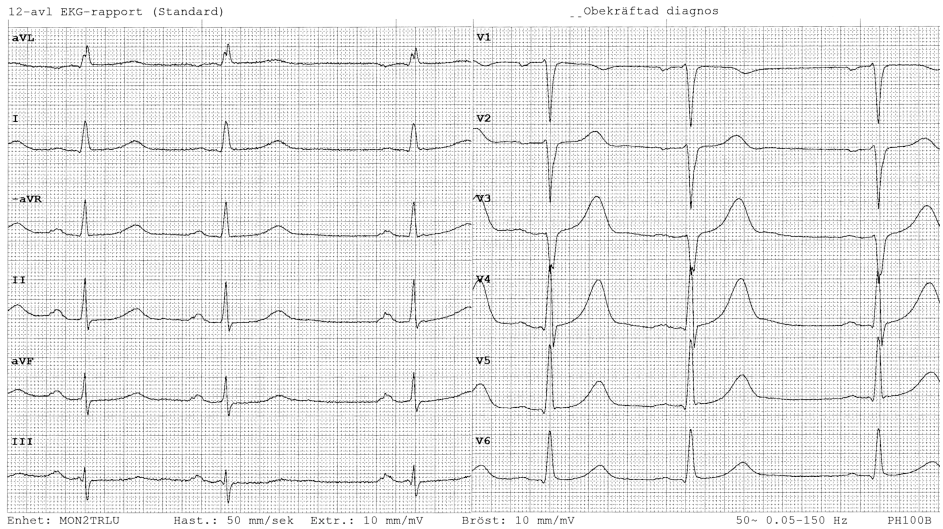

Figure 1.3: Typical 12-lead ECG, showing normal sinus rhythm. Only $5\,\mathrm{s}$ from each lead is plotted in this printout.

In order to speed up the treatment of patients with ACO, many countries, including Sweden, use early detection systems for ambulance patients. The ECG is collected and analyzed while the patient is in transit, and if STEMI is found, the patient is sent directly to the angiography lab, bypassing the ED and saving valuable minutes [16].

The overwhelmingly positive impact and undeniable clinical usefulness of STEMI notwithstanding, the correlation with ACO is not perfect. Studies have shown that about 35% of STEMI cases do not show angiographic evidence of occlusion [17, 18], and conversely, about 40% of all infarctions from ACO show no signs of STEMI on the ECG [19–21]. Although several modifications of the standard STEMI criteria exist, they still do not capture the full picture [22, 23]. The dichotomization of myocardial infarction into STEMI and NSTEMI has thus been challenged, in favor of the more pathophysiologically grounded, if not more easily identified, concept of acute coronary occlusion myocardial infarction (OMI) [23–26]. We will return to this subject when we discuss paper IV in section 3.4.

1.2.5 Treatment

Once a diagnosis of AMI has been determined, there's a number of treatment options, based on the severity of the event. Medications include thrombolytics to dissolve existing blood clots, anticlotting agents to prevent further clots, nitroglycerin to widen coronary vessels, and stimulants to regulate heart rhythm and relieve pain [7]. Un-

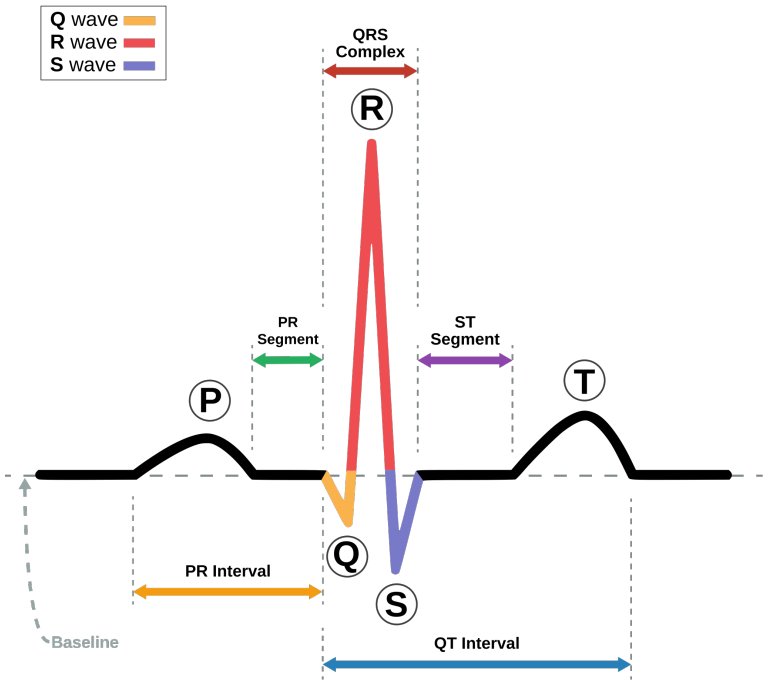

Figure 1.4: Idealized version of ECG heart-beat. Image by Anthony Atkielski, Wikimedia Commons.

stable patients and patients with STEMI are sent to the cardiac care unit as soon as possible, where a coronary angiography can be performed. Patients with NSTEMI are also usually scheduled for angiography within a few days [7]. Coronary angiography is a so-called minimally invasive procedure in which a catheter is used to inject contrast liquid into the heart, which allows a real-time visual inspection of the heart vessels by means of X-rays [27]. The catheter can then also be used to mechanically remove plaques and blockages, as well as insert a small metal mesh tube (stent) to physically keep an artery open. These procedures are referred to as percutaneous coronary intervention (PCI).

Another treatment option which is often preferred in the cases of multiple stenosis is the coronary artery bypass graft surgery (CABG), in which healthy blood vessels from the arm, chest, or leg is used to create an alternate path around a narrowed or blocked artery in the heart [7].

1.3 The emergency department

Eight days a week
Is not enough to show I care
— The Beatles, Eight days a week

The emergency department is special. It's open 24 hours a day, 7 days a week, 365 days per year. Patients arrive with all sorts of ailments, and the goal of the tending physician is not just to treat, but also to admit to the appropriate hospital ward where the patient can get expert care. This is an inherently time-constrained process, in which patients are prioritized at multiple stages based on the urgency of their conditions.

First up is a spot check, in which a nurse makes a judgment call to prioritize obviously severe cases. Then, after orderly awaiting their turn, the patient talks to a triage nurse, who admits the patient to the digital system, collects symptom descriptions, vital signs and sometimes blood samples and the electrocardiogram, and assigns a priority to the patient. This priority essentially dictates the urgency with which a doctor will see the patient: low priority patients may have to wait for hours, depending on how busy the ward is, allowing the highest priority patients to get medical attention as soon as possible.

Figuring out what is wrong can be an invasive process: X-rays, catheterizations and even blood samples can be costly not just in terms of equipment, but also in terms of harm to the patient, and the treatments for many conditions can have serious negative side effects [28].

The immediate goal of the clinician essentially boils down to the decision of *ruling in* (admitting to the hospital) or *ruling out* (discharging). Although AMI is the primary concern when it comes to chest pain (and this thesis), other serious outcomes, some of which are time-sensitive, must also be considered. If a patient is discharged from the ED, we don't want them to return the next day (or week, or month), at least not due to the same condition. A way to conceptualize this concern is to consider, besides a discharge diagnosis of AMI, any subsequent diagnosis exceeding some magnitude, within, say, 30 days of the initial visit. A commonly used composite endpoint for this purpose is that of major adverse cardiac events (MACE). The exact definitions vary, but commonly include unstable angina, AMI, pulmonary edema, invasive cardiac treatment such as PCI and CABG, and death.

Starting at the point when a patient arrives at the ED, there's typically a number of discrete points in time when a physician can make the decision to rule in or out, corresponding to the order in which new information becomes available, as illustrated in Fig. 1.5. Initially, nothing is known except for the medical history in the electronic

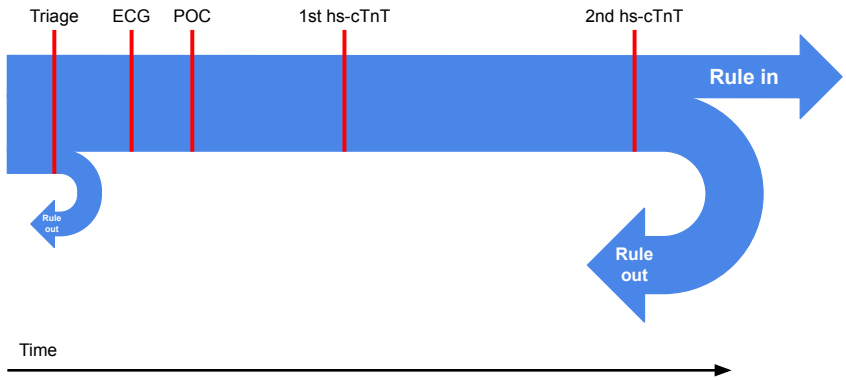

Figure 1.5: Schematic overview of different moments when new data becomes available (red lines), corresponding to logical decision points at the emergency department. POC = Point-of-care samples, hs-cTnT = high-sensitivity cardiac troponin T.

health record and the initial symptoms. Virtually all chest-pain patients will undergo an ECG, since it is both informative, quick, cheap, and non-invasive. The ECG is normally performed within 10 minutes of first medical contact (which may be in the ambulance). In most cases, the ECG is followed by drawing blood for analysis, with some tests analyzed in the emergency department (point-of-care), and others sent to the central laboratory. Point-of-care tests includes glucose, hemoglobin, lactate, and creatinine, and are quick to analyze, with results being available within minutes. The lab samples take about 30 to 60 minutes to process, but are able to measure important biomarkers, including troponin levels, at very low concentrations. If it was not possible to rule out after the initial lab-values have been analyzed, additional samples are collected after one to three hours.

1.3.1 Rule-based algorithms

A number of rule-based algorithms and guidelines exist to assist physicians in the management of chest-pain patients at the emergency department. One example is the HEART score [29] classifies patients as low, intermediate, and high risk based on history (H), ECG (E), age (A), risk factors (R), and troponin (T). Each category contributes between 0 and 2 points, and a total score of 0–3 corresponds to low risk, 4–6 to intermediate risk, and 7–10 corresponds to high risk. The HEART pathway [30] was developed as an extension of the HEART score to incorporate serial troponin testing, and the protocol continues to be updated and improved [31].

Another rule-based algorithm is the ESC 0h/1h rule-out and rule-in algorithm, summarized in Fig. 1.6, which relies on troponin levels measured at arrival and again after

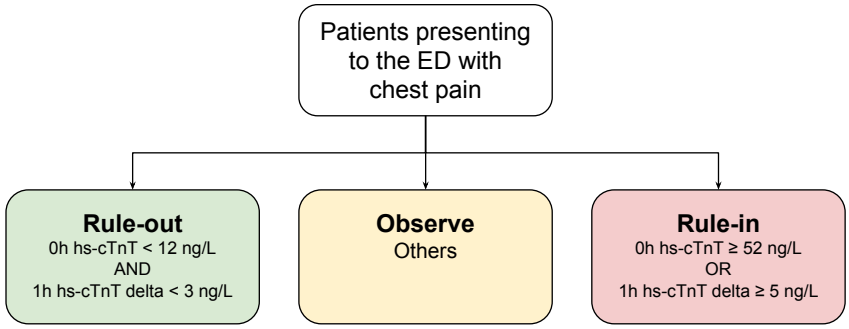

Figure 1.6: The first version of the 0h/1h rule-out and rule-in algorithm, by Reichlin *et al.* ED=Emergency Department, hs-cTnT=high-sensitivity cardiac troponin T.

one hour if the initial troponin was outside the decision boundary [32]. It has been validated and improved upon in several studies, incorporating additional timings for troponin, ECG symptoms and patient history, and has received a "class 1" recommendation by the European Society of Cardiology [7, 14, 33–35].

1.3.2 Decision support tools

Computerized clinical decision support systems have been around since the 1970s [36], and have been shown to play an important role in patient management, improving clinical outcomes, alleviating crowding, and decreasing healthcare costs. Advancements in machine learning, and deep learning in particular, which we will talk about in section 1.4, has the potential to revolutionize health care.

But despite all the promises and recent success stories in other areas of medicine, most impressively in radiology and pathology, studies specific to the emergency department are relatively few. So far, there has been practically no implementations of machine learning (ML) powered decision support systems into routine ED care. This thesis represents my work toward bridging this gap, by bringing us one step closer to the practical application of a new generation of decision support tools at the ED.

1.4 Machine learning

I'm just a spectator
An advocate documenting the loss
— Opeth, Blackwater Park

If artificial intelligence was a taxonomy of living things, machine learning might cover the animal kingdom, and this thesis would be concerned with the entomological study of a particular species of cranefly and their ecological impact as pollinators in organic farming. There are over 15 000 documented species of craneflies, divided into more than 500 genera [37]. The cranefly of this thesis is the artificial neural network, and the farming application is predicting adverse cardiovascular outcomes for chest-pain patients at the emergency department.[3]

A neural network is a type of function that takes an input and produces an output. It works essentially as a graph of interconnected computational nodes, sometimes called neurons, each of which can carry out some simple operation, such as addition or multiplication. A key point is that the neurons can have a state, i.e. a number (also referred to as parameters or weights), attached to them, which adjusts their behavior. The nodes can be grouped into modules, or layers, and the type and arrangement of these layers is referred to as the model *architecture*. Much like pieces of lego, new layers are invented at a rate that is difficult to keep up with, and they can be combined in ways constrained primarily by the imagination.[4]

In order for a neural network function to do anything useful, the parameters, often counting in the thousands, millions or even billions, need to be adjusted. This is done by a process called *training*. In the branch of supervised machine learning, the training relies on examples of correctly classified input-output pairs. The untrained model, in which all the parameters are randomly initialized, is then trained by an optimization algorithm (usually some variation of gradient descent) by observing the input-output pairs one by one, each time slightly adjusting the parameters of the model so that the model output for the input more closely resembles the correct output. The process is typically repeated so that the model "studies" each example multiple times. The method by which the parameters are adjusted is called back-propagation.

A number of problems must be addressed. If the model is complex enough, it is usually capable of "memorizing" all the training data, producing perfect predictions. But when such a model is presented with new, unseen data, the predictions can be

[3] A project hopefully more feasible than using craneflies for pollination, which I just made up.

[4] Also as with lego, the most impressive builds require a lot of money to build, or rather to train, in the case of neural networks.

Table 1.1: Summary of popular regularization techniques.

Technique	Pros	Cons
Add more data	Leads to better performance	Often costly. Slower training.
Reduce model complexity	Faster training	Reduced complexity. Can lead to lower performance, and underfitting
Data augmentation	Often better performance	Slower training.
Penalize large weights	Fast and simple	Can reduce performance.
Dropout (randomly ignore different nodes each update cycle)	Fast and simple	Reduces performance. May require reduced learning rate, leading to slower training.
Early stopping (stop training before overfitting occurs)	Easy in principle	Not always obvious when overfitting begins

arbitrarily bad. This phenomenon is known as overfitting. The converse, underfitting, is also possible, in which the model simplifies too much and fails to capture the complexity of the relationship between the inputs and the outputs. Both over- and underfitting are failures of the trained model to generalize to unseen examples. Several strategies exist to combat overfitting, collectively referred to as regularization. Table 1.1 summarizes some of the more common variations.

Another major problem is how to measure progress. If we use all the data for training, there is no way to know if the model is overfitting or not. It is thus typical to divide the data into two parts, often called training and validation[5] sets. The validation set is then used to regularly evaluate the model during training. But it is often necessary to train a model multiple times, to find an architecture and corresponding settings that work well for the problem at hand. The longer this development process continues, the more likely it is that the best performing model on the validation set will perform worse on unseen data, than what the validation set performance indicated. This is known as data leakage [38]. To address this, the best practice is to divide the original data into three, rather than two, sets: one for training, one for validation, and one for testing. It is then crucial not to evaluate the model on the test set until one is happy with the final results, or else information from the test set, too, may leak into the model and prevent an accurate assessment of the true generalization performance.

Machine learning algorithms, including neural networks, are typically designed to produce a fuzzy output (usually a real number between 0 and 1, which can be viewed as a probability), even when the target outcome is binary. From this perspective, all predictions are wrong, but some are less wrong than others. How wrong a prediction is can be formalized mathematically as a so-called loss function. The goal of the training

[5]Sometimes also referred to as tuning set.

process is for the optimization algorithm to minimize the loss. Different loss functions exist, and depending on the problem, some may be more appropriate than others. For binary classification, cross-entropy loss [39] is the most common choice.

Once a model has been trained, the predictions on the validation or test set can be evaluated. The loss, which is what the model has been trained to minimize, is not always easy to interpret, and might not correspond well to the application. In many situations, we need the model to produce a binary response rather than a probability, and in those cases, we are interested in quantifying the performance of the binary response. This is done in practice by selecting a threshold θ, and considering all predictions below θ to predict 0, or False, and the rest to predict 1, or True. How to select θ depends on the application, and what types of errors are worse.

When the predictions have been binarized, we can calculate the so-called confusion matrix by tallying each of the four possible combinations of predictions and outcomes: True positives (TP), where the prediction is "True" and the outcome is "True", false positives (FP) where the prediction is "True" and the outcome is "False", true negatives (TN) where both the prediction and outcome are "False", and finally the false negatives (FN) where the prediction is "False" and the outcome is "True". From these four numbers, a couple of dozen different metrics can be calculated. Appropriately confusingly, many of these metrics have multiple names: for instance, sensitivity, defined as $TP/(TP + FN)$ is also called recall, true positive rate, probability of detection, hit rate, and power.

Importantly, many of the binary performance metrics are trade-offs, and it is sometimes possible to neglect one in order to be great at another, and vice versa. Several metrics also depend directly on the balance of the dataset, i.e. the proportion of true cases in the outcomes. For instance, if we are predicting a disease with a prevalence of 0.1%, we can trivially achieve a 99.9% accuracy, defined as $(TP+TN)/(TP+FP+TN+FN)$, by predicting all examples as negative. To avoid such problems, it is often necessary to consider multiple metrics at once, such as sensitivity and specificity.

