Risk assessment in suspected acute coronary syndrome with decision support tools and artificial intelligence

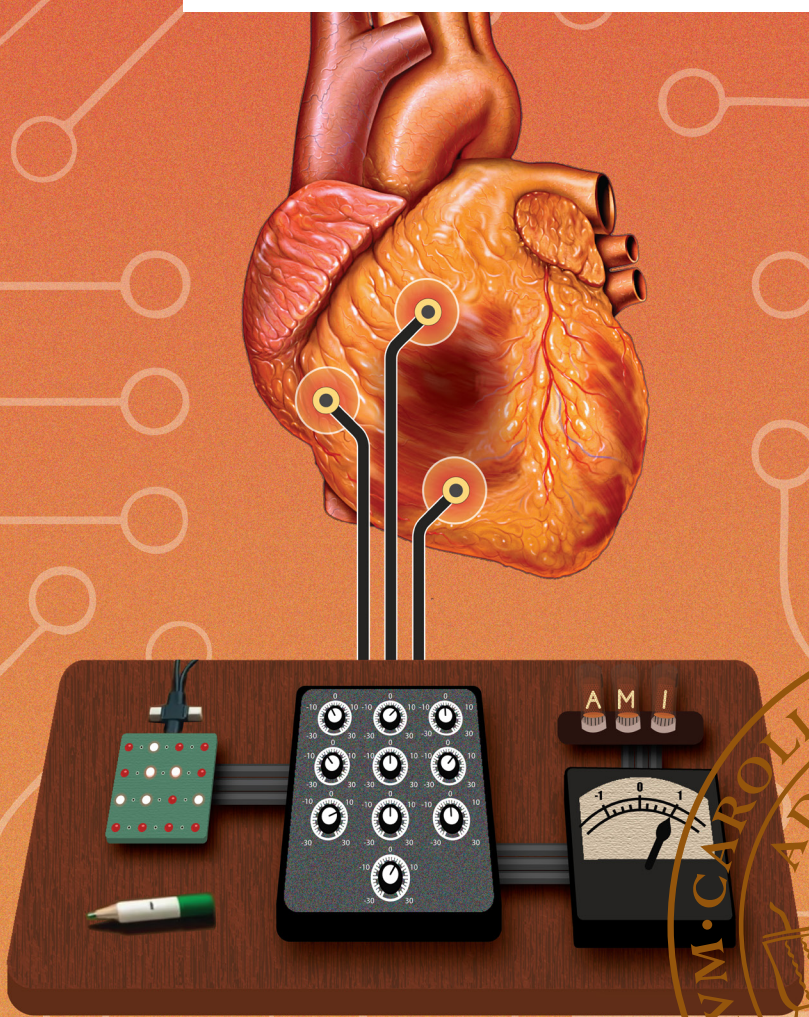
Olsson de Capretz, Pontus

2025

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

Risk assessment in suspected acute coronary syndrome with decision support tools and artificial intelligence

PONTUS OLSSON DE CAPRETZ
FACULTY OF MEDICINE | LUND UNIVERSITY



PONTUS OLSSON DE CAPRETZ is a physician specialized in emergency medicine, a student of electrical engineering and a proud father. He believes that the unexplored areas between disciplines is where the truly exciting stuff happens.

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

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2025:43
ISBN 978-91-8021-696-8
ISSN 1652-8220

Risk assessment in suspected acute coronary syndrome with decision support tools and artificial intelligence

Pontus Olsson de Capretz

LUNDS UNIVERSITET

DOCTORAL DISSERTATION
Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on the 2nd of May 2025 at 13.00 in Lecture Hall 5 at Skåne University Hospital, Entrégatan 7, Lund

Faculty opponent
Professor Johan Sundström, MD, PhD
Department of Medical Sciences, Clinical Epidemiology, Uppsala

Main supervisor
Professor Ulf Ekelund, MD, PhD
Lund University, Department of Clinical Sciences, Skåne University Hospital, Lund, Department of Emergency Medicine

Organization: LUND UNIVERSITY

Document name: DOCTORAL DISSERTATION **Date of issue** May 2nd.2025

Author(s): Pontus Olsson de Capretz **Sponsoring organization:**

Title and subtitle: Risk assessment in suspected acute coronary syndrome with decision support tools and artificial intelligene

Abstract:
Acute coronary syndrome (ACS), including acute myocardial infarction (AMI), are a leading cause of morbidity and mortality worldwide. Accurate risk evaluation in patients presenting with chest pain is essential for optimizing emergency department (ED) resource allocation and patient outcomes. Current clinical decision support tools rely on structured risk scores and biomarker measurements, such as high-sensitivity cardiac troponin T (hs-cTnT), but they may not fully capture the complexity of ACS presentation.

This thesis explores the potential of machine learning (ML) models as decision support tools to improve the early identification of AMI and reduce unnecessary diagnostic procedures.

The research consists of four studies. Study I evaluates whether combining glucose measurements with hs-cTnT improves the 0/1h hs-cTnT protocol in ED patients. Study II investigates theuse of prior electrocardiograms (ECGs) as inputs to ML-models. Study III compares ML models combining ECG and laboratory data against the 0-hour hs-cTnT rule-out protocol. Study IV examines the performance of a sequential ML approach that includes stepwise increasing patient information.

Results indicate that ML models can improve early risk stratification, with convolutional neural networks (CNNs) outperforming traditional logistic regression and rule-based protocols in predicting AMI or death within 30 days. However, prior ECG data provided limited additional value to ML models, and the sequential use of models resulted in a decline in sensitivity. Feature importance analysis showed that troponin and ECG features remain the dominant predictors used by the ML models.

These findings suggest that ML-based decision support could improve ED efficiency by safely reducing unnecessary testing and admissions, but further external validation and implementation research are required before clinical deployment.

Key words:
Chest pain, Emergency Department, Machine learning, Deep learning, ECG, Neural Network, troponin

Language English **Number of pages**: 84

ISSN and key title: 1652-8220

ISBN: 978-91-8021-696-8

Recipient's notes Price Security classification

Signature Date 2025-03-25

Risk assessment in suspected acute coronary syndrome with decision support tools and artificial intelligence

Pontus Olsson de Capretz

LUNDS UNIVERSITET

Cover photo by Pontus Olsson de Capretz:
Image of Heart by Patrick J. Lynch
https://commons.wikimedia.org/wiki/File:Heart_anterior_exterior_view.png
Image of Hans Olivecrona
http://www.medicinhistoriskasyd.se/smhs_bilder/displayimage.php?pid=6459&fullsize=1

Faculty of Medicine
Department of Clinical Sciences

ISBN 978-91-8021-696-8
ISSN 1652-8220

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

Till mormor, för möjligheter som du aldrig fick,
och önskningar som nu gått i uppfyllelse

Table of Contents

List of papers

Paper I
Glucose and high-sensitivity troponin T predict a low risk of major adverse cardiac events in emergency department chest pain patients.
de Capretz. PO, Khoshnood, A., Mokhtari, A., & Ekelund, U. (2021). Scandinavian Cardiovascular Journal, 55(6), 354–361.
https://doi.org/10.1080/14017431.2021.1987512

Paper II
Prior electrocardiograms not useful for machine learning predictions of major adverse cardiac events in emergency department chest pain patients.
Nyström A, **de Capretz PO**, Björkelund A, Lundager Forberg J, Ohlsson M, Björk J, Ekelund U.
J Electrocardiol. 2024 Jan-Feb;82:42-51.
doi: 10.1016/j.jelectrocard.2023.11.002. Epub 2023 Nov 20. PMID: 38006763.

Paper III
Machine learning for early prediction of acute myocardial infarction or death in acute chest pain patients using electrocardiogram and blood tests at presentation.
de Capretz PO, Björkelund A, Björk J, Ohlsson M, Mokhtari A, Nyström A, Ekelund U.
BMC Med Inform Decis Mak. 2023 Feb 2;23(1):25.
doi: 10.1186/s12911-023-02119-1. PMID: 36732708; PMCID: PMC9896766.

Paper IV
Stepwise increasing input to machine learning models predicting 30-day AMI or death in emergency department chest-pain patients.
de Capretz PO, Nyström A, Björkelund A, Björk J, Ohlsson M, Ekelund U.

Manuscript awaiting submission to medical journal

Abstract

Acute coronary syndrome (ACS), including acute myocardial infarction (AMI), are a leading cause of morbidity and mortality worldwide. Accurate risk evaluation in patients presenting with chest pain is essential for optimizing emergency department (ED) resource allocation and patient outcomes. Current clinical decision support tools rely on structured risk scores and biomarker measurements, such as high-sensitivity cardiac troponin T (hs-cTnT), but they may not fully capture the complexity of ACS presentation.

This thesis explores the potential of machine learning (ML) models as decision support tools to improve the early identification of AMI and reduce unnecessary diagnostic procedures.

The research consists of four studies. Study I evaluates whether combining glucose measurements with hs-cTnT improves the 0/1h hs-cTnT protocol in ED patients. Study II investigates the use of prior electrocardiograms (ECGs) as inputs to ML-models. Study III compares ML models combining demographics, ECG- and laboratory data against the 0-hour hs-cTnT rule-out protocol. Study IV examines the performance of a sequential ML approach that includes stepwise increasing patient information.

Results indicate that ML models can improve early risk stratification, with convolutional neural networks (CNNs) outperforming traditional logistic regression and rule-based protocols in predicting AMI or death within 30 days. However, prior ECG data provided limited additional value to ML models, and the sequential use of models resulted in a decline in sensitivity. Feature importance analysis showed that troponin and ECG features remain the dominant predictors used by the ML models.

These findings suggest that ML-based decision support could improve ED efficiency by safely reducing unnecessary testing and admissions, but further external validation and implementation research are required before clinical deployment.

Abbreviations

ACS	Acute Coronary Syndrome
AI	Artificial Intelligence
AMI	Acute Myocardial Infarction
ANN	Artificial Neural Network
AUROC	Area Under the Receiver Operating Characteristic Curve
CABG	Coronary Artery Bypass Grafting
CDSS	Clinical Decision Support System
CNN	Convolutional Neural Network
cTn	Cardiac Troponin
cTnI	Cardiac Troponin I
cTnT	Cardiac Troponin T
ECG	Electrocardiogram
ED	Emergency Department
EDACS	Emergency Department Assessment of Chest pain Score
ESC	European Society of Cardiology
GDPR	General Data Protection Regulation
GRACE	Global Registry of Acute Coronary Events
HEART	History, Electrocardiogram, Age, Risk factors, and Troponin
hs-cTnT	High-Sensitivity Cardiac Troponin T
ICD	International Classification of Diseases
ML	Machine Learning
MACE	Major Adverse Cardiac Events
MLP	Multilayer Perceptron
NLR	Negative Likelihood Ratio
NN	Neural Network
NPV	Negative Predictive Value
NSTEMI	Non-ST-Elevation Myocardial Infarction
OMI	Occlusion Myocardial Infarction

PCA	Principal Component Analysis
PCI	Percutaneous Coronary Intervention
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Value
RCT	Randomized Controlled Trial
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
ROC	Receiver Operating Characteristic
SA	Sinoatrial
SEM	Skåne Emergency Medicine
SHAP	Shapley Additive Explanations
STEMI	ST-Elevation Myocardial Infarction
TIMI	Thrombolysis in Myocardial Infarction
UA	Unstable Angina
USMLE	United States Medical Licensing Examination

Popular Science Summary

Inom varje människa finns en muskel vars absoluta syfte är att kontinuerligt förse kroppen med näringsämnen, syre och mineraler. Samma maskin används också för att dra ut slaggprodukter för transport ut ur kroppen. Vi kallar denna apparat för hjärtat. Varje dag slår hjärtat runt hundra tusen slag, och under en livstid nästan 300 miljoner slag. Det är alltså en muskel som ska hålla länge.

Ibland tar hjärtat stryk. En tryckande känsla i bröstet, som strålar ut mot vänster arm är klassiska tecken på en hjärtinfarkt. Just hjärtinfarkter är för tillfället en av de vanligaste orsakerna till död och lidande i världen och i Sverige. Tekniskt sett kan en hjärtinfarkt drabba vem som helst, men vissa faktorer är starkt knutna till en ökad risk. Ålder och kön, tidigare sjukdomar som högt blodtryck och diabetes är viktig information att inhämta om en hjärtinfarkt misstänks.

Under många år har vi varit mycket duktiga på att informera samhället om hjärtinfarktens klassiska kännetecken. Så duktiga att bröstsmärta nu är en av de vanligaste orsakerna till besök på landets akutmottagningar. Men lyckligtvis har de flesta med bröstsmärta inte en hjärtinfarkt. Här finns det ännu en utmaning. En komplett utredning för att utesluta en hjärtinfarkt innefattar mätning av hjärtats elektriska aktivitet med ett elektrokardiogram (EKG) och halten av troponiner i blodet. Troponiner är proteiner som finns i hjärtmuskeln och vars värde stiger kraftigt i blodet vid hjärtinfarkt.

Utöver detta sker eventuellt en bedömning av hjärtat vid ansträngning eller en angiografi, en form av röntgen där hjärtats egna blodkärl kan undersökas för tecken på trånga förhållanden. Dessa undersökningar kan både göra ont och medföra komplikationer så som allergiska reaktioner eller skador på blodkärlen. Så helst vill vi undvika att undersöka i onödan.

Med hjälp av provsvar, symptombeskrivning och kända riskfaktorer har läkare under lång tid skapat olika system för att kunna bedöma risk för allvarlig sjukdom. Kriterier och träddiagram har utarbetats för att lättare kunna sortera patienter i grupper som baserat på sannolikheten för hjärtinfarkt.

Användandet av artificiell intelligens och maskininlärning har bokstavligen exploderat under det senaste årtiondet och används nu i allt fler branscher. Även inom sjukvården börjar sådana metoder få fäste, men implementeringen går långsamt. Detta beror sannolikt på flera faktorer, men viktigast är behovet av system som är både säkra, effektiva och pålitliga.

Och det är just här denna forskning kommer in. I ett flertal arbeten har vi undersökt möjligheten att säkert utesluta hjärtinfarkt i olika skeden, helst så tidigt i processen

som möjligt. Tanken är att det kan finnas grupper av patienter där sannolikheten för hjärtinfarkt är så låg att det är större risk att utreda dem än att låta bli.

I studie I undersökte vi om det finns ett tilläggsvärde av blodsocker till de regelbaserade system som används i den kliniska vardagen idag. Studien visade att en av fyra patienter kunde skickas hem med mycket låg risk för hjärtinfarkt om både troponinvärde och blodsocker var normala.

I studie II introducerade vi flera ai-modeller och upptäckte att deras prestanda inte förbättrades när vi gav dem tillgång till patienters tidigare EKG, något kliniker ofta använder vid bedömning av huruvida EKG-förändringar är nya.

Studie III bygger på resultaten från de två tidigare studierna. Här lät vi modeller lära sig sambandet mellan hjärtinfarkt och patienters ålder, kön, EKG och flertalet blodprovsvärden. Vi testade modeller med olika tillgång till data och komplexitet och såg att mer avancerade modeller med tillgång till mer data kunde överträffa resultaten från våra vanliga regelbaserade system.

I Studie IV lade vi till en tidsmässig komponent. Vi utgick från att en del patienter från början hade mycket låg risk för hjärtinfarkt och att dessa potentiellt kunde undvika provtagning och EKG. Vi byggde därför modeller för flera steg i beslutsprocessen, där de första enbart hade tillgång till patenters ålder och kön, medan den sista modellen hade tillgång till all information om patienten upp till det första troponinvärdet. Resultaten från studien pekar på att patienter med låg risk för hjärtinfarkt kan identifieras väldigt tidigt, och att troponin och EKG är de klart dominerande komponenterna vid modellernas bedömning av risk för hjärtinfarkt.

Sammantaget tyder resultaten från dessa studier på att det finns en betydande del av patienter med bröstsmärta där hjärtinfarkt kan uteslutas på ett säkert sätt i ett tidigt skede och att onödiga procedurer därmed kan undvikas.

Preface

The seeds of this work were planted in 2015, during a traditional Swedish fika on a cold autumn afternoon. A dear friend of mine asked me: *"Pontus, did you know that doctors are wrong in one out of ten cases? One day, computers will replace doctors. And patients will be better off for it."*

The implications of that statement were both startling and exhilarating. I knew he was onto something. As a systems engineer by training, he had extensive experience in decision-making and understood the power of automated systems. I asked him, *"What would you do in my shoes?"*

The result of that question can be found in the pages of this thesis. With little to no knowledge of computer science, programming, or machine learning, I set out on a challenging path filled with frustration, setbacks, and the occasional precious *Aha!* moment.

Machine learning is not an easy field to enter without prior knowledge or formal training. All the math I had learned in high school had long been forgotten. Even the simple act of dividing fractions seemed to have vanished from my memory. Say what you will about the medical profession, but mathematical prowess is not a decisive factor for success.

When I finally began to grasp the fundamental mathematics behind these algorithms, I saw the beauty in their simultaneous simplicity and power. I fell in love with the idea that a simple derivative, the concept that high school students always question, asking *"When will I ever need to know this?"*, could be used to train a complex system in the same way one learns to ride a bicycle. Using a machine trained on examples, rather than painstakingly hardcoding rules, felt like steering a bike without using your hands, and it still amazes me to this day.

