

Aus der Klinik der Kardiologie, Pneumologie, Angiologie
der Heinrich-Heine-Universität Düsseldorf

Direktor der Klinik:

Prof. Dr. med. Malte Kelm

Diagnostische Effizienz neuronaler Netzwerke von
Elektrokardiogrammen im Bildformat

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Moritz Höckmann

2025

Als Inaugural-Dissertation gedruckt mit Genehmigung der Medizinischen Fakultät der
Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Univ.-Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Univ.-Prof. Dr. Hisaki Makimoto

Zweitgutachter: Univ.-Prof. Dr. med. Hug Aubin

Teile dieser Arbeit wurden veröffentlicht:

Hisaki Makimoto, Moritz Höckmann, Tina Lin, David Glöckner, Shqipe Gerguri, Lukas Clasen, Jan Schmidt, Athena Assadi-Schmidt, Alexandru Bejinariu, Patrick Müller, Stephan Angendohr, Mehran Babady, Christoph Brinkmeyer, Asuka Makimoto & Malte Kelm, *Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction*. Sci Rep, 2020. **10**(1): p. 8445

Versicherung an Eides statt:

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf sowie der Richtlinien der Medizinischen Fakultät zur Sicherung guter wissenschaftlicher Praxis erstellt worden ist. Die aus fremden Quellen direkt oder indirekt übernommenen Inhalte wurden als solche kenntlich gemacht. Ich bin mir darüber klar, dass der Bruch der obigen eidesstattlichen Versicherung in jedem Fall zum Nichtbestehen der betreffenden Promotionsleistung führt und die weitere Folge hat, dass die Fakultät über die Entziehung des Doktorgrades entscheidet (§ 16 Promotionsordnung). Die strafrechtlichen Konsequenzen einer falschen eidesstattlichen Versicherung sind mir bekannt (§156 StGB). Des Weiteren kann gemäß § 63 Absatz 5 HG eine Zuwiderhandlung mit einer Geldbuße geahndet werden.

Düsseldorf, den 20.05.2024



Zusammenfassung (in Deutsch und Englisch)

Deutsch:

Künstliche Intelligenz (KI) entwickelt sich in den letzten Jahren rasant weiter, auch in der Medizin. Insbesondere sind Fortschritte in der Bildanalyse zu vermerken. Die EKG-Diagnostik ist in dem Sinne eine Bildanalyse, als dass Kardiologen die in einem zweidimensionalen Bild dargestellte Wellenformen bewerten. Die hier vorliegende Arbeit stellt daher die Hypothese auf, dass eine KI, die ein sogenanntes Convolutional Neuronal Network (CNN) verwendet, auch EKG-Bilder und -Muster genau erkennen kann und dass ein Vorderwandinfarkt besser erkannt wird, als ein „Nicht-Vorderwandinfarkt“. Hierfür wurde eine EKG-Datenbank der Physikalisch-Technischen Bundesanstalt (PTB) mit 289 EKGs verwendet, darunter 148 Fälle mit Myokardinfarkt (MI), um ein CNN zur Erkennung von MI im EKG zu entwickeln. Unser CNN-Modell, das mit einer 6-Layer-Architektur ausgestattet ist, wurde mit Trainings-EKGs angeleitet. Danach wurden unser CNN und 10 Ärzte mit separaten, den beiden Gruppen unbekanntem EKGs getestet und ihre Fähigkeit zur Erkennung von Herzinfarkten anhand der Metriken F1 (harmonisches Mittel aus Sensitivität und positiv prädiktivem Wert) und Genauigkeit verglichen. F1 und Genauigkeit unseres CNN waren signifikant höher ($83\pm 4\%$, $81\pm 4\%$) im Vergleich zu den Ärzten ($70\pm 7\%$, $67\pm 7\%$, $P < 0,0001$). Auch die Eliminierung von Goldberger-Leitungen oder die Komprimierung des EKG-Bildes bis auf ein Viertel der Auflösung führte nicht zu einer signifikanten Verringerung der Erkennungsleistung.

Die vorliegende Arbeit konnte ebenfalls zeigen, dass eine KI mit *Deep Learning* Methoden mit einem einfachen CNN für die Bildanalyse bei der Erkennung von Herzinfarkten im EKG Ärzten im Fachgebiet Kardiologie nicht unterlegen ist. Zukünftige weitere Studien zum Einsatz von KI in der EKG-Bildauswertung sind notwendig, um den potentiellen Nutzen dieser Technik auch im klinischen Alltag weiter zu beleuchten.

Artificial intelligence (AI) has been developing increasingly and rapidly in medical technology, especially in image analysis, in recent years. ECG diagnosis is an image analysis in this sense, as cardiologists evaluate waveforms displayed in a two-dimensional image. Therefore, the work presented here hypothesizes that an AI using a so-called Convolutional Neural Network (CNN) can also accurately recognize ECG images and patterns. We used a ECG database by the Physikalisch-Technische Bundesanstalt (PTB) of 289 ECGs, including 148 cases of myocardial infarction (MI), to develop a CNN for detecting MI in the ECG. Our CNN model, which is equipped with a 6-layer architecture, was trained with training ECGs. Then, our CNN and 10 physicians were tested with test ECGs and their ability to detect MI was compared using the metrics of F1 (harmonic mean of precision and recall) and accuracy. F1 and accuracy of our CNN were significantly higher ($83\pm4\%$, $81\pm4\%$) compared with those of physicians ($70\pm7\%$, $67\pm7\%$, $P<0.0001$). Also, elimination of Goldberger leads or compression of the ECG image down to one quarter of the resolution did not significantly reduce recognition performance. Thus, the present work shows that an AI using Deep Learning methods with a simple CNN for image analysis may be not inferior to physicians in the specialty of cardiology in the detection of myocardial infarction in ECG. Future further studies on the use of AI in ECG image analysis are needed to further illuminate the potential utility of this technique in clinical practice.

Abkürzungsverzeichnis

Abkürzung	Erklärung
AI	Artificial intelligence
AUC	Area under the Curve (Statistisches Maß) Fläche unter der Kurve
CAD	Computer-assisted diagnosis
CNN	Convolutional neural Network
DL	Deep Learning
EKG	Elektrocardiogramm
F1	harmonisches Mittel aus Sensitivität und positiv prädiktivem Wert
Flip	Spiegeln
KI	Künstliche Intelligenz
Loss	Verlust
MI	Myokardinfarkt
NLP	Natural Language Processing – Sprachverarbeitung mittels Künstlicher Intelligenz
PTB	Physikalisch-Technische Bundesanstalt
ReLU	Rectified Linear Unit
ROC	Receiver-Operator-Characteristic
STEMI	ST-Hebungsinfarkt

Inhalt

Versicherung an Eides statt:.....	I
Zusammenfassung (in Deutsch und Englisch).....	II
Deutsch:	II
Englisch.....	III
Einleitung.....	- 1 -
Aspekte zum EKG	- 1 -
Aspekte zur künstlichen Intelligenz	- 2 -
Ziele der Arbeit:.....	- 5 -
Material und Methoden.....	- 7 -
EKG-Bearbeitung.....	- 11 -
Ableitungs-Set und Bildqualitätsvergleich	- 13 -
Training eines Convolutional Neural Network zur Erkennung eines MI im EKG	- 15 -
Beurteilungen der Genauigkeit.	- 15 -
Visualisierung der durch Deep Learning identifizierten Merkmale.....	- 16 -
Statistische Analyse.	- 16 -
Ergebnisse	- 17 -
Erkennungsfähigkeit der EKGs mit verschiedenen Ableitungseinstellungen.....	- 17 -
Erkennungsfähigkeit bei komprimierter Bildqualität.....	- 20 -
Lokalisationsdetektion des akuten ST-Hebungsinfarktes durch eine CNN-KI.....	- 22 -
KI-Erkennungsalgorithmus des STEMI (Grad-CAM).....	- 25 -
Diskussion.....	- 26 -
Schlussfolgerung und Ausblick.....	- 32 -
Verfügbarkeit der Daten.....	- 32 -
Quellenverzeichnis.....	- 33 -
Anhang.....	- 35 -
Danksagung.....	- 1 -

Einleitung

Aspekte zum EKG

Das EKG ist ein medizinisches Gerät, welches die Potenzialunterschiede, die durch die Elektronenverschiebungen bei Muskelkontraktionen entstehen, in Wellenform aufschreiben kann. Willem Einthoven entwickelte 1903 die erste Ableitung eines menschlichen EKGs mittels eines Saitengalvanometers. Die Probanden tauchten ihre Arme in eine wässrige Lösung oder erhielten feuchte Umschläge mit gesättigter Natriumchloridlösung, um einen Kontakt zu erstellen und die Potentialdifferenz mit einem Spannungsmesser aufzuzeichnen [1]. Die abgeleiteten Potentiale entstehen als Summenvektor der Herzmuskelzellen, die unter Depolarisation von einem negativen Membranpotential in eine positive Ladung übergehen. Durch das EKG können die an der Haut abgeleiteten Summationspotentiale abgeleitet und auf Papier in über die Zeit aufgezeichnet werden.

Zu Beginn der Erforschung des EKG gab es heftige Diskussionen welche Anteile der EKG-Welle welchen Muskelkontraktionen des Herzens entspricht [2]. Erste Rhythmusanalysen konnten jedoch auch schnell mit nur einer Ableitung durchgeführt werden [2]. Einthoven entwickelte seine Erfindung weiter und implementierte die Ableitungen Einthoven II (rechter Arm, linker Fuß) und Einthoven III (linker Arm, linker Fuß). Nun konnten Aussagen über die Herzachse, den Lagetyp und Ausbreitungs- oder Rückbildungsstörungen getroffen werden [3].

Eine Ableitung des EKGs über in Salzlösung getauchte Elektroden ermöglichte zwar einen guten elektrischen Kontakt, war jedoch in der Praxis aufwendig. 1925 konnten von F. Wilson die Elektroden weiterentwickelt werden. Zunächst wurden in gesättigte Salzlösung getauchte Binden um die Extremitäten gewunden – in einem weiteren Schritt wickelte Wilson Kupferdrähte über die Binden, sodass auch hier ein guter Kontakt entstand, ohne die Testperson an die Salzlösungsfässer zu fixieren.

Erst 1934 wurden von Frank Wilson die Brustwandableitungen entdeckt und 1942 die weiteren Extremitätenableitungen nach Goldberger, was dazu führte, dass erst im Jahre 1954 eine standardisierte Empfehlung in der Fachzeitschrift „Circulation“ veröffentlicht wurde, wie ein EKG geschrieben werden sollte [4].