When developing prediction models, it is convenient to be able to rank models based on a single number. It is perfectly possible to be in a situation where one model is better with respect to some metrics at some given threshold, and another model to be better at another threshold. This makes model selection and training very difficult. Ideally, we want to consider a metric that matches our real-world application, and then try to optimize for that. Failing at that, another common approach is to consider metrics where the selection of a threshold is not necessary. One way to do that is to plot the receiver operating characteristic (ROC), which is the curve obtained by calculating the true positive rate (TPR) and false positive rate (FPR) for all possible thresholds, and plot them against each other. The integral of this curve, called the

area under the receiver operating characteristic (AUC),[6] can be used as a threshold-independent metric, and it has some useful properties making it a popular choice among data scientists. The AUC is always a number between 0 and 1, where 0.5 corresponds to predictions made at random, 1 corresponds to perfect predictions, and 0 correspond to perfectly bad predictions. Amusingly, a model with an AUC below 0.5 is "so bad it's good", in that we can do the opposite of what the model tells us to get an AUC score above 0.5 instead. The AUC is independent of the class balance, which makes it easier to compare model performance across datasets.

A thing to note about all these metrics, is that while we might decide that we care about the AUC, or, say, the sensitivity so long as the specificity is sufficiently high, or some other metric, the minimizing of the loss function is not necessarily the same as maximizing the metric of choice. In particular, metrics like the AUC which depend on the entire dataset, are difficult to optimize with a traditional loss function.

Machine learning is a very broad category of algorithms that includes both tree-based models, neural networks, and many other completely different approaches. Deep learning is a subset of machine learning, typically defined as artificial neural networks (ANNs) with multiple layers.[7] The exact threshold at which a traditional ANN becomes "deep" is not well defined. However, a unifying characteristic is in the shift away from the manual crafting of input features, instead letting the model learn its own features automatically from the raw data. This *end-to-end* approach is of course attractive, because creating good features is usually difficult and time-consuming, requiring a deep understanding of the problem domain. Furthermore, the process of crafting features from complex inputs is almost by definition a lossy operation, and useful patterns can easily be overlooked, resulting in less capable models. The main disadvantage of deep learning end-to-end models is that they tend to require extremely large datasets. Techniques such as transfer learning and data augmentation can help to alleviate this "data hunger", but often there is no obvious limit to the benefits that can be realized by simply adding more data and using larger models.

An important caveat to the more-is-merrier dictum is that data *quality* is also essential. It's usually not enough to have a lot of data, if the data is unlabeled, noisy, corrupt or otherwise unstructured or unreliable. It is a common misconception that one can simply collect everything and throw it into some magic machine that figures out the rest. Perhaps that will be possible in the future, but for now, if you put garbage in, you'll get nothing but garbage in return [41–43].

[6] Sometimes also AUROC, to distinguish it from similar metrics, like the area under the precision-recall curve. In characteristically confusing fashion, the AUC is also sometimes referred to as the C-statistic.

[7] Other, more esoteric interpretations exist, such as the one favored by the excellent book "Deep Learning" by Goodfellow *et al.* [40].

1.5 Previous research

Of the nature of Mr. Knott himself Watt remained in particular ignorance. Of the many excellent reasons for this, two seemed to Watt to merit mention: on the one hand the exiguity of the material propounded to his senses, and on the other the decay of these. What little there was to see, to hear, to smell, to taste, to touch, like a man in a stupor he saw it, heard it, smelt it, tasted it, touched it.

— Samuel Beckett, Watt

When it comes to emergency care and chest pain, AMI is perhaps the most obvious outcome to consider, and of all the things that we can use to predict it, the electrocardiogram shines like a grail-shaped beacon in the night. There are several reasons to focus on the ECG: it is quick, non-invasive, cheap, available almost everywhere, and it provides a detailed, if sometimes difficult to interpret, view of the cardiovascular health of the patient.

One of the "top tips" in the book "ECGs made easy" by Hampton *et al.* [15] is to not panic, because "the ECG really is very easy!". The book, suspiciously, is several hundred pages long. The ECG clearly contains a lot of information, but extracting that information and turning it into clinical insight is a non-trivial task.

For a long time, the main challenge of automatic ECG analysis has revolved around feature-engineering [44]. A standard 12-lead ECG, sampled for 10 s at 500 Hz, is essentially a large table with 12 columns and 5000 rows, containing 60 k numbers.[8] Building a model that can (successfully) operate directly on such a large input has only recently become feasible with the advent of deep-learning algorithms, and even so, require datasets with tens of thousands of ECGs together with modern computer hardware that are several orders of magnitude faster than what was available even 20 years ago.

For these reasons, much of the history of computerized ECG analysis, which date back to at least the 1960's [45], rely on ways to extract and condense the most important parts of the ECG into a more manageable number of features [44]. Once extracted, the features can then be used to train machine learning models such as logistic regression, random forests, or neural networks.

There is no shortage of imagination in the various approaches to ECG feature engineering. Morphological features, including amplitudes and durations of different parts of the heart beat, are natural targets for extraction, since they correspond well to

[8] Though the standard ECG has 12 leads, 4 of them are linear combinations of the others, which means they do not contain additional information. Thus it is common in ML applications to only use the 8 linearly independent leads, to reduce the dimensionality of the input.

patterns recognized by humans. Other ideas include principal component analysis, hidden markov chains, transforms such as the discrete wavelet transform, the fourier wavelet transform, and the Karhunen-Loéve transform [44].

Many of these approaches to ECG feature engineering are sensitive to noise, such as baseline wander, power line interference, noise due to muscle activity, and radio frequency noise from other equipment. Additional pre-processing steps are thus often required, such as band-pass filters, discrete wavelet transforms, and beat averaging. These techniques can help to get rid of noise, but they also risk destroying important information.

Then, in the 2010s, major breakthroughs in computer vision ushered in a new age of machine learning, promising unparalleled performance with less engineering requirements: no more difficult and time-consuming hand-crafting of features that rely on hard-won human expert knowledge [46, 47]. Let the algorithm engineer its own features! The only problem is the data. And computational power. But mostly data.

A limiting factor for the application of machine learning, and particularly deep learning algorithms to ECG analysis, is the availability of large, public ECG datasets. For a long time, the only available databases were quite small, with ECGs numbering in the low hundreds or fewer. It was not until the Physionet Computing in Cardiology challenge of 2017 that larger ECG databases started to become publically available [48]. A summary of datasets can be found in Table 1.2.

Since 2016, the number of papers describing deep-learning approaches for ECG analysis has increased rapidly [47, 69, 70]. Although the majority of the publications focus on atrial fibrillation or arrhythmia detection, with a notable early example in the ResNet implementation by Rajpurkar *et al.* in 2017 [71], all sorts of other applications have been explored as well, including pulse detection, pediatric heart murmurs, mortality prediction, sleep apnea detection, P-wave detection, left bundle branch block detection, hypoglycemic event detection, data compression, emotion detection, biometric applications, and many others [47].

Applications of deep-learning to predict AMI using ECGs are less common, and many of the earlier examples are based on beat-segmented ECGs, use only morphological features, or are trained on small datasets [44]. In 2020, Ribeiro *et al.* [72] published a seminal paper describing one of the first successful applications of a deep-learning (ResNet) model to predicting AMI, using a dataset of over 2 million ECGs. Part of the dataset was subsequently released publicly, as the CODE 15% database [67].

In 2020, Wagner *et al.* [65] published the PTB-XL database, with a companion paper [73] exploring a variety of deep-learning methods applied to the dataset, including one for predicting AMI, among other outcomes. In 2022, Gustafsson *et al.* [74]

Table 1.2: Summary of public ECG datasets.

Dataset	Release year	Channels	Sample rate	Duration	Records
MIT-BIH Noise stress test database [49]	1984	2	360	30min	15
UCI Machine Learning Repository Arrhythmia Dataset [50]	1998	12	N/A	N/A	452
MIT-BIH normal sinus rhythm [51]	1999	2	N/A	24h	18
MIT-BIH malignant ventricular ectopy database [52]	1999	2	250	30min	22
The QT dataset [53]	1999	2	250	15min	105
MIT-BIH Atrial fibrillation [54]	2000	2	250	10h	23
BIDMC Congestive heart failure [55]	2000	2	250	20h	15
Fantasia [56]	2003	12	250	2h	40
PTB diagnostic ECG database [57]	2004	15	1000	2 min	549
MIT-BIH arrhythmia database [58]	2005	2	360	30min	48
Creighton university ventricular tachyarrhythmia database [59]	2007	1	250	8min	35
St Petersburg INCART 12-lead arrhythmia [60]	2008	12	257	30 min	75
European ST-T [61]	2009	2	250	2h	90
Chinese CVD database [62]	2012	12	500	10s	90
Physionet/CinC [48] 2017	2017	1	300	9-61s	12186
ICBEB2018 [63]	2018	12	500	6-60s	9831
Icentia11k [64]	2019	1	250	3-14 days	11000
PTB-XL ECG dataset [65]	2020	12	500	10s	21837
Chapman university and Shaoxing people's hospital dataset [66]	2020	12	500	10s	10646
CODE-15% [67]	2021	12	400	7-10s	345779
Large-scale multi-label ECG database [68]	2022	12	500	10s	25770

published a ResNet model trained on 500 k ECGs from unselected ED patients, to predict AMI. The model was able to predict STEMI and NSTEMI with an AUC of 98.5% and 83.2%, respectively.

In 2023, Al-Zaiti *et al.* [75] published an ML model for predicting OMI using the ECG. This was quickly followed by the publication of the Queen-of-Hearts model by Herman *et al.* [76] in 2024. Both models achieved excellent results, outperforming the STEMI criteria. We return to this topic in section 3.4.

Several studies have explored machine learning models predicting AMI using one or several high-sensitivity cardiac troponin (hs-cTn) samples, with and without the addition of other features. In 2019, Than *et al.* [77] used gradient boosting to construct the myocardial-ischaemic-injury-index (MI3) algorithm, predicting index visit AMI using paired hs-cTnI measurements combined with age and sex. In contrast to rule-based protocols like the HEART pathway or the ESC 0/1h protocol, the MI3 algorithm allows for different and flexible timings between the troponin measurements, which is useful in the ED setting. The MI3 algorithm has been externally validated in several studies and found to perform well [78, 79].

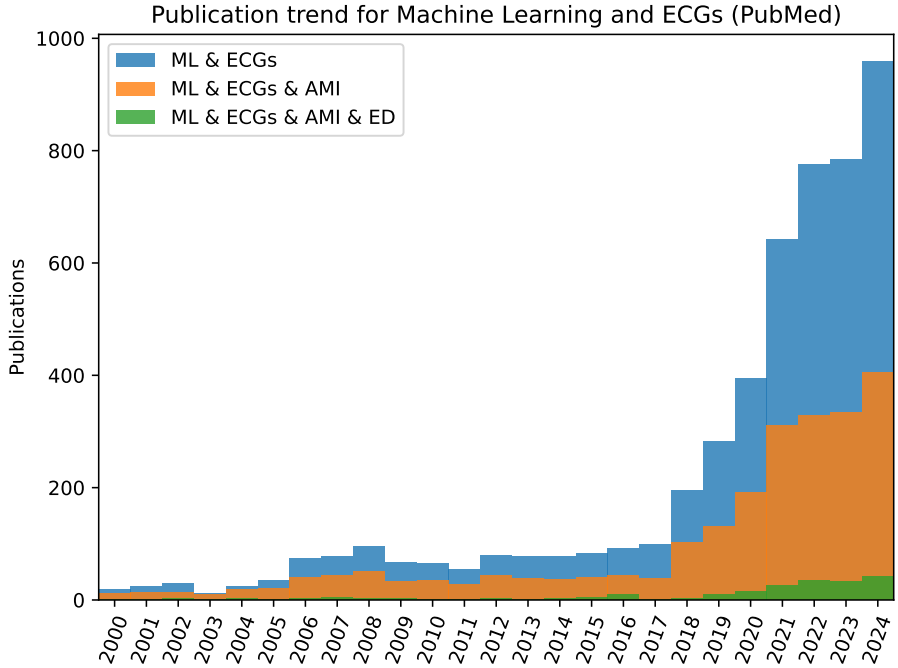

Figure 1.7: Number of publications on PubMed on the topic of machine learning (or deep learning, or artificial intelligence) and ECGs, between the years 2000 and 2024. The blue bars indicate ML and ECGs, orange further specifies AMI (or MACE), and the green bars are papers specific to the ED setting. ML = machine learning, ECGs = Electrocardiograms, AMI = Acute myocardial infarction, ED = Emergency department, MACE = Major adverse cardiac events.

In 2023, Doudesis *et al.* [80] developed a similar model, called CoDE-ACS, which classified fewer patients as intermediate compared to the ESC 0/1h protocol. Also in 2023, Neumann *et al.* [81] developed and validated the ARTEMIS model, capable of using single or serial hs-cTn samples from multiple different assays. The model was developed with heterogeneous global data.

In 2020, Zhang *et al.* [82] published the first machine learning model for predicting AMI that was actually integrated as a decision support tool in real-time clinical practice. It used a collection of features derived from medical history, together with the initial hs-cTnI value. ECG features were not included.

A thoroughly unthorough publication search on PubMed reveals a striking trend, visualized in Fig. 1.7. It would appear that more has been published on machine learning applications for ECGs during the first four years of my PhD, than was published *in total* when I first started.[9]

[9]As much as I would like to take credit for this, I have (so far) only contributed a single published paper to this veritable mountain of research.

Despite considerable interest in the development of deep-learning based prediction models for AMI and similar outcomes, very few such models have been prospectively evaluated in clinical trials, and fewer still are available as actual support systems for use in clinical practice. This speaks both of the difficulty of bringing such systems to their usable conclusion, but also of the opportunities that remain unfulfilled.

1.6 Aim of this thesis

The overall aim of this thesis was to further the development of decision support tools at the emergency department, by exploring ways to utilize digital health-care records for building deep-learning prediction models. The primary focus has been on predicting short-term adverse cardiac events among patients presenting with chest pain at the emergency department.

Paper I

In paper I, we aimed to quantify the benefit of prior electrocardiograms when predicting major adverse cardiac events using machine learning.

Paper II

In paper II, the aim was to quantify the improvements from transfer learning when predicting acute myocardial infarctions with electrocardiograms.

Paper III

In paper III, the aim was to develop a model for early rule-out of acute myocardial infarction, at multiple decision points, with step-wise increasing information.

Paper IV

In paper IV, we aimed to retrospectively identify and subsequently build a deep-learning model to predict the occurrence of acute coronary occlusion myocardial infarction among patients with chest pain at the emergency department.

Chapter 2

Data and Methods

2.1 Data sources

My work is based primarily on two large datasets containing health records from chest-pain patients at the ED: ESC-TROP (Effectiveness and Safety of a Clinical assessment and 0h/1h Troponin Rule-Out Protocol)[14, 35] and SEM (Skåne Emergency Medicine)[1]. ESC-TROP was the first to become available, and although the bulk of the data from SEM was delivered in 2021, it wasn't until the fall of 2023 that the ECGs were finally ready.

2.1.1 ESC-TROP

ESC-TROP contains information from 26 545 consecutive patients with non-traumatic chest pain at five different EDs in Skåne, Sweden, between February 2017 and November 2018. The exclusion criteria of the original study were previous enrollment (i.e. only the first visit in the time-period was included), non-Swedish residence, age under 18, discharge against medical advice, no hs-cTnT ordered, and STEMI diagnosis at the ED. Importantly however, the data I was ultimately given access to still included patients with STEMI diagnosis at the ED.

The dataset contains information from the relevant ED visit, including lab-values, diagnoses, and interventions. Diagnoses and interventions were included for a five year period prior to the visit, and one year after. All ECGs from any health-care visit were included as far back as possible, with the earliest being recorded in 1970, for a total of 536 135 ECGs. Additional information included cause of death for a follow-up period of 1 year, medications, socioeconomic data, and information about health-

care costs. Data sources included excerpts from the regional hospital records (Melior), the Swedish national patient register, the Swedish prescribed drug register, the Swedish Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA), SECTRA, and the SWEDEHEART registers, including RIKSHIA, SEPHIA, and SCAAR.

All the ECGs were pre-processed to a standardized format, and the Uni-G ECG analysis program [83] was applied to extract diagnoses, rhythm statements, and morphological features.

2.1.2 SEM

The SEM database includes everything that's in the ESC-TROP database (with some minor exceptions), and more. It includes all ED visits, including re-visits, for all causes, from all eight EDs in Skåne, for the full two-year period of 2017 to 2018. In total, SEM contains information from 325 539 patients and 630 275 ED visits, of which 51 351 visits were due to chest pain. A total of 1 223 642 ECGs are included. The only meaningful disadvantage to SEM compared to ESC-TROP is that the ECGs in SEM only cover a five year history, and were not initially processed by the Uni-G program.

2.2 Ethics

The large print giveth, and the small print taketh away
— Tom Waits, Step right up

The ESC-TROP study was approved by the Regional Ethics Review Board in Lund, Sweden (Dnr 2017/831 and 2018/708). The creation of the SEM cohort and its use for ML research has been approved by the Swedish Ethical Review Authority (Dnr 2019–05783) and Region Skåne (KVB 302-19). Both studies received ethical approval without the need for written informed consent. Written information was available online and posted visibly at the emergency departments. All patients had the option to decline participation at any time, for any reason.

The datasets provided to Lund University from Region Skåne were pseudonymized, removing all names, addresses, dates of birth, personal identification numbers and other sensitive information of no clinical value. The data is still considered highly sensitive, and is stored on encrypted servers within Lund University, without open access from the Internet.

2.3 Data processing

At every level, vital instructions were missing, and the instructions about what to do in the event of discovering that vital instructions were missing, were also missing.
— Douglas Adams, The hitch-hiker's guide to the galaxy

It is worth noting that the data received, both for ESC-TROP and SEM, was unstructured, in the sense that explicit relations between visits, diagnoses and such were no longer present. Part of the difficulties in processing the data was therefore simply to match events between the data sources using timestamps and other information. The pre-processing work also revealed several problems and inconsistencies with the data, ranging from duplicate entries, "impossible" events (such as when patient is supposedly admitted within minutes to hospitals in different cities), incomplete rows (missing timestamps, diagnoses, etc), duplicated entries where part of the information is in one row and some other part of the information is on a different row, and inconsistent diagnoses.