The five years I've spent working on this project have been some of the most exciting and challenging of my life. Becoming a father twice during this time has only made life more interesting. This thesis stands as proof that it is possible to learn something entirely new, that life will sometimes get in the way, but that, with persistence, you can still reach the finish line.

AI Statement

The author of this dissertation has used OpenAI's ChatGPT-4o for translation purposes and language quality improvements. All text has subsequently been reviewed and edited to ensure that the authors points of view have not been changed by the translation process.

AI is not about replacing us,
but making us better versions of ourselves.

- Rana el Kaliouby

Introduction

Part 1: The Problem

Acute coronary syndrome (ACS) including acute myocardial infarction (AMI) is one of the leading causes of death worldwide. [1] Chest pain, the classic symptom of an AMI, has been widely publicized through information campaigns and is now well known among the general population. In Swedish emergency departments (EDs), chest pain is one of the most common presentations, second only to abdominal pain. [2] However, most patients with chest pain do not have ACS.

Diagnosing AMI often requires invasive tests such as coronary angiograms, which carry a risk of severe complications in approximately one in a thousand patients. [3]

Subjecting many patients to unnecessary diagnostics also risks using too many resources, leading to backlogs and crowding in the ED, a condition that is associated with increased mortality rates for patients. [4,5]

Selecting the correct patient for admission and invasive procedures thus becomes an optimization problem, both in terms of reducing unnecessary risks to patients, and for improving ED logistics.

The search for methods to rapidly rule out ACS and AMI have been ongoing for decades. [6] Today, machine learning methods hold a potential to significantly improve the performance of ED decision makers, both in terms of correct diagnostics and in making decisions at earlier timepoints.

This thesis explores the use of machine learning methods as decision support tools for the evaluation of chest pain patients in the ED.

Cardiac anatomy and physiology

The heart is in essence a set of two muscular pumps. It contains four chambers: two ventricles responsible for pumping blood throughout the body, and two atria responsible for pre-filling the ventricles before their contractions. The right side of the heart pumps deoxygenated blood into the pulmonary circulation for oxygenation. This blood subsequently returns to the left side of the heart, which delivers oxygenated blood to the aorta and the systemic circulation. This circulation

is completed when deoxygenated blood returns to the right atrium through the vena cava.

The myocardium, responsible for cardiac contractions, receives oxygenated blood through the coronary arteries, which are the first arteries to branch off from the aorta.

The contractile effort of the heart is governed by the sinoatrial (SA) and atrioventricular (AV) nodes, which along with the his-purkinje fibre system generate and transmit electrochemical impulses to the myocardial cells of the atria and ventricles, thus setting the pace of myocardial contractions. [7]

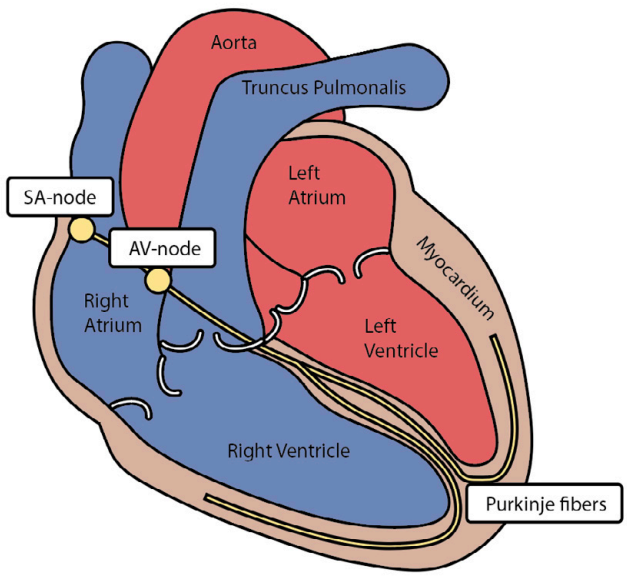


Figure 1: Anatomy of the human heart

Examining the Heart: History, Biomarkers and Diagnostic Tests

Patient history and symptoms

A fundamental step in evaluating patients presenting with chest pain involves obtaining a thorough patient history and physical examination. Patients experiencing AMI are often diaphoretic, anxious and typically describe central, compressive chest pain radiating to the left arm, back, or neck, commonly known as angina pectoris. However, myocardial infarctions can also present atypically with symptoms such as shortness of breath or abdominal pain.

An assessment of the patient's medical history is crucial since certain demographic and health-related risk factors increase the likelihood of AMI. Notable risk factors include diabetes, hypertension, hypercholesterolemia, smoking, and a family history of myocardial infarction in a close relative, such as a parent or sibling, occurring before the age of 65.

Physical evaluation includes palpation of the thorax and upper abdomen, with resulting pain indicating non-cardiac causes of chest pain.

Despite the importance of patient history, physical and known risk factors, relying solely on these aspects is typically insufficient for conclusively identifying or ruling out myocardial infarction. Therefore, objective data regarding cardiac function is gathered through two essential diagnostic tests: the electrocardiogram (ECG) and measurement of cardiac troponins in the blood. [8–11]

ECG

Invented by Einthoven in 1923, the electrocardiogram (ECG) is a non-invasive method to measure the electrical activity of the heart. This is done by placing electrodes on the skin and measuring the electrical potential between two points.

Einthoven's work led to the development of the 12-lead ECG, which is the standard ECG format used today. [12]

A normal ECG waveform has three distinct phases. The P-wave represents atrial depolarisation and the QRS-complex ventricular depolarisation while the T-wave represents ventricular repolarization. [7]

Typical ECG findings associated with AMI include ST-segment elevation (suggesting ST-elevation myocardial infarction, STEMI), ST-segment depression and/or T-wave inversions, and pathological q-waves (a sign of prior infarction). [13]

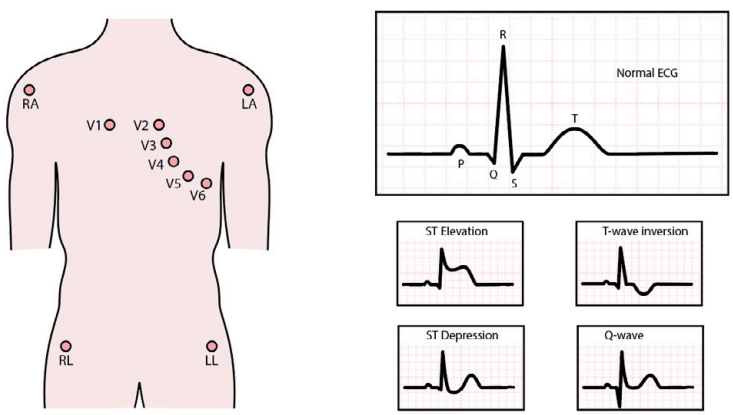


Figure 2: Electrode placement and typical ECG changes during an AMI

Cardiac biomarkers: Troponin

Cardiac Troponins (cTn) are proteins found in high concentrations as an integral part of the myocardial cell contractile apparatus. Troponin I (cTnI) is the inhibitory subunit that prevents contraction in the absence of calcium while Troponin T (cTnT) is the tropomyosin-binding subunit that anchors the complex to the contractile apparatus. cTnI and cTnT have clinical value for the diagnosis of AMI. cTnT is the biomarker used in the studies included in this thesis. [14]

cTnT receives its diagnostic value by being released into the bloodstream when cardiomyocytes are injured, and elevated circulating levels of cTnT in the bloodstream can thus be used as a marker for damage to these cells.

The first cTnT assay was developed in the 1980s by German researchers, and high sensitivity cardiac Troponin T (hs-cTnT) assays have later been introduced as a more sensitive method for detecting myocardial damage. [15]

Troponin levels typically rise within a few hours and peak around 10-50 hours after an infarction, depending on whether blood flow to the infarcted area was restored. [16] Myocardial infarction is not the only reason for elevated troponin levels, however, with myocarditis, tachyarrhythmias, pulmonary embolism, heart failure and impaired kidney function as common differential diagnoses. [17] There are also situations where proteins released after skeletal musculature injury can be erroneously detected by hs-cTnT assays. [13]

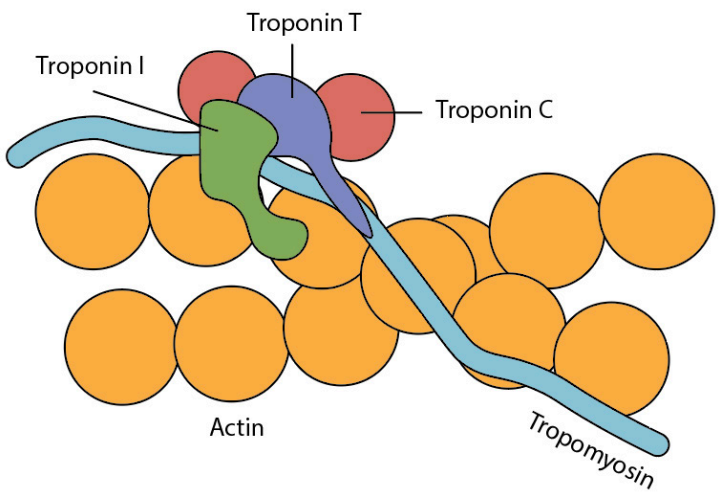


Figure 3: The troponin complex

Point of care tests

Creatinine

Creatinine is a breakdown product of creatine phosphate, a compound used by muscles for energy. It is released into the bloodstream and filtered out by the kidneys. Measuring creatinine levels in the blood is commonly used to assess kidney function and elevated creatinine levels may indicate impaired renal function. In patients with kidney failure, elevated hs-cTnT levels has traditionally been thought secondary to impaired renal clearance but there is likely a multifactorial relationship between kidney function and cardiac injury. [18–20]

Glucose

Glucose is a simple sugar (monosaccharide) and an essential source of energy. Blood glucose levels are tightly regulated by hormones such as insulin and glucagon. Elevated glucose levels may indicate diabetes mellitus, but some studies have also linked high glucose levels to an increased risk of myocardial infarction. [21,22]

Hemoglobin

Hemoglobin is a protein found in red blood cells, responsible for transporting oxygen from the lungs to the body's tissues and carrying carbon dioxide back to the lungs for exhalation. It has been suggested as a predictor for AMI with a potential rationale being that low hemoglobin levels reduce the amount of oxygen delivered to the myocardium. [23]

Defining acute coronary syndrome and myocardial infarction

AMI occurs when cardiomyocytes perish as a result of lack of flow of oxygenated blood to the myocardium. By definition, there are five types of AMI of which type 1 is the primary concern of this thesis.

Type 1 AMI is caused by acute atherothrombotic plaque rupture or ulceration leading to coronary artery occlusion while type 2 AMI occurs due to an imbalance between myocardial oxygen supply and demand such as in patients with anemia.

Type 3 AMI is sudden cardiac death with presumed AMI while types 4 and 5 are related to percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG), respectively.

According to the fourth international definition of myocardial infarction the diagnosis of type 1 AMI requires: [13]

1. Evidence of myocardial injury. This is detected as elevated levels of cardiac biomarkers such as hs-cTnT in the blood stream, with a rising or falling

pattern on serial testing. Myocardial damage takes time to develop, and the rise of hs-cTnT can be slow during the initial hours following an infarction.

2. Clinical- or ECG signs consistent with ischemia, or imaging evidence of infarction.

Acute coronary syndrome (ACS) includes three conditions related to reduced blood flow to the heart: ST-elevation myocardial infarction (STEMI), non-ST-elevation myocardial infarction (NSTEMI), and unstable angina (UA).

The distinction between STEMI and NSTEMI is based on the characteristic ECG findings giving these conditions their names, while UA is diagnosed based on clinical presentation and the absence of significant troponin elevation.

The distinction between STEMI and NSTEMI plays an important role when deciding optimal patient treatment, as patients with STEMI are directed to the angiography suite for immediate revascularization, while NSTEMI patients initially are given medical treatment. [13]

While STEMI and NSTEMI are well-established categories of ACS, the classification of unstable angina (UA) is more difficult. UA is defined as clinical signs of myocardial ischemia at rest or with little exertion combined with normal troponin levels. [24] Despite the introduction of high-sensitivity troponin assays, which have improved the detection of small myocardial injuries, UA continues to be a frequently assigned diagnosis. [25] This complicates evaluation as troponin values by definition have very little diagnostic value for this condition, and these patients are often admitted and undergo cardiac stress testing procedures.

Simply ruling out ongoing AMI or ACS is seldom enough for clinicians. The term Major Adverse Cardiac Events (MACE) has been used in research to define a range of consequences of an AMI. Common components of MACE include, AMI, death, stroke, need for percutaneous coronary intervention (PCI) and heart failure. [26,27]

Studies I and II use MACE as the target outcome, while studies III and IV use 30-day AMI or death.

Treatment of ACS

For patients with STEMI, the preferred treatment is primary percutaneous coronary intervention (PCI), which involves mechanically restoring blood flow through a catheter-based procedure. PCI should be performed as soon as possible, ideally within 90 minutes of first medical contact. [28]

Patients with ACS typically also receive antiplatelet medications to prevent clot formation and often statins, betablockers and ACE-inhibitors, in order to minimize risks of further infarctions and for complication reduction. [24]

Past procedures and medications might therefore contain information about past diseases such as AMI and might possibly even be protective factors when determining the risk of future AMI. This is explored in study IV.

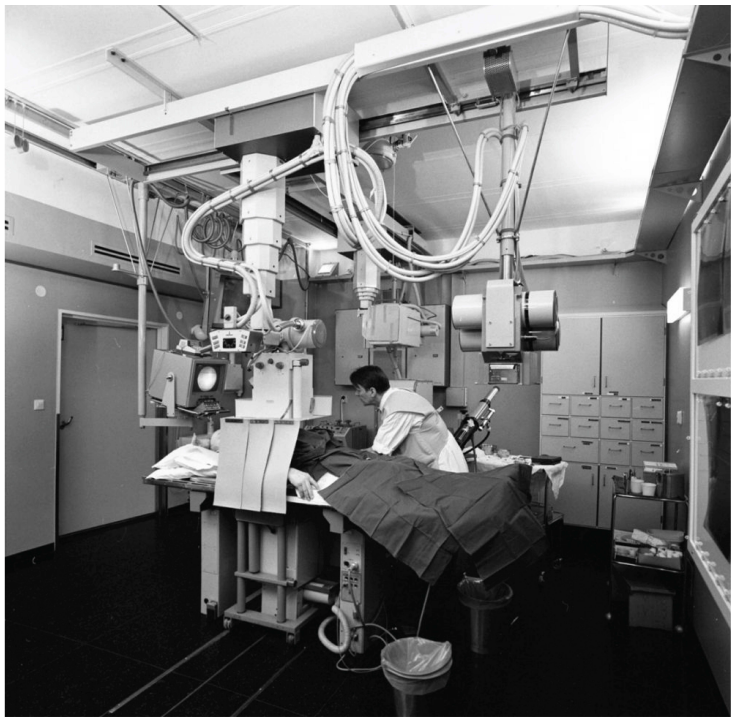


Figure 4: Hans Olivecrona angiography, Lund 1969

Part 2: The Setting

Epidemiological features of chest pain in the ED

In 2023, nearly two million patients visited Swedish EDs with a median length of stay exceeding four hours. Region Skåne reported some of the longest waiting times, with a median ED stay of almost five hours, and waiting times of more than 60 minutes before a patient was being seen by a physician. [29] This is worrying, as AMI is a time-sensitive condition where prompt care can reduce the effects of the infarction, leading to vastly different outomes down the line.

Chest pain has a wide variety of causes, from benign conditions such as musculoskeletal pain and gastroesophageal reflux to severe conditions such as AMI,

pulmonary emboli and acute aortic dissections. Data from 2006 in Lund showed that only 7.5% of patients presenting to the ED with chest pain had ACS. [30] This trend has been relatively stable despite changes in lab analyses and diagnostic criteria, known as diagnostic drift. In 2017, the prevalence of AMI in two hospitals in Skåne were 6-7%. [31]

A similar pattern can be seen internationally, with 12% of Norwegian chest pain patients and 11% of patients in a Spanish sample having ACS. [32,33] In a Tunisian sample, 22% had ACS [34] while a large systematic review by Fanaroff described a median rate of ACS around 13-14% across multiple studies. [9]

Admission rates of chest pain patients in Region Skåne have fallen from almost 50% in 2006 to 30-33% in 2018 while the proportion of admitted patients with AMI has increased slowly from 11% in 2008 to 15 - 20% in 2018. [30,31]

The diagnostic evaluation of ACS among chest pain patients can be a resource-intensive process. An Australian study found that patients diagnosed with ACS had an average hospital stay of 8 days, with a mean cost of $13,509, whereas patients with non-cardiovascular conditions had a shorter average stay of 2 days and a lower cost of $3,331. Reducing unnecessary evaluations and associated costs can be important for optimization of healthcare resources. [35]

The anatomy of an ED

Much like the heart, the ED has a distinct system of flow. Instead of transporting erythrocytes for oxygenation, the currency of an ED is the patient and the diseases the ED intends to manage. Patients with chest pain principally enter the ED in two ways, by walking in or by ambulance. In many places, ambulances with STEMI-patients completely bypass the ED and set off for the angiography suite instead.