Die Erfolgsgeschichte des EKGs ist also in der heutigen 12-Kanal-Version keine 100 Jahre alt, entwickelte sich in der Kardiologie jedoch schnell als Basisdiagnostikum. Die Vorteile, die es bietet, sind vielfältig: Es ist nichtinvasiv, indolent, schnell verfügbar und kostengünstig. Erfassen kann das EKG die Herzfrequenz, Herzrhythmus, den Lagetyp, sowie Erregungsausbreitungs- oder Rückbildungsstörungen [4]. Für das Schreiben eines EKGs werden heutzutage meist 10 Elektroden mit einem Unterdruckmechanismus am Patienten angebracht und binnen weniger Sekunden ein EKG abgeleitet [5]. Der Untersuchende erhält zügig eine Information über die oben aufgeführten Punkte. Besonders bei Verdacht auf Herzinfarkt ist ein EKG noch immer das Diagnostikum der Wahl um einen ST-Hebungs-Infarkt zu diagnostizieren oder auszuschließen. Weitere Pathologien, wie Schenkelblöcke, Arrhythmien oder Hypertrophiezeichen können somit in der Praxis, Notaufnahme oder im Notarzteeinsatz effizient erfasst werden[6].

In dieser Arbeit wurden als Ausgangsmaterial 12-Kanal-Oberflächen-EKGs verwendet, die mit einem Vorschub von 50 Millimetern pro Sekunde geschrieben wurden. In der Rhythmologie sind durchaus auch Schreibgeschwindigkeiten mit 25 Millimetern pro Sekunde üblich[6].

Der Vollständigkeit halber sei auch das Holter- oder Langzeit-EKG, ein Belastungs-EKG und das bei einer Elektrophysiologischen Untersuchung abgeleitete intrakardiale EKG erwähnt. Seit einiger Zeit sind zudem sogenannte „Wearables“ verfügbar, die ebenfalls die Möglichkeit bieten, neben der eigentlichen Funktion (z.B. als SmartWatch) die Möglichkeit bieten, ein EKG abzuleiten. Diese Art von EKG-Erfassung findet in dieser Arbeit keine Anwendung.

Aspekte zur künstlichen Intelligenz

Künstliche Intelligenz ist der Sammelbegriff für Anwendungen, bei denen Computer menschenähnliche Intelligenzleistungen erbringen. Darunter fallen unter anderem das maschinelle Lernen oder Machine Learning, das Verarbeiten natürlicher Sprache (NLP – Natural Language Processing) oder auch Deep Learning [7].

Künstliche Intelligenz kann man in starke- und schwache künstliche Intelligenz unterteilen [8]. Starke künstliche Intelligenz kann Dinge erlernen, verstehen und durchführen. Mit Dingen ist hier gemeint, dass sie nicht nur den Schach- oder Go-Weltmeister schlagen kann, sondern beispielsweise auch erlernen kann autonom einen

PKW zu fahren oder ein EKG zu interpretieren. Einer starken KI wäre es möglich ein eigenes Bewusstsein zu entwickeln, Probleme selbst zu erkennen und Lösungen für die Probleme zu suchen. Bisher existiert eine starke künstliche Intelligenz nicht [9].

Was es jedoch gibt sind schwache künstliche Intelligenzen. – Sie verfügen auf einem sehr eingeschränkten Feld über eine teilweise hohe Kompetenz. Die oben genannten Felder „Schach“, „Go“, „EKG-Befundung“ oder auch das „autonome Fahren“ können allesamt einzeln durch eine KI gelöst werden. Eine KI, die alle Aspekte abdeckt, existiert jedoch aktuell nicht. In der sehr einfachen Definition der KI wird nicht darauf eingegangen, wie eine KI an die Handlungsfähigkeit gelangt. Selbstlernen ist demnach nicht zwingend hierfür erforderlich – das vorherige Programmieren der KI ohne Lernzuwachs ist möglich.

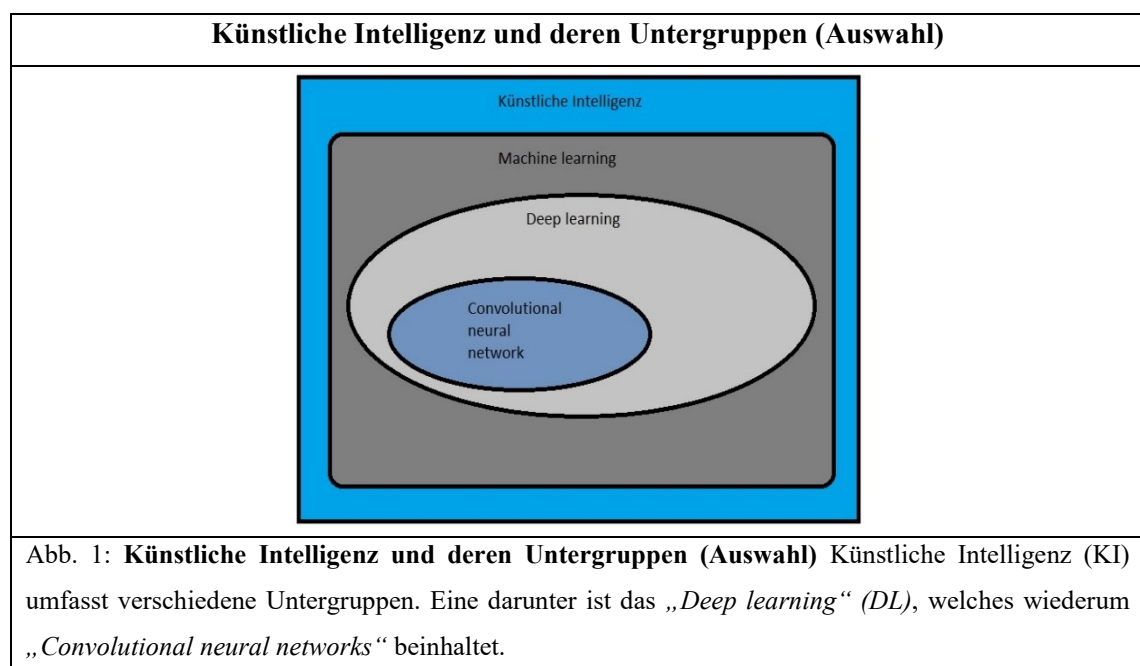
KI kann man in weitere Untergruppen unterteilen. Eine weitere Untergruppe der Künstlichen Intelligenz ist das Maschine Learning (Abb. 1). Es beinhaltet vor allem Konzepte wie eine Maschine lernen kann was gewünscht ist oder welche Dinge erlernt werden sollen. Vor allem ist hier das sogenannte „Supervised Learning“, „Unsupervised Learning“ und „Reinforcement-Learning“ zu nennen. Beim Supervised Learning wird der KI ein zu beurteilender Input und die Lösung gestellt [10]. Es wird beispielsweise also gelernt, Strukturen in Dreiecke und Kreise zu unterteilen – ausschließlich durch das Präsentieren von Datensätzen mit Dreiecken und Kreisen mit der passenden Lösung. Was einen Kreis oder ein Dreieck ausmacht wird aber nicht erklärt: Die KI muss dies selbst lernen. [10]

Beim unsupervised Learning werden nur Eingangsdaten verwendet, bei denen das Ergebnis noch nicht feststeht. Dieses Verfahren kann eingesetzt werden, um zu Kategorisieren und Gruppen zu bilden. Ergebnisse dieser Kategorien werden bewusst nicht vorgegeben, ergeben nicht immer Sinn oder kristallisieren Untergruppen unzureichend heraus – es ist jedoch eine gute Möglichkeit große Datenmengen zu bearbeiten und zu sortieren. Ein Beispiel wären Analysen in der Datascience um Ausreißer zu finden oder Vorgruppierungen zu treffen. [10]

Beim Reinforced learning kann durch Belohnen oder Bestrafen gelernt werden. Anwendungsgebiete wären das Optimieren von Arbeitsschritten oder das Erlernen des Spiels „Go“. Beim Verlieren kann eine Bestrafung oder bei Sieg eine Belohnung

ausgesprochen werden, bis die KI gegebenenfalls sogar in der Lage ist mit Weltmeistern zu konkurrieren.

Innerhalb des Deep learnings wird als Teilgebiet des Maschine learnings versucht, dem menschlichen Lernen noch näher zu kommen. Oft wird hierfür das Convolutional Neural Network (CNN) benutzt. Es enthält immer einige sogenannte Layer/Schichten, die an menschliche Neurone erinnern sollen. Üblich sind ein Inputlayer, welche Informationen aufnimmt, mindestens ein „Deep-“ oder Hidden-Layer welches die Lern und Verarbeitungsprozesse gestaltet – und ein Outputlayer, welcher für die Ausgabe zuständig ist.



Beim CNN werden die einzelnen Layer hintereinandergeschaltet und die Gewichtung der Vernetzungen bei jedem Lernvorgang neu gewichtet. Diese Art der Architektur empfindet das Gehirn am ehesten nach [11].

Anders als bei herkömmlichen Programmierungen wird beim Deep-Learning und bei Neuronalen Netzwerken kein Lösungsweg vorgegeben, vielmehr wird einer KI durch häufiges Wiederholen antrainiert, das Problem lösen zu können. Soll eine KI z.B. auf Bildern Dreiecke und Kreise auseinanderhalten können, so wird eine große Anzahl an Bildern von je Dreiecken und Kreisen benötigt, die schon korrekt beantwortet wurden [12].

Es werden nicht Programme mit Entscheidungsbäumen ausgestattet, um entscheiden zu können, um welche Figur es sich handelt. Vielmehr lernt die KI anhand der Bilder selbst, die Bilder voneinander zu differenzieren. Woran die KI ihre Entscheidung festlegt, ist am Ende nicht immer nachzuvollziehen [13]. Daher ist eine KI grundsätzlich anfällig für Confounding. Sollten alle Fotos mit Dreiecken mit blauem Hintergrund und alle Fotos mit Kreisen mit rotem Hintergrund ausgestattet werden, so ist die Wahrscheinlichkeit groß, dass die KI die Hintergrundfarbe mit hohem Gewicht in die Entscheidungsfindung integriert. Ähnliches gilt für die Häufigkeitsverteilung der einzelnen Fälle. Alle möglichen Entscheidungsmöglichkeiten müssen während des Trainings einer KI in ähnlicher Stückzahl vorliegen. Sollte man von 100 Fotos nur 2 von Dreiecken für die Trainingsphase nutzen, so würde eine starke Veranlagung der KI geschaffen sich meistens für „Kreis“ zu entscheiden, da statistisch gesehen diese Antwort meist richtig war. Es kann zusammengefasst werden, dass für das Training einer KI hohe Anforderungen an die Trainingsdaten gestellt werden. Es müssen Confounder möglichst geringgehalten und die einzelnen Fälle gleichmäßig verteilt werden [14]. Positiv ist zu vermerken, dass dadurch, dass nicht zwischen mehreren Lösungsmöglichkeiten entschieden, sondern der Lösungsprozess offen ist, auch Lösungen für bisher ungelöste Aufgaben gefunden werden können.

Auch im medizinischen Bereich wurden diagnostische und therapeutische Fähigkeiten der KI rasch entwickelt und integriert [15]. Im Bereich der Kardiologie wurde kürzlich über den Nutzen der KI für die Echokardiografie und auch für die Elektrokardiografie (EKG) unter Verwendung von Big Data berichtet [16, 17]. Praktisch ist der Einsatz dieser Technologien aufgrund lokaler Rechenleistung und begrenzter menschlicher Ressourcen noch im Ausbau.