The diagnoses in particular presented us with a dilemma. There were three primary sources of information: the regional database system Melior, the national patient register, and the RIKSHIA register from SWEDEHEART (only for heart-related diagnoses). Counting the number of patients with an AMI diagnosis (ICD-10 code starting with I21) within 30 days of the index visit, resulted in the venn diagram in Fig. 2.1, suggesting a concerning inconsistency between the registers for this very important diagnosis. Upon closer examination, we made the judgment call to "trust" the national register sources, and simply ignore the diagnoses from Melior. Of the 161 AMI cases only recorded by Melior, about 30 turned out to be prior diagnoses, another 30 were from outpatient clinics, and around 100 were I21-diagnoses that were revised after ED discharge, sometimes to R074 (unspecified chest pain), sometimes to other I-diagnoses (diseases of the circulatory system). These patients were typically old, with multiple diseases.

2.4 Machine learning methods

You're not even close to baseline
— Blade Runner 2049

The algorithmic foundation of this thesis has been the artificial neural network (ANN), which comes in many different shapes and forms. The ANN can be conceptualized as a computational directed graph (i.e. a *network*), where the nodes of the graph

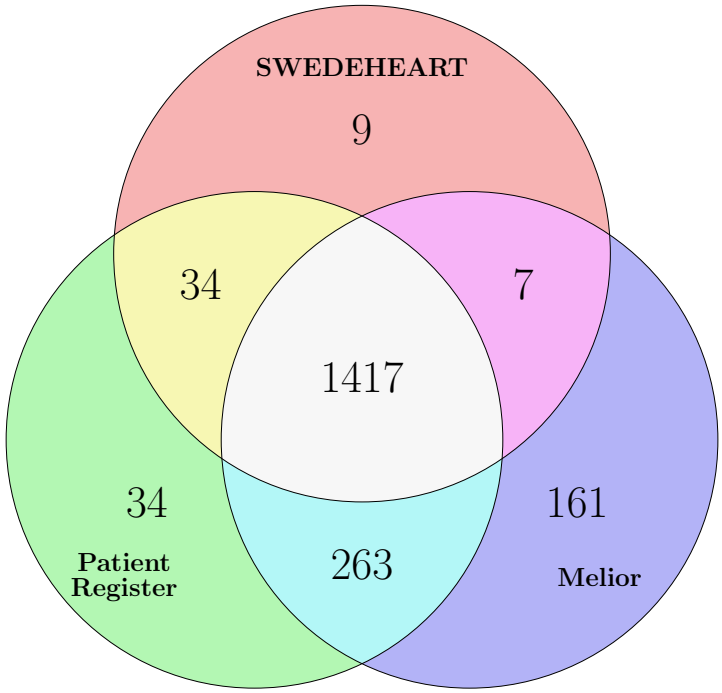

Figure 2.1: Venn-diagram of AMI diagnoses (I21) according to different sources. Melior = the regional electronic health record database, Patient register = the Swedish national inpatient register, SWEDEHEART = National quality register for heart disease.

correspond to calculations and manipulations of data, and the edges connecting the nodes determine the order of the calculations. We refer to the specification of this computational graph as the architecture of the ANN. Although the computational nodes, or layers, in the graph could in principle be almost anything, most realizations of ANNs are built from a pool of standard layers. In the following subsections, I will describe some of the building blocks (layers) used in my work, as well as the different structures (architectures) used. For a more in-depth introduction to deep learning, I warmly recommend the book aptly titled Deep Learning, by Goodfellow *et al.* [40].

2.4.1 Building blocks

Here is a brief and non-exhaustive overview of the most important layers that I have been using.

Dense layer

One of the simpler and more straightforward building blocks is the *dense* layer, sometimes called *fully-connected* layer. This is essentially just a multiplication of the input tensor by some weights, optionally followed by the addition of a constant. A feature of the dense-layer is that it can reduce the dimension of the input.

The rectified linear unit

The rectified linear unit (ReLU) is a popular choice to introduce non-linearities into the computational network [84]. Despite its complicated sounding name, it simply lets all positive inputs through without change, and sets any negative numbers to zero.

The convolutional layer

The convolutional layer applies a mathematical convolution operation to the input. The parameters of the convolutional layer correspond to a kernel, that essentially acts as a filter on the input. The convolutional layer is a popular way to reduce the number of parameters in a network that works on images or time-series data. Rather than connecting every data point in the input with some weight, as is done in the dense layer, the filter is applied in a sweeping fashion across the entire input. The resulting output is the same size as before (it can also be reduced or increased by adjusting the step-size of the sweeping), but the number of necessary parameters is limited to the size of the filter. In contrast, the dense layer would require at least as many parameters as there are inputs, which is costly when the input is a large image, or, say, an ECG. Typically, a convolutional layer applies not just one, but multiple filters at once. How many, and the size of the filters are part of the hyper-parameters of that layer [85].

Pooling Layers

Pooling is way to summarize and down-sample information across multiple dimensions, commonly used in conjunction with the convolutional layer. There are several variations, including max-pooling and average-pooling, and one of the use-cases is to reduce the dimensionality to obtain a more compact feature representation, which increases the receptive field of subsequent layers.

Flatten

Flatten is a simple operation that just reduces a high-dimensional tensor into a single vector. Sort of like turning all the words in a book into a single (very long) line. It is typically used at the end of a convolutional network.

The dropout layer

The dropout layer, as discussed briefly in section 1.4, is a simple but effective way to reduce overfitting [86]. It picks a random subset of its inputs and turns them to zero, and lets everything else pass unaffected. The set of inputs that are thus dropped out is different for each training sample. The dropout is only applied during training, and not during inference. The intuition is that by removing parts of the data at random, the network is forced to become more robust, and not rely entirely on an individual feature.

Batch Normalization

One of the problems that can occur when building deep neural networks is that the loss-gradient becomes smaller and smaller, leading to numerical instability and poor performance. This is known as the vanishing gradient problem. A way to counter this is to scale and shift the layer inputs so they are approximately zero mean and unit variance (i.e. they are normalized). This is what the batch normalization does, by keeping track of batches of input data and gradually adjusting the normalization parameters during training [87]. The layer is commonly used in deeper networks.

2.4.2 Architectures

Logistic regression

Arguably the simplest type of ANN is that consisting of just a single dense layer that compresses the input to a single number. If we plug this number into the logistic function, we obtain a neural network that is equivalent to logistic regression. The structure is visualized in Fig. 2.2.

Feed-forward neural network

A feed-forward neural network, sometimes called a multilayer perceptron, is any ANN in which the computational graph is just a straight line: the data passes through each layer one at a time, sequentially and without any loops. See Fig. 2.2 for an example.

Convolutional neural networks

A convolutional neural network (CNN) is an ANN that makes use of convolutional layers. Typically, convolutional layers are staggered in such a way as to gradually reduce the dimensionality of the input, with the idea of capturing features at different levels of detail. The convolutional layers are usually paired with pooling layers, non-linearities like ReLU, and sometimes regularization and normalization layers as well. It is common to end a sequence of convolutional layers with a flattening layer, and follow with dense layers. Fig. 2.2 shows an example.

Residual neural networks

The residual neural network (ResNet), introduced by He *et al.* in 2016 [88], is a type of CNN that is characterized by the use of shortcut connections in the computational graph. Network depth were shown to be of critical importance for complex tasks [89, 90], but attempts to build very deep networks proved to be difficult [91, 92]. The shortcut connections solved a problem with degrading training accuracy, which allowed for substantially increased network depths. The idea has sparked many variations and modifications [93, 94], and enjoyed widespread popularity in many applications, including for ECGs [71, 72, 74, 76]. One variation is visualized in Fig. 2.2.

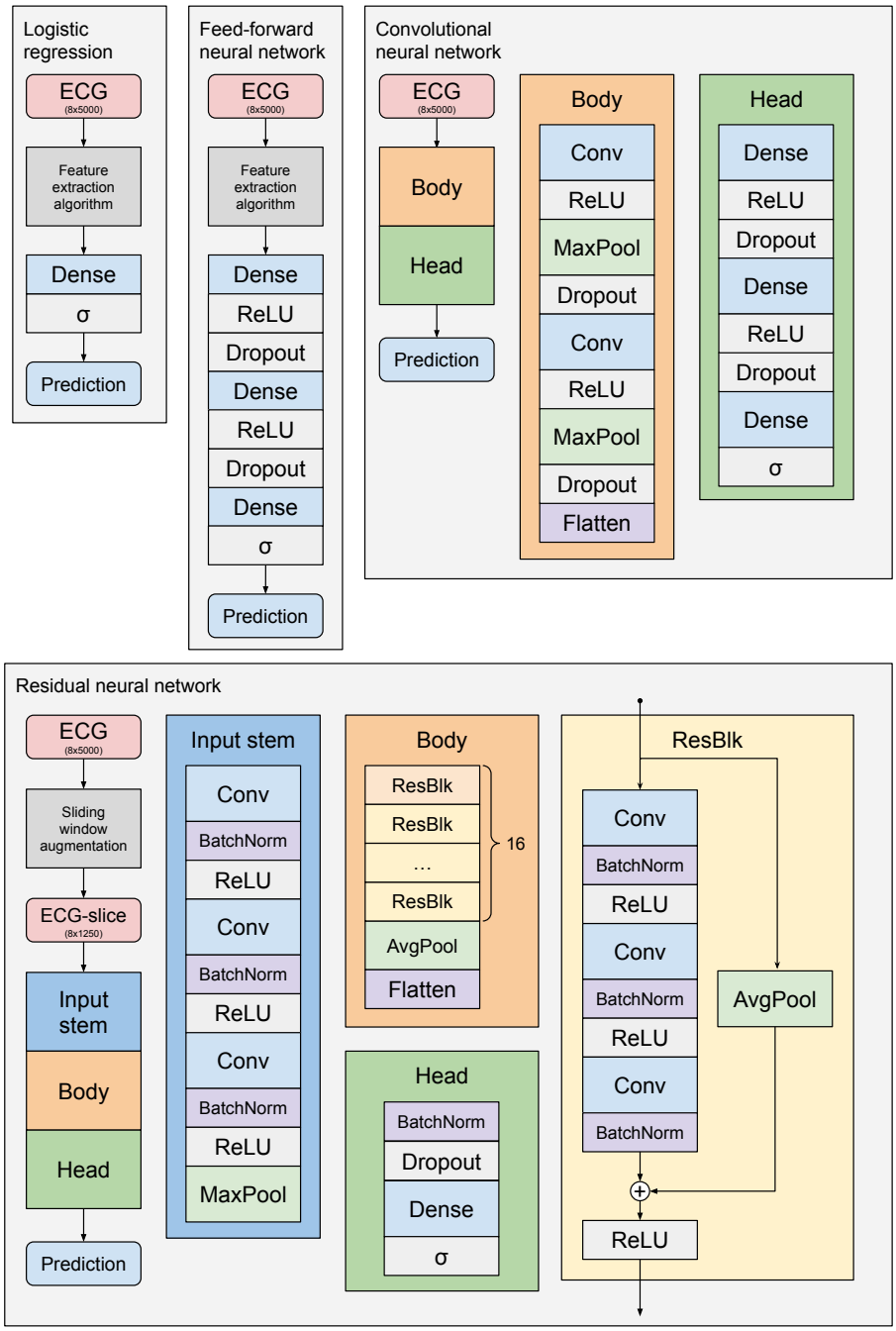

Figure 2.2: Examples of neural network model architectures for ECGs. Top left: logistic regression realized as a one-layer neural network. Top middle: a feed-forward network with two hidden layers. Top right: a CNN with two convolutional layers. Bottom: a ResNeXt-50 architecture. σ corresponds to the logistic function.

2.4.3 Hyper-parameter optimization

The implementation of most machine learning algorithms require certain choices to be made with regards to settings and parameters of the model. These *hyper-parameters* affect (sometimes profoundly) the workings of the model, and are not learned automatically during the training process. Instead, they have to be adjusted to match the specifics of the application. Hyper-parameters that work well for one task on one data set are unlikely to be optimal for any other.

Finding good hyper-parameters can be quite difficult and time-consuming. Many strategies exist. The perhaps most obvious one, called grid-search, is to exhaustively try all combinations in a pre-specified space of settings (the grid), and pick the one that performed best. This can work well when the parameter space is small, but quickly becomes prohibitively expensive as the number of hyper-parameters increase. A popular and simple alternative that often works well in practice is called random search: pick hyper-parameters at random, train the model, and repeat. This slightly moronic-sounding approach has been shown to find good settings faster than grid-search [95], and because of the ease of implementation, it is widely used. Modifications exist to stop training early, or to automatically adjust the search space after each iteration.

It is not always well-defined what counts as a hyper-parameter. It is possible to construct models and networks where the number of layers, for instance, or even choice of ANN architecture, becomes a hyper-parameter. I used this approach in paper IV, for instance, when determining which type of ResNet structure to use.

2.5 Reproducibility

Applying machine learning methods to real-world data comes with some challenges. Many tests and experiments are typically required, in which different models are trained and compared across a whole range of settings, which can be difficult to specify in advance. An inordinate amount of time is usually spent trying to understand, and fix, the data, wrangling it to conform with some expected model format, slicing it this way and that. The iterative nature of the process makes it difficult to keep track of results, especially if you have to go backwards (the stuff you did two weeks ago worked better than this, but now you can't remember how you did it exactly). Cleaning and manipulating the data – a process I have come to call *massaging* – is difficult and error-prone, with no obvious way to know when it's done, or that what you did was correct. Sometimes you might find out weeks, months or even years later that some subtle (and, we pray, insignificant) mistake was made in the processing of some part of the dataset, which might affect the results of some past project. But now

the code-base has evolved to the point where it's difficult or impossible to know how the data was processed exactly, if it is even affected by the newly discovered mistake, and re-running anything would require a lot of effort. If I can't reproduce what I did in the past, can I still claim to be a data *scientist*?

These are some of the questions, observations and fears that motivated me to build my own framework, or laboratory, as it were, for conducting data science experiments. I call it *mim*, pretentiously after the Norse god of wisdom and knowledge.[1] The source code for mim, including all the projects described in this thesis, warts and all, can be found at `https://www.github.com/Tipulidae/mim`.

The guiding principle for the mim project was reproducibility and traceability. The main idea is that rather than trying to keep track of tests or experiments by manual documentation, I use the code itself, together with version control (git), to organize experiments. An experiment in this context is a piece of executable code, specifying exactly the data to use, together with how to process it, and what model and parameters and other settings to use. Executing the experiment automatically saves all the potentially relevant metadata, along with the results of the model, history of training, and other things, to disk. The results of a particular experiment can then be loaded at any point, with a reference to the exact conditions under which it was executed, even if the code-base itself is subsequently changed. Moreover, re-running an experiment after correcting a mistake in the data-processing, for instance, is ideally just a key-press away. Each experiment thus designed is therefore reproducible, and we can trace any bugs in the code back to any previously run experiment to figure out what has or has not been affected. In theory.

The data, especially when it's sensitive, requires additional attention. On the one hand, it's important to be able to reuse and continuously develop and improve the data processing and massaging. But how to keep track of changes to the data? Traditional version control systems don't work well with large datasets, and saving the data for all intermediate steps presents further challenges in organization, as well as additional storage space, and concerns regarding sensitive data. There's also a risk that we fix mistakes in the processing code, but forget to actually update the relevant data and its dependencies. The resulting mismatch between the data and the code that produced the data can be incredibly difficult to identify and resolve.[2] The alternative is to keep only the "raw" form of the data, and track only the code that is used to process it. This works well, as long as we remember to commit all code changes before running an experiment, and as long as the data processing isn't computationally expensive. In mim, I have mostly relied on the latter approach, with only the most expensive pre-processing performed "offline".

[1] Plausibly etymologically linked to the word "memory".

[2] Ask me how I know.

I firmly believe that something like mim is necessary to organize and maintain some sort of structure when exploring data-science questions. A different question is whether my solution was particularly good, and what I could have done differently. In retrospect, I think me "doing it myself" may have resulted in a certain reluctance toward adapting novel or experimental methods, especially when they would require the adaptation of new libraries, such as pytorch [96], to work with an API that was designed mostly with scikit-learn [97] and tensorflow [98] in mind. On the other hand, once things were working, I felt more confident in a lot of the reproducibility, as well as correctness of many of the intermediate steps, than I probably would have if I for instance had used someone else's code as a starting point.

I should have probably tried harder to decouple my own interface, especially the one for constructing data sets, from other libraries (tensorflow). It's a tough line to draw, which has to be balanced against "reinventing the wheel". In contrast, I never regretted my reliance on the scikit-learn library for dealing with anything related to data splits, cross-validation, pre-processing and "old school" ML methods.

Because each experiment is a separate unit of code, and because you naturally end up testing *a lot* of things throughout the lifetime of a research project, some of the experiment listings became somewhat unwieldy, even when using fancy code-folding technology. The experiments for paper I, for instance, is around 5k lines of code, and paper II is over 11k lines. This unwieldy listing of experiments prompted some possibly misguided attempts at refactoring, and also lead to less-than-rigorous re-running of experiments, with the almost instant regret of being unable to access previous versions. A more sophisticated way of versioning old experiments, or otherwise keeping track of possibly useful experimentation without creating an unholy mess of a code listing, would be a necessary addition to mim, moving forward.

At the end of the day, it is interesting to consider the current status of computational science, with respect to the tools and technologies being used. Other disciplines take pride in their highly advanced laboratory equipment, and employ entire teams of engineers to help them build it. We may not require rockets or particle accelerators to do our job, but many of the daily challenges of data science could likely be alleviated by better, professionally developed software. Just like how the cookbook of Marie Curie is still radioactive a hundred years later, I think a lot of my own work, to not mention others, bear the hallmarks of being developed in a kitchen, and not a lab. That's not to say the work is bad – Curie won the Nobel prize not once but twice – but it is, perhaps, similarly nascent.

Chapter 3

Results and discussion

In this chapter, I discuss some of the background and motivation, challenges, results, and limitations for each of the four papers. My hope is to provide a bit of nuance and behind-the-scenes commentary while also summarizing the most important results and discoveries.