Patients in the ED typically start out in the triage area, where a triage nurse or physician get the initial information and perform diagnostic testing including vital parameters, ECG recordings and blood tests. A decision then must be made whether the patient should be accepted into the ED or sent home or to another instance of care, such as a general practicioners office. To standardize this process, many EDs utilize structured triage systems as RETTs (Rapid emergency triage and treatment system). [36]

Critically ill patients are directed to the resuscitation room for stabilization and emergency treatments, while lower acuity patients who require further evaluation are directed to the main ED area for additional workup and finally a decision on whether to discharge them or to admit them to the hospital.

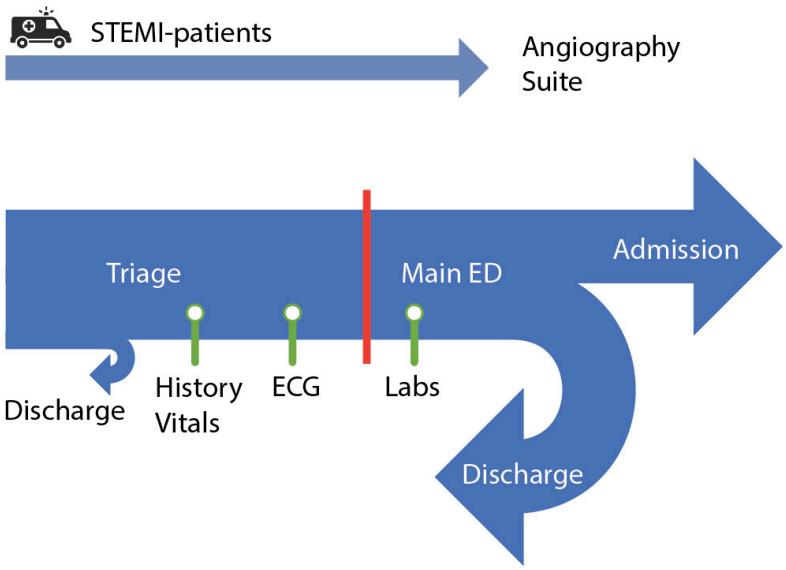


Figure 5: Flow through the ED. Patients with STEMI typically bypass the ED completely.

Decisions in the ED

The ED is often staffed by a combination of junior and senior physicians, as well as nurses with varying levels of experience. The decision-making process in this environment can be thought of as a layered decision network, where different clinicians act as interconnected decision nodes.

Initially, junior doctors evaluate patients, gathering information on symptoms, medical history, and initial test results. This forms the first node of decision-making. Simultaneously, triage personnel contribute by recording vital signs such as blood pressure, heart rate, and oxygen saturation, along with a brief summary of the patient's condition representing another node in the decision network.

If uncertainty remains, particularly in cases of suspected acute coronary syndrome (ACS), a more senior physician is consulted. The senior physician can be thought of as the second layer of decision-making, receiving filtered representations of the patient's condition from the junior physician and triage personnel. Based on all this information, the senior physician determines whether the patient is experiencing a myocardial infarction.

For complex cases, additional steps may be required in the decision-making process, necessitating deeper layers in the network. Specialist cardiologists may be consulted, and in certain cases, the decision might be deferred to a multidisciplinary

conference the following day. In this way, clinical decision-making can extend well beyond the ED and contain many layers.

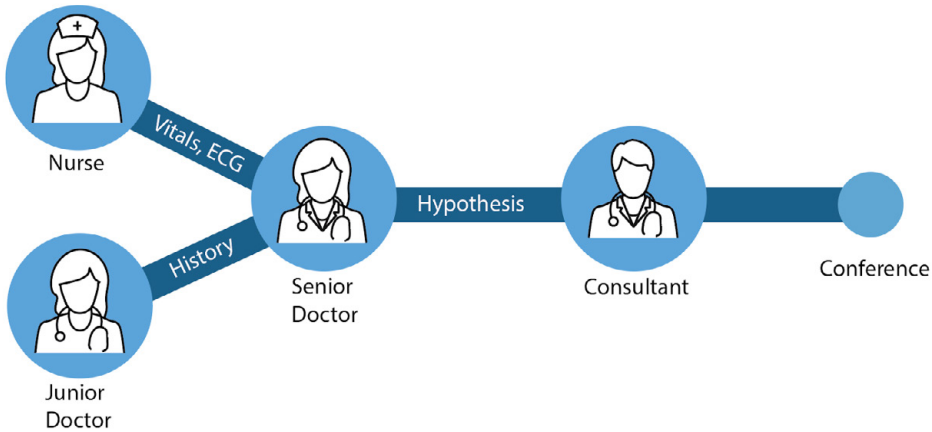


Figure 6: The ED decision network

When new data becomes available, decisions might change. What seemed like a complex problem at the onset, might seem trivial at later stages. This is especially important in the triage area, where important decisions are made with very little information.

Despite structured triage systems, human cognitive biases remain a significant challenge in emergency medicine. Studies suggest that clinicians often struggle with estimating pretest probability, leading to overtesting in some cases and missed diagnoses in others. Indeed, the low prevalence of ACS among admitted patients is a testament to this overestimation of risk. [37,38] Decision-making is further complicated by time constraints and cognitive overload, all of which are inherent to the high-pressure ED environment.

Clinical decision support systems (CDSS) may help mitigate some of these challenges by providing more accurate risk stratification.

Part 3: The Tools

Decision support tools and big data

CDSS:es have long been in use for aiding clinicians make safe decisions.

Interesting work on computerized pretest probability was done by Kline et al, who developed a tree-based protocol matching clinical attributes from a given patient to a database of almost 15 000 patient records with known outcomes and compared it to a logistic regression model. The attribute matching protocol attained a slightly higher AUROC than the logistic regression model and assigned 24% of patients into a low risk category. [39]

For chest pain patients, TIMI (Thrombolysis in Myocardial Infarction), EDACS (Emergency Department Assessment of Chest pain Score), and GRACE (Global Registry of Acute Coronary Events) scores are some of the best-known risk scores used today. All of these were developed using multiple logistic regression techniques on a subset of the ED chest pain population, namely ACS patients for TIMI and GRACE-scores, and patients where chest pain was not obviously non-cardiac, for the EDACS-score. The HEART-score (History, Electrocardiogram, Age, Risk factors and Troponin) deserves a special mention, as it was developed on undifferentiated ED chest pain patients using clinical experience instead of statistical methods. [40–46] Updated protocols such as EDACS-ADP and the HEART pathway continue to improve the performance of these tools. [47–49]

A 0/1-hour combination algorithm using serial troponin testing at 0hrs and 1hrs has recently been developed to safely rule out AMI and cardiac death in the ED. This algorithm stratifies patients as low, intermediate or high risk depending on clinical factors, ECG patterns and serial troponin results, with a sensitivity of over 99% and a positive predictive value of over 70%. It has been validated in several studies, including a randomized controlled trial, and has received a class 1 recommendation in the European Society of Cardiology (ESC) guidelines for the management of acute coronary syndrome. The 0-hour arm of this ESC 0/1h algorithm served as the baseline we compared our models to in study III. [24,31,50–52]

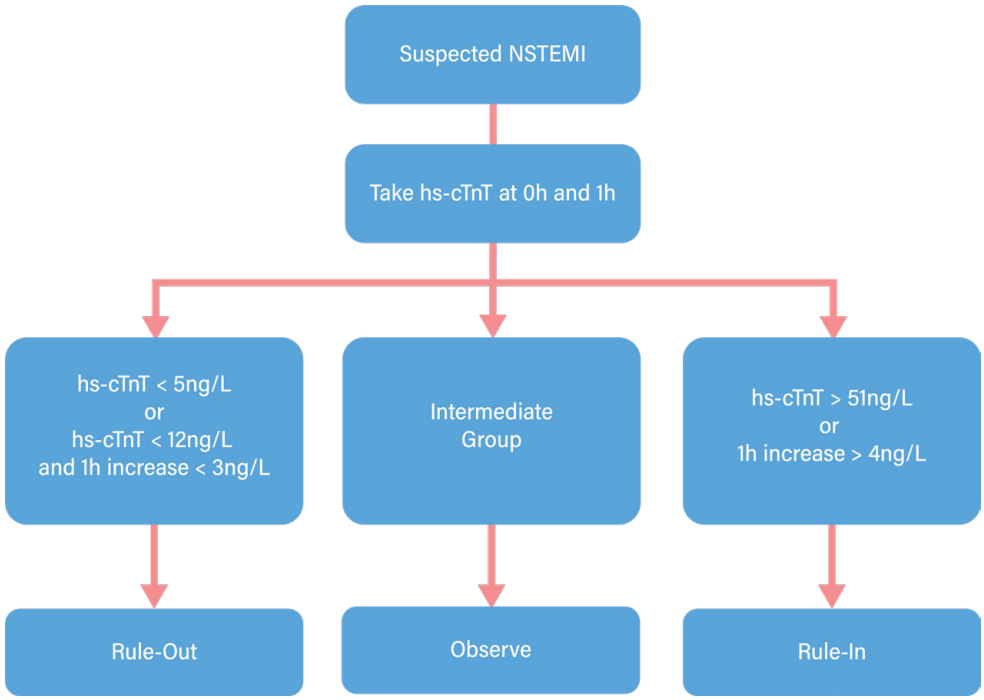


Figure 7: The ESC 0/1 hour algorithm

In recent years, increasingly large datasets have allowed researchers to create new diagnostic tools with superior performance, but a common drawback of such datasets is that they seldom contain information on patient symptoms at presentation. This creates an issue for CDSS creators, as symptom description is a crucial component used by doctors in clinical practice. Nevertheless, important and hitherto unknown clinical patterns could possibly be hiding in these very large datasets. Finding such patterns could help clinicians further risk-stratify patients before important decisions such as invasive angiography. [53,54]

Machine learning tools

Over the past decade, machine learning has exploded as a field, and is now used across a wide range of industries, including healthcare. The primary focus has often been on the diagnostic interpretation of imaging data, particularly radiology images, but automated ECG interpretation has also been studied extensively. [55] One of the strengths of these models is that they can operate in a data-driven manner, free from preconceived notions. Special attention has been given to a specific type of model, artificial neural networks, which have the capability to learn nonlinear relationships.

The first description of neural networks was introduced by Warren McCulloch and Walter Pitts in their 1943 paper A Logical Calculus of the Ideas Immanent in Nervous Activity. [56] Fifteen years later, psychologist Frank Rosenblatt successfully constructed a working model of such a network, which came to be known as the perceptron. [57] Rosenblatt's model demonstrated that neural networks could learn linearly separable patterns. A later publication however, Perceptrons (1969) by Marvin Minsky and Seymour Papert, showed that this type of model was incapable of learning non-linearly separable functions, such as the XOR function. [58] Many years later, researchers discovered that by using multiple perceptrons across several layers, these limitations could be overcome. It was eventually proven that such a multilayer perceptron could approximate any continuous function, given a sufficient number of nodes, an idea now known as the Universal Approximation Theorem. [59]

In 2024, Geoffrey Hinton and John Hopfield received the Nobel Prize "for foundational discoveries and inventions that enable machine learning with artificial neural networks". [60] These breakthroughs, and others, laid the foundations for today's deep neural networks, which feature architectures with hundreds of layers and trillions of parameters.

This thesis has focused on a small subset of the machine learning field, namely the use of artificial neural networks for classification tasks using supervised learning principles.

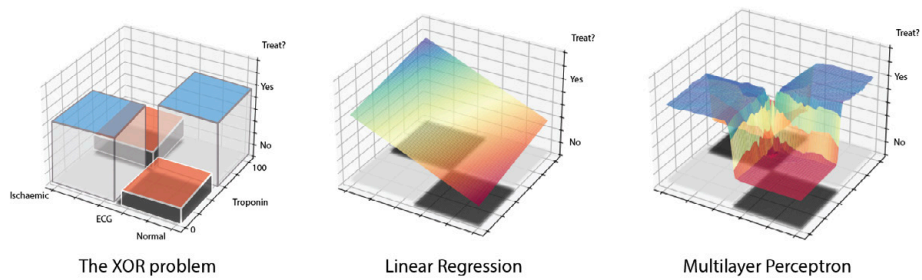


Figure 8: Solving the XOR problem. Consider a situation where we want to offer medical treatment to patients with ischaemic ECG changes **or** high troponin values (white areas). If a patient has both, we instead want to perform immediate angiography, and if the patient has none of these, we don't want to do anything (black areas). The linear regression model cannot create a plane such that all cases above the plane correctly receive treatment, but a multilayer perceptron can.

Logistic Regression and the Multilayer Perceptron

Logistic regression is a method of fitting data to a specific outcome. It can be split up into two parts, namely a linear regression and a logistic, nonlinear transformation. The linear regression procedure consists of adding a specific weight to each input feature, adding a bias term and summing these contributions to generate a final output. The weights and bias are selected such that the line produced by the regression model best fits the data. To create a classification system, the output needs to be transformed from a linear scale into a range between 0 and 1 for predicting the probabilities of the output classes. This is done using the logistic function. [61]

A fully connected multilayer perceptron (MLP) has the same structure as the ED decision network described above. Each node in a layer receives input from all nodes in the previous layer and propagates its output to all nodes in the next layer. Each node in a MLP is in essence a logistic regression model, but nodes in intermediate layers often use functions other than the logistic function, such as ReLU. [62] These activation functions are essential to the performance of the models, as stacking linear nodes on top of each other otherwise do not yield additional benefits. However, these activation functions also complicate the relationships between inputs and outputs, leading to difficulties understanding model predictions. These issues have made neural networks colloquially known as "black box models".

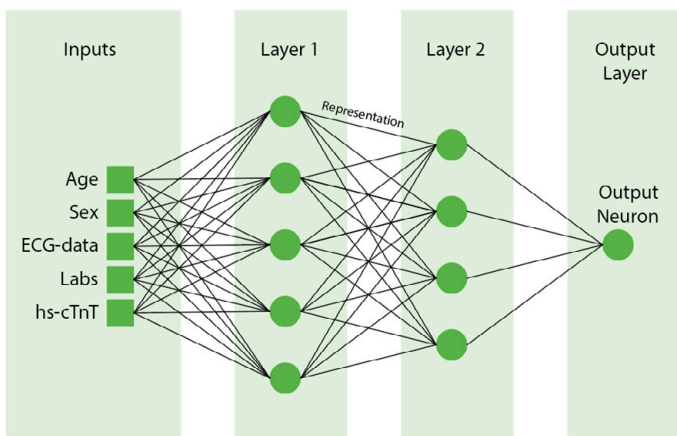


Figure 9: A Multilayer Perceptron. All inputs are fed into each neuron of layer 1. Each neuron then outputs a transformation of these inputs, a representation. All layer 1 representations are then fed into the nodes in the next layer, until the data reaches the output layer, where the classification procedure is performed.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are inspired by the structure and function of the visual cortex in biological systems. In the human brain, neurons in the visual cortex respond to specific regions of an image, known as receptive fields, allowing for the hierarchical processing of visual information. Similarly, CNNs use convolutional layers composed of small, trainable filters that slide over an input image, detecting local patterns such as edges, textures, and shapes. These detected features are then passed through successive layers, progressively capturing more complex structures.

Fundamental milestones in the development of CNN architecture was the Neocognitron developed by Fukushima and the usage of backpropagation in the LeNet developed by LeCun et al in the 1980s. [63,64] Arguably, the current boom in artificial intelligences can be credited to AlexNet, developed by Alex Krizhevsky et al. [65]

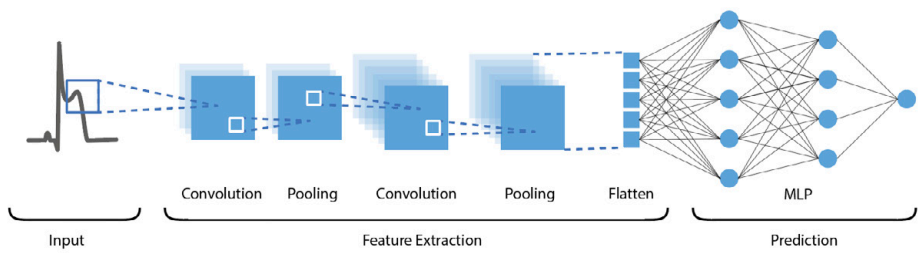


Figure 10: A Convolutional Neural Network

Residual Neural Networks (ResNet)

ResNets are a relatively recent architecture designed to overcome challenges in very deep networks, particularly the vanishing gradient problem. [66] This issue arises when a network has many layers, causing the signal to degrade as it propagates through the network.

As an example, consider an ECG signal that contains noise. In a deep neural network, each layer modifies the input signal slightly, which can ultimately weaken the original signal and make it difficult for the model to distinguish relevant information from noise. ResNets solve this by focusing on learning the residual error rather than the entire output. The problem is essentially turned on its head. Instead of modifying the signal at each layer, the network preserves the original signal and learns only the necessary corrections, allowing deep networks to maintain strong predictive performance.