Ziele der Arbeit:

In aktuellen kardiologischen Leitlinien ist die Interpretation eines EKGs elementarer Bestandteil [18]. Auch die zeitliche Relevanz einer solchen Befundung ist beispielsweise in der Leitlinie zum akuten ST-Hebungsinfarkt (STEMI) hinterlegt – es soll hier ein EKG binnen 10 Minuten nach Eintreffen des Patienten vorliegen [18]. Die Befundung eines EKGs kann jedoch komplex sein und stellt vor allem bei schrittmachergeführten Patienten oder schon vorbekannten Leitungsstörungen wie Schenkelblöcken die ärztlichen Kollegen vor Herausforderungen. Die künstliche Intelligenz, welche in den letzten Jahren zunehmend auch im medizinischen Bereich an Bedeutung zunimmt, kann,

nach entsprechendem Training, ebenfalls EKGs befunden [19]. Auf Basis einer EKG-Datenbank soll in dieser Arbeit eine KI erstellt werden, die aus einer einfachen neuronalen Netzwerkarchitektur besteht und anhand der EKGs als "Bilder" und nicht anhand elektrischer Signale (EKG-Rohdaten) darauf trainiert wurde, ST-Hebungsinfarkte (STEMIs) zu erkennen. Die Leistung der KI soll bewertet werden, indem erfasst wird, wie verlässlich diese Myokardinfarkte detektieren kann. Auch der Einfluss der EKG-Ableitungsreduktion und der Bildkompression auf die Erkennungsleistung der KI werden evaluiert. Zusätzlich werden die unterschiedlichen Lokalisationen eines Herzinfarkts verglichen. Als Vergleichsgruppe wird eine Gruppe von 10 Ärzten herangezogen, die sich aus 5 Assistenzärzten und 5 Fachärzten zusammensetzt.

Diese Arbeit soll daher insgesamt die Machbarkeit der praktischen Einführung dieser KI-Systeme bei mäßigen Rechenkapazitäten und Ressourcen in der EKG-Diagnostik des Myokardinfarktes helfen zu bewerten.

Die Hypothese dieser Arbeit ist, dass eine KI zumindest gleich gut oder besser als in der EKG-Diagnostik erfahrene Ärzte einen akuten ST-Hebungsinfarkt erkennt, selbst wenn die KI mit einer einfachen neuronalen Netzwerkarchitektur ausgestattet ist und mit einer kleinen Datenbank trainiert wird.

Als zusätzliche Hypothese im Sinne einer untergeordneten Fragestellung wird untersucht, ob die Lokalisation des Myokardinfarktes eine Auswirkung auf die Genauigkeit der Befundung der KI im Vergleich zu der Gruppe der Ärzte aufweist.

Material und Methoden

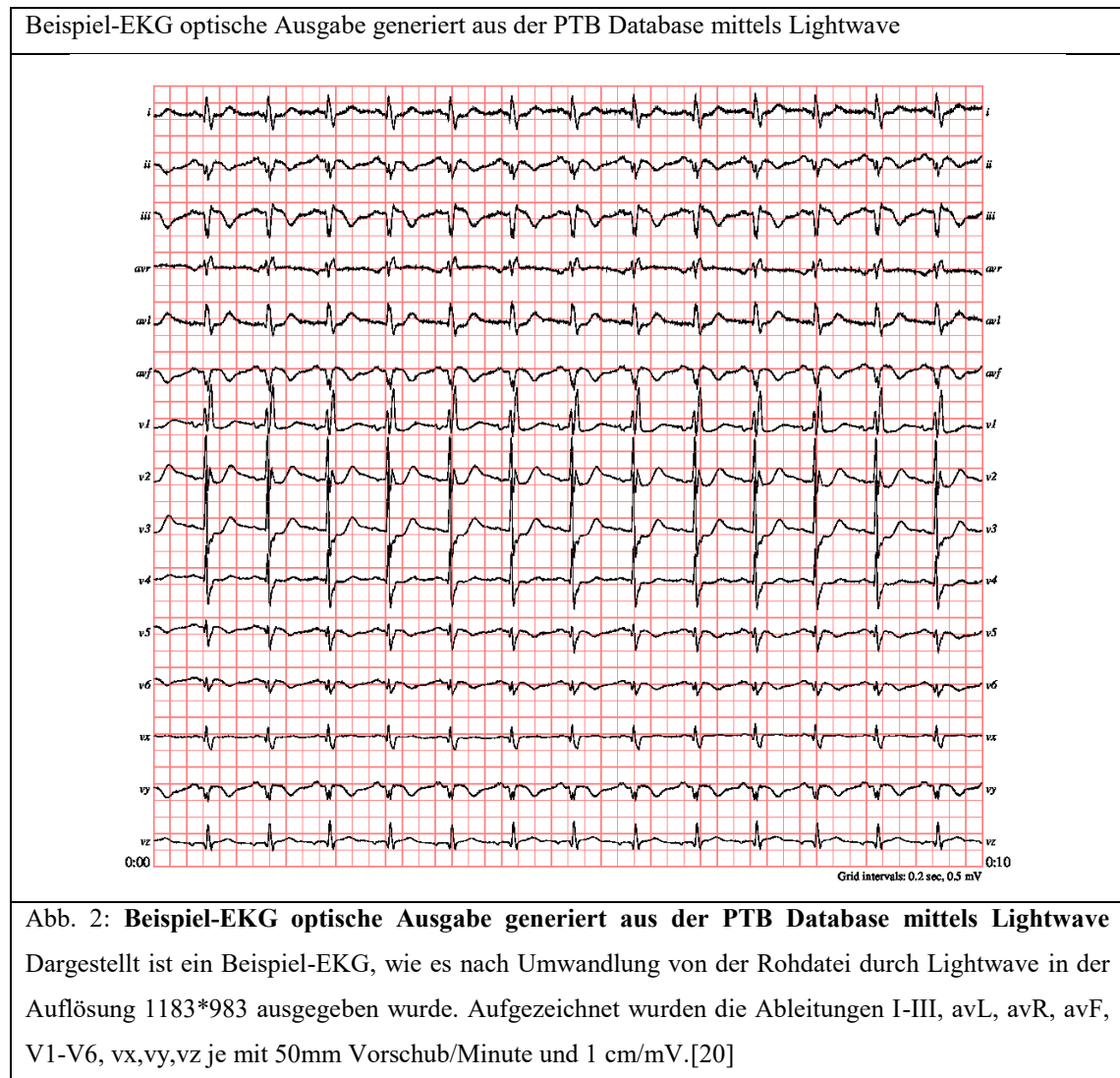
Für die Durchführung unseres Projekts benötigten wir einen möglichst großen Datensatz von zuverlässig befundeten 12-Kanal-EKGs mit und ohne Herzinfarktlokalisierung. Es sollten weiterhin Grunderkrankungen, sowie eine Infarktlokalisierung erhoben worden sein. Wir fanden entsprechende Daten im PTB-ECG-Dataset [20]. Die Datenbank enthält 549 Datensätze von 290 Probanden (im Alter von 17 bis 87 Jahren, Durchschnittsalter 57,2 Jahre; 209 Männer, Durchschnittsalter 55,5 Jahre, und 81 Frauen, Durchschnittsalter 61,6 Jahre; das Alter von einer Frau und 14 Männern wurde nicht erfasst). Jeder Proband ist mit einem bis fünf Datensätzen vertreten. Jeder Datensatz enthält 15 gleichzeitig gemessene Signale: die konventionellen 12 Ableitungen (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) sowie die 3 Frank-Ableitungs-EKGs (vx, vy, vz). Jedes Signal wurde mit 1000 Abtastungen pro Sekunde und einer Auflösung von 16 Bit über einen Bereich von $\pm 16,384$ mV digitalisiert. Ein Beispieldatensatz bestehend aus einem EKG und einer dazugehörigen Textdatei befindet sich im Anhang.

Die dazugehörige Textdatei (auch Header-Datei (.hea) genannt) der meisten dieser EKG-Aufzeichnungen enthält eine detaillierte klinische Zusammenfassung, einschließlich Alter, Geschlecht, Diagnose und gegebenenfalls Daten zur Krankengeschichte, Medikation und Interventionen, Koronararterienpathologie, Ventrikulographie, Echokardiographie und Hämodynamik. Die klinische Zusammenfassung ist für 22 Probanden nicht verfügbar, diese Patienten wurden exkludiert. Die Diagnoseklassen der übrigen 268 Probanden werden im Folgenden zusammengefasst (Tab. 1).

Patientenkollektiv der genutzten Datenbank	
Diagnose	Anzahl
Myokardinfarkt	148
Kardiomyopathie	18
Schenkelblock	15
Herzrhythmusstörungen	14
Myokardiale Hypertrophie	7
Valvuläre Herzkrankheit	6
Myokarditis	4
Sonstiges	4
Gesunde Kontrollgruppe	52

Tab. 1: **Patientenkollektiv der genutzten Datenbank "PTB ECG Database"** [20] Aufgezeigt wird Erkrankungen der eingeschlossenen Patienten

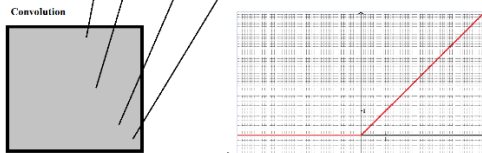
Zu jedem in Tabelle 1 erwähnten Patientenfall existiert ein EKG, welches wie Abbildung 2 zu sehen ist, exportiert wurde. Für den Export der Rohdaten in Bildmaterial nutzen wir das Programm Lightwave, welches von Physionet.org zur Verfügung gestellt wird [21, 22].