In general, it is fair to say that the first two papers are mostly concerned with generic ways to improve predictions of AMI at the emergency department by introducing additional data: prior ECGs in paper I and pre-training on unrelated ECGs in paper II. In contrast, the last two papers focus on more specific applications. The goal of paper III was to systematically investigate early points of rule-out of AMI, while paper IV identifies and predicts OMI.

3.1 Paper I – The case of prior electrocardiograms

Title	Prior electrocardiograms not useful for machine learning predictions of major adverse cardiac events in emergency department chest-pain patients
Aim	To quantify the benefit of prior electrocardiograms when predicting MACE using machine learning.
Data	ESC-TROP (19 499 ED chest-pain patients)
Results	No added benefit from prior ECGs was found. This was consistent across several models of varying complexity, both with and without additional clinical input variables, for different subgroups of patients, and for different subsets of the outcome.

3.1.1 Motivation

In paper I we explored options for utilizing prior ECGs to predict major adverse cardiac events (MACE) within 30 days of ED chest-pain visits.

For a layperson, the idea that previous ECGs should be useful might seem somewhat far-fetched. But it makes sense both from a practical and pathophysiological perspective: infarctions are known to sometimes permanently alter the appearance of the ECG, so when looking at an ECG in isolation, it is not necessarily obvious if an indication of infarction is new or if it originates from a prior event. International guidelines recommend the consultation of prior ECGs [99], based primarily on results from a study by Lee *et al.* in 1990 [100], which associated the availability of prior ECGs with an increased specificity for triage of AMI. As a result, it is not unusual for an emergency department physician to consult prior ECGs.

3.1.2 Previous research

There are some recent studies analyzing prior ECGs in a machine-learning setting, though they rely on "old-school" ML techniques and primarily hand-crafted morphological features [101–103]. Probably the most similar previous study was that of Ohlsson *et al.* [104] from 2001, in which a neural network was trained on morphological features from 4691 ECGs to predict AMI. The authors found that the addition of

a prior ECG lead to an improvement in AUC from 0.85 to 0.88. As far as I'm aware, ours is the first study to evaluate the usefulness of prior ECGs for deep-learning models working on the full ECG signal.

3.1.3 Challenges

Defining MACE turned out to be a challenge in and of itself. Apparently there is no real consensus on what a major adverse cardiovascular event really entails, with different authors using different sets of diagnoses and intervention codes. Even internally, within our research group, there was no clear answer: at one point I had four different lists of diagnose codes to choose from, from different people. These problems were ultimately resolved by my supervisors, but in hindsight, I think it might have been better to simplify and focus on AMI, which makes up the vast majority of all MACEs anyway.

There were two primary technical challenges in this paper: How do we process the ECGs, and how do we convince ourselves (and the reader) that any improvement (or lack thereof) can be attributed to the prior ECG? The simplest solution to the first question seemed to be to begin by extracting features from each ECG, and then use those features in a down-stream model. We opted for two basic types of feature extraction: a set of hand-picked morphological features mirroring those of Forberg *et al.* [105], optionally followed by a dimensionality reduction, and automatic feature extraction based on convolutional neural networks operating on the full ECG signal.

To address the second challenge, we decided to build (and tune) two versions of each model: one where features from only the index ECG was used, and one where the features from a prior ECG was included. This model structure is illustrated in Fig. 3.1. We chose four different models of varying complexity, and repeated the process both for ECGs only, and for ECGs including additional clinical features (most importantly, the initial hs-cTnT lab-value). Thus, a total of 16 different models were developed (four model types, two types of input, with and without prior ECGs).

3.1.4 Results

The expectation was overwhelmingly that the prior ECGs should be of some help. To our surprise, this turned out not to be the case, as illustrated by Fig. 3.2, which shows similar AUC scores for models with and without prior ECGs, for all four model types, both with and without additional clinical variables.

So we started looking for other patterns: perhaps there's a subgroup where the prior

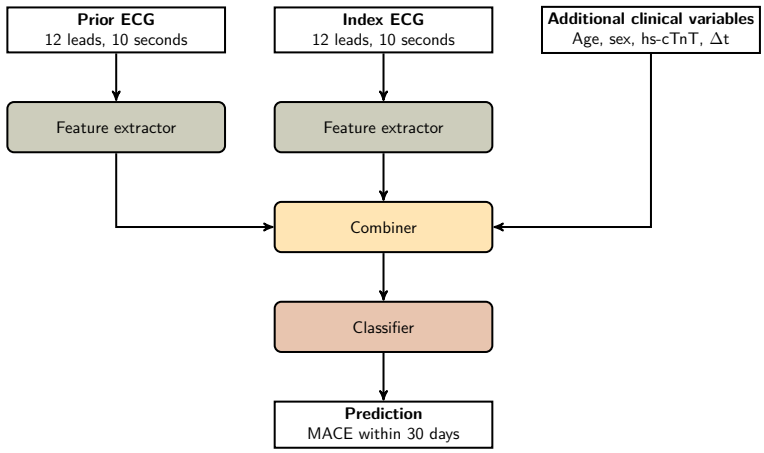

Figure 3.1: Abstract overview of the basic structure of the machine learning models. Each ECG is passed through a feature extractor function before being combined with additional clinical variables. The feature extractor is the same within a model, but can differ between models. All models make use of the index ECG (middle column), but the prior ECG (left column) and additional clinical variables (right column) are only used for some of the models. Δt is the time between index and prior ECG.

ECG is helpful, but it is somehow drowned out in the full population? We divided the test set into different groups, and evaluated the models on the groups separately to see if there was any group where prior ECGs helped more. We tried doing this for patients where the index ECG was "pathological", according to the Uni-G algorithm. We stratified on age quartiles, on previous AMI, and we predicted the different main subsets of the MACE outcome separately. Although there were some variations across models, none of the groups showed any improvement from prior ECGs that was either substantial or consistent between models.

For all our unsuccessful efforts to find any way at all to make use of prior ECGs in this context, or indeed any context, it seems to me that if there is a useful signal, it is at least non-trivial to exploit it. To be sure, human beings work differently from the algorithms explored in this study, but I am nevertheless left wondering about the origin of the clinical practice of consulting prior ECGs. How much does it actually help the clinicians? The conclusions about increased specificity by Lee *et al.* is based on data that is 40 years old, and a lot has happened since then. The very definition of the AMI diagnosis has changed multiple times, the population has naturally shifted, and we are much better at detecting (and treating) minor infarctions that would have previously gone undetected. It would be very interesting to see a modern study of the impact of prior ECGs for physicians, and how often they actually alter the clinical decision on the floor.

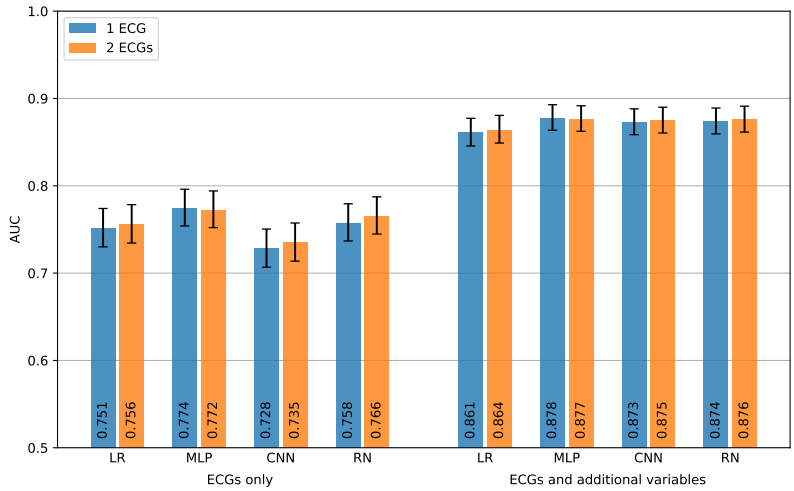

Figure 3.2: Main results showing the test-set AUC, including 95% confidence intervals approximated with bootstrapping. The blue bars correspond to models not using prior ECGs, and the orange bars correspond to models using both the index ECG and the prior ECG. The additional variables used by models on the right hand side of the figure are patient age, sex, troponin T and time since the previous ECG. LR = Logistic Regression, MLP = Multi-Layer Perceptron, CNN = Convolutional Neural Network, RN = ResNet.

3.1.5 Limitations

The choice of MACE as the target outcome might seem odd in hindsight. It was perhaps a combination of over-confidence (it was my first paper after all), and a belief that MACE, being more complex and a better approximation of what we actually care about (compared to predicting AMI alone), might somehow be more strongly coupled to prior ECGs.

3.2 Paper II – Transfer learning

Title	Transfer learning for predicting acute myocardial infarction using electrocardiograms
Aim	To quantify the improvements from transfer learning when using ECGs to predict AMI.
Data	SEM (44 k ED chest-pain visits, and 800 k ECGs from non-chest-pain visits)
Results	Pre-training to predict age and sex lead to a considerable improvement in the downstream task of predicting AMI. The improvements were consistent across several state-of-the-art ECG models, with the best performing model improving from 79% AUC to 85% AUC.

3.2.1 Motivation

The experience of working with ECGs so far was somewhat disappointing. It seemed difficult to outperform traditional ML methods using traditional ECG-features, and any models we trained from scratch tended to do better when they were smaller and used median-beats instead of the full signal. In a way it is perhaps not surprising: the traditional ECG features represent, more or less, a century's worth of human pattern recognition applied to ECGs. It's a powerful set of expert features, hard-won through clinical trials and real-world experience.

My experience with more complex models, such as the ResNet model in study I, was disheartening. Training from scratch seemed outright impossible, and even the pre-trained version from Ribeiro *et al.* [72] was slow, prone to overfitting, and ultimately didn't perform any better than much smaller counterparts. Then again, it was trained on 2 million ECGs, whereas we used only around 10 k in our training sets. That's two orders of magnitude less data, in a field where data generally is known to be decisive for performance.

Then, in the fall of 2023, we finally gained access to the long-awaited ECGs from the full SEM cohort. SEM was a super set of the ESC-TROP database, and included not just the chest-pain visits, but *all* ED visits in that time-period. The total ECG count was just over 1.2 M. Unfortunately, most of those ECGs can't be used directly

for supervised learning, because they were not from patients with chest pain, and so training to predict the AMI outcome would be very misleading. However, it opened up the door to transfer learning in a whole new way.

3.2.2 Previous research

Transfer learning, in a nutshell, is a way to improve the performance of a machine learning model by transferring information from another domain [106]. This can of course be done in many ways, but a popular variation is to pre-train a model on one dataset (called the source), and fine-tune on a different dataset (the target). Transfer learning has been used to great success in many applications, including for ECGs. The most relevant to this paper is Strodthoff *et al.* [73], who showed a significant improvements on PTB-XL [65] for using ECGs to predict 71 different outcomes, including AMI. Jang *et al.* [107] used a large single lead ECG dataset to pre-train an autoencoder, which was subsequently fine-tuned to predict ECG rhythms. Weimann *et al.* [108] found an improved performance for detecting atrial fibrillation after pre-training on 630 k hours of single lead ECG data. Mehari *et al.* [109, 110] explored novel self-supervised methods for ECGs, and obtained large improvements on the PTB-XL outcomes. To my knowledge, we are the first to investigate the benefits of transfer learning explicitly for predicting AMI.

3.2.3 Challenges

Initial experiments were very promising. But even though it was useful to my own efforts to build ECG-based prediction models, it wasn't instantly obvious how to convert the results into a useful paper. At first I thought about comparing my simple transfer learning strategy (pre-training on age and sex) with other approaches, but was discouraged by the complexity of the alternatives that I found, and was scared that trying to implement them would require too much time and effort.

Ideally, a thorough evaluation of a model with and without transfer learning would require proper tuning of the model hyper-parameters. However, such tuning is computationally expensive, and the pre-training for the largest model already took several days to converge. Even with multiple GPUs running in parallel, back-of-the-envelope calculations suggested the tuning process might take half a year, if not more. This seemed excessive, so we started to consider other options. That's how we came up with the idea of picking a handful of published, well-performing ECG models, and evaluating them without any further parameter tuning.

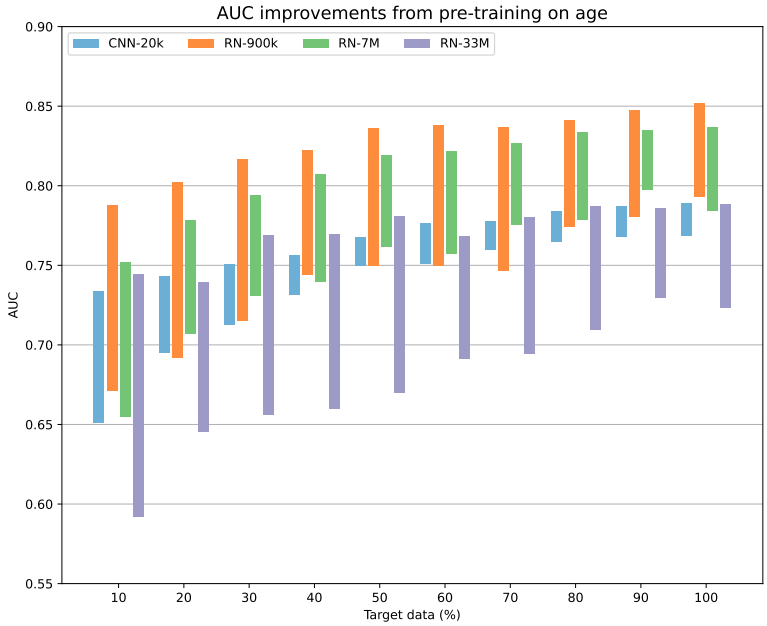

Figure 3.3: Bar chart showing the AUC for predicting AMI. The bottom of each bar shows the AUC for models without pre-training, and the top of each bar shows the AUC for models pre-trained on age, using the full source dataset. The size of each bar thus corresponds to the improvement from transfer learning.

3.2.4 Results

We ended up comparing four different models, evaluating the effect of transfer learning for different amounts of pre-training and fine-tuning data. All models turned out to benefit our transfer learning strategy, with the best overall model (a type of ResNet model described by Mehari *et al.* [109]) improving from 79% AUC to 85% AUC. Fig. 3.3 shows the improvement for each model when pre-training on the full source dataset, for different sizes of the target dataset. The larger models tended to benefit more from pre-training, and the benefit was greater when the target dataset was smaller. The full results for the best model is shown in a heat-map in Fig. 3.4, illustrating the benefit of transfer learning across different combinations of source and target dataset sizes. Remarkably, comparing the top-left and bottom-right cells in the heat-map, it can be seen that with on 10%, i.e., fewer than 3000 training examples, the model utilizing transfer learning performs on par with a non-pre-trained model that has 10 times as much training data on the target task.

I had previously, and I would now say erroneously, been under the impression that transfer learning was primarily a way to speed up training, as opposed to directly improving predictions. Perhaps this view makes sense in theory: whatever patterns

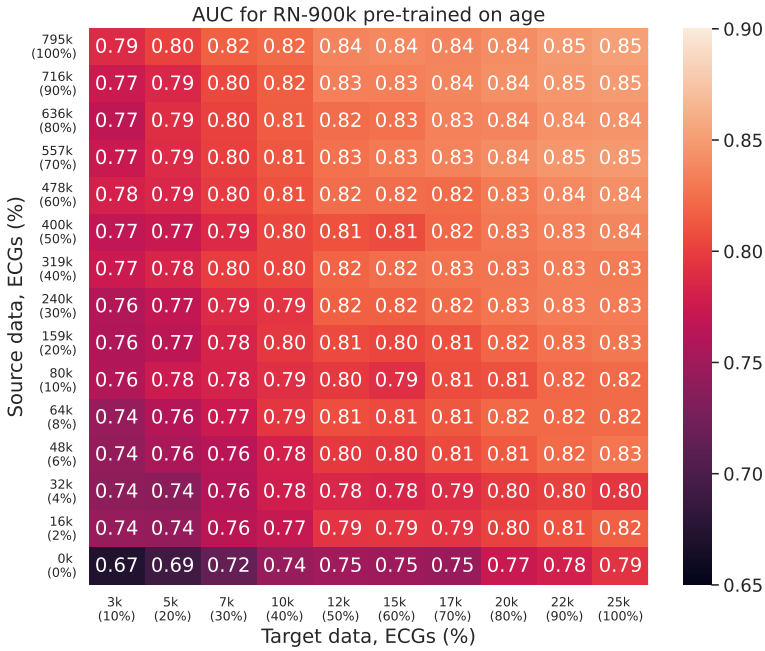

Figure 3.4: Heatmap showing the AUC for predicting AMI with the RN-900k model, for different amounts of source and target training data. The x-axis shows the number of ECGs in the target training dataset, and the y-axis shows the number of ECGs in the source training dataset. The bottom row shows results without pre-training.

and information is in your source dataset is not going to change just because you did some pre-training on a different task. But in practice, all the models I've worked with so far have been helped by transfer learning, some more than others. But it's better than that: not only does it seem that any given model performs better when it is first pre-trained, the improvement from pre-training is bigger for more complex models, essentially "unlocking" that technology for use in the downstream task. Thus it seems that, at least when dealing with ECGs and AMI, the benefit is two-fold: there's a benefit from training on additional ECGs, and a benefit from being able to use a bigger or more complex model that would otherwise perform poorly.

3.2.5 Limitations

Perhaps the biggest limitation of paper II, in my opinion, is that we didn't end up trying any other options for transfer learning. Instead of saying "here's the best out of several popular transfer learning option for ECGs", we settled for the more modest "this simple transfer learning method seems to work well". Another limitation is the choice to not explicitly tune the models. This naturally introduces a certain amount of variance in our estimates of the transfer learning benefits. But, as discussed earlier,

parameter tuning is computationally expensive, and it's likely that tuning attempts would have led to either testing fewer models, or testing on fewer combinations of dataset sizes.