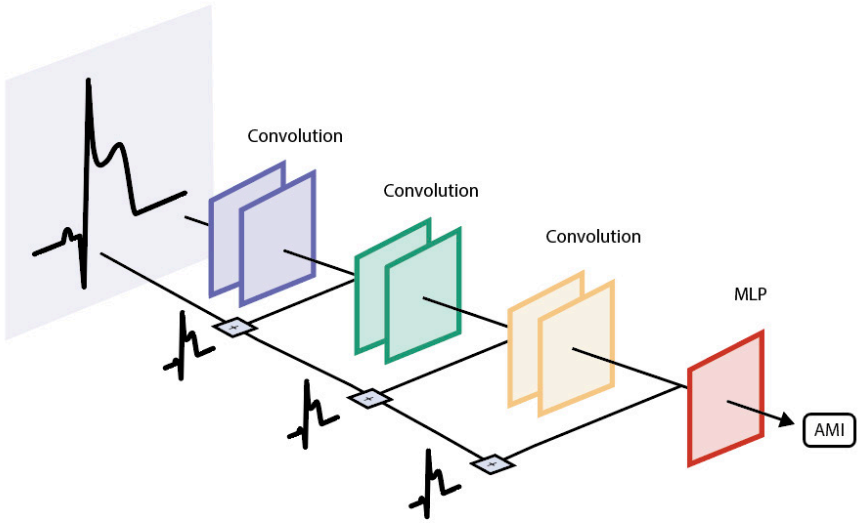


Figure 11: ResNet architecture. The output of each convolution layer group is combined with its input signal.

How networks learn

Supervised learning relies on models learning to map the relationship between a set of input features and a known target outcome or label. The standard approach to supervised learning of neural networks can be split into three concepts: the cost function, its reduction, and its propagation.

As neural networks are created, their weights and biases are usually randomly initialized. The network is then presented with a number of cases and its total prediction error is calculated. The goal is to reduce this error by changing the individual weights and biases. The function describing the error is called a cost function. Cost functions can be calculated in different ways but are in essence a measurement of how far the models' predictions diverge from the true label.

Using the partial derivative of the cost function for a specific weight, we can calculate what direction we should change this weight in order to minimize this function. By moving a small step (termed the learning rate) in the direction opposite this gradient we will have reduced the cost function by a small amount. This is repeated many times until we have reached a bottom of the function, where the gradient is flat. This is the gradient descent algorithm.

Backpropagation is a technique that propagates the error backwards through the network in order to efficiently compute the gradients and update the weights using the gradient descent algorithm. [67]

Figure 12: Gradient descent. During each learning iteration the weight is adjusted slightly in the direction that reduces the model prediction error. When a minimum is found, the gradient approaches zero and the weight stops adjusting.

Transfer learning

Transfer learning is a way of using pretrained models in new ways, by retraining them on similar tasks and leveraging useful concepts already learned by the model. In this way we forgo using random weights and might find deeper local minima than those found using the standard approach.

The earliest record of transfer learning is arguably that by Bozinowski in 1976. [68] This method is increasingly being used for improving model training in the medical field. [69] One interesting study using this approach was performed by Bird et al and used transfer learning techniques for classifications based on electromyography (EMG) and electroencephalography (EEG) data. In this study transfer learning from EMG to EEG data yielded impressive performance increases compared to training models on EEG data alone. [70] Other studies have shown that age and sex can be predicted based on ECG data. In study IV, we used this knowledge to pretrain models to predict age and then retrained these models for AMI prediction. [71,72]

Explainability

As we have moved from linear, shallow models to more complex and deep architectures, the relationship between the input data and output labels have become increasingly hard to comprehend. In the European Union, AI explainability is regulated by the AI Act of 2021. [73] This act designates any AI system used in a medical setting as "high risk" and subjects these to three requirements according to Chapter 2 Articles 12 to 14:

1. That the performance of the system is logged and records are kept.
2. That the models are transparent and present information to users such that outputs can be interpreted and used correctly.
3. There is a human in the decision-loop.

Many systems have been developed to try to solve these issues. [74] A clear differentiation of such systems can be made between ante-hoc and post-hoc systems, where ante-hoc systems are explainable by design, by maintaining linear relationships between input features and the target. Examples include linear regression models and decision trees.

Post-hoc systems are used when models do not lend themselves to easy interpretation. This is where neural networks fit in. Post-hoc systems try to make sense of the results of these black box models by different means, often implementing surrogate models that map the decision outputs of the network sufficiently well, or by other means such as Shapley additive explanations (SHAP) for input feature importance and GRAD-CAM for image section importance. [75–77]

In study IV we used permutation feature importance, which is a comparatively simple technique for showing how much specific variables contributed to the output labels. [78] Initially developed for random forest models, this technique relies on randomly shuffling an input variable to break the relationship between the variable and the outcome. The reduction in model performance on shuffled data is then interpreted as the relative contribution of that variable to the model output.

Related work

Medical ai models are commonly used trying to predict the probability of a target disease. Once a model for prediction has been created, specific thresholds can be used for decision making, such as the rule out of AMI if the probability is below a certain threshold. Performance of these models is rapidly improving, deep neural networks trained on ECG data are already surpassing medical professionals on some diagnostic tasks such as arrhythmia detection. [79]

Table 1: Related work

Author (Year)	Method	Inclusion, n	Ruled out %	Sensitivity %
Lin, 2018	XGBoost	4049	46	90
Than, 2019	Gradient Boosting	11011	69	98
Björkelund, 2020	ANN	5695	57	97
Ibrahim, 2020	CNN	105758	61	93
Zhang, 2020	Random Forest	85254	Not reported	92
Duodensis, 2023	XGBoost	20324	72	98
Neumann, 2023	Multiple	27674	41-68	99 - 100

Using neural networks for predicting ACS goes back at least to the 90s, where Baxt et al used a neural network to diagnose AMI among 331 patients with chest pain. The model outperformed clinicians with a sensitivity of 97% compared to the 78% of the clinicians. [80] ACS prediction using ECG data is a well-researched topic and machine learning methods are reported to have higher sensitivity than clinicians for ACS prediction. [81] Simply using ECG data, however, does not take into consideration age, sex, risk factors and troponin test results. For model acceptance clinicians will likely require a holistic approach where all aspects of a patient's condition are taken into consideration.

Additionally, the choice of model architecture is important. With the drawback of reduced explainability, more complex models need to be significantly more performant to warrant their use.

In a small study in 2005, Harrison et al compared multilayer perceptrons to logistic regression on patients with suspected ACS. Variables included patient symptoms, risk factors and ECG data. Results were excellent, with models attaining AUROC of 0.96-0.98. Interestingly, the logistic regression models performed on par with the neural networks. The impressive results in this study likely relates to a combination of using patient symptoms as input variables and the diagnostic criteria of ACS at the time. [82]

Lin et al developed an XGBoost model on clinical and demographic data but excluding troponin levels in 2018. This model could rule out almost half of all patients of myocardial infarction, but sensitivity was relatively low, 90%. [83]

The question of sensitivity is important. Ideally model sensitivity should not be inferior to current practice, and the risk of consequences of a missed infarction should be lower than the risks associated with clinical workup. A survey by Than et al on acceptable risk of MACE among discharged patients concluded that clinicians accepted a test sensitivity of 99%. [84]

More recently, Than et al developed a gradient boosting algorithm (MI3) incorporating age, sex and serial hs-cTnI samples. This model outperformed the

European Society of Cardiology 0/3-hour pathway in terms of sensitivity and could select 69% of patient as low risk with a sensitivity of 98%. These results were later validated on a larger cohort of 20 000 patients in Scotland. McCord et al further improved on the MI3 algorithm by shortening the timespan of serial testing to 30 minutes, while maintaining similar performance. A study by Björkelund et al used a neural network trained on serial hs-cTnT samples showed similar results. Doudesis et al created a stepwise XGBoost-algorithm that could rule out 61% of patients using a single hs-cTnI sample, and an additional 11% using serial tests. This algorithm was later validated on a cohort from 5 countries. [85–90]

Zhang et al created a range of models for prediction of 30-day AMI or death. Input features included demographic factors, risk factors and troponin results but not ECG data. AUROC values for AMI reached 0.915 for the best performing model, a random forest classifier. [91]

In 2023, Neumann et al from the ARTEMIS (Artificial intelligence in suspected myocardial infarction study) group created super learner models using a combination of logistic regression, gradient boosting, elastic nets and random forests. Models were developed for a range of hs-cTn assays and had variable results, with the best model using serial hs-cTnI selecting 68% of patients for rule out with a sensitivity of almost 99%. [92] The strong performance of logistic regression and tree-based algorithms such as XGBoost suggests that simpler models may be just as effective when trained on tabular data.

Hand-crafted ECG features have traditionally been used as inputs to ML models. This makes models susceptible to the choice of these features. Recently, large ECG-datasets have allowed automated feature extraction techniques to be used. This is one area where neural networks shine. Model architectures such as CNNs and ResNets are able to find important features on their own and as such, are less susceptible to the biases of the model developer. In study II-IV we used such models to automate ECG feature selection, comparing performance to simpler models such as logistic regression in study II. Similar work by Xiao et al used a multimodal ResNet model incorporating ECG-data, age and sex. This model attained an AUROC of 92%. The study did not report numbers ruled out with high sensitivity values, but examination of the ROC-AUC curve reveals that few (if any) patients could be ruled out with a sensitivity above 95%. [93]

Another similar study was performed by Ibrahim et al, who created CNN, RNN, and XGBoost models trained on ECG data, age and sex as input features. Interestingly, the XGBoost model using selected ECG-features outperformed both CNN and RNN models in their work. Two potential flaws in study methodology could explain this. Firstly, minority oversampling was used before patients were split, potentially allocating copied patient cases to both train and test groups, a source of data-leakage. Secondly, extracted ECG features were used for the CNN and RNN models instead of ECG signal data. [94]

Aims

Overall aims

The overall aim of this thesis is to study the application of decision support systems and machine learning algorithms to improve the management of patients with chest pain in the ED.

Study I aims

The aim of this study was to evaluate whether the addition of glucose cutoffs to a rules-based system (ESC 0/1h hs-cTnT protocol) could identify more patients for safe rule out and rule in of MACE within 30 days.

Study II aims

To identify new ECG-changes, clinicians commonly compare the ECG with prior ECGs in the same patient. This study aimed to evaluate the value of prior ECGs to machine learning models predicting 30-day MACE.

Study III aims

In this study, we aimed to compare the performance of various machine learning algorithms against the 0-hour arm of the 0/1h ESC protocol. Our objective was to maximize the number of patients eligible for early rule-in or rule-out decisions for myocardial infarction when the first hs-cTnT result was available.

Study IV aims

Building on all previous work, in this study we aimed to explore if an even earlier rule-out was possible by investigating the predictive performance of machine learning models with stepwise increasing information up to the results of the first hs-cTnT blood test.

Methods

Overview

This thesis includes four diagnostic accuracy studies (I-IV). Study I and II explore the use of specific variables for prediction of AMI, while study III and IV examine more complex models with increasing amounts of inputs.

Table 2 provides an overview of materials and methods for all studies included in the thesis.

Table 2. Overview of studies

Paper	**I**	**II**	**III**	**IV**
Study design	Prospective	Retrospective	Retrospective	Retrospective
Cohort	SCORE	ESC-Trop	EXPECT	SEM
Study period	2013 - 2014	2017-2018	2013 - 2014	2017 - 2018
Participants	1031	19499	9519	40312
Split	None	Chronologic	Chronologic	Geographic
Models	Decision Rules	LogReg, MLP, CNN, ResNet	LogReg, ANN, CNN	ANN, ResNet
Outcome	30d MACE	30d MACE	30d AMI/Death	30d AMI/Death
Metric	N ruled out/in	ROC-AUC	N ruled out/in	ROC-AUC, N ruled out

Setting

All studies were performed using registers of patient visits to hospitals in Region Skåne, southern Sweden. Four databases were used: SCORE, EXPECT, ESC-TROP and SEM.

SCORE

The SCORE-database is based on a prospective observational study performed in the ED of Skånes Universitetssjukhus Lund between february 2013 and april 2014.

Data was collected by research assistants between 9 am and 9 pm during weekdays. Adult patients with chest pain were included if troponins were analyzed as part of patient workup and the patient gave written informed consent to participate. Patients who did not speak Swedish or English and patients with dementia were not included in the study. All patients with ST-elevation myocardial infarctions were excluded.

The database contains data on 1167 patients, including age, sex, previous diseases and medications, laboratory values such as troponin, glucose and creatinine, ECG-data, patient symptoms and vital parameters and final diagnoses. [51]

EXPECT

The EXPECT (Evaluation of Unknown Predictors of Electrocardiographic Changes – a Transnational Study) database consists of 300 826 records collected retrospectively from patients visiting the EDs of Lund and Helsingborg from January 1, 2010, to December 31, 2014, and patients visiting the ED in Odense and Esbjerg between March 13, 2013 and April 30, 2014. [95,96]

ESC-TROP

This database includes the first visit for 26545 patients with chest pain presenting to one of five EDs (Lund, Malmö, Helsingborg, Kristianstad, Ystad) between February 1, 2017 and November 30, 2018. [97]

SEM

The SEM (Skåne Emergency Medicine) database contains retrospective data from a cohort of 325539 unique patients with 630275 ED visits from January 1, 2017 to December 31, 2018. All EDs of eight hospitals in Region Skåne (SUS Malmö, Helsingborg, SUS Lund, Kristianstad, Ystad, Trelleborg, Landskrona, Hässleholm) were included in the database. [2]

Figure 13: Study Hospitals

Statistical Analysis

Data analyses were performed with IBM SPSS Statistics version 23 (IBM SPSS Statistics, IBM Corp, Armonk, New York), Medcalc version 19.0.7 (MedCalc Software Ltd, Ostend, Belgium) and the python programming language (Python Software Foundation, Wilmington, Delaware, USA) Models were created using Tensorflow (Google LLC, Mountain View, California, USA) and Pytorch. [98]

Categorical variables were compared using Pearson's Chi-squared- or Fisher's exact test, while continuous variables were compared using independent samples T-tests. Continuous variables were described using means and standard deviations, or median and interquartile range if distributions were skewed. Categorical variables were described using proportions.

Bootstrapping with 1000 resamplings was used to obtain 95% confidence intervals for percentages of rule-in and rule-out (Studies III-IV).

A p-value <0.05 was considered statistically significant.

Sensitivity, specificity, positive- and negative predictive values were used as predefined performance metrics, while ROC-AUC, number of patients ruled in and out were used to evaluate and compare performances of the models.

Paper 1

Introduction

This work was initiated as a result of prior studies on dual testing using troponins and other readily available blood tests such as glucose. [99–102] The rationale was that glucose might be a marker of diabetes, a risk factor of AMI, or possibly that glucose levels rise with infarction as a result of the stress response related to the condition. [21,22,103]

Setting

This was a secondary analysis based on the SCORE-database outlined previously. In addition to previously listed exclusion criteria, patients were excluded if troponin- or glucose blood samples were hemolyzed or missing.

The study had two primary outcomes: 30-day MACE was the primary outcome, and 30-day MACE without unstable angina was a secondary outcome. MACE was defined as an adjudicated diagnosis of AMI, unstable angina, cardiogenic shock, high-degree atrioventricular block requiring intervention, ventricular arrhythmias requiring intervention, cardiac arrest, or death from a cardiac or unknown cause.

Adjudication was performed independently by two cardiologists, adding a third cardiologist if there was disagreement.

The study was approved by the Regional Ethics Review Board in Lund, Sweden (Dnr 2013/5 and 2015/76)

Index test

Three rule-out strategies were analyzed with/without the addition of glucose <5.6mmol/L.

Strategy 1: hs-cTnT <5 ng/L
Strategy 2: hs-cTnT < 12ng/L with 1h change <3ng/L
Strategy 3: hs-cTnT ≤ 14ng/L

Three rule-in strategies were also analyzed with/without the addition of glucose ≥ 11 mmol/L.

Strategy 1: hs-cTnT ≥ 52ng/L
Strategy 2: 1h change ≥ 5ng/L
Strategy 3: hs-cTnT > 14ng/L

Additionally, for each rule-out strategy, the optimal glucose cut-off was determined as the cut-off where the maximum number of patients could be ruled out while maintaining a sensitivity for 30-day MACE of 100%.

Performance metrics

Acceptable negative predictive value for any strategy was set to 98% for MACE and 99.5% for MACE without unstable angina.

Paper 2

Introduction

To determine if ECG changes are new, guidelines recommend comparison of the current record with previous records, if available. Several studies have shown improved performance of machine learning algorithms when serial ECG data is used. [104–106] It is unclear whether more sophisticated machine learning algorithms are improved by using serial ECGs, which is what we wanted to explore in this study. We hypothesized that in some patients, prior ECG records might already show existing changes, making a new ECG finding less indicative of an acute myocardial infarction.

Setting

This retrospective study utilized data from the ESC-TROP study. This database included each patient's first ED visit within this timeframe, including ECGs, hs-cTnT measurements, and prior medical history up to five years before the ED visit. Patients were excluded if they had a STEMI diagnosis, no hs-cTnT data, or no quality ECG within two hours of arrival.

The study was approved by the Regional Ethics Review Board in Lund, Sweden (Dnr 2017/831 and 2018/708)

Input variables

Ten seconds long ECG records with 12 leads, sampled at 500Hz. Index visit ECG was determined as any ECG of sufficient quality within 2 hours of ED arrival. Previous ECGs were the record closest in time but at least one week prior to the index visit

A set of 228 features were extracted from each ECG using the UNI-G algorithm [107] Features used were Q, R, S, T+, T-, ST, ST2/8 and ST3/8 amplitudes, Q, R, S and QRS durations, QRS area and ST slope, STslope positive and STslope negative for each lead (19 features for 12 leads = 228 features). This feature selection was based on a previous study. [108]

Other features were patient age, sex, initial hs-cTnT values and time between ECGs. The last two features were logarithmized before use in order to normalize their skewed distributions.