Aufbau des Convolutional Neural Network (CNN)

Aufbau des Convolutional Neural Network (CNN) mit der Abfolge der einzelnen Layer

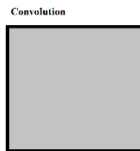
Eingabe (EKG, 1183 x 983 Pixel)



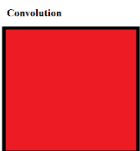
- Batch Normalisation
- ReLu
- Max Pooling



- ReLu
- Max Pooling

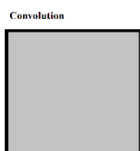


- ReLu
- Max Pooling

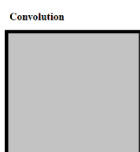


Der vierte Convolution- Schritt wurde bei niedrigen Auflösungen ausgelassen

- ReLu
- Max Pooling



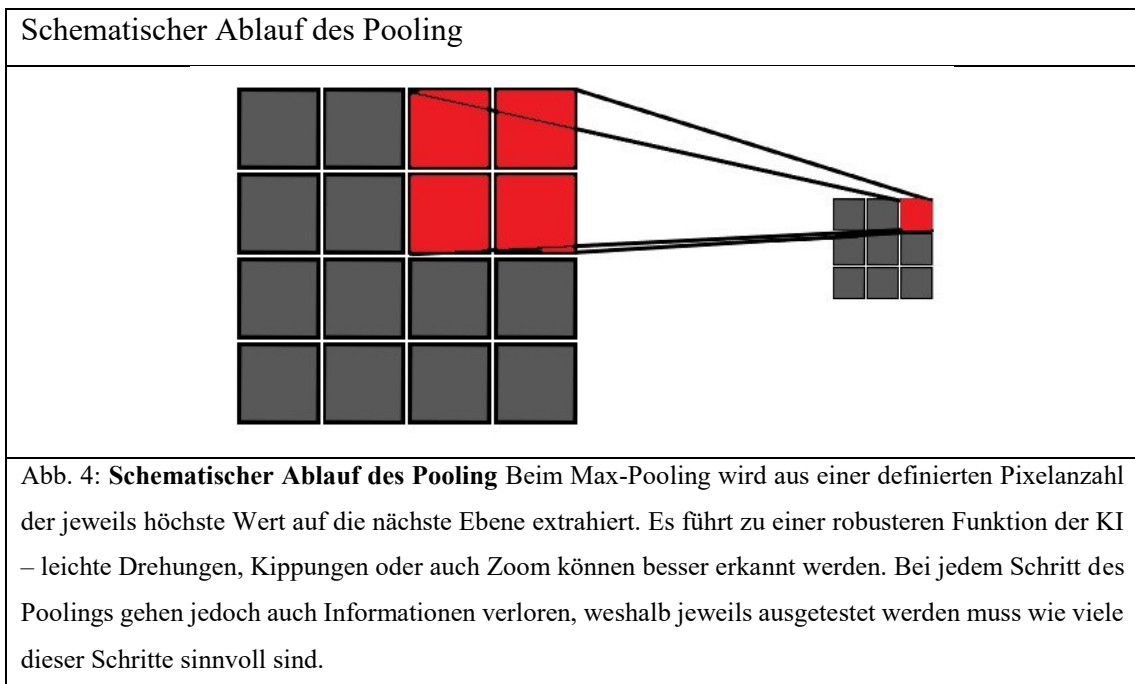
- ReLu
- Max Pooling



- Max Pooling
- Dropout
- Affine
- Batch Normalisation
- ReLu
- Dropout
- Affine
- Sigmoid

Beginnend mit dem Convolution-Schritt, werden bei unserem CNN die Eingabebilder im zu Beginn durch eine zu definierende gewichtete Faltungsmatrix neu gewichtet. Die Faltungsmatrix tastet jeden Pixel ab und gewichtet, je nach benachbarten Pixeln, den aktuell zu untersuchenden Pixel [23]. Das Prinzip erinnert so an die Verschaltung der Netzhaut im Auge (rezeptives Feld). [24] Durch diesen Schritt können Bildanteile deutlich besser analysiert und wiederkehrende Muster erkannt werden. Lokale, relative Bedeutungen, wie beispielsweise ein Kreis oder ein Dreieck oder eine ST-Streckenhebung – kann so unabhängig von der Position auf dem Bild erkannt werden.

Abb. 3: Schematischer Aufbau des Convolutional Neural Network. Zu Beginn eine Eingabedatei als Bild im PNG-Format. Für die Eingabedateien in niedrigerer Auflösung nutzten wir bei der Versuchsreihe mit unterschiedlichen Bildauflösungen einen Convolution-Schritt weniger (Convolution 4). Dieser wurde ersatzlos gestrichen.



Max-Pooling ist ein Verfahren, welches in CNN die Informationsdichte komprimieren kann [25]. Es werden in zweidimensionalen Layern benachbarte Pixelinformationen zusammengefasst, um weniger Rechenleistung zu benötigen und somit einen höheren Datendurchsatz zu ermöglichen. Zusätzlich führt es zu zuverlässigeren Ergebnissen bei Varianten in der Bilderkennung wie gekippten-, gestreckten- oder vergrößerten Bildern (Abb. 4).

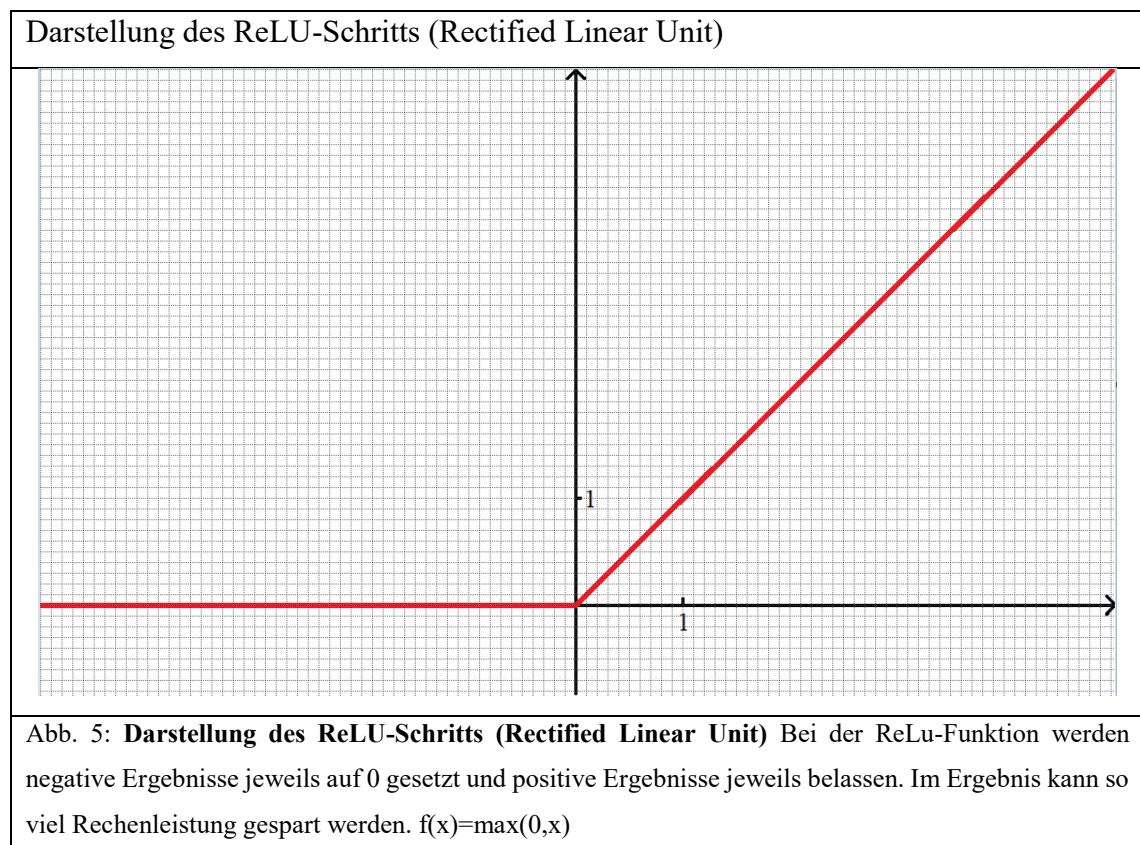
Der zentrale Schritt beim Max Pooling ist, dass aus einer definierten Anzahl an Pixeln eine Reduktion auf eine niedrigere Auflösung erzielt wird. Hierfür werden die jeweils höchsten Werte der definierten Pixelgruppe ausgewählt und auf einen neuen Datensatz übertragen (Abb. 4).

Andere Verfahren, wie Average-Pooling (Durchschnittswert von benachbarten Pixeln wird gemittelt) oder Global-Pooling lieferten in unserem, speziellen Fall schlechtere Ergebnisse – prinzipiell sind diese Verfahren aber nicht unterlegen [12].

Bei zu großer Reduktion der Daten durch Max-Pooling kann ein deutlich verschlechtertes Ergebnis auftreten. Dieses Phänomen manifestiert sich häufig, wenn zu viele Layer mit je Max-Pooling-Verfahren hintereinandergeschaltet werden. Bei uns trat dieses Phänomen auf, wenn wir als Input-Datei eine weiter in der Auflösung reduzierte Bilddatei nutzen. Für die reduzierten Ableitungen nutzten wir, wie unter Methoden beschrieben,

Ausschnitte der EKGs, welche die Auflösung entsprechend reduzierte. Für die komprimierten Auflösungen wurde bewusst die Auflösung vermindert, um zu testen, wie es die Erkennungsleistung verändert. Um das Problem zu lösen verwendeten, wir entweder eine Auflösung von 744*368 Pixeln mit 6 Max-Pooling Schritten oder bei den Durchläufen mit reduzierten Ableitungszahlen einen Max-Pooling-Schritt weniger.

Rectified Linear Unit-Funktionen (ReLU) begrenzen die neuronalen Aktivierungswerte auf positive Werte. Ursprünglich wurden sigmoidale Funktionen genutzt, die jedoch höhere Rechenkapazität kosten und weitere Probleme mit sich brachten und auch Maximalwerte veränderten. ReLU beschneidet negative Werte und belässt positive Werte unbearbeitet (siehe Abbildung 5): $f(x)=\max(0,x)$.



EKG-Bearbeitung

Genutzt wurde die PTB-EKG-Datenbank die 549 Datensätze von 12-Kanal-EKGs von 290 Patienten im PNG-Format (1183*983) enthält [20, 21]. Diese Datenbank wurde bereits durch eine unbeteiligte dritte Partei online veröffentlicht – daher gab es keine ethischen Bedenken bezüglich der Offenlegung der Daten. Nach Download aller EKGs

- 11 -

der Datenbank wurden diese für unsere Studie vorbereitet. Entnommen wurden 289 Patientendatensätze, wovon ein Datensatz ohne EKG vorlag. Die erschlossenen EKGs bestanden aus 148 EKGs mit der Diagnose eines Akuten Myokardinfarkts (MI), mit Lokalisationsprotokoll; weiter enthielt der Datensatz 141 EKGs ohne MI (klassifiziert als „Non-MI“) (Tab. 1). Um das Datenset für jeden Versuchsdurchlauf vorzubereiten, schnitten wir die EKG-Bilder (Auflösung 744*368) jeweils aus (Abstände in Pixeln: oben 36, links 202, unten 780, rechts 570) (Abb. 6a).

Für die Untersuchung der Haupthypothese sowie der Nebenfragestellung wurden unterschiedliche methodische Ansätze gewählt:

1) Detektionsgüte des akuten ST-Hebungsinfarkts durch eine CNN-KI

Von den Originaldaten wurden randomisiert ein Test-Set (n=25) und ein Validierungsset (n=25) gezogen. Die Stichprobengröße wurde durch Oversampling ausgeglichen zwischen MI und Non-MI, um ein Trainingsset zu erhalten (je n=108) (Bild 2b). Für das Oversampling wurden die Original-EKGs nochmals an einer bisher nicht genutzten Stelle ausgeschnitten (Abstände: oben 36, links 664, unten 780, rechts 1032) (Stelle „B“ in Bild 6 a). Um Reproduzierbarkeit zu erreichen, bereiteten wir 10 verschiedene Trainings-/Validations-/Test-Sets vor [14].