3.3 Paper III – Stepwise predictions and the super model

Title	Stepwise increasing input to machine learning models predicting 30-day AMI or death in emergency department chest-pain patients
Aim	To develop a model for early rule-out of AMI, at multiple decision points corresponding to when information becomes available at the ED. Also, to elucidate the feature importance at each step.
Data	SEM (40 312 ED chest-pain visits)
Results	16% could be safely ruled-out based only on age and sex. 49% could be ruled-out after the initial hs-cTnT lab-value was included. Medical history and point-of-care blood samples were not important when the ECG was available. Troponin, when available, was the strongest predictor of AMI. Sequential rule-out at each step resulted in a breach of the safety margins, despite each individual predictor being within the safety margins.

3.3.1 Motivation

One approach to improving predictions of AMI at the ED, besides obtaining more data, is to include more features. Lab-values, especially hs-cTnT measurements, are particularly good predictors. This should not be a surprise, especially considering that they are part of the diagnostic process, and so there's an obvious and strong causal connection between the measured troponin levels and an AMI diagnosis.

Although including additional lab-values can help to make better predictions, the clinical utility is not necessarily increased: once multiple hs-cTnT samples are available, we essentially already know if there was an infarction, and a decision support tool would be of limited impact.

An alternative to making *better* predictions, is to make *earlier* predictions. Information typically become available at discrete decision points during a patient visit, corresponding to when measurements such as the ECG or lab-values are collected and processed. Ever since I started as a PhD student, we have dreamed about the "ideal"

ED decision support model (a "super model", if you will), which makes predictions from what is available at the moment, and gradually refines the predictions as more information is collected.

In this paper, I collaborated with Pontus Olsson de Capretz to finally bring us a step closer to realizing this "super model", by creating and combining models for five different decision points at the ED. In step one, we use only age and sex as predictors, in step two we include variables relating to the patient medical history, in step three the ECG, in step four point-of-care blood samples (glucose, creatinine, and hemoglobin), and finally in step five the initial hs-cTnT lab variable.

My contribution as second author was mostly on the technical side, assisting with the data processing and model development. The work on paper III and paper IV was largely conducted in parallel.

3.3.2 Previous research

In previous studies, models have been developed for different decision points at the emergency department: Than *et al.* [77] and Björkelund *et al.* [111] developed models relying on two hs-cTn lab-values, and found improvements when compared to the ESC 0h/1h pathway. Olsson de Capretz *et al.* [112] found that a large number of patients could be safely ruled-out based only on the initial hs-cTnT and the ECG. In paper I, part of the analysis considered the index ECG alone for predicting MACE.

To the best of our knowledge, this is the first study to use the holistic approach of predicting AMI at multiple discrete decision points and estimating a cumulative rule-out by applying models sequentially.

3.3.3 Challenges

As always, combining ECGs with other features can be challenging. The approach used in this study mirrored that of paper IV, in which ECG features were extracted in a pre-processing step before being combined with other features. In an earlier draft, Olsson de Capretz used a model developed by Björkelund *et al.* in 2020 [111] to extract ECG features. But as I was making progress on paper IV, we realized that the same transfer learning approach would likely work well also in this project, with only minor changes. Replacing the model by Björkelund *et al.* with the one developed for paper IV resulted in a substantial improvement, as illustrated in Fig. 3.5. Curiously, however, the improvement from using a more powerful ECG feature extractor was largely canceled once the initial hs-cTnT values were included in the feature-set.

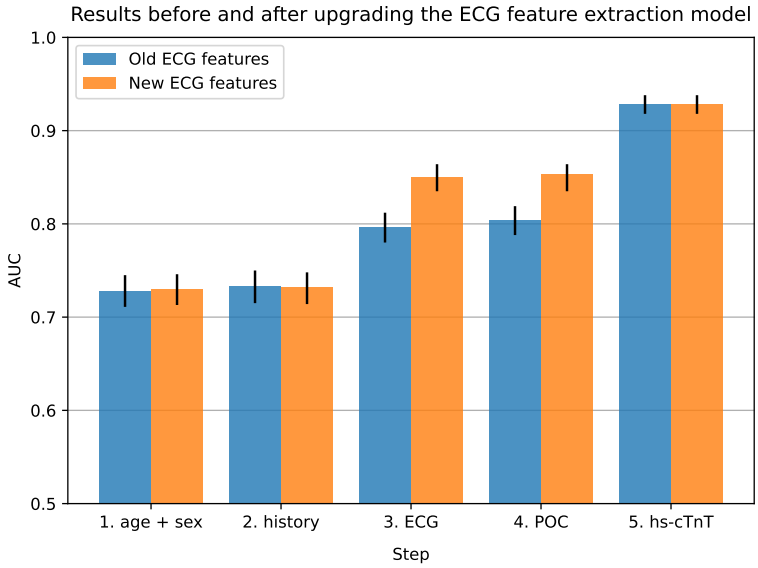

Figure 3.5: Results of the (combined) models, using features from all steps up to the current point. The models represented by the blue bars used the old ECG extraction model, whereas the orange bars used the new ECG extraction model. The black lines indicate 95% confidence intervals. POC=Point-of-care features (creatinine, hemoglobin, glucose), AUC=Area under the receiver operating characteristic curve.

The concept of a safe rule-out is intuitive and well-established among ED physicians. However, using the number of safely ruled-out patients as a model performance metrics comes with a few challenges. One of the challenges is how to define the safety constraints that determine the binary threshold.

A much-cited 2010 survey of 1029 emergency department physicians by Than *et al.* [113] provided some guidance. In the survey, participants were asked what they consider to be an acceptable risk of major adverse cardiac events after discharge from the ED. The results suggested a generally acceptable risk of between 0.1% and 1%. Based on this, we decided to use a minimum sensitivity of 99% in combination with a minimum negative predictive value of 99.5% as the main safety constraints. However, if we want to realistically discharge a patient without even observing an initial hs-cTnT lab-value, as is the goal in the first four decision steps, even stronger safety constraints may be warranted. For this reason, we enforced a 100% sensitivity constraints for the first four decision steps.

3.3.4 Results

We trained and evaluated models for each of the five feature groups individually, as well as in combination, where features from previous steps were included. The orange

Table 3.1: Rule out results for the combined models on the test set, evaluated for each step in isolation and sequentially. POC=Point-of-care blood samples (hemoglobin, glucose, and creatinine), hs-cTnT=high-sensitivity cardiac troponin T, NPV=negative predictive value.

	Rule out				Sensitivity		NPV		Missed
Step	Step-specific	(%)	Sequential	(%)	Step-specific	Sequential	Step-specific	Sequential	Sequential
1. Age+sex	1561	(16%)			0.991		0.996		6
2. +History	1589	(16%)	1816	(18%)	0.986	0.984	0.994	0.994	11
3. +ECG	549	(6%)	1865	(19%)	0.999	0.984	0.998	0.994	11
4. +POC	572	(6%)	1874	(19%)	0.999	0.984	0.998	0.994	11
5. +hs-cTnT	4843	(49%)	5039	(51%)	0.978	0.963	0.997	0.995	26

bars in Fig. 3.5 show the AUC for the combined models.

Rule-out, i.e. the greatest proportion of patients that could be ruled-out while satisfying the safety constraints, was evaluated for each model separately as well as cumulatively by applying the models in a step-wise fashion, such that only patients not ruled-out by previous models would proceed to the next step. The rule-out results are summarized in Table 3.1.

As many as 16% of chest-pain patients could be safely ruled-out based only on age and sex. Interestingly, steps three and four had a lower rule-out score than steps one and two, despite using more data and having a higher AUC. Applying the models sequentially, we could rule-out 51% of the patients at the final step, but the sensitivity was reduced below the acceptable level of 99%.

In terms of feature importance, this project confirmed some loosely held hypotheses, that essentially boil down to this: age is much more important than the medical history, but once the ECG is introduced, it's value is diminished. Point-of-care blood samples don't help that much, and once the initial hs-cTnT lab-value is introduced, the other variables are only marginally useful, with the ECG being the most important of those.

3.3.5 Limitations

One of the limitations of this paper is related to the sequential evaluation of the models across the different steps. The erroneous rule-out of a patient at an early step can't really be undone in subsequent steps. This means that false negatives are tallied up across all steps in a fashion that can easily breach the safety margins, even though each individual model on its own passes the test. I believe this is what happened at step 5 of the cumulative model, where the sensitivity unexpectedly dropped to 96%, even though the last model on its own had a sensitivity of 98%.

Another limitation is related to a lack of power owing to the very strong safety constraints of the rule-out metric. When combined with the relatively low incidence of

AMI, the safety constraints forces the threshold selection to be very sensitive to the predictions of individual patients. Not only does this make rule-out very difficult to optimize, it also causes a large uncertainty in the metric itself. Estimating the distribution of rule-out via bootstrapping reveals a tendency for scores to cluster into several modes, which are due to the presence or absence of individual patients in the bootstrap sample. The somewhat curious results of the step-specific rule-out numbers illustrates this effect, I think. The model at step 2, based on age, sex, and medical history, ruled out 16% of the patients, but at step 3, when ECGs were added, suddenly only 6% of the patients were ruled out, despite the AUC being 12% higher. Presumably a single "unlucky" AMI patient got a very low predicted probability, drastically shifting the threshold to conform to the 100% sensitivity constraint. In my view, this casts serious doubts on the validity of the rule-out percentages reported in the paper.

For these reasons, I think we should either relax the safety constraints of the rule-out metric, or abandon it all together, at least until we can figure out better ways to estimate its uncertainty. For now, though, it seems like the clinical appeal of the rule-out metric is valued higher than concerns around generalizability and discriminative power.

3.4 Paper IV – Predicting occlusion myocardial infarctions

Title	Predicting occlusion myocardial infarctions in the emergency department using artificial intelligence
Aim	To identify acute coronary occlusion myocardial infarction among patients with chest pain at the emergency department, and to build a deep-learning model to predict this outcome from the ECG and other health-care data available early in the ED process.
Data	ESC-TROP (24 511 ED chest-pain visits)
Results	A total of 467 patients (1.9% of all ED chest-pain visits, 29% of AMI cases) were identified, out of which only about one in ten received timely reperfusion therapy. The deep-learning model was able to predict the outcome with an AUC of 95.3% when using the ECG and the initial hs-cTnT lab-value, and an AUC of 87.2% when using ECGs only. The model had roughly twice the sensitivity of the STEMI criteria at the same specificity.

3.4.1 Motivation

I first read the "OMI-manifesto" by Meyers *et al.* [24] in 2021, and although I'm not really in a position to judge the voracity of their claims, it always seemed to me like they were on to something. The basic observation, like we covered in section 1.2.4, is that out of all myocardial infarctions, only some – the acute coronary occlusion myocardial infarctions (OMI) – require emergent revascularization treatment. But the method by which these patients are identified in practice depend entirely on the ECG and the presence of ST-segment elevations [4, 7].

There is a growing body of research indicating that STEMI is an imperfect predictor of OMI, with a sensitivity as low as 44% [114]. This suggests there is considerable room for improvement, where about 40% of OMI patients presenting without STEMI go largely undetected [21], and between 30% and 40% of patients undergoing emergency treatment as a result of ST-elevations on the ECG, show no angiographic evidence of an occlusion ever having taken place [18].

3.4.2 Previous research

There have been numerous attempts to create better ECG rules for identifying OMI [22, 115–117], but they tend to be complex and subtle, with unclear inter-evaluator reliability [118–120]. And as we saw in section 1.5, hand-crafting features is both difficult and time consuming, and is fundamentally limited to the pattern recognition skills of humans. In this day and age, surely we can do better with deep learning!

At the moment of writing, only two machine learning algorithms for predicting OMI have been published. The first was the ECG-SMART algorithm by Al-Zaiti *et al.* [75] in 2023, which obtained an AUC of 87%, outperforming both doctors and the STEMI criteria. The ECG-SMART model is a random forest model, using a collection of hand-crafted features. In 2024, Herman *et al.* [76] published the Queen-of-Hearts algorithm, which had an AUC of 93.8%, also beating the STEMI criteria, and performing similarly to human ECG-experts. The Queen-of-Hearts is currently being validated at multiple external sites [121–124], including on the ESC-TROP material.

3.4.3 Challenges

Meanwhile in Lund, we have this beautiful (all things considered) dataset of over 20 k mostly unselected chest-pain patients, complete with all the most important Swedish national quality register data. The only problem is that OMI is not an ICD-10 diagnosis, and the registers don't exhaustively document occlusions. Without the labels, we can't use supervised learning, and even if we could, we wouldn't be able to tell if the predictions were correct. The only solution, if we want on the OMI bandwagon, is to somehow find a good enough proxy for OMI, with the material that is available.

Our work in this regard started in earnest in the spring of 2023, and it would take a year and a half before we were finally satisfied that we had identified most of the OMI cases. During this time, we had analyzed and combined different register data, manually read over 600 free-text angiography summaries, watched angiography recordings, studied ECGs and performed manual reviews of full health records. My part in this process was mostly confined to the slicing and dicing, combining and restructuring of data, in order to answer questions from my supervisors about how many patients were in group so-and-so, carefully trying to patch together information from different sources so as not to miss anything. Fig. 3.7 shows an overview of the final flowchart, hinting at some of the complexity involved. Part of the difficulty of the process was due to inconsistencies between the data sources, and how to cross-reference them. For example, one of the sub-groups had 10 patients with intervention codes suggesting CABG or PCI, but no registered angiography. Similar corner-cases were found for

pretty much every other "impossible" situation imaginable. In the end, I estimate we spent 95% of the time identifying the final 5% of OMI cases. There are probably still cases missing, but we did the best we could with what we had, and I think it's about as good as it can get without manually examining all angiographies.[1]

With labels defined, I was excited to build on previous experiences with transfer learning for ECG models and to finally create a model to predict OMI. Although the results from the transfer learning was very promising, one of the remaining challenges was how to squeeze the last bits of performance by model selection and tuning. The main obstacle was the pre-training process, which was the most time consuming by far. It is tempting to optimize the pre-training and fine-tuning separately, but I was worried that the best model with respect to mean absolute error in predicting age, would not necessarily be the best choice for fine-tuning on OMI. The only way to know for sure is to test it empirically. Also, two different pre-trained model architectures may easily end up requiring different parameters for fine-tuning. In other words, the two steps are entangled.

I decided to begin with a round of random search over model architectures (mostly variations of ResNet) for pre-training. It took almost a month to train 500 models, using 6 high-end consumer-grade GPUs[2] in parallel. Next, I selected a static fine-tuning protocol for predicting AMI in SEM (I was avoiding the OMI task for fear of overfitting), and created the scatter-plot in Fig. 3.6 to visualize the correlation between the pre-training and fine-tuning performance metrics. Although there is a clear correlation, it confirmed my suspicions that we can't naively use the best model with respect to age predictions, and expect it to be the optimal choice for the downstream task.

In a second round of random search, I let the choice of pre-trained model be part of the search space, picking randomly from the top 50 models. The fine-tuning target was still AMI on SEM. Although the best model on the downstream task was dominated by a handful of pre-trained models, these were not the ones at the very top of the pre-training performance. But what do we even mean by best downstream model anyway? By what metric? The clinical perspective provides some constraints in this regard. In order to provide urgent angiography to suspected OMI cases, we need the positive predictive value (PPV) to be reasonably high. The low prevalence of OMI (around 1.9%) would thus force a very high specificity. As a result, even without an in-depth analysis (which would fall out scope for the paper), the clinical expectation was that we are more interested in the high-specificity region of the ROC space (i.e. the left-most part). For these reasons, besides the regular AUC metric, I also considered

[1] This is not a bad idea as such. Unfortunately, the angiography recordings were not part of the ESC-TROP database itself, and had to be reviewed separately for selected patients.

[2] Two each of Nvidia GeForce 3090, 4090, and A40 GPUs.

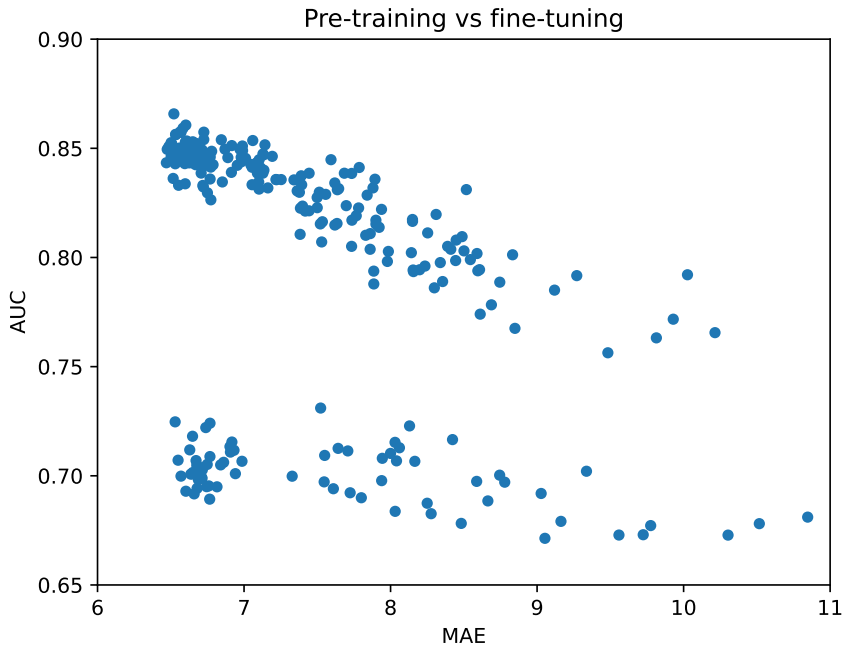

Figure 3.6: Scatter plot showing mean absolute error (MAE, years) for age-predictions on the x-axis, and AUC for predicting AMI on the y-axis. Each dot corresponds to a different pre-trained model architecture. Metrics are evaluated on the validation sets.

the partial AUC [125], where specificity is above 95% and 98%. I also evaluated the sensitivity at a few fixed thresholds, corresponding to a minimum specificity.

Annoyingly, the results of my fine-tuning experiments would find different ”winners” depending on the choice of metric. In other words, the best model with respect to overall AUC was different from the best model with respect to sensitivity at specificity 97%, for instance. Further scrutiny of the high-specificity metrics indicated a worrying trend: the uncertainty, as could be estimated for instance by bootstrapping, was very high. For the threshold specific metrics, the threshold also appeared to be extremely sensitive to individual data points, which in hindsight is natural when considering the low prevalence of OMI. This high variance made me wary of selecting parameters based on these metrics, as it seemed to me I wasn't really able to distinguish a good model from a mediocre one. After careful consideration, I thus made the decision to select model parameters based on overall AUC performance, rather than the other metrics discussed.