Models and training

Models used included logistic regression, multilayer perceptron, convolutional neural networks, and residual neural networks pre-trained on large external datasets. [109]

Models were trained with four different input sets: (1) index ECG alone, (2) index and prior ECG, (3) index ECG plus clinical variables (age, sex, and hs-cTnT), and (4) both ECGs plus clinical variables (including time between ECGs).

Data was split chronologically into training (50%), validation (25%) and test groups (25%). Model hyperparameters were determined by random search.

To aid the logistic regression model, principal component analysis (PCA) was used on the 228 Glasgow features for dimensionality reduction. Grid search was used to determine optimal output dimensions of the PCA. The MLP used the 228 Glasgow features without PCA.

CNN and ResNet models used the ECG data rather than the Glasgow features. In order to fit the ResNet architecture, ECG data was resampled and rescaled and leads were reordered. The convolutional part of the ResNet model was fixed, while the last output layer was replaced by a small MLP.

For each network, an ensemble of 10 identical models with different initial seeds, was used. The mean of all model outputs was used as the final output of the ensemble.

Outcome Metrics

The primary clinical outcome, 30-day MACE, was defined as unstable angina, atrioventricular block type 2 or 3, ventricular arrhythmia requiring acute intervention, cardiac arrest, pulmonary edema, cardiogenic shock, coronary artery bypass grafting, percutaneous coronary intervention, transvenous pacemaker insertion, temporary cardiac pacing or death from any cause.

The primary performance metric was the area under the receiver operating characteristic curve (AUROC), with a particular focus on the added value of prior ECGs in the four different input sets. Percentile bootstrapping with 10000 samples was used to obtain 95% confidence intervals for AUROCs.

Subgroup analysis

In routine care, prior ECGs are seldom reviewed if the index ECG is normal. A subgroup analysis was therefore done on all patients where index ECGs were flagged as pathological (AMI or myocardial ischemia) by the Uni-G algorithm [107] Subgroup analysis was also done on the following age subsets 18-50, 51-64, 65-75 and over 76 years. Analyses were also made on patients with/without prior AMI.

Paper 3

Introduction

When conceiving this study, we wanted to build upon the results of study I, where dual testing with a single troponin value could be used to exclude AMI safely in a substantial proportion of patients. We hypothesized that by including several demographic factors, ECG data and blood samples, we should be able to find a subgroup of patients where the likelihood of AMI was very low. We did not include prior ECG records, as we could not see any additional benefit in study II. Using a combination of information sources in this manner would also mimic the way clinicians work and could possibly improve model acceptance upon implementation.

Setting

This study was a retrospective study of data using a subset of the EXPECT database. This subset included all visits to two EDs, Skåne University Hospital at Lund (serving 320.000 inhabitants) and Helsingborg Hospital (serving 250.000 inhabitants) in Sweden between 2013 and 2014.

Inclusion criteria were adult (≥18 years) patients presenting with chest pain, where a hs-cTnT sample and ECG were taken as part of the workup. Only the first ED visit for each patient was included in the study. Patients were then excluded if initial test results for hs-cTnT, glucose, hemoglobin, or creatinine were missing or hemolyzed, or if ECG data quality was insufficient.

This study was approved by the Regional Ethics Review Board in Lund, Sweden (Dnr 2018-708) and the Swedish Ethics Review Authority (Dnr 2019-03523)

Input variables

Any test results within four hours after ED arrival were considered. ECGs were sampled using 12 leads and samples were 10 seconds long with a sampling rate of 1000 Hz. The Uni-G algorithm was used to filter out unusable ECG records and to obtain a 1.6 second long median ECG beat. [107]

Outcome Metrics

The primary clinical outcome was AMI or all-cause death within 30 days, while index event myocardial infarction was used as a secondary outcome. The rule-out performance of each model required a sensitivity of at least 99% and a negative predictive value (NPV) above 99.5%. Rule-in performance was defined by specificity over 90% and positive predictive value (PPV) above 70%. Models were evaluated by AUROC, rule-out proportion, and rule-in proportion.

Models and training

Multiple machine learning (ML) models were developed to predict AMI or death within 30 days, with the following inputs: age, sex, initial ECG, and first blood test values (hs-cTnT, glucose, creatinine, and hemoglobin).

Models varied by complexity: logistic regression, artificial neural network (ANN) which in this study used the MLP architecture described earlier, and convolutional neural networks (CNN) for raw and median ECG beats. The models' performances were compared to the European Society of Cardiology's 0-hour protocol, which uses a single hs-cTnT measurement (cutoffs <5 ng/L for rule-out and ≥52 ng/L for rule-in).

Patients were split chronologically into training (50%), tuning (25%) and testing (25%) groups. The tuning group was used to find a cutoff for each model that would allow the maximum amount of patient to be either ruled in or out while maintaining prespecified performance metrics.

All models used age, sex and blood sample results as inputs, while the CNN models additionally included ECG-data either as the raw ECG signal (CNN-raw) or the median beat (CNN-MB). 8 of the 12 ECG channels were used as inputs to these models, as the remaining four (III, aVL, aVR, aVF) are linear combinations of others and present no new information.

Paper 4

Introduction

In clinical practice, patient risk of AMI is continually evaluated, and some patients are sent home without hs-cTnT sampling. The decision to rule out AMI in these instances is primarily at the discretion of the clinician. We wanted to create a tool that could estimate the relative risk of AMI for patients with stepwise increasing information as in routine care, such that clinicians could make informed decisions to the benefit of patients. Building on the results of study III, we wanted to explore what decisions could be made at specific points in time.

Setting

This study was a retrospective multicenter analysis based on data from the SEM database described earlier.

Patients were eligible for inclusion if they presented to the ED with chest pain and had a high-sensitivity cardiac troponin T (hs-cTnT) sample drawn within four hours of arrival. All patient visits were included in the study. Patients were excluded if initial test results for hs-cTnT, glucose, hemoglobin, or creatinine were missing or hemolyzed, or if ECG data quality was insufficient.

This study was approved by the Swedish Ethical Review Authority (Dnr 2019–05783) and Region Skåne (KVB 302−19).

Data Splitting and Model Development

Patient visits were split by location into a training-tuning (75%) and a test group (25%). Visits to Skåne University Hospital in Malmö were selected as the test group. If a patient had visits in both the training and testing groups, all that patient's visits in the testing group were excluded to prevent data leakage.

The training-tuning group was further divided into training (80%) and tuning (20%) groups by randomly splitting visits on the patient level. The tuning group was used to find a cutoff for each model that would allow the maximum number of patients to be either ruled in or out while maintaining prespecified sensitivity and negative predictive values.

Input Variables

Models utilized a stepwise approach, adding new clinical information at each step to simulate the diagnostic process. A step was defined as a point in time when new information would be available. Input variables included:

Step 1: Patient age and sex.

Step 2: Past medical history and coronary angiography results (up to five years before the index visit), redeemed medications (up to one year prior).

Step 3: ECG features extracted from 10-second, 500 Hz, 12-lead ECG recordings (excluding leads III, aVL, aVR, and aVF due to redundancy).

Step 4: Laboratory results (glucose, hemoglobin, creatinine).

Step 5: First hs-cTnT measurement.

Any test results within four hours after ED arrival were considered. ECGs were sampled using 12 leads and samples were 10 seconds long with a sampling rate of 500 Hz. Eight of the 12 ECG channels were used as inputs to these models, as the remaining four (III, aVL, aVR, aVF) are linear combinations of others and present no new information.

Outcome Metrics

The primary clinical outcome was AMI or all-cause death within 30 days, identified through the Swedish National Inpatient Register and Cause of Death Register. AMI was defined based on the Third Universal Definition of Myocardial Infarction.

Model performance was assessed using the area under the receiver operating characteristic curve (AUROC), rule-out proportion, defined as the percentage of patients safely ruled out using predefined sensitivity and NPV constraints, and sequential rule-out performance, evaluating whether a stepwise ML approach could maintain high sensitivity when used in practice.

Model Training and Validation

A transfer-learning-based ResNet model was used for ECG feature extraction. Neural networks (NNs) were then developed for each step independently and combined into a sequential rule-out framework in a three-stage process:

1. Models were created to evaluate whether the variables in a specific step were informative. Variables were deemed informative if the model AUROC had a lower bound of the 95% Confidence Interval (CI) above 0.5. All steps

with informative variables were used for creating the final combined models.

2. Combined models were created for each step, with all variables available up to that step. Model performance was evaluated using AUROC and the proportion of patients ruled out using that model.
3. Models were evaluated sequentially, using all models available at a specific step, in order to test if the usage of multiple models could maintain the prespecified sensitivity and NPV thresholds.

Feature Importance Analysis

Feature permutation analysis was performed to determine the relative importance of input variables at each step. The AUROC changes from shuffling individual features were used to quantify variable contributions.

Results

Demographics

As table 3 shows, studies varied in number of participants, but the age and sex distributions remained similar. Percentages with the target outcome also remained similar, with more outcomes in studies I and II where the outcome was more broadly defined (MACE vs AMI/death).

Table 3, Overview of study patient demographics

Study	**I**	**II**	**III**	**IV**
Participants, n	1031	19499	9519	40312
Age (mean years)	60.7	62.6	59.1	60.4
Female (%)	46.2	49.8	47.3	47.3
With outcome (%)	11.6	10.6	8.4	6.9
Previous AMI (%)	20.2	7.4	11.1	8.4
hs-cTnT (median ng/L)	6	8	6	7

Paper I

1031 patients were left after exclusion according to the predefined exclusion criteria. Of these, 120 patients (11.6%) had MACE within 30 days.

As seen in Table 4 below, sensitivity increased while the number of patients eligible for rule-out decreased when glucose was added as a diagnostic criterion.

A combination of a single hs-cTnT value ≤14 ng/L and glucose < 5.6 mmol/L identified 252 (24.4%) of patients for rule-out with a 99.6% NPV for MACE excluding unstable angina. Using serial troponin samples, an initial value of hs-cTnT ≤12 ng/L with a rise/fall below 3ng/L combined with a glucose < 5.6 mmol/L allowed the rule out of 240 (23.2%) of patients with a NPV of 100% for MACE excluding unstable angina.

No dual rule-in strategy performed better than using hs-cTnT alone.

When searching for optimal glucose cutoff values, a combination of strategy 2 and 3 was found to maximize the number of patients eligible for rule out of 30-day MACE (Table 5). This strategy identified 431 patients (41.8%) for rule-out with a sensitivity of 100% for MACE without unstable angina.

Table 4: Rule-out strategies.

30-day MACE	**Sensitivity**	**Specificity**	**NPV**	**NLR**	**N**
0h Hs-cTnT <5	96.7	33.5	98.7	0.10	309
0h Hs-cTnT <5 + glucose < 5.6	99.2	15.6	99.3	0.05	143
0h Hs-cTnT < 12 and 1h delta < 3	87.5	73.2	97.8	0.17	682
0h Hs-cTnT <12 and 1h delta < 3 + glucose < 5.6	98.3	26.1	99.2	0.06	240
0h Hs-cTnT ≤ 14	74.1	80.4	96.0	0.32	763
0h Hs-cTnT ≤ 14 + glucose < 5.6	97.5	27.3	98.8	0.09	252
30-day MACE without UA					
0h Hs-cTnT <5	97.6	32.4	99.4	0.07	309
0h Hs-cTnT <5 + glucose < 5.6	98.8	15.0	99.3	0.08	143
0h Hs-cTnT < 12 and 1h delta < 3	98.8	71.9	99.9	0.02	682
0h Hs-cTnT <12 and 1h delta < 3 + glucose < 5.6	100	25.3	100	0.0	240
0h Hs-cTnT ≤ 14	85.7	79.3	98.4	0.18	763
0h Hs-cTnT ≤ 14 + glucose < 5.6	98.8	26.5	99.6	0.04	252

Table 5: Combined rule-out strategies.

30-day MACE	**Sensitivity**	**Specificity**	**NPV**	**NLR**	**N**
Hs-cTnT <5 or hs-cTnT < 12 + delta < 3	81	70.5	97.7	0.27	684
Hs-cTnT ≤ 14 + glucose < 5.6 or hs-cTnT < 12 + delta < 3 + glucose < 5.6	96.4	26.3	98.8	0.14	252
Hs-cTnT ≤ 14 + glucose < 5.2 or hs-cTnT < 12 + delta < 3 + glucose < 6.3	90.5	44.7	98.1	0.21	431
30-day MACE without UA					
Hs-cTnT <5 or hs-cTnT < 12 + delta < 3	97.6	72	99.7	0.03	684
Hs-cTnT ≤ 14 + glucose < 5.6 or hs-cTnT < 12 + delta < 3 + glucose < 5.6	98.8	26.5	99.6	0.05	252
Hs-cTnT ≤ 14 + glucose < 5.2 or hs-cTnT < 12 + delta < 3 + glucose < 6.3	100	45.5	100	0	431

Table 6: Rule-in strategies.

30-day MACE	Sensitivity	Specificity	PPV	PLR	N
0h Hs-cTnT ≥ 52	38.3	97.9	70.8	18.4	65
0h Hs-cTnT ≥ 52 + glucose ≥ 11	6.7	99.7	72.7	20.2	11
1h delta ≥ 5	46.7	97.8	73.7	21.3	76
1h delta ≥ 5 + glucose ≥ 11	11.9	99.7	76.9	37.6	13
0h Hs-cTnT > 14	74.2	80.4	33.2	3.8	268
0h Hs-cTnT > 14 + glucose ≥ 11	12.5	97.9	44.1	6.0	34
30-day MACE without UA					
0h Hs-cTnT ≥ 52	53.6	97.9	69.2	25.4	65
0h Hs-cTnT ≥ 52 + glucose ≥ 11	9.5	99.7	72.7	30.1	11
1h delta ≥ 5	65.5	97.8	72.4	29.5	76
1h delta ≥ 5 + glucose ≥ 11	10.7	99.6	69.2	25.4	13
0h Hs-cTnT > 14	85.7	79.3	26.9	4.1	268
0h Hs-cTnT > 14 + glucose ≥ 11	11.7	97.8	41.2	5.31	34

Paper II

19499 patients were included in the study. Of these, patients in the training group were more likely to have a prior history of AMI, congestive heart failure and pulmonary disease. Mean age was 62.6 years, 49.8% were female and 30-day MACE was observed in 10.6% of cases.

While most models had a slightly higher AUROC when prior ECG data was added, there was no significant improvement, as the 95% confidence intervals overlapped.

When models used only ECG data, the AUROC averaged around 0.77, with the best model being the MLP with a single ECG with AUROC 0.774 (95% CI: 0.754–0.796).

When all variables were used the AUROC improved drastically to around 0.87. In this group the MLP again performed the best, reaching an AUROC of 0.878 (95% CI: 0.864–0.893).

These results were maintained in the subgroup with/without pathological index ECG. When different age groups were examined, a slight improvement in AUROC was seen for some models in younger age groups, while some models performed slightly worse for older patients when a previous ECG was added. These performance changes were not significant. Models performed better for patients without prior AMI.

When evaluating best model performance on subsets of the clinical outcome with only ECG data, unstable angina (AUROC 0.71) was harder to predict than AMI (AUROC 0.8) or death (AUROC 0.83).

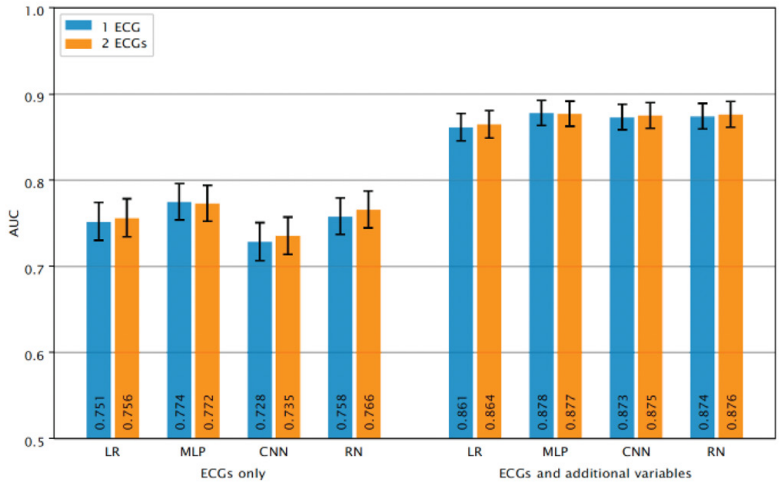


Figure 14: Main results, AUROC values for each model on the test set for predicting 30-day MACE. LR, Logistic Regression; MLP, Multilayer Perceptron; CNN, Convolutional Neural Network; RN, Residual Neural Network

Paper III

A total of 12381 patients were included, and patients were then excluded if exclusion criteria were met, leaving 9519 patients for the final analysis, of which 2379 patients were selected as the test group. Among included patients the mean age was 59.1 years, 47.3% were female and 8.4% experienced AMI or died within 30 days. Patients in the training group had a slightly higher prevalence of AMI (9.0%) compared to the testing group (7.8%).