2) Lokalisationsdetektion des akuten ST-Hebungsinfarktes durch eine CNN-KI

Für die additive Fragestellung legten wir den Fokus auf die stärkere Vergleichbarkeit zwischen den Ärzten und der KI und den Infarktlokalisationen „Vorderwand“ und „Nicht-Vorderwand“, sodass wir ein einheitliches Test-Set etablierten. Wir nutzen die insgesamt 289 Patientendatensätze und teilten wieder in die 148 Datensätze mit der Diagnose eines MI und 141 Datensätze ohne Diagnose eines MI und zogen per Randomisierung 38 Datensätze mit MI und 34 Datensätze ohne MI als Test-Set. Ein Validierungs-Set wurde analog mit je 20 Datensätzen erzeugt. Für das Trainings-Set verblieben noch 88 Datensets mit- und 89 Datensets ohne MI. Für dieses Set nutzten wir ebenfalls Oversampling, nach oben beschriebenem Muster und erreichten ein Trainingsset mit 224 Datensätzen.

Ableitungs-Set und Bildqualitätsvergleich

Wir erstellten die Varianten der Trainings-/Validations-/Test-Sets aus unseren Original-EKG-Sets; es wurden jeweils die Ableitungen (Abb. 6 c) und/oder die Auflösung (Abb. 6d) verändert. Für das Erstellen der EKG-Sets mit reduzierter Anzahl an Ableitungen haben wir 5 Varianten in voller Bildqualität vorbereitet; Original (12 Ableitungen), 9-, 7-, Extremitäten- und präkordiale Ableitungen. Das 9-Ableitungs-EKG bestand aus den Ableitungen I, II, III und V1-6. Das 7-Ableitungs-EKG bestand aus den Ableitungen I, II, III, V1-2 und V5-6.

Für unsere Bildqualitätsvariationen haben wir das 9-Kanal-EKG übernommen und seine Bildqualität in 4 verschiedenen Verhältnissen komprimiert; Verhältnisse: volle Qualität (744*368), halbe Qualität (398*260), Viertelqualität (282*184) und quadratische Form (224*224)

EKGs aus dem Datensatz mit entsprechender Datenverarbeitung

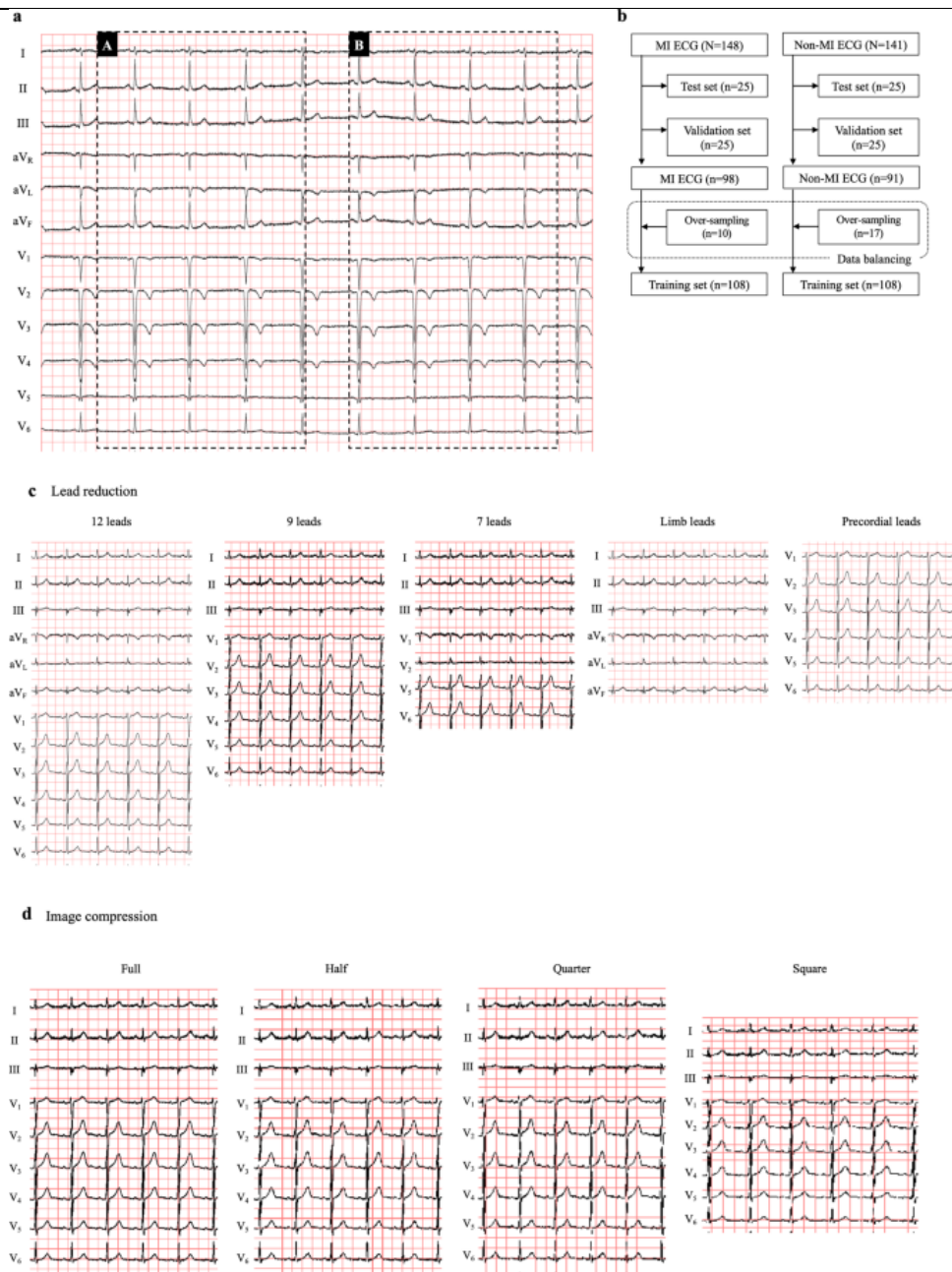


Abb. 6: EKG mit entsprechender Datenverarbeitung. a. Ausschnitte, die für die einzelnen EKGs gewählt wurden. „A“ wurde standardmäßig gewählt, „B“ für das Generieren von zusätzlichen Datensätzen für das Oversampling. b. Generierung der jeweiligen Test-, Validations- und Trainingssets. c. Exemplarische Darstellung der Ausschnitte der EKGs, die für reduzierte Ableitungen gewählt wurden. d. Auflösungen in voller (full), halber (half), viertel (quarter) und komprimierter/gestauchter (square). Reproduktion der Abbildung mit Genehmigung des Rechteinhabers [26].

Training eines Convolutional Neural Network zur Erkennung eines MI im EKG

Wir entwickelten ein 6-Layer neuronales Netzwerk als CNN-Modell für die vorliegende Studie mit Deep-Learning-Architektur, was eine Form des maschinellen Lernens ist, die entwickelt wurde, um das visuelle System zu imitieren [27]. Für jedes Trainings-/Validierungs-/Test-EKG-Set wurde dieses CNN-Modell anhand eines Trainingsatzes von 216 EKG-Bildern (MI: non-MI=108:108) mit dem Ergebnis mit dem Ergebnis "Ja (MI)" oder "Nein (Non-MI)" trainiert und mit einem Validierungsset (MI: Non-MI=25:25) validiert. Diese Verwendung des Validierungssets sollte ein Overfitting während des Trainings des CNN vermeiden. Nur während des Trainings des CNN wurde Datenaugmentation verwendet (zufällige Drehung [-4 bis 4 Grad], Breitenverschiebung [$\pm 5\%$], Scherung [Scherwinkel gegen den Uhrzeigersinn: bis zu 0,05 Grad]), um die Lernfähigkeit zu erhöhen. Die Randbereiche, die außerhalb der Bildgrenze lagen, wurden nach dem "nearest"-Modus ausgefüllt. Als Methoden zur Datenaugmentation wurden Flip (Spiegeln), Zoom und Helligkeitsvariationen nicht verwendet. Nach dem Training testeten wir die Leistung des CNN-Modells mit einer Stichprobengröße von 50 EKGs (MI: non-MI=25:25). Die Ausgabe dieses Modells war: „Ja (MI)“ oder „Nein (non-MI)“ zu unbekanntem EKGs, d. h. EKG-Bildern, die beim Training nicht verwendet wurden (Abb. 6b). Wir wiederholten diesen Trainings- und Testprozess 10 Mal für 10 verschiedene Trainings-/Validierungs-/Test-EKG-Sets, um seine Reproduzierbarkeit zu gewährleisten. Das Training des CNN wurde unabhängig für jedes Trainings-/Validierungs-/Test-EKG-Set und jede Variante des Trainings-/Validierungs-/Test-EKG-Sets (reduzierte Ableitungen, Bildkomprimierung) auf der Grundlage des Validation-Loss durchgeführt, um ein Modell mit dem geringsten Validation-Loss zu wählen. (Vgl. Tab. 5)

Beurteilungen der Genauigkeit.

Für die Metriken zur Bewertung der Fähigkeit des CNN oder der Ärzte, EKGs zu erkennen, haben wir das F1-Maß (harmonisches Mittel aus Sensitivität und positiv prädiktivem Wert) und die Genauigkeit in Prozent angenommen. Die Fläche unter der Kurve (AUC) der Receiver-Operator-Characteristic (ROC)-Kurve wurde ebenfalls bewertet. Diese Metriken wurden für jedes Trainings-/Validierungs-/Test-EKG-Set (n=10) berechnet.

Für das Hauptprojekt beantworteten zehn Ärzte je 2 Test-Sets á 50 12-Kanal-EKGs (insg. 100 EKGs). Es gab 10 verschiedene Test-Sets, die auch bei den Tests für CNN verwendet

wurden. Die Ärzte setzten sich zusammen aus 5 Fachärzten (darunter 3 Oberärzte) und 5 Assistenzärzten (darunter 2 Ärzte mit weniger als 3 Jahren Berufserfahrung). Alle 10 Ärzte waren mindestens 1 Jahr in der medizinischen Notfallversorgung tätig, einschließlich der Behandlung des akuten Koronarsyndroms.

Jeder Testsatz wurde jeweils einem Facharzt und einem Assistenzarzt randomisiert zugewiesen. Die Ärzte beurteilten, ob der Patient eines jeden EKGs einen MI erlitten hatte oder nicht. Die Antworten der Ärzte für jeden Datensatz wurden erhoben und die Mehrheitsmeinung als Konsens der Ärzte betrachtet. Wenn die Meinungen gleichmäßig verteilt waren, wurde die Meinung der zertifizierten Kardiologen mit doppeltem Gewicht gewertet.

Für das Zusatzprojekt wurden 72 EKGs nach dem Zufallsprinzip ausgewählt und alle zehn Kardiologen beurteilten, ob bei diesen 72 Patienten ein MI vorlag oder nicht. Diese Bewertung diente dazu, herauszufinden, bei welchen EKG alle Kardiologen richtig oder falsch geurteilt haben und um die Vergleichbarkeit zwischen KI und Ärzten zu erhöhen.

Die Daten zu den Infarktlokalisationen wurden ebenfalls aus der PTB-Datenbank extrahiert. Die Erkennungsgenauigkeit der Infarktlokalisation wurde ebenfalls bewertet. Für das Zusatzprojekt entschieden wir uns in “Vorderwand-” und “Nicht-Vorderwandinfarkt” zu unterscheiden.