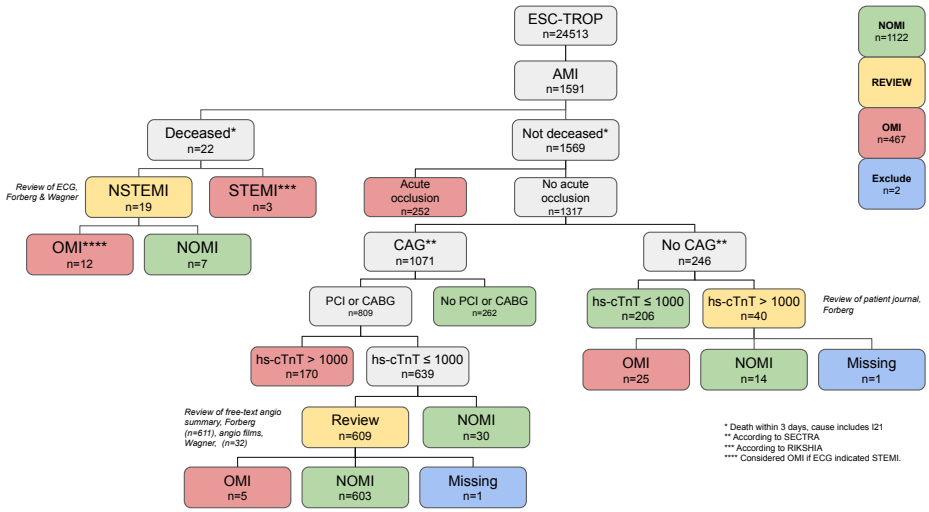

Figure 3.7: Overview of the annotation process for acute coronary occlusion myocardial infarction (OMI). AMI = Acute myocardial infarction, PCI = Percutaneous coronary intervention, CAG = Coronary angiography, CABG = Coronary artery bypass graft surgery, hs-cTnT = high-sensitivity cardiac troponin T.

3.4.4 Results

In the end, we identified 467 cases of OMI, corresponding to 1.9% of all ED visits, and 29% of AMI cases. Compared to non-OMI infarctions (NOMI), the patients with OMI had twice as high 30-day mortality rate, despite being younger and with fewer comorbidities. These numbers are largely in line with that of previous research [126, 127]. Furthermore, only 11% of the OMI patients received angiography within 2 hours, which further confirms the need for improved routines for this patient group.

The prediction model turned out well, predicting OMI with an AUC of 87.2% using ECGs only. When combined with medical history and lab-values, it reached an AUC of 95.3%. ROC-curves for a few of the input feature group combinations are plotted in Fig. 3.8, together with the decision points corresponding to the universal STEMI criteria (machine interpreted), and STEMI-equivalents calculated by the Uni-G ECG program [83].

It turned out that the inital hs-cTnT lab-value scored really well on its own. In fact, even without any logistic regression shenanigans, just straight-up using the hs-cTnT lab-value as a predictor for OMI, we end up with an AUC of 91.1%. A hard threshold of 110 ng/L on the initial hs-cTnT leads to a sensitivity of 48%, compared to 27% for the STEMI criteria. This information alone seems to me could be employed almost instantly in clinical practice, without the need for complex AI models and all the implementation difficulties associated with it. Although of course a more thorough

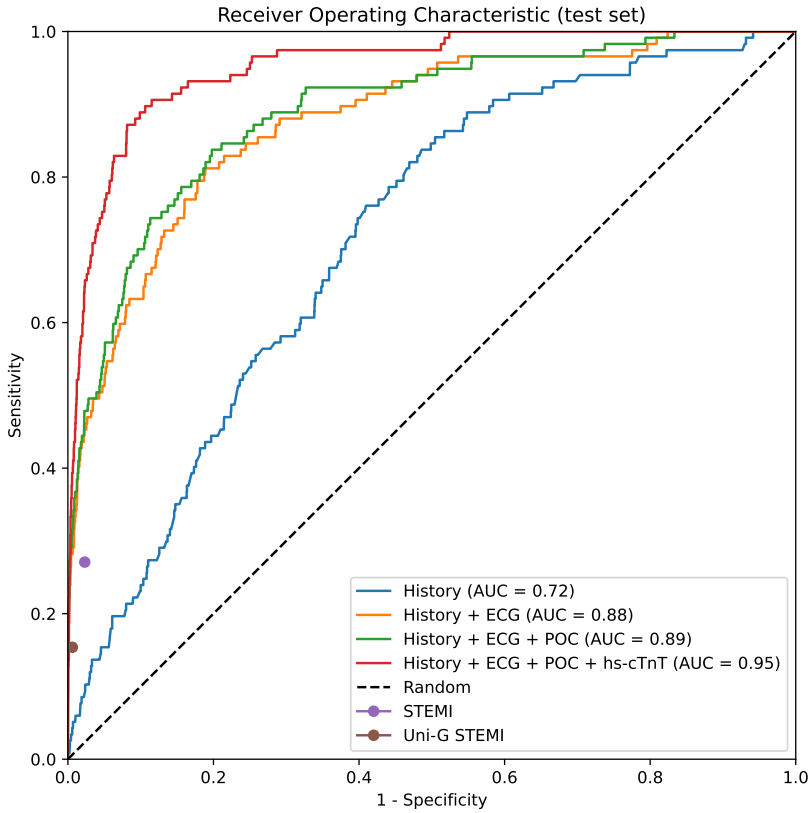

Figure 3.8: Receiver Operating Characteristic (ROC) curves for the AI model with different inputs, evaluated on the test set. The positive class is OMI, the negative class is everyone else. Also plotted are the STEMI criteria and the Uni-G algorithm statements corresponding to STEMI. The dashed line corresponds to randomly guessing the outcome.

analysis would still be required to determine the appropriate tradeoff between sensitivity and specificity, as well as clinical trials to evaluate the safety and efficiency of such an algorithm. And at that point, we might as well go the extra mile of evaluating an AI model too.

Out of curiosity, I also evaluated the AUC for the maximum recorded hs-cTnT values for each patient, even though they are obviously not known when the patient arrives at the ED, and are thus ”cheating”. The AUC turned out to be 99.3%, which means the maximum recorded hs-cTnT are essentially perfect proxies for risk of OMI (a threshold of 250 ng/L results in a sensitivity of 92.5% at a specificity of 97.7%). Note also

that these are only the highest *recorded* hs-cTnT measurements, not a systematically measured peak. Even though it is not directly useful from a clinical perspective, it is a potentially valuable proxy outcome to train supervised models against. After all, from an ML perspective, the lack of labeled datasets currently represents the greatest obstacle to building stronger models, and by simply predicting the maximum hs-cTnT instead, we might have a very strong candidate pre-training target in existing as well as future datasets.

Ultimately, our contributions to the OMI game is two-fold: our annotation efforts confirm previous research, and highlight the need for change both for pre-hospital screening and ED management of suspected OMI. The dataset will continue to be of value for future studies, including as external validation of other models: a collaboration with Herman *et al.* to validate the Queen-of-Hearts algorithm is already well underway, with very promising results. The second contribution is in the development of our own OMI model, which is only the third of its kind, and the first to be non-proprietary and open source. We are also first to explore the benefit of additional clinical variables for predicting OMI using deep learning.

3.4.5 Limitations

Despite a large initial cohort of over 26 000 chest-pain patients, the final number of identified OMI cases was only 467. Although this is of course a good from the perspective of the general health and well-being among patients, it is a bit on the low side from a machine learning perspective. Class imbalance is a challenging problem, and the development of strategies suitable for deep-learning methods is an area of active research [128]. Exploring how recent advancements in area might be useful to our application would be an interesting future research direction.

Chapter 4

Conclusions

Why is this thus? What is the reason for this thusness?
— Artemus Ward

4.1 Summary

In this thesis, we have explored different ways to apply deep-learning algorithms to predict adverse cardiac events among chest-pain patients at the emergency department. In paper I, we found – surprisingly, and contrary to clinical expectations – that prior electrocardiograms appear to be of limited value when predicting MACE. Besides calling into question the general recommendation for physicians to consider prior ECGs, the main benefit of our discovery lie in the opportunity to simplify the development of future prediction models, which in light of our results would be unlikely to extract much benefit from such efforts. In paper II, we confirmed that transfer learning is a valuable technique when training deep-learning models to predict AMI using ECGs. We found that a simple-to-implement strategy of pre-training on age and sex can lead to a considerable increase in the downstream classification performance. The fact that age and sex are available alongside most ECG recordings increases the utility of our approach, thus forming a valuable baseline before implementing more complex approaches.

In paper III, we built models for early rule-out of AMI, at multiple clinical decision points. We found that 16% of patients could be safely ruled out based on age and sex alone, and 50% could be ruled out when incorporating the ECG, medical history and initial lab-values. This suggests that a rule-out decision could be possible very early in the clinical pathway, possibly already at ED presentation.

In paper IV, we identified patients with acute coronary occlusion myocardial infarctions, finding that although 29% of all AMI cases fall into this high-risk category, only 11% receive treatment within the recommended timeframe. We constructed a deep-learning model capable of predicting OMI with an AUC of 95.6% when utilizing both the electrocardiogram and initial hs-cTnT lab-sample, resulting in more than twice as high sensitivity as the international STEMI criteria at the same specificity.

Much work remains to be done, including clinical validation studies, before we can use any of these algorithms in clinical practice. Nevertheless, the contributions in this thesis brings us forward in several important regards: we discovered strategies that considerably enhance the predictive power of deep-learning models using ECGs, and identified multiple areas where such models could potentially be used as decision support tools to streamline the clinical pathway for chest-pain patients at the emergency department. We hope this work will ultimately contribute to improved patient outcomes for this important and vulnerable patient group.

4.2 Future work

The driving force behind the work in this thesis was always the possibility of improved health-care practices through the use of decision support tools. The road to deploying such tools in clinical practice is understood to be both long and winding, with many uncertainties, and this thesis can only take us so far. Among the most important milestones yet to cross is the conduction of a prospective clinical study in which the impact of the prediction models we have constructed can be evaluated in terms of patient outcomes. Before such a study can even begin, software must be developed by, or in collaboration with the hospital, to connect the various required systems for electronic health records and other data sources. The prediction model must be prospectively validated, a graphical user interface must be designed, study protocols must be written, ethical and legal permits must be applied for and received. Much of this work is happily already in progress, and I'm hopeful that the first prospective validation study can be started as early as the fall of 2025.

When it comes to data collection, good things are expected: The construction of the SEM2 database is in full progress, and will fill the information gap between 2018 and 2024. We estimate an additional 2 million ECGs, together with all the other health-care records, to be available as early as this summer.

In terms of model development, we have barely scratched the surface of what can be done with the data sources currently at hand, to not even mention what we hope to get. The technical field is advancing at a staggering pace, with new and exciting

papers being published on a daily basis. There are so many ideas for novel model architectures, and when combined with the ever increasing amount of available data, the opportunities are seemingly endless. Large public ECG databases are published left and right, with perhaps the most exciting one that I've seen so far containing well over 10M ECGs [129], coupled with rich patient metadata, including hundreds of diagnostic outcomes. At this point, the development of a foundation model [130] for ECGs may be within reach, with efforts in this direction already on-going [131].

As for the OMI paradigm shift, I think it is already well underway. The Queen-of-Hearts model is currently on-going external validation at multiple sites, including a collaboration with us. My impression is that the model is extremely capable, and I would be surprised if it does not become widely used. And as more data is pooled and bigger and better models can be applied, it seems to me that the gaps of the STEMI era might finally be entirely closed.

Acknowledgments

I am counted amongst the legions of the unrighteous
Who dread not being immersed in the pits of fire
We are they whom the gods detest
— Nile, Those Whom the Gods Detest

Contrary to my subjective experience in certain moments, a dissertation is not produced in a vacuum. Many people have helped, directly and indirectly, throughout the years, and I would never have made it this far if it weren't for you. Perhaps more than anyone, I want to thank Anders Björkelund, who has been my informal mentor and ever loyal ally, co-author, room mate, skeptic, tech-support, rubber duck, moral support, proofreader, and friend. You have truly been invaluable.

I was blessed with no fewer than four supervisors, all of whom have been instrumental to my studies: Jonas Björk, Ulf Ekelund, Mattias Ohlsson, and Jakob Lundager Forberg. Thank you for all the support, feedback, discussions, and for the freedom to pursue my own interests. Your skills and knowledge complement each other in a way that has been really helpful in this inter-disciplinary work. I am particularly grateful to Jakob: your support, especially in the last two years, made all the difference.

I started my PhD studies in the middle of a global pandemic, with the University recommending people to work from home. The contrast between the first and second halves of my studies makes me appreciate my colleagues at the office all the more, particular my fellow PhD students, with whom I've been fortunate to share office space. Thank you Thomas for the philosophical conversations, and for keeping us all informed about the latest LLM developments. Thank you Daqu, for always seeing the bright side of everything, and for organizing the badminton group. Thank you Suze for being my PhD guinea pig, showing the way and being an inspiration during tough times. Thank you Lucas, for all the laughs and lunches. Thank you Malin, for being down to earth and a good listener. Thanks to Emil, Alexander, Eric, and to all the ~~inmates~~ office mates in COSHE, cellblock B (and A, despite the distance). I feel

like we all became more of a family after the move, and I really appreciate all of you!

I want to smash additional gratitude to all the people I got to play badminton with over the years, especially the regulars: Bo, Jarkko, Tomás, Emil, Leif, Eric, Dibyajyoti, Alexander, Loui, Daqu – It's been a workout and a delight to challenge you on the court, week after week!

I have gotten to know several people in the AIR-Lund network, and in so doing expanded my perspective and learned a great deal. I am grateful to all of you! Thank you Pontus, for all the conversations, ideas and collaborations. It's been a real treat to have your expertise available so close at hand. You may be one of the most curious people I know, and your ability to engage with the technology have really helped to lower the interdisciplinary barrier and made it easier to communicate. Speaking of polyglots, thank you Johanna Berg, for the lunches, pep-talks and Spirit Island games. Just like Pontus, your fluency in techno-babble have brought a unique perspective and led to many fascinating conversations and ideas. Thank you Ellen and Ali, for bringing sanity to otherwise crazy-making circumstances. Thank you André for always having interesting ideas, and for introducing me to the OMI manifesto. Thank you Charlotte and Stefan, for all the intriguing conversations and collaborations about explainability and interpretability. I hope the stains of solutionism will come off eventually.

Tack pappa, för allt stöd, såväl mentalt som ekonomiskt. Jag har *nästan* förlåtit dig för att du lurade med mig till Sälen när allt var som allra stressigast. Tack mamma, för att du alltid lyssnar, stöttar och är på min sida. Jag älskar er båda!

Tack till mina fantastiska syskon Olle och Stina, och era växande familjer. Tack alla bonusfamiljemedlemmar, partners och hela den Ytterbrinkska klanen. Och ett särskilt tack till Anna Visala, som alltid varit där, oavsett vad som hänt eller hur mycket tid som förflutit.

Ester Daniel, Judit och Yoshi: Tack för att jag får vara familj med er. Det är en ynnest att få dela vardagen med er, även när allt annat är kämpigt. Ni är bäst, och jag älskar er!

Bibliography

[1] U. Ekelund, B. Ohlsson, O. Melander, J. Björk, M. Ohlsson, J. Lundager Forberg, P. Olsson de Capretz, A. Nyström, and A. Björkelund, "The Skåne Emergency Medicine (SEM) cohort," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 32, no. 1, p. 37, Apr. 2024, ISSN: 1757-7241. DOI: 10.1186/s13049-024-01206-0.

[2] M. A. Kohn, E. Kwan, M. Gupta, and J. A. Tabas, "Prevalence of acute myocardial infarction and other serious diagnoses in patients presenting to an urban emergency department with chest pain," *The Journal of Emergency Medicine*, vol. 29, no. 4, pp. 383–390, 2005, ISSN: 0736-4679. DOI: 10.1016/j.jemermed.2005.04.010.

[3] M. J. Holzmann, T. Andersson, M. L. Doemland, and S. Roux, "Recurrent myocardial infarction and emergency department visits: A retrospective study on the Stockholm Area Chest Pain Cohort," *Open Heart*, vol. 10, no. 1, e002206, Jun. 2023, ISSN: 2053-3624. DOI: 10.1136/openhrt-2022-002206. (visited on 02/21/2025).

[4] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, H. D. White, and ESC Scientific Document Group, "Fourth universal definition of myocardial infarction (2018)," *European Heart Journal*, vol. 40, no. 3, pp. 237–269, Jan. 2019, ISSN: 0195-668X, 1522-9645. DOI: 10.1093/eurheartj/ehy462.

[5] Y. Sandoval, K. Thygesen, and A. S. Jaffe, "The universal definition of myocardial infarction," *Circulation*, vol. 141, no. 18, pp. 1434–1436, 2020. DOI: 10.1161/CIRCULATIONAHA.120.045708.

[6] W. H. Organization *et al.*, "Hypertension and coronary heart disease: Classification and criteria for epidemiological studies, first report of the expert committee on cardiovascular diseases and hypertension [meeting held in geneva from 13 to 18 october 1958]," in *Hypertension and coronary heart disease: classification and criteria for epidemiological studies, first report of the Expert Committee*

on Cardiovascular Diseases and Hypertension [meeting held in Geneva from 13 to 18 October 1958], 1959.

[7] R. A. Byrne, X. Rossello, J. J. Coughlan, E. Barbato, C. Berry, A. Chieffo, M. J. Claeys, G.-A. Dan, M. R. Dweck, M. Galbraith, *et al.*, "2023 ESC Guidelines for the management of acute coronary syndromes," *European Heart Journal: Acute Cardiovascular Care*, vol. 13, no. 1, pp. 55–161, Feb. 2024, ISSN: 2048-8726, 2048-8734. DOI: 10.1093/ehjacc/zuad107.

[8] G. Meisenberg and W. H. Simmons, *Principles of Medical Biochemistry*. Elsevier Health Sciences, 2016.