The CNN model using median ECG beats (CNN-MB) showed the best performance, ruling out 55% of patients with a sensitivity of 99.5% and NPV above 99.9%, outperforming the ESC protocol, which ruled out 47.2% with a sensitivity of 98.9%. The 95% CI between the CNN-MB model (53.1-57) did not overlap with that of the ESC 0h model (45.1-49.3) which indicates that the difference in performance is significant.

For rule-in, CNN-MB maintained a PPV above 70% and identified 5.3% of patients for rule-in, whereas the ESC protocol identified 6.6% but failed to reach the target PPV.

These results were consistent for the secondary outcome, index event AMI.

Table 7: Rule out and rule in performance for all models.

Rule Out

	Sensitivity	NPV	Ruled out %	Ruled out (n)	Missed AMI or Death
ESC 0h	98.9	99.8	47.2	1123	2
LogReg	97.8	99.6	38.5	915	4
ANN	99.5	99.9	46.6	1109	1
CNN-MB	99.5	99.9	55	1309	1
CNN-RAW	99.5	99.9	50.8	1208	1

Rule In

	Specificity	PPV	Ruled in %	Ruled in (n)	Incorrectly ruled in
ESC 0h	97.4	63.9	6.6	158	57
LogReg	98.4	65.1	4.3	103	36
ANN	98.2	69.8	5.4	129	39
CNN-MB	98.5	73.6	5.3	125	33
CNN-RAW	98.2	70.5	5.5	132	39

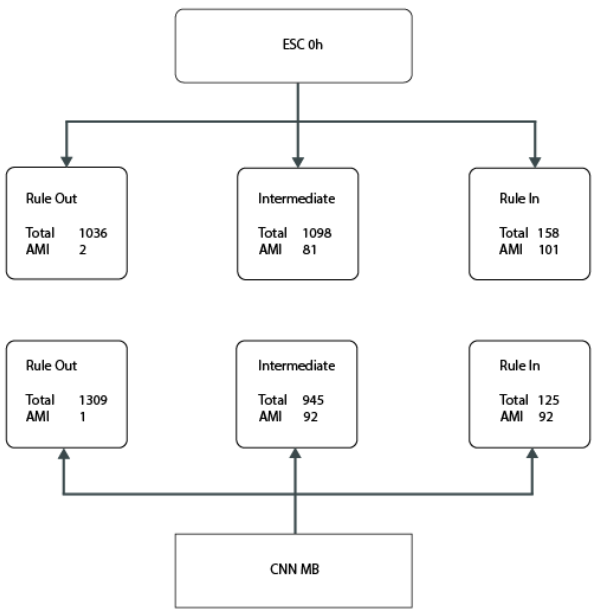


Figure 15: Comparisons between the baseline model (ESC 0h) and the best performing model (CNN-MB).

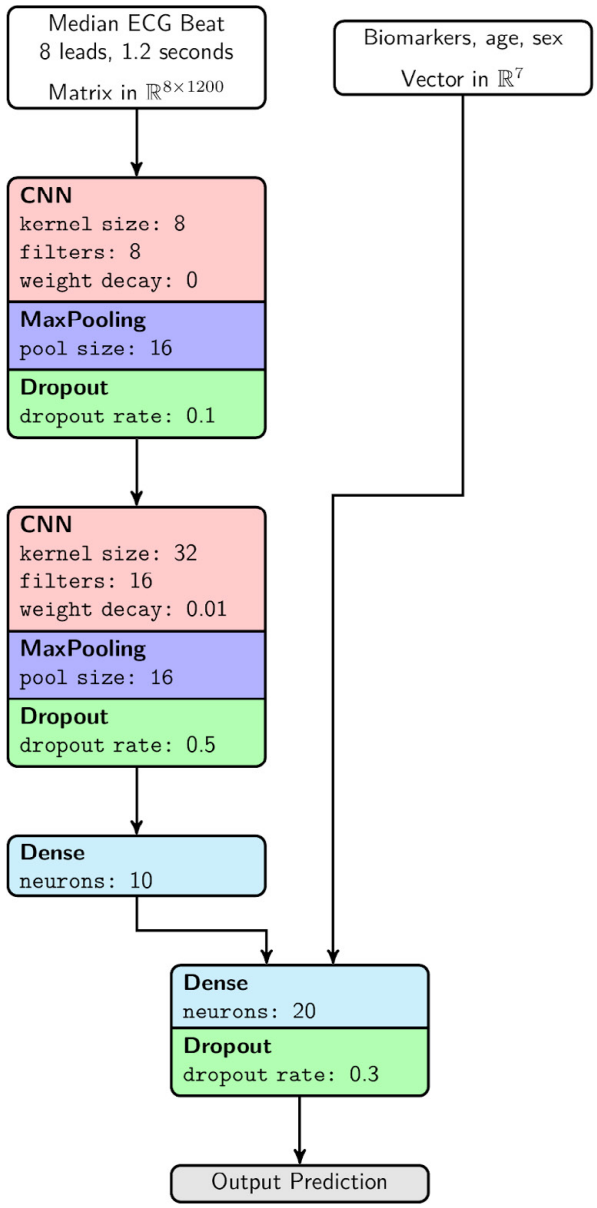


Figure 16: Architecture of the CNN-MB model.

Paper IV

After applying the predefined exclusion criteria, a total of 40,312 patient visits were included in the analysis. Among these, 2,797 (6.9%) experienced AMI or all-cause death within 30 days.

As shown in Table 8, model performance improved at each step, with ECG (Step 3) and hs-cTnT (Step 5) providing the most significant AUROC gains. Sequential application of the models resulted in a cumulative rule-out of 51% of patients, but sensitivity dropped below the prespecified 99% threshold.

Table 9 shows that the highest-performing individual model used all input features, achieving an AUROC of 0.928 and ruling out 49% of patients while maintaining an NPV above 99.5% but sensitivity was below 99% (97.8%). A combination of age and sex alone ruled out 15% of patients with an AUROC of 0.73.

Feature importance analysis revealed that ECG and hs-cTnT were the dominant contributors to model performance, with medical history and other laboratory results providing minimal additional predictive value.

Table 8: Model performance on test-set, AUROC values

Step	**In isolation**	**Combined**
1. Age + sex	0.730	0.730
2. Medical History	0.608	0.732
3. ECG	0.844	0.850
4. Glucose, Hb, Creatinine	0.695	0.853
5. hs-cTnT	0.909	0.928

Table 9: Patients ruled out at each step vs sequentially

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

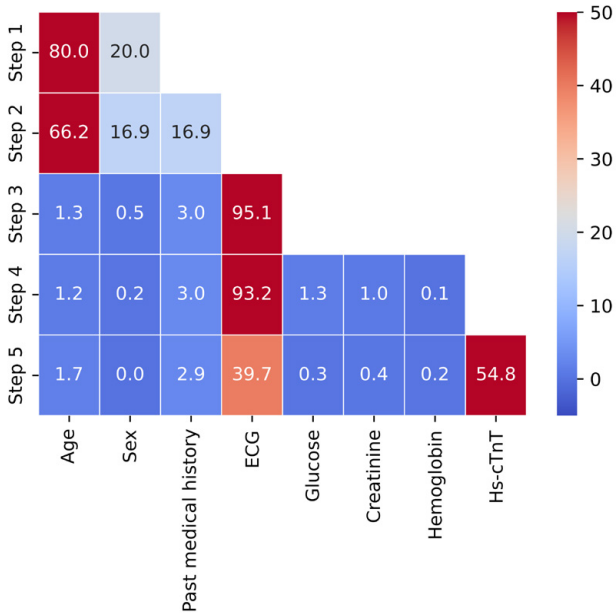


Figure 17: Feature importance values

Discussion

Studies

Paper I

There might be an additional value in a combination of high sensitivity troponins and glucose when evaluating patients for rule out of MACE. The results of this study should be seen as an indication that glucose could be an interesting candidate variable for more complex machine learning models.

There is a range of other molecules currently being investigated for use in patients with suspected AMI. [110] Among these copeptin, a glycopeptide released during the first hours after an infarction, is an interesting candidate for dual testing with a single hs-cTnT value for early rule out of AMI. [111] However, copeptin is not routinely measured, while glucose is readily available at any ED.

We also tested using past diabetes instead of glucose values, but this did not improve the performance of the troponin algorithms. Therefore, the improved performance of dual testing with glucose cannot be solely attributed to a past diagnosis of diabetes.

Regrettably, we could not see improved performance for rule-in. This might be due to the high cutoff selected, 11 mmol/L. Most rule-in strategies already selected few patients for rule-out, but PPV remained high, around 70% for most strategies. The PPV did increase somewhat, but not significantly. This is due to the wide confidence intervals related to the small sample size. Very few patients were ruled in with the dual testing strategy, however, and it's unlikely that significant improvements would be seen using a larger dataset.

Paper II

The results of this study suggest that prior ECGs may not significantly enhance ML models for 30-day MACE prediction in chest pain patients. This implies that models without prior ECGs could still achieve high accuracy and is surprising given that guidelines recommend consideration of prior ECGs, and the importance of ECG changes being new in the definition of AMI. It is important to remember that absence of a result is not a proof that there is no value to prior ECGs. There might

still be value in these, but results indicate that this is not easy to find. The additional complexity of getting access to prior ECG data, and the fact that some patients do not have prior ECG records suggests that focus should be on other model inputs than prior ECGs.

Interestingly the results were consistent regardless of the use of other clinical variables such as age, sex and hs-cTnT values, and even in subgroups where one would expect prior ECGs to be important such as patients with pathological index ECGs.

Another interesting aspect of this study was that simpler models such as the MLP performed on par or even better than more complex models such as the ResNet. We would expect the ResNet to be able to capture more complex representations than those learned by the MLP presented with engineered features. The likely reason this didn't happen is that there was insufficient data. 20 000 records might seem like a lot, but for these models it's barely scratching the surface of the problem. Large ECG datasets used today contain millions of records and this is probably what is needed for these complex architectures to show their worth. We did explore this in study IV where we could see significant model improvements when trained on more data.

Paper III

This study demonstrated that ML models could potentially replace serial hs-cTnT measurements, allowing for quicker patient management decisions with only one set of blood tests and ECG. Our CNN models showed promise in identifying patients for early rule-out and rule-in, potentially reducing ED crowding. Specifically, the CNN-MB model outperformed the baseline 0h-ESC protocol, although the difference in number of ruled-out patients was small. This could be an effect of the temporal data split, where a substantial proportion of patients had very low hs-cTnT values, exaggerating the performance of the baseline model. One issue with this study is that it's hard to disentangle whether more complex models outperformed simpler models due to their architecture or if it's simply that these models had access to more data.

In this study we had access to more data than in study II, but it's likely still not enough. The selected decision thresholds of the models were based on the patients with 30-day AMI or death in the tune set and a very high sensitivity threshold of 99%. With these criteria, only one patient was allowed as a false negative in the tune group. This resulted in conservative rule-out decision thresholds. With access to more data the models could possibly select a larger proportion of patients for rule-out.

The results of our study are on par with similar work done by other research groups using other model architectures and input variables. [85–89] We did manage to achieve

a slightly higher sensitivity in our test group compared to these other studies, but this remains to be externally validated. Two strengths of the models in this study are the use of a single hs-cTnT value instead of serial tests, and that models do not require any patient history or physical examination, giving a probability score solely based on objective data.

Paper IV

This was a challenging study to perform, both in terms of collecting and preprocessing data and model creation. It's hard to imagine that the initial scope of the study was even wider. In addition to the five diagnostic steps, I also wanted to explore different model architectures such as random forests, and different hyperparameter search techniques. The number of permutations would have been staggering. As it stands the paper is still quite complex and possibly a bit overreaching.

A continuation of previous studies, study IV delved into the specific input features used when predicting AMI and death. We could see in study III that our models only had small performance improvements compared to the baseline, which only used a troponin cutoff. Therefore, I wanted to know exactly how much difference factors such as age and sex made when troponin results were introduced to the models. The intuition was that since troponins were a part of the diagnostic test while at the same time being a criterion for the target diagnosis, incorporation bias would play a significant part in the performance of the models.

The models developed in this study did not reach the stringent predefined sensitivity threshold, but they can still be useful simply as tools for estimating the pretest probability of disease for a patient before a complete history is taken. This could be important in triage, where a decision must be made to take an ECG and blood tests or to send the patient home. By understanding the individual patient risk, informed decisions can be taken, and patients can be more involved in decisionmaking.

Additionally, sequential model use led to deteriorating sensitivity. I have not seen this type of analysis done previously, but clinical risk scores are often used in sequence in this manner for various conditions. As these scores are tested in isolation of each other, this raises the question if our current risk scores are as performant as we think they are.

Another theme I wanted to explore was explainability. Studies II - III did not contain any explainability measurement. To justify the use of these models, I had to be able to peek inside.

Interestingly, feature importance analysis indicated that most other variables were not very informative once ECG data was introduced. This was surprising given that the clinical intuition is that a normal ECG does not rule out infarction, and that age

plays an important role for risk prediction. Troponin was still by far the most informative variable, but even when this was added, the model continued to use the ECG data.

Using a wide range of input variables mimics clinical practice and might improve model acceptance if users see that the models consider all aspects of the workup. If models are simply using the hs-cTnT values "behind the scenes" however, users might loose confidence in model performance. Visualising what the model "sees" using explainability techniques such as feature importance could improve model transparency and ultimately improve user trust and adoption.

General Discussion

The data, the model and the noise

In any form of modeling, we must consider that a perfect model requires either having a complete dataset or a perfect causal relationship between our features and the outcome. The first scenario is not realistic for clinical research, but we are getting closer by building increasingly large databases of patient cases. The second scenario is more interesting to discuss.

The question, therefore, is: How close can we come to a perfect relationship between input features and the target?

The first step is the selection of features. We could include everything we have and let the model find what's important. This is an accepted method when constructing predictive models. However, this can lead to models having difficulties locating important relationships as data becomes increasingly sparse when the number of feature dimensions grow. This is known as the curse of dimensionality. [112] Models can be overfit on patterns related to the specific training data instead of learning meaningful representations.

One example of overfitting can be seen in Study I. Reviewers wanted the optimal glucose cutoffs for all rule out decisions in addition to the prespecified cutoffs. Using the optimal values enabled a small increase in the proportion of patients ruled out, but these results would likely not generalize to other settings.

When selecting features, we also face another question. Could we modify our features in some way to extract more information? Feature engineering has historically been an important method to help models learn. With the development of deep neural networks, we have moved away from this approach, assuming that models will generate these intermediate features, or representations, on their own. In study IV we did use some feature engineering and feature selection but generally

tried to include as many variables as possible, using feature importance analysis to see what actually did improve the models.

Another crucial aspect to consider is noise. We are searching for a signal that represents the relationship between our features and the outcome, but multiple types of noise complicate this task.

Firstly, how confident should we be regarding the outcome labels? With medical data (and in our studies), we typically look at ICD-codes as a surrogate for patient diagnoses. Mislabelling, i.e., that the recorded diagnoses do not match the patient's actual condition, could create significant issues by breaking the relationship between the input variables and the outcome. This is particularly true if different coding procedures at different locations lead to a discrepancy between the training and test group outcomes.

Secondly, there is noise in our features. Specifically for lab data and ECGs, measurement errors can cause the data to deviate from the true values. Additionally, there is a risk that a patient may have been mistakenly assigned another patient's results. Other factors, such as the timing of patient presentation to the ED at different locations, could also affect results.

In our Study IV, we observed that the test group more frequently had hs-cTnT < 5 ng/L compared to the training-tuning group. It takes time for troponin levels to rise in the blood following a myocardial infarction. Could it be that some patients in the test group arrived at the hospital very quickly, and their initial troponin measurement was falsely low?

All of this contributes to noise in the signal, which can hinder models from learning optimally.

We may also encounter issues related to the model architecture and training. Traditional models, such as logistic regression, are linear and risk failing to capture nonlinear relationships. It remains an open question what the best model architecture is for a given problem. A common method, which we also employ, is to test a wide range of architectures and select the best performing one. However, this does not guarantee that the chosen architecture is truly optimal. On top of this, model training algorithms can get stuck in suboptimal minima. Had we chosen a different error minimization method or simply different initial weights for our networks, we might have been able to reduce errors further.

One should also ask, is the model even using the data? This is an important point. In Study III, we had two models that received the same input data: CNN-RAW and CNN-MB. Interestingly, CNN-MB performed slightly better than CNN-RAW. Both models had access to the same information, with the only major difference being that one model received raw ECG data (CNN-RAW), while the other received processed data (CNN-MB), specifically median beats.

When we started studying sequential models in Study IV, our initial plan was to use one of these models to extract ECG features from our dataset. For some reason, we first chose CNN-RAW, but the results were unexpected: all patients ended up with identical values for their ECG features.

After some investigation, we found an explanation: all feature weights were extremely small. The dominant factor in the network layers was the bias, meaning that in the end, the only thing being passed through was the bias term. In other words, CNN-RAW was not actually using the ECG data when making its predictions! This might explain why CNN-RAW did not perform as well as we had anticipated. When we analyzed the CNN-MB model, it became clear that this model was indeed using the ECG data.

What can we learn from this? It highlights the importance of looking inside the models rather than just evaluating their overall performance. In Study IV, we specifically examined the relative contribution of each variable to the model's final decisions. To do this, we used feature importance analysis to assess how different inputs influenced the model's predictions.

All this leads to the conclusion that creating good predictive models is a really difficult process. Often initial results are promising, but diminishing returns make it increasingly difficult to improve models as we move closer to perfect predictions.