Visualisierung der durch Deep Learning identifizierten Merkmale. Wir erstellten Aktivierungskarten der letzten CNN-Layer mit Grad-CAMs, die die relative positive Aktivierung dieses Layers in Bezug auf die Eingabe darstellt[28]. Über diesen Weg kann gezeigt werden, welche Bereiche durch die KI besonders betrachtet worden sind, um das EKG zu befunden.

Statistische Analyse.

Normalverteilte Daten wurden als Mittelwert \pm Standardabweichung angegeben. Bei nicht normalverteilten Daten wurde der Medianwert mit dem unteren - und oberen Quartil angegeben. Anteile wurden als Zahlen und Prozentsätze angegeben. Gegebenenfalls wurden ein Chi-Quadrat-Test, ein Kruskal-Wallis-Test, ein Student-t-Test, ein exakter Test nach Fisher oder eine einseitige Varianzanalyse durchgeführt. Wenn ein multivariater Test statistische Signifikanz ergab, wurde eine Post-hoc-Analyse unter Verwendung des HSD-Tests von Turkey durchgeführt. Für globale Teststatistiken

wurde ein Signifikanzniveau von 5 % angenommen. Die Analysen wurden mit Python 3.6.0 oder JMP (SAS, Version 11) sowie Prism und Excel durchgeführt.

Ergebnisse

Erkennungsfähigkeit der EKGs mit verschiedenen Ableitungseinstellungen. Die Erkennungsleistung unseres CNN und der Ärzte ist für das 12-Kanal-Standard-EKG in hoher Auflösung in Abb. 7a dargestellt. Das CNN erreichte im Vergleich zu den Ärzten eine signifikant höhere F1 und Genauigkeit ($83\pm 4\%$ vs. $70\pm 7\%$, $p < 0,0001$; $81\pm 4\%$ vs. $67\pm 7\%$, $p < 0,0001$). Die Sensitivität ($86\pm 7\%$ vs. $67\pm 10\%$, $p < 0,0001$), die Spezifität ($76\pm 7\%$ vs. $67\pm 10\%$, $p = 0,018$) und der negative Prädiktive Wert ($85\pm 7\%$ vs. $60\pm 10\%$, $p < 0,0001$) waren bei unserem CNN signifikant höher als bei den Ärzten. Der positive prädiktive Wert unseres CNN war tendenziell höher als der der Ärzte ($79\pm 5\%$ vs. $74\pm 6\%$, $p = 0,053$). Auch wenn wir die Leistung des CNN mit der von Fachärzten verglichen, waren die Ergebnisse des CNN im Vergleich zu denen von Fachärzten höher oder zumindest nicht schlechter.

Genauigkeit der KI im Erkennen von ST-Hebungsinfarkten im Vergleich zu Ärzten und bei unterschiedlichen EKG-Ableitungen, sowie Auflösungen

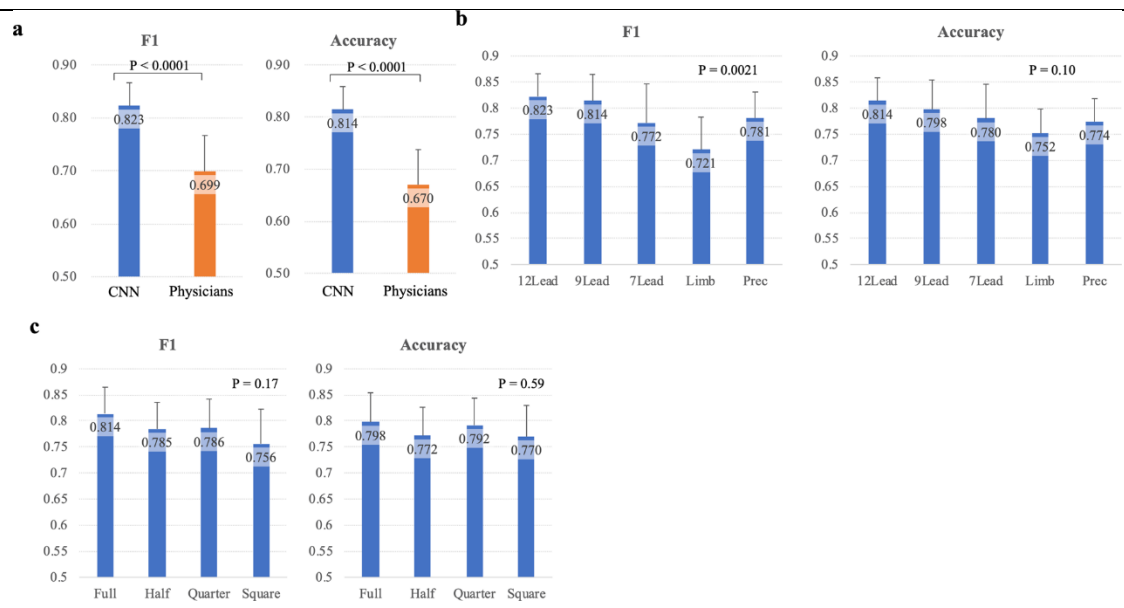


Abb. 7: Genauigkeit der KI im Erkennen von Herzinfarkten im Vergleich zu Ärzten und bei unterschiedlichen EKG-Ableitungen, sowie Auflösungen Vergleich der Erkennungsleistung von Herzinfarkten im EKG von einem Convolutional Neural Network (CNN) und Ärzten (Physicians). Verglichen wird über zwei unterschiedliche Statistiken: Dem F1 Wert (Mittelwert aus Sensitivität und positiv prädiktivem Wert) sowie der Genauigkeit (Verhältnis der richtig Beantworteten Fälle zu allen Fällen). Die Erkennungsleistung des CNN war signifikant höher, als die der Ärzte. (F1 Wert 0.788 ± 0.056 vs 0.699 ± 0.068 , $P < 0.0001$; Genauigkeit 0.788 ± 0.052 vs 0.67 ± 0.067 , $P < 0.0001$). (b) Der Einfluss der Reduktion der EKG-Ableitungen wird gezeigt. Das 6-Layer CNN mit Extremitätenableitungen zeigte einen signifikant niedrigeren F1 Wert als mit 12- oder 9-Kanal-EKG. (0.721 ± 0.062 vs 0.823 ± 0.043 , $P = 0.0022$; 0.721 ± 0.062 vs 0.814 ± 0.051 , $P = 0.0058$). Die Reduktion von 12- auf 9- und 7- EKG-Ableitungen führten zu keiner relevanten Reduktion der Genauigkeit. (c) Auf eine Reduktion von der vollen (Full) über die halbe Auflösung (half) bis auf eine viertel-Auflösung führten zu keiner Signifikanten Reduktion der Genauigkeit oder des F1-Werts. Die Kompression auf eine Quadratische Form führte zu keiner signifikanten Reduktion der Erkennungsleistung. Reproduktion der Abbildung mit Genehmigung des Rechteinhabers [26]

Der F1-Wert und die Genauigkeit des Ärztekonsenses waren dem Durchschnitt der Einzelärzte nicht überlegen (0,67 bzw. 0,69). Im Hinblick auf die Erkennung von MI nach Infarktort wurde die Genauigkeit der MI-Fälle in den Test-EKG-Sets bewertet. Es gab keine signifikanten Unterschiede in der Genauigkeit in Abhängigkeit von der Infarktlokalisierung ($P=0,27$).

Erkennungsraten der KI bei unterschiedlichen EKG-Ableitungen [26]						
	12 Ableitungen	9 Ableitungen	7 Ableitungen	Körper- ableitungen	Brustwand- ableitungen	P Wert
Sensitivität	0.86 ± 0.07	0.88 ± 0.07	0.75 ± 0.10	0.65 ± 0.09	0.81 ± 0.09	<0.0001
Spezifität	0.76 ± 0.07	0.71 ± 0.10	0.81 ± 0.10	0.86 ± 0.07	0.74 ± 0.09	0.0047
PPV	0.79 ± 0.05	0.76 ± 0.07	0.80 ± 0.08	0.82 ± 0.06	0.76 ± 0.06	0.13
NPV	0.85 ± 0.07	0.87 ± 0.08	0.77 ± 0.07	0.71 ± 0.05	0.80 ± 0.06	<0.0001
F1-Wert	0.82 ± 0.04	0.81 ± 0.05	0.77 ± 0.07	0.72 ± 0.06	0.78 ± 0.05	0.0021
Genauigkeit	0.81 ± 0.04	0.80 ± 0.06	0.78 ± 0.07	0.75 ± 0.05	0.77 ± 0.04	0.10
<i>Loss</i>	1.56 ± 0.31	1.50 ± 0.19	1.41 ± 0.23	1.44 ± 0.18	1.41 ± 0.20	0.55
AUC (Area under the Curve)	0.88 ± 0.05	0.88 ± 0.04	0.86 ± 0.05	0.85 ± 0.05	0.87 ± 0.04	0.45
Genauigkeit (Training)	0.89 ± 0.06	0.91 ± 0.05	0.88 ± 0.06	0.88 ± 0.04	0.88 ± 0.03	0.39
AUC (Training)	0.97 ± 0.03	0.98 ± 0.02	0.96 ± 0.03	0.96 ± 0.03	0.97 ± 0.01	0.29
Genauigkeit (Validation)	0.85 ± 0.06	0.85 ± 0.06	0.82 ± 0.05	0.82 ± 0.03	0.82 ± 0.03	0.40
AUC (Validation)	0.93 ± 0.05	0.93 ± 0.05	0.91 ± 0.06	0.92 ± 0.05	0.92 ± 0.05	0.89
Tab. 2: Erkennungsraten der KI bei unterschiedlichen EKG-Ableitungen						

Die Erkennungsleistung unseres CNN mit reduzierten EKG-Ableitungen ist in Abb. 7b und Tabelle 3 dargestellt. Das F1-Maß des CNN, das mit Extremitäten-EKGs trainiert wurde (72±6%), war signifikant niedriger als das mit 12-Kanal-EKGs (82±4%, p=0,0022) und 9-Kanal-EKGs (81±5%, p=0,0058). Es gab jedoch keinen signifikanten Unterschied in der Genauigkeit zwischen der Anzahl der verwendeten EKG-Ableitungen (p=0,10). Alle CNN befundeten EKGs effizienter als Ärzte, außer wenn nur Extremitäten-EKGs verwendet wurden (höhere F1 und Genauigkeit als menschliche Ärzte (p<0,05)).

Zwischen den CNNs mit unterschiedlicher Anzahl von EKG-Ableitungen gab es keine signifikanten Unterschiede bei den Test-Loss und der AUC von EKG-Ableitungen (Tab. 2). Es gab auch keine signifikanten Unterschiede in der Genauigkeit und AUC der Trainings- und Validierungssets am Ende des Trainings.