[9] H. Patil, O. Vaidya, and D. Bogart, "A review of causes and systemic approach to cardiac troponin elevation," *Clinical Cardiology*, vol. 34, no. 12, pp. 723–728, 2011. DOI: 10.1002/clc.20983.

[10] S. G. Goodman, P. G. Steg, K. A. Eagle, K. A. Fox, J. López-Sendón, G. Montalescot, A. Budaj, B. M. Kennelly, J. M. Gore, J. Allegrone, *et al.*, "The diagnostic and prognostic impact of the redefinition of acute myocardial infarction: Lessons from the Global Registry of Acute Coronary Events (GRACE)," *American Heart Journal*, vol. 151, no. 3, pp. 654–660, 2006, ISSN: 0002-8703. DOI: 10.1016/j.ahj.2005.05.014.

[11] K. Thygesen, J. Mair, E. Giannitsis, C. Mueller, B. Lindahl, S. Blankenberg, K. Huber, M. Plebani, L. M. Biasucci, M. Tubaro, *et al.*, "How to use high-sensitivity cardiac troponins in acute cardiac care," *European Heart Journal*, vol. 33, no. 18, pp. 2252–2257, Jun. 2012, ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehs154.

[12] A. Saenger, R. Beyrau, S. Braun, R. Cooray, A. Dolci, H. Freidank, E. Giannitsis, S. Gustafson, B. Handy, H. Katus, *et al.*, "Multicenter analytical evaluation of a high-sensitivity troponin t assay," *Clinica Chimica Acta*, vol. 412, no. 9, pp. 748–754, 2011, ISSN: 0009-8981. DOI: 10.1016/j.cca.2010.12.034.

[13] J. W. Pickering, L. R. Joyce, C. M. Florkowski, V. Buchan, L. Hamill, and M. P. Than, "Emergency department use of a high-sensitivity point-of-care troponin assay reduces length of stay: An implementation study preliminary report," *European Heart Journal. Acute Cardiovascular Care*, vol. 13, no. 12, pp. 838–842, Oct. 2024, ISSN: 2048-8734. DOI: 10.1093/ehjacc/zuae114.

[14] A. Mokhtari, J. L. Forberg, J. Sandgren, C. Hård Af Segerstad, C. Ellehuus, U. Ekström, J. Björk, B. Lindahl, A. Khoshnood, and U. Ekelund, "Effectiveness and Safety of the ESC-TROP (European Society of Cardiology oh/1h Troponin Rule-Out Protocol) Trial," *Journal of the American Heart Association*, e036307, Oct. 2024, ISSN: 2047-9980. DOI: 10.1161/JAHA.124.036307.

[15] J. Hampton and J. Hampton, *The ECG Made Easy, Ninth edition*. Elsevier, 2019.

[16] K. Huber, B. J. Gersh, P. Goldstein, C. B. Granger, and P. W. Armstrong, "The organization, function, and outcomes of ST-elevation myocardial infarction networks worldwide: Current state, unmet needs and future directions," *European Heart Journal*, vol. 35, no. 23, pp. 1526–1532, Apr. 2014, ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehu125.

[17] D. M. Larson, K. M. Menssen, S. W. Sharkey, S. Duval, R. S. Schwartz, J. Harris, J. T. Meland, B. T. Unger, and T. D. Henry, ""false-positive" cardiac catheterization laboratory activation among patients with suspected ST-segment elevation myocardial infarction," *JAMA*, vol. 298, no. 23, pp. 2754–2760, Dec. 2007, ISSN: 0098-7484. DOI: 10.1001/jama.298.23.2754.

[18] J. M. McCabe, E. J. Armstrong, A. Kulkarni, K. S. Hoffmayer, P. D. Bhave, S. Garg, A. Patel, J. S. MacGregor, P. Hsue, J. C. Stein, *et al.*, "Prevalence and Factors Associated With False-Positive ST-Segment Elevation Myocardial Infarction Diagnoses at Primary Percutaneous Coronary Intervention–Capable Centers: A Report From the Activate-SF Registry," *Archives of Internal Medicine*, vol. 172, no. 11, Jun. 2012, ISSN: 0003-9926. DOI: 10.1001/archinternmed.2012.945.

[19] H. P. Meyers and S. W. Smith, "Prospective, real-world evidence showing the gap between ST elevation myocardial infarction (STEMI) and occlusion MI (OMI)," *International Journal of Cardiology*, vol. 293, pp. 48–49, Oct. 2019, ISSN: 01675273. DOI: 10.1016/j.ijcard.2019.07.043.

[20] P. Hillinger, I. Strebel, R. Abächerli, R. Twerenbold, K. Wildi, D. Bernhard, T. Nestelberger, J. Boeddinghaus, P. Badertscher, D. Wussler, *et al.*, "Prospective validation of current quantitative electrocardiographic criteria for ST-elevation myocardial infarction," *International Journal of Cardiology*, vol. 292, pp. 1–12, Oct. 2019, ISSN: 01675273. DOI: 10.1016/j.ijcard.2019.04.041.

[21] H. P. Meyers, A. Bracey, D. Lee, A. Lichtenheld, W. J. Li, D. D. Singer, J. A. Kane, K. W. Dodd, K. E. Meyers, H. C. Thode, *et al.*, "Comparison of the ST-Elevation Myocardial Infarction (STEMI) vs. NSTEMI and Occlusion MI (OMI) vs. NOMI Paradigms of Acute MI," *The Journal of Emergency Medicine*, vol. 60, no. 3, pp. 273–284, Mar. 2021, ISSN: 07364679. DOI: 10.1016/j.jemermed.2020.10.026.

[22] H. Pendell Meyers, A. Bracey, D. Lee, A. Lichtenheld, W. J. Li, D. D. Singer, Z. Rollins, J. A. Kane, K. W. Dodd, K. E. Meyers, *et al.*, "Accuracy of OMI ECG findings versus STEMI criteria for diagnosis of acute coronary occlusion

myocardial infarction," *IJC Heart & Vasculature*, vol. 33, 2021, ISSN: 2352-9067. DOI: 10.1016/j.ijcha.2021.100767.

[23] J. McLaren, J. N. De Alencar, E. K. Aslanger, H. P. Meyers, and S. W. Smith, "From ST-Segment Elevation MI to Occlusion MI," *JACC: Advances*, vol. 3, no. 11, p. 101 314, Nov. 2024, ISSN: 2772963X. DOI: 10.1016/j.jacadv.2024.101314.

[24] P. Meyers, S. Weingart, and S. Smith, *The OMI Manifesto*, 2018. [Online]. Available: http://hqmeded-ecg.blogspot.com/2018/04/the-omi-manifesto.html (visited on 01/23/2025).

[25] E. K. Aslanger, H. P. Meyers, Department of Emergency Medicine, Carolinas Medical Center, Charlotte; North Carolina-United States of America, S. W. Smith, and Department of Emergency Medicine, University of Minnesota, Hennepin County Medical Center, Minneapolis; Minnesota-United States of America, "Time for a new paradigm shift in myocardial infarction," *The Anatolian Journal of Cardiology*, vol. 25, no. 3, pp. 156–162, Apr. 2021, ISSN: 21492263, 21492271. DOI: 10.5152/AnatolJCardiol.2021.89304.

[26] J. T. McLaren, M. El-Baba, V. Sivashanmugathas, H. P. Meyers, S. W. Smith, and L. B. Chartier, "Missing occlusions: Quality gaps for ED patients with occlusion MI," *The American Journal of Emergency Medicine*, vol. 73, pp. 47–54, Nov. 2023, ISSN: 07356757. DOI: 10.1016/j.ajem.2023.08.022.

[27] R. Mehran and G. D. Dangas, "Braunwald's heart disease: A textbook of cardiovascular medicine," in Elsevier Health Sciences, 2019, ch. 20, pp. 374–400.

[28] C. R. Carpenter, A. S. Raja, and M. D. Brown, "Overtesting and the downstream consequences of overtreatment: Implications of "preventing overdiagnosis" for emergency medicine," *Academic Emergency Medicine*, vol. 22, no. 12, pp. 1484–1492, 2015. DOI: 10.1111/acem.12820.

[29] A. J. Six, B. E. Backus, and J. C. Kelder, "Chest pain in the emergency room: Value of the HEART score," *Netherlands Heart Journal*, vol. 16, pp. 191–196, 2008.

[30] S. A. Mahler, R. F. Riley, B. C. Hiestand, G. B. Russell, J. W. Hoekstra, C. W. Lefebvre, B. A. Nicks, D. M. Cline, K. L. Askew, S. B. Elliott, *et al.*, "The HEART Pathway Randomized Trial," *Circulation: Cardiovascular Quality and Outcomes*, vol. 8, no. 2, pp. 195–203, 2015. DOI: 10.1161/CIRCOUTCOMES.114.001384.

[31] T. Nilsson, E. Johannesson, J. Lundager Forberg, A. Mokhtari, and U. Ekelund, "Diagnostic accuracy of the HEART Pathway and EDACS-ADP when combined with a 0-hour/1-hour hs-cTnT protocol for assessment of acute chest

pain patients," *Emergency Medicine Journal*, vol. 38, no. 11, pp. 808–813, 2021, ISSN: 1472-0205. DOI: 10.1136/emermed-2020-210833.

[32] T. Reichlin, C. Schindler, B. Drexler, R. Twerenbold, M. Reiter, C. Zellweger, B. Moehring, R. Ziller, R. Hoeller, M. Rubini Gimenez, *et al.*, "One-hour rule-out and rule-in of acute myocardial infarction using high-sensitivity cardiac troponin t," *Archives of Internal Medicine*, vol. 172, no. 16, pp. 1211–1218, Sep. 2012, ISSN: 0003-9926. DOI: 10.1001/archinternmed.2012.3698.

[33] A. Mokhtari, C. Borna, P. Gilje, P. Tydén, B. Lindahl, H.-J. Nilsson, A. Khoshnood, J. Björk, and U. Ekelund, "A 1-h combination algorithm allows fast rule-out and rule-in of major adverse cardiac events," *JACC*, vol. 67, no. 13, pp. 1531–1540, 2016. DOI: 10.1016/j.jacc.2016.01.059.

[34] D. P. Chew, K. Lambrakis, A. Blyth, A. Seshadri, M. J. Edmonds, T. Briffa, L. A. Cullen, S. Quinn, J. Karnon, A. Chuang, *et al.*, "A randomized trial of a 1-hour troponin t protocol in suspected acute coronary syndromes," *Circulation*, vol. 140, no. 19, pp. 1543–1556, 2019. DOI: 10.1161/CIRCULATIONAHA.119.042891.

[35] A. Mokhtari, A. Khoshnood, J. Lundager Forberg, C. Hård af Segerstad, U. Ekström, T. Schyman, M. Akbarzadeh, B. Lindahl, and U. Ekelund, "Effectiveness and Safety of the European Society of Cardiology 0-/1-h Troponin Rule-Out Protocol: The Design of the ESC-TROP Multicenter Implementation Study," *Cardiology*, vol. 145, no. 11, pp. 685–692, 2020, ISSN: 0008-6312, 1421-9751. DOI: 10.1159/000509390.

[36] E. H. Shortliffe and B. G. Buchanan, "A model of inexact reasoning in medicine," *Mathematical Biosciences*, vol. 23, no. 3, pp. 351–379, 1975, ISSN: 0025-5564. DOI: 10.1016/0025-5564(75)90047-4.

[37] P. Oosterbroek, "Catalogue of the craneflies of the world (CCW)," *Cranefly Recording Scheme Newsletter*, vol. 16, pp. 9–10, 2019. [Online]. Available: https://ccw.naturalis.nl/ (visited on 02/18/2025).

[38] S. Kaufman, S. Rosset, and C. Perlich, "Leakage in data mining: Formulation, detection, and avoidance," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 6, Jan. 2011, pp. 556–563. DOI: 10.1145/2020408.2020496.

[39] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 472–482, 1981. DOI: 10.1109/TIT.1981.1056373.

[40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org.

[41] M. F. Kilkenny and K. M. Robinson, "Data quality: "garbage in – garbage out"," *Health Information Management Journal*, vol. 47, no. 3, pp. 103–105, 2018. DOI: 10.1177/1833358318774357.

[42] S. E. Whang and J.-G. Lee, "Data collection and quality challenges for deep learning," *Proc. VLDB Endow.*, vol. 13, no. 12, pp. 3429–3432, Aug. 2020, ISSN: 2150-8097. DOI: 10.14778/3415478.3415562.

[43] S. E. Whang, Y. Roh, H. Song, and J.-G. Lee, "Data collection and quality challenges in deep learning: A data-centric ai perspective," *The VLDB Journal*, vol. 32, no. 4, pp. 791–813, Jul. 2023, ISSN: 0949-877X. DOI: 10.1007/s00778-022-00775-9.

[44] S. Ansari, N. Farzaneh, M. Duda, K. Horan, H. B. Andersson, Z. D. Goldberger, B. K. Nallamothu, and K. Najarian, "A Review of Automated Methods for Detection of Myocardial Ischemia and Infarction Using Electrocardiogram and Electronic Health Records," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 264–298, 2017, ISSN: 1937-3333, 1941-1189. DOI: 10.1109/RBME.2017.2757953.

[45] H. V. Pipberger, R. J. Arms, and F. W. Stallmann, "Automatic Screening of Normal and Abnormal Electrocardiograms by Means of a Digital Electronic Computer.," *Experimental Biology and Medicine*, vol. 106, no. 1, pp. 130–132, Jan. 1961, ISSN: 1535-3702, 1535-3699. DOI: 10.3181/00379727-106-26260.

[46] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14539.

[47] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Computers in Biology and Medicine*, vol. 122, 2020, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2020.103801.

[48] G. D. Clifford, C. Liu, B. Moody, L.-w. H. Lehman, I. Silva, Q. Li, A. E. Johnson, and R. G. Mark, "AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4. DOI: 10.22489/CinC.2017.065-469.

[49] G. Moody, W. Muldrow, and R. Mark, "The MIT-BIH noise stress test database," *Computers in cardiology*, vol. 11, pp. 381–384, 1984.

[50] H. Guvenir, B. Acar, H. Muderrisoglu, and R. Quinlan, *Arrhythmia*, UCI Machine Learning Repository, 1997. DOI: 10.24432/C5BS32.

[51] G. Moody, *MIT-BIH Normal Sinus Rhythm Database*, 1999. DOI: 10.13026/C2NK5R.

[52] P. Albrecht, G. Moody, and R. Mark, *MIT-BIH Malignant Ventricular Ectopy Database*, 1999. DOI: 10.13026/C22P44.

[53] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Computers in cardiology 1997*, IEEE, 1997, pp. 673–676. DOI: 10.1109/CIC.1997.648140.

[54] G. Moody and R. Mark, *MIT-BIH Atrial Fibrillation Database*, 2000. DOI: 10.13026/C2MW2D.

[55] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, 1986, ISSN: 0735-1097. DOI: 10.1016/S0735-1097(86)80478-8.

[56] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, R1078–R1084, 1996. DOI: 10.1152/ajpregu.1996.271.4.R1078.

[57] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das internet," *Biomedizinische Technik / Biomedical Engineering*, vol. 40, pp. 317–318, Jan. 1995.

[58] G. Moody and R. Mark, *MIT-BIH Arrhythmia Database*, 2005. DOI: 10.13026/C2F305.

[59] F. Nolle, F. Badura, J. Catlett, R. Bowser, and M. Sketch, *Creighton university ventricular tachyarrhythmia database*, 2007. DOI: 10.13026/C2X59M.

[60] E. Yakushenko, *St Petersburg INCART 12-lead Arrhythmia Database*, 2008. DOI: 10.13026/C2V88N.

[61] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The european ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *European Heart Journal*, vol. 13, no. 9, pp. 1164–1172, Sep. 1992, ISSN: 0195-668X. DOI: 10.1093/oxfordjournals.eurheartj.a060332.

[62] J.-w. Zhang, L.-p. Wang, X. Liu, H.-h. Zhu, and J. Dong, "Chinese cardiovascular disease database (ccdd) and its management tool," in *2010 IEEE International Conference on BioInformatics and BioEngineering*, 2010, pp. 66–72. DOI: 10.1109/BIBE.2010.19.

[63] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018, ISSN: 2156-7018. DOI: 10.1166/jmihi.2018.2442.

[64] S. Tan, G. Androz, A. Chamseddine, P. Fecteau, A. Courville, Y. Bengio, and J. P. Cohen, "Icentia11k: An unsupervised representation learning dataset for arrhythmia subtype discovery," *arXiv preprint arXiv:1910.09570*, 2019. DOI: 10.48550/arXiv.1910.09570.

[65] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, p. 154, May 2020, ISSN: 2052-4463. DOI: 10.1038/s41597-020-0495-6.

[66] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific Data*, vol. 7, no. 1, p. 48, Feb. 2020, ISSN: 2052-4463. DOI: 10.1038/s41597-020-0386-x.

[67] A. H. Ribeiro, G. M. Paixao, E. M. Lima, M. Horta Ribeiro, M. M. Pinto Filho, P. R. Gomes, D. M. Oliveira, W. Meira Jr, T. B. Schon, and A. L. P. Ribeiro, *CODE-15%: A large scale annotated dataset of 12-lead ECGs*, Jun. 2021. DOI: 10.5281/zenodo.4916206.

[68] H. Liu, D. Chen, D. Chen, X. Zhang, H. Li, L. Bian, M. Shu, and Y. Wang, "A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements," *Scientific Data*, vol. 9, no. 1, p. 272, Jun. 2022, ISSN: 2052-4463. DOI: 10.1038/s41597-022-01403-5.

[69] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107 187, Sep. 2021, ISSN: 09507051. DOI: 10.1016/j.knosys.2021.107187.

[70] Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. b. M. Pauzi, Q. Zhang, and P. Y. Lim, "Deep learning-based ecg arrhythmia classification: A systematic review," *Applied Sciences*, vol. 13, no. 8, 2023, ISSN: 2076-3417. DOI: 10.3390/app13084964.

[71] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," Jul. 2017, arXiv: 1707.01836. [Online]. Available: http://arxiv.org/abs/1707.01836.

[72] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira, *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications*, vol. 11, no. 1, p. 1760, Dec. 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-15432-4.