Random, chronological or geographic splits

Throughout these studies, we have used different types of data splitting for model training and evaluation. Random splitting is probably the most common approach in machine learning applications, allowing for an even distribution of the target labels.

A chronological split, on the other hand, can be used to test models in a way that more closely resembles real-world conditions. Diagnostic procedures and criteria evolve over time. This means that models developed on a specific patient dataset risk predicting outcomes that no longer fully align with the current reality. This introduces a challenge for models, as the problem they aim to solve has now changed.

A geographic split, which was used in Study IV, confronts another common issue, namely that a model trained and tested in one location should perform equally well in another, where the patient population may differ.

Random splitting may produce overly optimistic results, as the test group does not experience diagnostic or geographic drift. Ideally, a model should be developed and then tested in a different location and at a later time, which is typically done in external validation.

In our studies, we primarily used chronological splits (Studies II and III) and geographic splits (Study IV). Our demographic descriptions show that when using a chronological split, the training dataset contains patients with more pre-existing conditions than the test dataset. However, we do not see a clear trend in the distribution of outcome labels across our studies. It is statistically probable that patients that seek often for the same issue (frequent flyers) disproportionally end up in the training set.

In Study II, the proportion of patients with MACE was slightly higher in the test group (11% vs 10%) than in the training group. In Study III, the proportion of AMI/death was slightly lower in the test group compared to the training group (8% vs 9%). However, in Study III, the test group contained significantly more patients with hs-cTnT < 5 ng/L than the tuning group (47% vs 36%). Interestingly, this trend also appears in the geographic split of Study IV, where 38% of patients in the test set had hs-cTnT < 5 ng/L, compared to 29% in the train-tune set.

The reason for these differences is not entirely clear, but they pose a significant challenge for the models, as the patient population they were trained on differs noticeably from the one they are tested on.

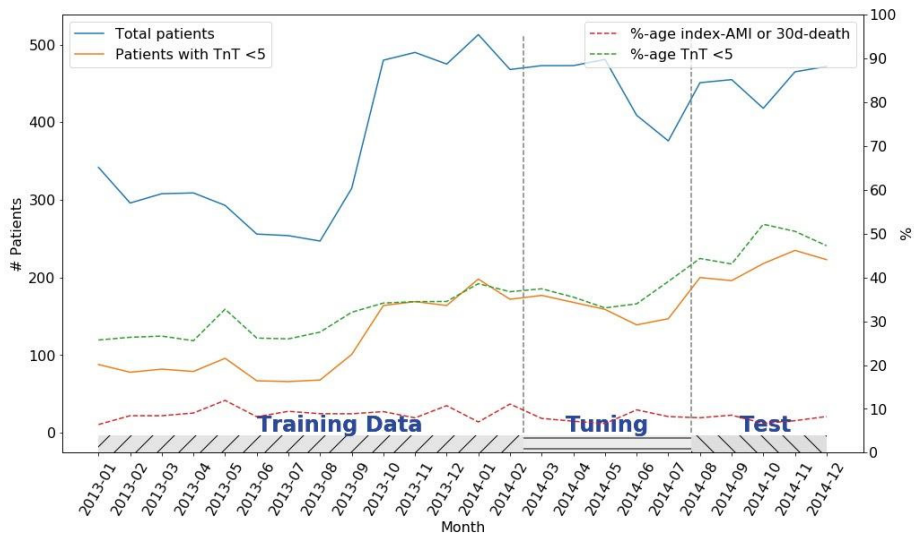


Figure 18: The distribution of troponin values in the dataset of study III.

The Issues with Feature Importance

In Study IV, we used feature importance to analyze our models and explore which variables they considered in their decision-making process. This procedure is not entirely unproblematic.

If a feature is not important to the model, it only means that this particular model is not using it, not that the feature itself is useless. As an example, we included diagnosis codes for previous diseases as input features. As soon as ECG data was introduced, the models barely used these diagnosis codes anymore. However, had we included additional details about past diagnoses such as the date of diagnosis or the number of times it was recorded, we might have obtained different results.

When certain variables are strongly correlated, the model can still make predictions even when some variables are missing, because the missing information is encoded in the remaining correlated variables. As a result, it may appear that each individual correlated variable contributes less to the model than it actually does. The ECG-features were developed using a transfer learning method based on predicting age. When these were added, age lost practically all its predictive power. Could this mean that this information now resides within the ECG features? During Study IV we first use the CNN-MB model from study III as an ECG feature extractor. When we compare the results of the feature importance analysis using the CNN-MB model and the final ResNet model, there is a striking difference in feature importances. Models trained with these new ResNet-based ECG features significantly outperformed the ones trained on CNN-MB-based features. More interestingly, age, sex, and previous diseases were barely used by the model anymore.

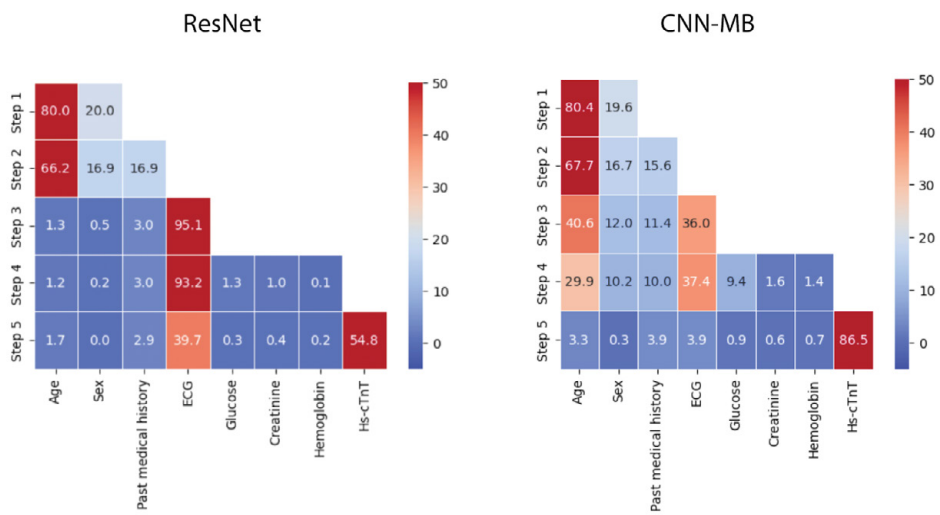


Figure 19: ResNet vs CNN-MB feature importance

Finally, when we mix feature values in this way, we may create patient cases that cannot exist in the real world. Since the model learns from real patient patterns, it may make unusual decisions when these patterns are broken. For instance, we might pair ECG features from a very old patient with the age of a very young patient, a combination that never occurs in real-world data. This introduces a form of synthetic distribution bias, as we deviate from real-world cases in an artificial way.

There still is some value in visualisation techniques such as feature importance. Had we used this in study III we would have immediately understood that the CNN-raw model did not use the ECG data, and reconfiguration of this model might have improved its performance.

Obstacles to model implementation

All the studies in this thesis have been based on register data. The real test of the capabilities of the models is introducing them to clinical practice. This presents yet another layer of issues. Firstly, lets consider current developments related to diagnostic drift.

The traditional STEMI/NSTEMI distinction is based on ECG findings, but this classification does not always correlate with the presence of an occluded coronary artery. Studies have shown that some STEMI diagnoses do not correspond to an acute coronary occlusion, while some NSTEMIs do. The role of STEMI and NSTEMI definitions have recently been up for discussion, with a new definition, OMI vs NOMI gaining traction. This could mean changes to which patients are selected for direct reperfusion therapy in the future, changing the composition of chest pain patients that present to the ED. [113–115] Studies on machine learning methods using this outcome are already emerging. [116]

Even the definition of AMI is under debate, with proposals to change the types of AMI to spontaneous, secondary and procedural infarctions. [117]

Further, changes in diagnostic procedures and available assays might make developed models unusable. A pertinent example is that recently Region Skåne decided to change from hs-cTnT to hs-cTnI assays. As all our current registers and models are based on hs-cTnT, we now need to collect completely new data and train new models instead. What does this mean for models such as those developed in study III and IV? These could still be used by centers that use hs-cTnT assays, and such an external validation of the models is important, but these changes preclude their use in our own clinic.

Lastly, a range of issues regarding the ethical use of models needs to be discussed. This is a big topic best left to its own section.

Computer says no.

- *Carol Beer, Little Britain*

Ethical issues

Patient informed consent was only required for study I, while the other studies gave patients access to the possibility of opting out, i.e., erasing their data from the records upon request. This is in accordance with the General Data Protection Regulation (GDPR) and Patientdatalagen, which stipulates among other things the right of data erasure. [118,119]

When a model is created, however, it is not possible to disentangle certain data points from the underlying data distribution the model relies on. The model cannot unlearn the information gained from the person wishing to withdraw consent.

We also need to consider where the responsibility of model predictions lies. Once a model makes a decision that is incorrect, who is to blame? Is it the developer who created the model? Or is it the clinician acting on the decision? When model predictions are correct and followed there is no issue, but what if the model is incorrect? Patients and their relatives will require somebody to take responsibility for incorrect diagnoses, workups and treatments. And what if the clinician chooses to go against the suggestion of the model? These are questions without a clear answer.

Another issue is that models are currently being trained on data labeled by clinicians. Bias and discriminative behaviour during diagnostic procedures will propagate to the models, which will learn to pick up on the same biases.

Additionally, models are trained using a specific subsample of the entire patient population. In this thesis we have chosen chest pain patients presenting to EDs in Skåne. If these models were to be deployed users might, and are probably likely to, apply them to completely different patient groups such as any patient where a troponin was taken, regardless of whether the patient experienced chest pain. Models might be applied on children or patients admitted to wards. The list goes on. None of these groups were the intended group for these models, and performance is not guaranteed.

Finally, users might try to game the model into providing the answers they want by changing the inputs to figure out when a decision is changed.

All these issues are still open problems, worthy of exploration in the future.

Conclusions

This thesis demonstrates the potential of ML models to improve the early risk evaluation of patients with suspected AMI in the ED. By combining hs-cTnT, ECG data, and other clinical variables, ML-based decision support systems improved diagnostic accuracy and allowed for a reduction in unnecessary testing. However, the results also showed the limited added value of prior ECGs and the decline in sensitivity when using a sequential ML approach.

ML models were also shown to outperform traditional rules-based systems in AMI prediction. These findings suggest that ML could improve ED flow and improve patient outcomes. Before clinical implementation, further external validation and evaluation of real-world usage are important. Future research should focus on optimizing ML models for reliability, interpretability, and usability in ED settings.

Future Perspective

Artificial intelligence is currently progressing at a pace that at times is hard to keep up with. The release of ChatGPT [120,121] has seen incredible interest in the public space. In the medical field, Google recently published Med-Palm2 [122,123], a large language model capable of answering USMLE questions on par with physicians. AMIE, another medical ai model, outperformed general practitioners not only in finding the correct diagnosis, but it also had better bedside manners. [124]

Where does this research fit into this larger narrative? One interesting research direction could be tool creation for these generalist systems. These models can work on many tasks, but they are not specialists, i.e., they do not know the local setting. Models such as the ones created in this thesis could possibly be used as tools by more general models to get access to information on the local patient distribution, without giving them direct access to sensitive patient data.

Another interesting topic would be exploring the effects of releasing the models into the real world. Studies are needed not only for finding real barriers to successful implementation, but also to get a better understanding of how the models are used and misused.

Acknowledgements

The contents of this book have not been easy to produce, and many people have helped me on the way.

Ahmad Game, if it hadn't been for that conversation we had over coffee so many years ago, perhaps none of this would have happened. Thank you for all the wonderful conversations and for always inspiring me to try harder and reach further.

Mattias Ohlsson, **Axel Nyström**, and **Anders Björkelund**, without you, I wouldn't even have been able to begin understanding the field of machine learning. You have helped, corrected, and challenged my understanding of countless aspects throughout this intense learning journey. I would like to thank **Anders** especially for all your hard work on Study III, and **Mattias** for your incredibly educational course in machine learning. **Axel**, it goes without saying that Study II would have been impossible without your help, but beyond that, you've walked alongside me throughout this entire PhD journey. We've faced similar challenges and frustrations, and it truly would have been a lonely road without you.

Jonas Björk, **Arash Mokhtari**, and **Ardavan Khoshnood**, you have spent countless hours meticulously reading my articles, testing my hypotheses, and challenging my methodological choices. For this, I am incredibly grateful. **Arash**, you did an enormous job with Article I and from the beginning, you've been a great source of inspiration for me. I have read your articles over and over again to understand how solid research should be conducted.

Eric Dryver and **Tsvetelina Nilsson**, you have broadened my research ambitions and been incredible sources of inspiration, helping me through the many moments when I doubted whether research was the right path for me.

Ulf Ekelund, throughout all these years you have been a rock-solid presence to lean on. Whenever I needed help or guidance, you were there.

Erica, **Juni**, and **Saga**, much of the time I spent on this work was time I could have spent with you. **Erica**, I am forever grateful for your patience, your sacrifices, and all your hard work during this period. Without your support, none of this would have been possible.

Finally, I would like to thank my brother **Ola** and my dear grandmother **Siv**.

Ola, you have always believed in me and cheered me on in all the countless projects I have started. You've taken incredible responsibility for grandma, especially in recent years as age has begun to take its toll. I wouldn't have had the strength for any of this without you.

Grandma, ever since I was very young, you've done everything to ensure I received a good education and kept me on track, making sure I understood that I cannot waste the gift I've been given. You were denied the opportunity to pursue higher education yourself, even though you had the potential. You promised that your children would have the chance you never got. That didn't happen, but you didn't give up. Now, as I write the final words of this thesis, I am also finishing what you started so many years ago. You did it, Grandma. I did it. Thank you. I love you.

References

1. Organization, W. H. *World Health Statistics 2024: Monitoring Health for the SDGs, Sustainable Development Goals*. (World Health Organization, Geneva, 2024).
2. Ekelund, U. *et al.* The skåne emergency medicine (SEM) cohort. *Scand. J. Trauma Resusc. Emerg. Med.* **32**, 37 (2024).
3. Al-Hijji, M. A. *et al.* Safety and Risk of Major Complications With Diagnostic Cardiac Catheterization. *Circ. Cardiovasc. Interv.* **12**, e007791 (2019).
4. Af Ugglas, B., Djärv, T., Ljungman, P. L. S. & Holzmann, M. J. Emergency department crowding associated with increased 30-day mortality: a cohort study in Stockholm Region, Sweden, 2012 to 2016. *J. Am. Coll. Emerg. Physicians Open* **1**, 1312–1319 (2020).
5. Wessman, T. *et al.* The association between length of stay in the emergency department and short-term mortality. *Intern. Emerg. Med.* **17**, 233–240 (2022).
6. Hollander, J. E. The Continuing Search to Identify the Very-low-risk Chest Pain Patient. *Acad. Emerg. Med.* **6**, 979–981 (1999).
7. Boron, W. F. & Boulpaep, E. L. *Medical Physiology*. (Elsevier, 2016).
8. Eriksson, D. *et al.* Diagnostic Accuracy of History and Physical Examination for Predicting Major Adverse Cardiac Events Within 30 Days in Patients With Acute Chest Pain. *J Emerg Med* (2019) doi:10.1016/j.jemermed.2019.09.044.
9. Fanaroff, A. C., Rymer, J. A., Goldstein, S. A., Simel, D. L. & Newby, L. K. Does This Patient With Chest Pain Have Acute Coronary Syndrome?: The Rational Clinical Examination Systematic Review. *JAMA* **314**, 1955–1965 (2015).
10. Canto, J. G. *et al.* Association of age and sex with myocardial infarction symptom presentation and in-hospital mortality. *JAMA* **307**, 813–822 (2012).
11. Swap, C. J. & Nagurney, J. T. Value and limitations of chest pain history in the evaluation of patients with suspected acute coronary syndromes. *JAMA* **294**, 2623–2629 (2005).
12. EINTHOVEN, W. THE GALVANOMETRIC REGISTRATION OF THE HUMAN ELECTROCARDIOGRAM, LIKEWISE A REVIEW OF THE USE OF THE CAPILLARY-ELECTROMETER IN PHYSIOLOGY. *Ann. Noninvasive Electrocardiol.* **2**, 93–93 (1997).
13. Thygesen, K. *et al.* Fourth universal definition of myocardial infarction (2018). *Eur. Heart J.* **40**, 237–269 (2018).
14. Meisenberg, G. & Simmons, W. H. *Principles of Medical Biochemistry*. (Elsevier, 2017).