Erkennungsraten der KI bei unterschiedlichen Auflösungen des EKGs [26]					
	Volle Auflösung	Halbe Auflösung	Viertel Auflösung	Quadratisch	P Wert
Sensitivität	0.88 ± 0.07	0.84 ± 0.10	0.77 ± 0.10	0.72 ± 0.10	0.0021
Spezifität	0.71 ± 0.10	0.71 ± 0.12	0.82 ± 0.10	0.82 ± 0.09	0.022
PPV	0.76 ± 0.07	0.75 ± 0.08	0.82 ± 0.09	0.81 ± 0.09	0.19
NPV	0.87 ± 0.08	0.82 ± 0.09	0.79 ± 0.07	0.75 ± 0.07	0.016
F1-Wert	0.81 ± 0.05	0.78 ± 0.05	0.79 ± 0.06	0.76 ± 0.07	0.17
Genauigkeit	0.80 ± 0.06	0.77 ± 0.05	0.79 ± 0.05	0.77 ± 0.06	0.59
Loss	1.50 ± 0.19	1.77 ± 0.28	1.48 ± 0.18	1.46 ± 0.22	0.0082
AUC (Area under the Curve)	0.88 ± 0.04	0.87 ± 0.05	0.87 ± 0.05	0.85 ± 0.05	0.48
Genauigkeit (Training)	0.91 ± 0.04	0.92 ± 0.04	0.86 ± 0.05	0.83 ± 0.06	0.0003
AUC (Training)	0.98 ± 0.02	0.99 ± 0.02	0.95 ± 0.04	0.93 ± 0.05	0.0004
Genauigkeit (Validation)	0.85 ± 0.06	0.81 ± 0.07	0.80 ± 0.07	0.79 ± 0.07	0.17
AUC (Validation)	0.93 ± 0.05	0.92 ± 0.04	0.87 ± 0.07	0.85 ± 0.08	0.016

Tab. 3: **Erkennungsraten der KI nach Auflösung** gezeigt ist die statistische Auswertung der Erkennungsraten der KIs mit je reduzierten Auflösungen der Ausgangs-EKGs.

Erkennungsfähigkeit bei komprimierter Bildqualität.

Die Erkennungsleistung des CNN beim Lesen von EKGs mit verschiedenen Bildqualitäten ist in Abb. 7c dargestellt. Die Verringerung der Bildqualität um bis zu einem Viertel oder sogar eine Änderung des Seitenverhältnisses des EKGs hatten keinen Einfluss auf F1 oder die Genauigkeit. Wie in Abb. 7 c gezeigt, war die Sensitivität des CNN beim Lesen von EKGs in voller oder halber Bildqualität deutlich höher als bei quadratischen EKGs ($88 \pm 7\%$, $84 \pm 10\%$ vs. $72 \pm 10\%$, $p=0,0022$ bzw. $0,042$). Andererseits war die Spezifität des CNN beim Lesen von quadratischen EKGs tendenziell höher als bei EKGs mit halber Bildqualität ($82 \pm 9\%$ vs. $71 \pm 12\%$, $p=0,091$). Der Test-Loss des CNN war bei EKGs mit halber Bildqualität signifikant höher als bei den anderen 3 Modellen. Es gab keinen signifikanten Unterschied in der Test-AUC zwischen den Modellen mit den verschiedenen Bildqualitäten. Die Leistung in den Validierungssets, insbesondere die AUC, war bei EKGs mit quadratischer Form signifikant niedriger.

Genauigkeit in jedem EKG Test-Set (KI) bezogen auf die Lokalisation [26]					
	Anteriorer MI Genauigkeit (richtig/gesamt)	Inferiorer MI Genauigkeit (richtig/gesamt)	Posteriorer MI Genauigkeit (richtig/gesamt)	Septaler MI Genauigkeit (richtig/gesamt)	Lateraler MI Genauigkeit (richtig/gesamt)
Set 01	0.88 (7/8)	0.88 (14/16)	1.00 (3/3)	0.50 (1/2)	0.90 (9/10)
Set 02	0.91 (10/11)	0.79 (11/14)	1.00 (2/2)	1.00 (5/5)	0.86 (6/7)
Set 03	0.93 (13/14)	0.73 (8/11)	0.50 (1/2)	1.00 (6/6)	0.67 (6/9)
Set 04	0.80 (8/10)	0.64 (9/14)	0.67 (2/3)	0.60 (3/5)	0.80 (8/10)
Set 05	0.67 (6/9)	1.00 (16/16)	1.00 (1/1)	1.00 (1/1)	1.00 (10/10)
Set 06	1.00 (10/10)	0.79 (11/14)	1.00 (4/4)	1.00 (5/5)	0.91 (10/11)
Set 07	1.00 (11/11)	1.00 (13/13)	1.00 (1/1)	1.00 (6/6)	1.00 (6/6)
Set 08	0.92 (12/13)	0.80 (8/10)	0.50 (1/2)	1.00 (6/6)	0.67 (6/9)
Set 09	0.93 (13/14)	0.82 (9/11)	1.00 (1/1)	1.00 (5/5)	0.89 (8/9)
Set 10	1.00 (13/13)	0.73 (8/11)	0.86 (6/7)	1.00 (5/5)	1.00 (11/11)
Total	0.91	0.82	0.85	0.94	0.87

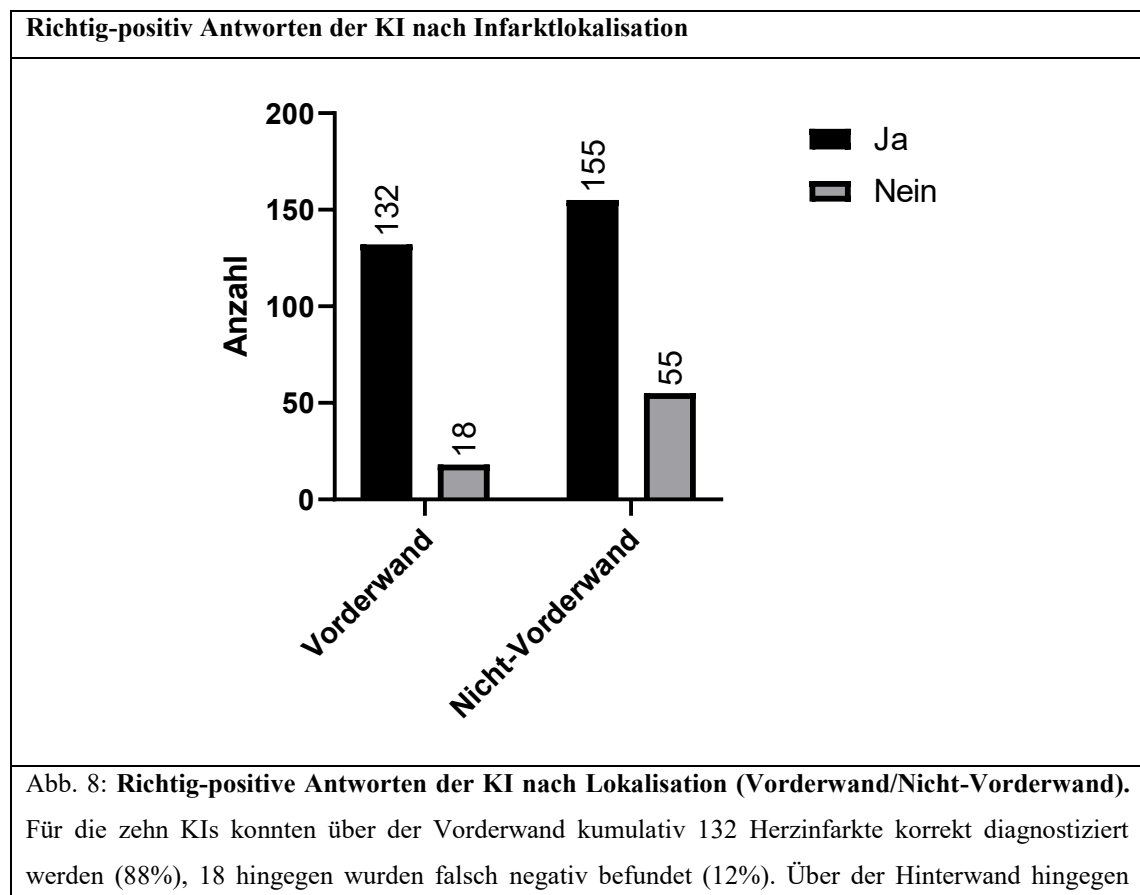
Tab. 4. Genauigkeit in jedem EKG Test-Set bezogen auf die Lokalisation. Gezeigt ist die Genauigkeit der jeweiligen Künstlichen Intelligenz (KI) mit den dazugehörigen Ergebnissen aufgeschlüsselt nach

Lokalisation. In den Klammern steht die erste Zahl für die richtigen Antworten, die zweite Zahl für die gesamt möglichen richtigen Antworten[26].

Im Hauptprojekt wurden die einzelnen KIs je mit einer anders ausgewählten EKG-Auswahl angelernt und getestet. Es erschwert die Vergleichbarkeit der KIs untereinander stark, bietet jedoch die Möglichkeit eine Wiederholung aufzuzeigen, was ein “Proof of Concept ermöglicht”. Wir konnten wiederholt und verschiedene KIs anlernen, die alle vergleichbare Resultate erzielten (Tab. 4).

Lokalisationsdetektion des akuten ST-Hebungsinfarktes durch eine CNN-KI

Für die Nebenfragestellung verglichen wir die Erkennungsleistung der KI gegen die Ärzte nach Lokalisation. Bei Lokalisationen unterschieden wir zwischen „Vorderwandinfarkt“ und „Nicht-Vorderwandinfarkt“. Die Aufteilung ist vor allem der klinischen Relevanz geschuldet. Eine Hebung über der Vorderwand ist oft mit einer „culprit lesion“ im RIVA vergesellschaftet. Bei mehreren Stenosen soll im Rahmen eines Herzinfarkts zuerst diese infarktverursachende Stenose behandelt werden[18]. Die Erkennungsleistungen der KI und der Ärzte ist im Folgenden dargestellt.