[73] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, 2021. DOI: 10.1109/JBHI.2020.3022989.

[74] S. Gustafsson, D. Gedon, E. Lampa, A. H. Ribeiro, M. J. Holzmann, T. B. Schön, and J. Sundström, "Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients," *Scientific Reports*, vol. 12, no. 1, p. 19 615, Nov. 2022, ISSN: 2045-2322. DOI: 10.1038/s41598-022-24254-x.

[75] S. S. Al-Zaiti, C. Martin-Gill, J. K. Zègre-Hemsey, Z. Bouzid, Z. Faramand, M. O. Alrawashdeh, R. E. Gregg, S. Helman, N. T. Riek, K. Kraevsky-Phillips, *et al.*, "Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction," *Nature Medicine*, vol. 29, no. 7, pp. 1804–1813, Jul. 2023, ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-023-02396-3. (visited on 12/12/2024).

[76] R. Herman, H. P. Meyers, S. W. Smith, D. T. Bertolone, A. Leone, K. Bermpeis, M. M. Viscusi, M. Belmonte, A. Demolder, V. Boza, *et al.*, "International evaluation of an artificial intelligence–powered electrocardiogram model detecting acute coronary occlusion myocardial infarction," *European Heart Journal - Digital Health*, vol. 5, no. 2, pp. 123–133, Mar. 2024, ISSN: 2634-3916. DOI: 10.1093/ehjdh/ztad074.

[77] M. P. Than, J. W. Pickering, Y. Sandoval, A. S. Shah, A. Tsanas, F. S. Apple, S. Blankenberg, L. Cullen, C. Mueller, J. T. Neumann, *et al.*, "Machine Learning to Predict the Likelihood of Acute Myocardial Infarction," *Circulation*, vol. 140, no. 11, pp. 899–909, Sep. 2019, ISSN: 0009-7322, 1524-4539. DOI: 10.1161/CIRCULATIONAHA.119.041980.

[78] D. Doudesis, K. K. Lee, J. Yang, R. Wereski, A. S. V. Shah, A. Tsanas, A. Anand, J. W. Pickering, M. P. Than, and N. L. Mills, "Validation of the myocardial-ischaemic-injury-index machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogenous population: A prespecified exploratory analysis," *The Lancet Digital Health*, vol. 4, no. 5, e300–e308, May 2022, ISSN: 2589-7500. DOI: 10.1016/S2589-7500(22)00025-5.

[79] P. Lopez-Ayala, J. Boeddinghaus, T. Nestelberger, L. Koechlin, T. Zimmermann, P. Bima, J. Glaeser, C. C. Spagnuolo, A. Champetier, O. Miro, *et al.*, "External validation of the myocardial-ischaemic-injury-index machine learning algorithm for the early diagnosis of myocardial infarction: A multicentre cohort study," *The Lancet Digital Health*, vol. 6, no. 7, e480–e488, Jul. 2024, ISSN: 2589-7500. DOI: 10.1016/S2589-7500(24)00088-8.

[80] D. Doudesis, K. K. Lee, J. Boeddinghaus, A. Bularga, A. V. Ferry, C. Tuck, M. T. H. Lowry, P. Lopez-Ayala, T. Nestelberger, L. Koechlin, *et al.*, "Machine learning for diagnosis of myocardial infarction using cardiac troponin concentrations," *Nature Medicine*, vol. 29, no. 5, pp. 1201–1210, May 2023, ISSN: 1546-170X. DOI: 10.1038/s41591-023-02325-4.

[81] J. T. Neumann, R. Twerenbold, F. Ojeda, S. J. Aldous, B. R. Allen, F. S. Apple, H. Babel, R. H. Christenson, L. Cullen, E. Di Carluccio, *et al.*, "Personalized diagnosis in suspected myocardial infarction," *Clinical Research in Cardiology*, vol. 112, no. 9, pp. 1288–1301, Sep. 2023, ISSN: 1861-0692. DOI: 10.1007/s00392-023-02206-3.

[82] P.-I. Zhang, C.-C. Hsu, Y. Kao, C.-J. Chen, Y.-W. Kuo, S.-L. Hsu, T.-L. Liu, H.-J. Lin, J.-J. Wang, C.-F. Liu, *et al.*, "Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 28, no. 1, p. 93, Sep. 2020, ISSN: 1757-7241. DOI: 10.1186/s13049-020-00786-x.

[83] P. Macfarlane, B. Devine, and E. Clark, "The university of glasgow (Uni-G) ECG analysis program," in *Computers in Cardiology, 2005*, Lyon, France: IEEE, 2005, pp. 451–454. DOI: 10.1109/CIC.2005.1588134.

[84] P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for activation functions*, 2017. arXiv: 1710.05941 [cs.NE]. [Online]. Available: https://arxiv.org/abs/1710.05941.

[85] K. O'shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[86] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014, ISSN: 1532-4435.

[87] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015. arXiv: 1502.03167 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1502.03167.

[88] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[89] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[90] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015.

[91] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 5353–5360.

[92] R. K. Srivastava, K. Greff, and J. Schmidhuber, *Training very deep networks*, 2015. DOI: 10.48550/arXiv.1507.06228. eprint: 1507.06228.

[93] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, *Aggregated residual transformations for deep neural networks*, 2017. arXiv: 1611.05431 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1611.05431.

[94] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 558–567. DOI: 10.1109/CVPR.2019.00065.

[95] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012. [Online]. Available: http://jmlr.org/papers/v13/bergstra12a.html.

[96] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *33rd Conference on Neural Information Processing Systems*, 2019.

[97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[98] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/.

[99] J.-P. Collet, H. Thiele, E. Barbato, O. Barthélémy, J. Bauersachs, D. L. Bhatt, P. Dendale, M. Dorobantu, T. Edvardsen, T. Folliguet, *et al.*, "2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the european society of cardiology (ESC)," *European Heart Journal*, vol. 42, no. 14, pp. 1289–1367, Aug. 2020, ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehaa575.

[100] T. H. Lee, E. F. Cook, M. C. Weisberg, G. W. Rouan, D. A. Brand, and L. Goldman, "Impact of the availability of a prior electrocardiogram on the triage of the patient with acute chest pain," *Journal of General Internal Medicine*, vol. 5, no. 5, pp. 381–388, Sep. 1990, ISSN: 0884-8734, 1525-1497. DOI: 10.1007/BF02599421.

[101] A. Sbrollini, M. C. De Jongh, C. C. Ter Haar, R. W. Treskes, S. Man, L. Burattini, and C. A. Swenne, "Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach," *BioMedical Engineering OnLine*, vol. 18, no. 1, p. 15, Dec. 2019, ISSN: 1475-925X. DOI: 10.1186/s12938-019-0630-9.

[102] C. C. ter Haar, R. J. G. Peters, J. Bosch, A. Sbrollini, S. Gripenstedt, R. Adams, E. Bleijenberg, C. J. H. J. Kirchhof, R. Alizadeh Dehnavi, L. Burattini, *et al.*, "An initial exploration of subtraction electrocardiography to detect myocardial ischemia in the prehospital setting," *Annals of Noninvasive Electrocardiology*, vol. 25, no. 3, e12722, 2020. DOI: 10.1111/anec.12722.

[103] Z. Bouzid, Z. Faramand, C. Martin-Gill, S. M. Sereika, C. W. Callaway, S. Saba, R. Gregg, F. Badilini, E. Sejdic, and S. S. Al-Zaiti, "Incorporation of serial 12-lead electrocardiogram with machine learning to augment the out-of-hospital diagnosis of non-ST elevation acute coronary syndrome," *Annals of Emergency Medicine*, vol. 81, no. 1, pp. 57–69, 2023, ISSN: 0196-0644. DOI: 10.1016/j.annemergmed.2022.08.005.

[104] M. Ohlsson, H. Öhlin, S. M. Wallerstedt, and L. Edenbrandt, "Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction," *The American Journal of Cardiology*, vol. 88, no. 5, pp. 478–481, Sep. 2001, ISSN: 00029149. DOI: 10.1016/S0002-9149(01)01722-2.

[105] J. L. Forberg, M. Green, J. Björk, M. Ohlsson, L. Edenbrandt, H. Öhlin, and U. Ekelund, "In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department," *Journal of Electrocardiology*, vol. 42, no. 1, pp. 58–63, Jan. 2009, ISSN: 00220736. DOI: 10.1016/j.jelectrocard.2008.07.010.

[106] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, Dec. 2016, ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6.

[107] Y. D. Jang Jong-Hwan Kim Tae Young, "Effectiveness of transfer learning for deep learning-based electrocardiogram analysis," *Healthc Inform Res*, vol. 27, no. 1, pp. 19–28, 2021. DOI: 10.4258/hir.2021.27.1.19.

[108] K. Weimann and T. O. F. Conrad, "Transfer learning for ECG classification," *Scientific Reports*, vol. 11, 2021, ISSN: 2045-2322. DOI: 10.1038/s41598-021-84374-8.

[109] T. Mehari and N. Strodthoff, "Self-supervised representation learning from 12-lead ECG data," *Computers in Biology and Medicine*, vol. 141, pp. 105–114, 2022, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2021.105114.

[110] T. Mehari and N. Strodthoff, "Towards quantitative precision for ECG analysis: Leveraging state space models, self-supervision and patient metadata," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 11, pp. 5326–5334, 2023. DOI: 10.1109/JBHI.2023.3310989.

[111] A. Björkelund, M. Ohlsson, J. Lundager Forberg, A. Mokhtari, P. Olsson De Capretz, U. Ekelund, and J. Björk, "Machine learning compared with rule-in/rule-out algorithms and logistic regression to predict acute myocardial infarction based on troponin T concentrations," *JACEP Open*, vol. 2, no. 2, e12363, Apr. 2021, ISSN: 26881152. DOI: 10.1002/emp2.12363.

[112] P. Olsson de Capretz, A. Björkelund, J. Björk, M. Ohlsson, A. Mokhtari, A. Nyström, and U. Ekelund, "Machine learning for early prediction of acute myocardial infarction or death in acute chest pain patients using electrocardiogram and blood tests at presentation," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 25, Feb. 2023, ISSN: 1472-6947. DOI: 10.1186/s12911-023-02119-1.

[113] M. Than, M. Herbert, D. Flaws, L. Cullen, E. Hess, J. E. Hollander, D. Diercks, M. W. Ardagh, J. A. Kline, Z. Munro, *et al.*, "What is an acceptable risk of major adverse cardiac event in chest pain patients soon after discharge from the Emergency Department?" *International Journal of Cardiology*, vol. 166, no. 3, pp. 752–754, Jul. 2013, ISSN: 01675273. DOI: 10.1016/j.ijcard.2012.09.171.

[114] J. N. de Alencar Neto, M. K. Scheffer, B. P. Correia, K. G. Franchini, S. P. Felicioni, and M. F. N. De Marchi, "Systematic review and meta-analysis of diagnostic test accuracy of ST-segment elevation for acute coronary occlusion," *International Journal of Cardiology*, vol. 402, p. 131 889, 2024, ISSN: 0167-5273. DOI: 10.1016/j.ijcard.2024.131889.

[115] M. C. Kontos, J. A. de Lemos, S. B. Deitelzweig, D. B. Diercks, M. O. Gore, E. P. Hess, C. P. McCarthy, J. K. McCord, P. I. Musey, T. C. Villines, *et al.*, "2022 ACC Expert Consensus Decision Pathway on the Evaluation and Disposition of Acute Chest Pain in the Emergency Department," *Journal of the American College of Cardiology*, vol. 80, no. 20, pp. 1925–1960, 2022. DOI: 10.1016/j.jacc.2022.08.750.

[116] E. K. Aslanger, Ö. Yıldırımtürk, B. Şimşek, E. Bozbeyoğlu, M. A. Şimşek, C. Yücel Karabay, S. W. Smith, and M. Değertekin, "DIagnostic accuracy oF electrocardiogram for acute coronary OCClUsion resuLTing in myocardial infarction (DIFOCCULT Study)," *IJC Heart & Vasculature*, vol. 30, p. 100 603, Oct. 2020, ISSN: 23529067. DOI: 10.1016/j.ijcha.2020.100603.

[117] H. P. Meyers, A. Bracey, D. Lee, A. Lichtenheld, W. J. Li, D. D. Singer, Z. Rollins, J. A. Kane, K. W. Dodd, K. E. Meyers, *et al.*, "Ischemic ST-segment depression maximal in V1–V4 (versus V5–V6) of any amplitude is specific for occlusion myocardial infarction (versus nonocclusive ischemia)," *Journal of the American Heart Association*, vol. 10, no. 23, 2021. DOI: 10.1161/JAHA.121.022866.

[118] A. Sharma, D. F. Miranda, H. Rodin, B. A. Bart, S. W. Smith, and G. R. Shroff, "Interobserver variability among experienced electrocardiogram readers to diagnose acute thrombotic coronary occlusion in patients with out of hospital cardiac arrest: Impact of metabolic milieu and angiographic culprit," *Resuscitation*, vol. 172, pp. 24–31, 2022, ISSN: 0300-9572. DOI: 10.1016/j.resuscitation.2022.01.005.

[119] R. E. Gregg, T. Yang, S. W. Smith, and S. Babaeizadeh, "ECG reading differences demonstrated on two databases," *Journal of Electrocardiology*, vol. 69, pp. 75–78, 2021, ISSN: 0022-0736. DOI: 10.1016/j.jelectrocard.2021.09.005.

[120] D. A. Cook, S.-Y. Oh, and M. V. Pusic, "Accuracy of physicians' electrocardiogram interpretations: A systematic review and meta-analysis," *JAMA Internal Medicine*, vol. 180, no. 11, pp. 1461–1471, Nov. 2020, ISSN: 2168-6106. DOI: 10.1001/jamainternmed.2020.3989.

[121] P. O. Baker, S. R. Karim, S. W. Smith, H. P. Meyers, A. E. Robinson, I. Ibtida, R. M. Karim, G. A. Keller, K. A. Royce, and M. A. Puskarich, "Artificial intelligence driven prehospital ECG interpretation for the reduction of false positive emergent cardiac catheterization lab activations: A retrospective cohort study," *Prehospital Emergency Care*, vol. 0, no. 0, pp. 1–9, 2024. DOI: 10.1080/10903127.2024.2399218.

[122] J. Choi, V. Torelli, S. Diaz, E. Vaish, L. Katic, A. Nagourney, Z. Khan, A. Silverman, and S. Farhan, "Abstract 4144910: AI-enhanced recognition of occlusion in acute coronary syndrome (AERO-ACS): A retrospective review," *Circulation*, vol. 150, no. Suppl_1, A4144910–A4144910, 2024. DOI: 10.1161/circ.150.suppl_1.4144910.

[123] S. Harris, F. Brown, A. May, and R. Bach, "Abstract 4144603: Artificial intelligence tool accurately predicts occlusion myocardial infarction and may reduce false-positive cath lab activations," *Circulation*, vol. 150, no. Suppl_1, A4144603–A4144603, 2024. DOI: 10.1161/circ.150.suppl_1.4144603.

[124] S. R. Karim, H. C. Helseth, P. O. Baker, G. A. Keller, H. P. Meyers, R. Herman, and S. W. Smith, *Artificial Intelligence Detection of Occlusive Myocardial Infarction from Electrocardiograms Interpreted as "Normal" by the Conventional Algorithm*, Feb. 2025. DOI: 10.20944/preprints202502.0281.v1.

[125] L. E. Dodd and M. S. Pepe, "Partial AUC estimation and regression," *Biometrics*, vol. 59, no. 3, pp. 614–623, Aug. 2003, ISSN: 0006-341X. DOI: 10.1111/1541-0420.00071.

[126] A. R. Khan, H. Golwala, A. Tripathi, A. A. Bin Abdulhak, C. Bavishi, H. Riaz, V. Mallipedi, A. Pandey, and D. L. Bhatt, "Impact of total occlusion of culprit artery in acute non-ST elevation myocardial infarction: A systematic review and meta-analysis," *European Heart Journal*, vol. 38, no. 41, pp. 3082–3089, Nov. 2017, ISSN: 0195-668X, 1522-9645. DOI: 10.1093/eurheartj/ehx418.

[127] C.-S. Hung, Y.-H. Chen, C.-C. Huang, M.-S. Lin, C.-F. Yeh, H.-Y. Li, and H.-L. Kao, "Prevalence and outcome of patients with non-ST segment elevation myocardial infarction with occluded "culprit" artery – a systemic review and meta-analysis," *Critical Care*, vol. 22, no. 1, p. 34, Feb. 2018, ISSN: 1364-8535. DOI: 10.1186/s13054-018-1944-x.

[128] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," *Machine Learning*, vol. 113, no. 7, pp. 4845–4901, Jul. 2024, ISSN: 1573-0565. DOI: 10.1007/s10994-022-06268-8.

[129] Z. Koscova, Q. Li, C. Robichaux, V. Moura Junior, M. Ghanta, A. Gupta, J. Rosand, A. Aguirre, S. Hong, D. E. Albert, *et al.*, "The Harvard-Emory ECG Database," *medRxiv*, pp. 2024–09, 2024.

[130] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, *et al.*, *On the opportunities and risks of foundation models*, 2022. arXiv: 2108.07258. [Online]. Available: https://arxiv.org/abs/2108.07258.

[131] J. Li, A. Aguirre, J. Moura, C. Liu, L. Zhong, C. Sun, G. Clifford, B. Westover, and S. Hong, *An electrocardiogram foundation model built on over 10 million recordings with external evaluation across multiple domains*, 2024. arXiv: 2410.04133 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2410.04133.

Typeset March 13, 2025. No large language models were involved in the creation of any part of this thesis.

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

Predicting adverse cardiac events at the emergency department

If artificial intelligence was a taxonomy of living things, machine learning might cover the animal kingdom, and this thesis would be concerned with the entomological study of a particular species of cranefly. There are over 15000 documented species of craneflies, divided into more than 500 genera. The cranefly of this thesis is the artificial neural network, and the entomological study explore ways to apply modern deep-learning algorithms for predicting adverse cardiovascular outcomes for patients with chest pain at the emergency department.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2025:38
ISBN 978-91-8021-691-3
ISSN 1652-8220