15. Katus, H. A. *et al.* Enzyme linked immuno assay of cardiac troponin T for the detection of acute myocardial infarction in patients. *J. Mol. Cell. Cardiol.* **21**, 1349–1353 (1989).
16. Katrukha, I. A. & Katrukha, A. G. Myocardial Injury and the Release of Troponins I and T in the Blood of Patients. *Clin. Chem.* **67**, 124–130 (2020).
17. Giannitsis, E. & Katus, H. A. Cardiac troponin level elevations not related to acute coronary syndromes. *Nat. Rev. Cardiol.* **10**, 623–634 (2013).
18. Aviles, R. J. *et al.* Troponin T levels in patients with acute coronary syndromes, with or without renal dysfunction. *N Engl J Med* **346**, 2047–52 (2002).
19. Fridén, V. *et al.* Clearance of cardiac troponin T with and without kidney function. *Clin. Biochem.* **50**, 468–474 (2017).
20. Sarnak, M. J. & Levey, A. S. Cardiovascular disease and chronic renal disease: a new paradigm. *Am J Kidney Dis* **35**, S117-31 (2000).
21. Lippi, G., Cervellin, G. & Targher, G. Random plasma glucose measurement may improve the diagnostic specificity of highly sensitive troponin in the emergency department. *Int. J. Cardiol.* **155**, 172–173 (2012).
22. Altabas, V. *et al.* Glucose metabolism disorders in patients with acute coronary syndromes. *Acta Clin. Croat.* **51**, 71–77 (2012).
23. Feng, Q. Z., Zhao, Y. S. & Li, Y. F. Effect of haemoglobin concentration on the clinical outcomes in patients with acute myocardial infarction and the factors related to haemoglobin. *BMC Res Notes* **4**, 142 (2011).
24. Byrne, R. A. *et al.* 2023 ESC Guidelines for the management of acute coronary syndromes. *Eur. Heart J. Acute Cardiovasc. Care* **13**, 55–161 (2024).
25. Eggers, K. M., Jernberg, T. & Lindahl, B. Unstable Angina in the Era of Cardiac Troponin Assays with Improved Sensitivity-A Clinical Dilemma. *Am. J. Med.* **130**, 1423-1430.e5 (2017).
26. Poudel, I., Tejpal, C., Rashid, H. & Jahan, N. Major Adverse Cardiovascular Events: An Inevitable Outcome of ST-elevation myocardial infarction? A Literature Review. *Cureus* **11**, e5280 (2019).
27. Bosco, E., Hsueh, L., McConeghy, K. W., Gravenstein, S. & Saade, E. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. *BMC Med. Res. Methodol.* **21**, 241 (2021).
28. Lawton, J. S. *et al.* 2021 ACC/AHA/SCAI Guideline for Coronary Artery Revascularization: Executive Summary: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **145**, e4–e17 (2022).
29. Socialstyrelsen. *Statistik Om Akutmottagningar, Väntetider Och Besök 2023*. https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikamnen/vantetider-och-besok-vid-sjukhusbundna-akutmottagningar/ (2024).
30. Ekelund, U., Akbarzadeh, M., Khoshnood, A., Björk, J. & Ohlsson, M. Likelihood of acute coronary syndrome in emergency department chest pain patients varies with time of presentation. *BMC Res. Notes* **5**, 420 (2012).

31. Mokhtari, A. *et al.* Effectiveness and Safety of the ESC-TROP (European Society of Cardiology 0h/1h Troponin Rule-Out Protocol) Trial. *J. Am. Heart Assoc.* **13**, e036307 (2024).
32. Bjørnsen, L. P., Naess-Pleym, L. E., Dale, J., Grenne, B. & Wiseth, R. Description of chest pain patients in a Norwegian emergency department. *Scand. Cardiovasc. J. SCJ* **53**, 28–34 (2019).
33. Martínez-Sellés, M. *et al.* Chest pain in the emergency department: incidence, clinical characteristics and risk stratification. *Rev. Esp. Cardiol.* **61**, 953–959 (2008).
34. Baccouche, H. *et al.* Acute coronary syndrome among patients with chest pain: Prevalence, incidence and risk factors. *Int. J. Cardiol.* **214**, 531–535 (2016).
35. Cullen, L. *et al.* Cost and outcomes of assessing patients with chest pain in an Australian emergency department. *Med. J. Aust.* **202**, 427–432 (2015).
36. Wireklint, S., Elmqvist, C. & Göransson, K. An updated national survey of triage and triage related work in Sweden: a cross-sectional descriptive and comparative study. *Scand. J. Trauma Resusc. Emerg. Med.* **29**, 89 (2021).
37. Bobbio, M. *et al.* Diagnostic accuracy of predicting coronary artery disease related to patients' characteristics. *J. Clin. Epidemiol.* **47**, 389–395 (1994).
38. Richardson, W. S. Five uneasy pieces about pre-test probability. *J. Gen. Intern. Med.* **17**, 882–883 (2002).
39. Kline, J. A. *et al.* Pretest probability assessment derived from attribute matching. *BMC Med. Inform. Decis. Mak.* **5**, 26 (2005).
40. Six, A. J., Backus, B. E. & Kelder, J. C. Chest pain in the emergency room: value of the HEART score. *Neth. Heart J. Mon. J. Neth. Soc. Cardiol. Neth. Heart Found.* **16**, 191–196 (2008).
41. Backus, B. E. *et al.* Chest pain in the emergency room: a multicenter validation of the HEART Score. *Crit. Pathw. Cardiol.* **9**, 164–169 (2010).
42. Backus, B. E. *et al.* A prospective validation of the HEART score for chest pain patients at the emergency department. *Int. J. Cardiol.* **168**, 2153–2158 (2013).
43. Than, M. *et al.* 2-Hour accelerated diagnostic protocol to assess patients with chest pain symptoms using contemporary troponins as the only biomarker: the ADAPT trial. *J. Am. Coll. Cardiol.* **59**, 2091–2098 (2012).
44. Than, M. *et al.* Development and validation of the Emergency Department Assessment of Chest pain Score and 2 h accelerated diagnostic protocol. *Emerg. Med. Australas. EMA* **26**, 34–44 (2014).
45. Antman, E. M. *et al.* The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA* **284**, 835–842 (2000).
46. Fox, K. A. A. *et al.* Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *BMJ* **333**, 1091 (2006).
47. Mahler, S. A. *et al.* The HEART Pathway Randomized Trial. *Circ. Cardiovasc. Qual. Outcomes* **8**, 195–203 (2015).

48. Stopyra, J. P. *et al.* Performance of the EDACS-accelerated Diagnostic Pathway in a Cohort of US Patients with Acute Chest Pain. *Crit. Pathw. Cardiol.* **14**, 134–138 (2015).
49. Nilsson, T., Johannesson, E., Lundager Forberg, J., Mokhtari, A. & Ekelund, U. 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. *Emerg. Med. J. EMJ* (2021) doi:10.1136/emermed-2020-210833.
50. Neumann, J. T. *et al.* Diagnosis of Myocardial Infarction Using a High-Sensitivity Troponin I 1-Hour Algorithm. *JAMA Cardiol.* **1**, 397–404 (2016).
51. Mokhtari, A. *et al.* A 1-h Combination Algorithm Allows Fast Rule-Out and Rule-In of Major Adverse Cardiac Events. *J. Am. Coll. Cardiol.* **67**, 1531–1540 (2016).
52. Chew, D. P. *et al.* A Randomized Trial of a 1-Hour Troponin T Protocol in Suspected Acute Coronary Syndromes: The Rapid Assessment of Possible Acute Coronary Syndrome in the Emergency Department With High-Sensitivity Troponin T Study (RAPID-TnT). *Circulation* **140**, 1543–1556 (2019).
53. Thygesen, L. C. & Ersbøll, A. K. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur. J. Epidemiol.* **29**, 551–558 (2014).
54. Rubinger, L., Ekhtiari, S., Gazendam, A. & Bhandari, M. Registries: Big data, bigger problems? *Injury* **54 Suppl 3**, S39–S42 (2023).
55. Martínez-Sellés, M. & Marina-Breysse, M. Current and Future Use of Artificial Intelligence in Electrocardiography. *J. Cardiovasc. Dev. Dis.* **10**, (2023).
56. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
57. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386 (1958).
58. Minsky, M. & Papert, S. *Perceptrons: An Introduction to Computational Geometry*. (MIT Press, Cambridge, MA, USA, 1969).
59. Cybenko, G. V. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2**, 303–314 (1989).
60. The Royal Swedish Academy of Sciences. Press release: The Nobel Prize in Physics 2024. (2024).
61. Peng, J., Lee, K. & Ingersoll, G. An Introduction to Logistic Regression Analysis and Reporting. *J. Educ. Res. - J EDUC RES* **96**, 3–14 (2002).
62. Nair, V. & Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. in *Proceedings of the 27th International Conference on Machine Learning (ICML)* 807–814 (2010).
63. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).
64. LeCun, Y. *et al.* Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).

65. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. in *Advances in Neural Information Processing Systems* vol. 25 (2012).
66. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. (2015).
67. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
68. Bozinovski, S. Reminder of the First Paper on Transfer Learning in Neural Networks, 1976. *Informatica* **44**, (2020).
69. Ebbehoj, A., Thunbo, M. Ø., Andersen, O. E., Glindtvad, M. V. & Hulman, A. Transfer learning for non-image data in clinical research: A scoping review. *PLOS Digit. Health* **1**, e0000014 (2022).
70. Bird, J. J., Kobylarz, J., Faria, D. R., Ekárt, A. & Ribeiro, E. P. Cross-Domain MLP and CNN Transfer Learning for Biological Signal Processing: EEG and EMG. *IEEE Access* **8**, 54789–54801 (2020).
71. Attia, Z. I. *et al.* Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. *Circ. Arrhythm. Electrophysiol.* **12**, e007284 (2019).
72. Lima, E. M. *et al.* Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nat. Commun.* **12**, 5117 (2021).
73. Commission, E. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. (2021).
74. Vilone, G. & Longo, L. Explainable Artificial Intelligence: a Systematic Review. *CoRR* **abs/2006.00093**, (2020).
75. Andreas Holzinger, C. B., Constantinos S. Pattichis, Douglas B. Kell. What do we need to build explainable AI systems for the medical domain? *arXiv* (2017).
76. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst. NeurIPS* **30**, (2017).
77. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* 618–626 (2017). doi:10.1109/ICCV.2017.74.
78. Leo Breiman. Random forests. *Mach. Learn.* **45**, 5--32 (2001).
79. Hannun, A. Y. *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
80. Baxt, W. G. Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann. Intern. Med.* **115**, 843–848 (1991).
81. Zworth, M. *et al.* Machine learning for the diagnosis of acute coronary syndrome using a 12-lead ECG: a systematic review. *Can. J. Emerg. Med.* **25**, 818–827 (2023).
82. Harrison, R. F. & Kennedy, R. L. Artificial Neural Network Models for Prediction of Acute Coronary Syndromes Using Clinical Data From the Time of Presentation. *Ann. Emerg. Med.* **46**, 431–439 (2005).

83. Lin, Y., Dong, S., Chu, W., Kwauk, X. & Liu, Y. Applications of Machine Learning Algorithms in Predicting Coronary Artery Disease and Myocardial Infarction. **1**, 81–93 (2018).
84. Than, M. *et al.* What is an acceptable risk of major adverse cardiac event in chest pain patients soon after discharge from the Emergency Department? *Int. J. Cardiol.* **166**, 752–754 (2013).
85. Doudesis, D. *et al.* 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. *Lancet Digit. Health* **4**, e300–e308 (2022).
86. Than, M. P. *et al.* Machine Learning to Predict the Likelihood of Acute Myocardial Infarction. *Circulation* **140**, 899–909 (2019).
87. McCord, J. *et al.* Machine Learning to Assess for Acute Myocardial Infarction Within 30 Minutes. *Crit. Pathw. Cardiol.* **21**, (2022).
88. Björkelund, A. *et al.* Machine learning compared with rule-in/rule-out algorithms and logistic regression to predict acute myocardial infarction based on troponin T concentrations. *J. Am. Coll. Emerg. Physicians Open* **2**, e12363 (2021).
89. Doudesis, D. *et al.* Machine learning for diagnosis of myocardial infarction using cardiac troponin concentrations. *Nat. Med.* **29**, 1201–1210 (2023).
90. Boeddinghaus, J. *et al.* Machine Learning for Myocardial Infarction Compared With Guideline-Recommended Diagnostic Pathways. *Circulation* **149**, 1090–1101 (2024).
91. Zhang, P.-I. *et al.* Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. *Scand. J. Trauma Resusc. Emerg. Med.* **28**, 93 (2020).
92. Neumann, J. T. *et al.* Personalized diagnosis in suspected myocardial infarction. *Clin. Res. Cardiol. Off. J. Ger. Card. Soc.* **112**, 1288–1301 (2023).
93. Xiao, R. *et al.* Integrating multimodal information in machine learning for classifying acute myocardial infarction. *Physiol. Meas.* **44**, 044002 (2023).
94. Ibrahim, L., Mesinovic, M., Yang, K.-W. & Eid, M. A. Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values. *IEEE Access* **8**, 210410–210417 (2020).
95. Hansen, T. G. *et al.* New-Onset Atrial Fibrillation Among Patients With Infection in the Emergency Department: A Multicenter Cohort Study of 1-Year Stroke Risk. *Am. J. Med.* **133**, 352-359.e3 (2020).
96. Schade Hansen, C. *et al.* Association between QTc prolongation and mortality in patients with suspected poisoning in the emergency department: a transnational propensity score matched cohort study. *BMJ Open* **8**, (2018).
97. Mokhtari, A. *et al.* 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* **145**, 685–692 (2020).
98. Paszke, A. *et al.* PyTorch: an imperative style, high-performance deep learning library. in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2019).

99. Haller, P. M. *et al.* The association of anaemia and high-sensitivity cardiac troponin and its effect on diagnosing myocardial infarction. *Eur. Heart J. Acute Cardiovasc. Care* (2021) doi:10.1093/ehjacc/zuab066.

100. Greenslade, J. H. *et al.* Combining presentation high-sensitivity cardiac troponin I and glucose measurements to rule-out an acute myocardial infarction in patients presenting to emergency department with chest pain. *Clin. Biochem.* **48**, 288–291 (2015).

101. Shortt, C. *et al.* Rule-In and Rule-Out of Myocardial Infarction Using Cardiac Troponin and Glycemic Biomarkers in Patients with Symptoms Suggestive of Acute Coronary Syndrome. *Clin. Chem.* **63**, 403–414 (2017).

102. Kavsak, P. *et al.* Clinical chemistry score versus high-sensitivity cardiac troponin I and T tests alone to identify patients at low or high risk for myocardial infarction or death at presentation to the emergency department. *Can. Med. Assoc. J.* **190**, E974–E984 (2018).

103. Timmer, J. R. *et al.* Hyperglycemia is an important predictor of impaired coronary flow before reperfusion therapy in ST-segment elevation myocardial infarction. *J. Am. Coll. Cardiol.* **45**, 999–1002 (2005).

104. Gregg, R. E., Deluca, D. C., Chien, C. S., Helfenbein, E. D. & Ariet, M. Automated serial ECG comparison improves computerized interpretation of 12-lead ECG. *J. Electrocardiol.* **45**, 561–565 (2012).

105. Ohlsson, M., Ohlin, H., Wallerstedt, S. M. & Edenbrandt, L. Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction. *Am. J. Cardiol.* **88**, 478–481 (2001).

106. Ariet, M., Zhou, S., DeLuca, D. C. & Greenfield, J. C. Computerized serial comparison of electrocardiograms: program performance in myocardial infarction. *J. Electrocardiol.* **40**, 147–154 (2007).

107. Macfarlane, P. W., Devine, B. & Clark, E. *The University of Glasgow (Uni-G) ECG Analysis Program*. *Computers in Cardiology* vol. 32 (2005).

108. Forberg, J. L. *et al.* In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department. *J. Electrocardiol.* **42**, 58–63 (2009).

109. Ribeiro, A. H. *et al.* Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **11**, 1760 (2020).

110. Tilea, I., Varga, A. & Serban, R. C. Past, Present, and Future of Blood Biomarkers for the Diagnosis of Acute Myocardial Infarction-Promises and Challenges. *Diagn. Basel Switz.* **11**, (2021).

111. Reichlin, T. *et al.* Incremental value of copeptin for rapid rule out of acute myocardial infarction. *J. Am. Coll. Cardiol.* **54**, 60–68 (2009).

112. Bellman, R. *Dynamic Programming*. (Princeton University Press, Princeton, NJ, 1957).

113. Meyers, P., Weingart, S. & Smith, S. The OMI Manifesto. *Dr. Smith's ECG Blog* (2018).

114. Meyers, H. P. *et al.* Comparison of the ST-Elevation Myocardial Infarction (STEMI) vs. NSTEMI and Occlusion MI (OMI) vs. NOMI Paradigms of Acute MI. *J. Emerg. Med.* **60**, 273–284 (2021).
115. Meyers, H. P. *et al.* Accuracy of OMI ECG findings versus STEMI criteria for diagnosis of acute coronary occlusion myocardial infarction. *IJC Heart Vasc.* **33**, 100767 (2021).
116. Herman, R. *et al.* International evaluation of an artificial intelligence-powered electrocardiogram model detecting acute coronary occlusion myocardial infarction. *Eur. Heart J. Digit. Health* **5**, 123–133 (2024).
117. Lindahl, B. & Mills, N. L. A new clinical classification of acute myocardial infarction. *Nat. Med.* **29**, 2200–2205 (2023).
118. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation, GDPR). (2016).
119. Sveriges Riksdag. Patientdatalag (2008:355). (2008).
120. Vaswani, A. *et al.* Attention Is All You Need. in *Advances in Neural Information Processing Systems (NeurIPS)* vol. 30 (2017).
121. OpenAI. ChatGPT. (2024).
122. Singhal, K. *et al.* Large Language Models Encode Clinical Knowledge. (2022).
123. Singhal, K. *et al.* Towards Expert-Level Medical Question Answering with Large Language Models. (2023).
124. Tu, T. *et al.* Towards Conversational Diagnostic AI. (2024).