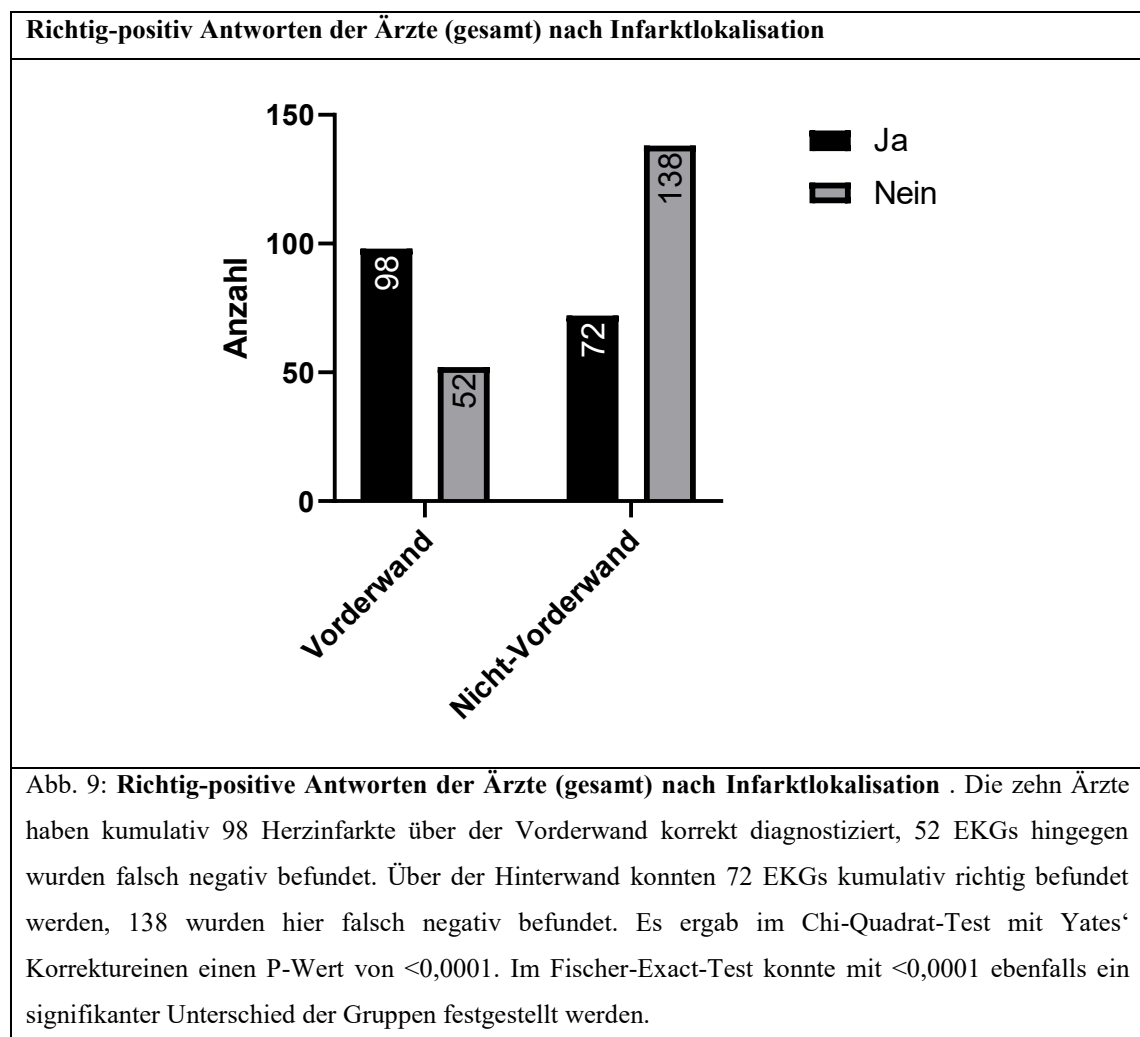


konnten 155 EKGs kumulativ richtig befundet werden (73,8%), 55 wurden hier hingegen falsch negativ befundet (26,1%). Es ergab im Chi-Quadrat-Test mit Yates' Korrektur einen P-Wert von 0,0015. Im Fischer-Exact-Test konnte mit 0.009 ebenfalls ein signifikanter Unterschied der Gruppen festgestellt werden.

Auffällig zeigte sich in der Auswertung der KI (Abb. 8), dass 88% der Vorderwandinfarkte erkannt wurden, jedoch nur 73,8% der Nicht-Vorderwandinfarkte. Ein noch stärkeres Bild zeigt sich in der Analyse der richtig befundenen EKGs durch die Ärzte, 65% der Vorderwand-Herzinfarkte wurden als richtig positiv befundet, über der Nicht-Vorderwand waren es nur 34% (Abb. 9).

Bei den Fachärzten konnten 64% der Vorderwandinfarkte korrekt befundet werden, über der Nicht-Vorderwand waren es 26% (Abb. 10).

Bei den Assistenzärzten zeigte sich: 66% der Vorderwandinfarkte wurden korrekt befundet, jedoch nur 44% der Nicht-Vorderwand-Infarkte (Abb. 11).



Richtig-positiv Antworten der Ärzte (Fachärzte) nach Infarktlokalisierung

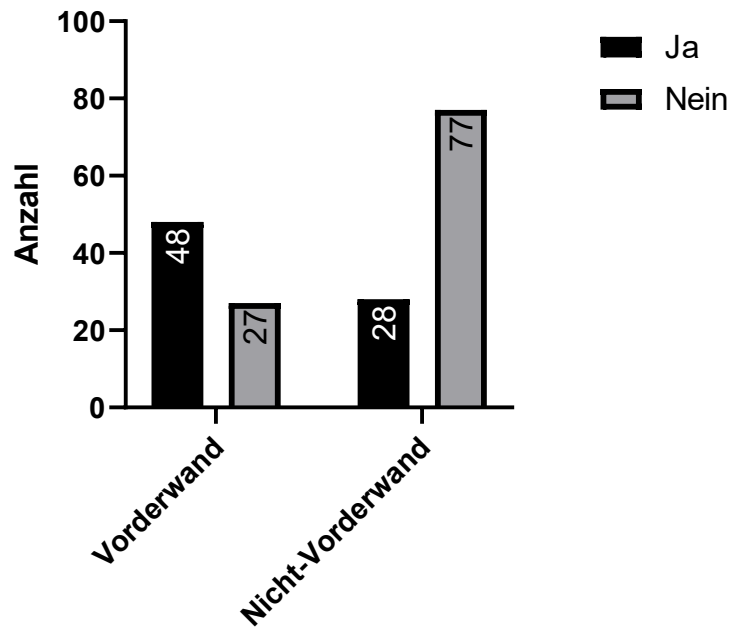


Abb. 10: **Richtig-positiv Antworten der Ärzte (Fachärzte) nach Infarktlokalisierung.** Für die zehn Ärzte konnten kumulativ 48 Herzinfarkte über der Vorderwand korrekt diagnostiziert werden, 27 hingegen wurden falsch negativ befundet. Über der Hinterwand konnten 28 EKGs kumulativ richtig befundet werden, 77 wurden hier falsch negativ befundet. Es ergab im Chi-Quadrat-Test mit Yates' Korrekturreinen einen P-Wert von $<0,0001$. Im Fischer-Exact-Test konnte mit $<0,0001$ ebenfalls ein signifikanter Unterschied der Gruppen festgestellt werden.

Richtig-positiv Antworten der Ärzte (Assistenzärzte) nach Infarktlokalisierung

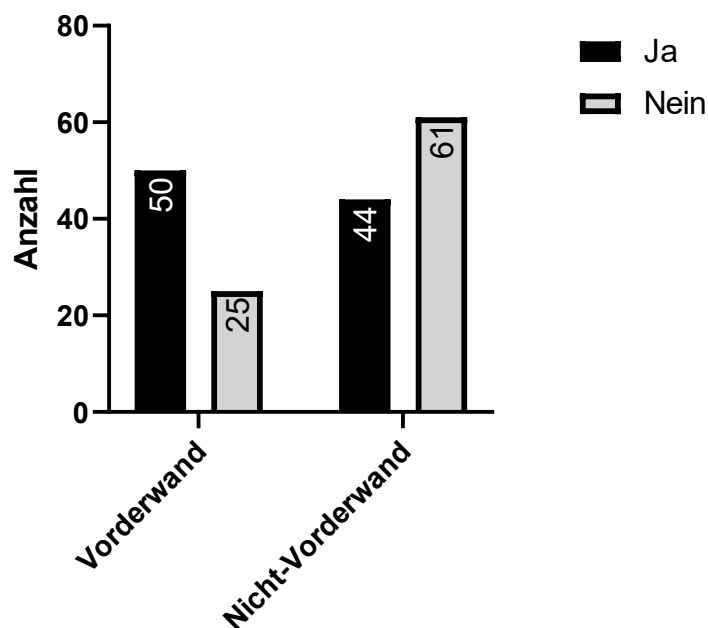
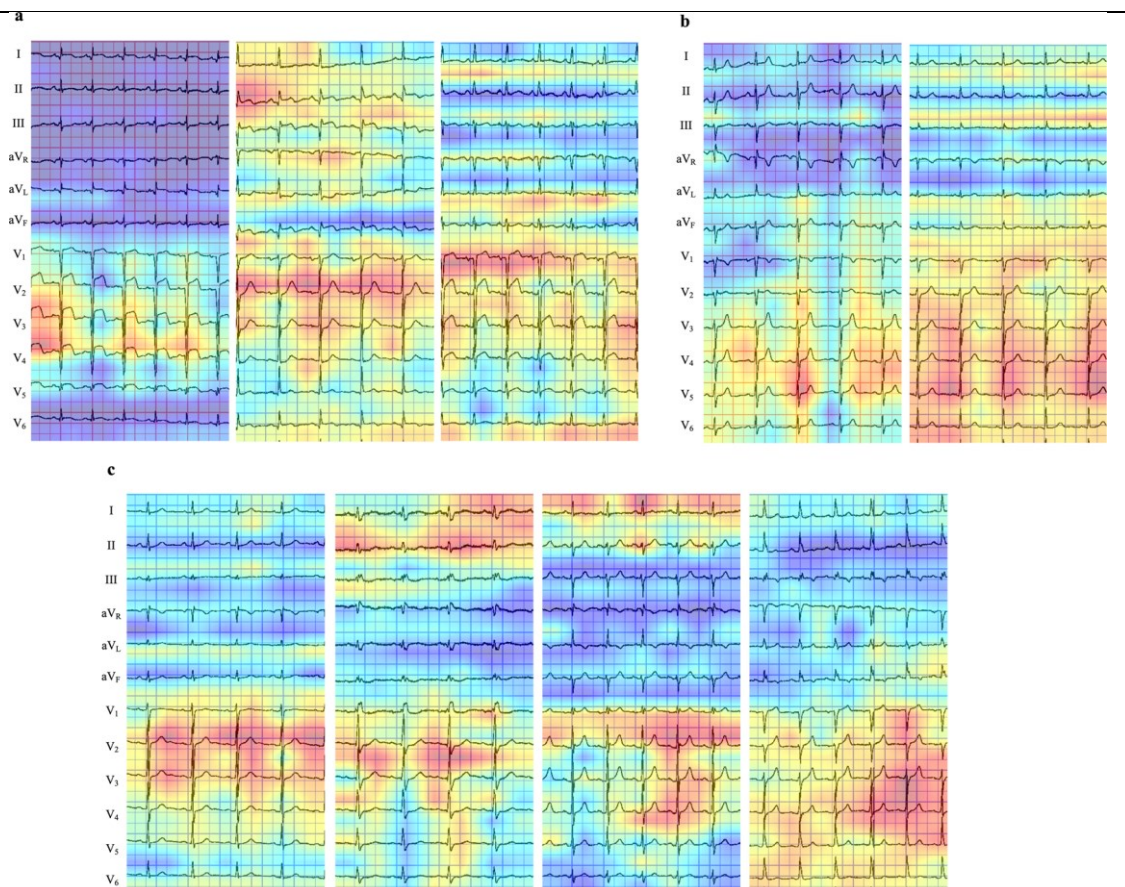


Abb. 11: **Richtig-positiv Antworten der Ärzte (Assistenzärzte) nach Infarktlokalisierung.** Für die fünf Assistenzärzte konnten kumulativ 50 Herzinfarkte über der Vorderwand korrekt diagnostiziert werden, 25 hingegen wurden falsch negativ befundet. Über der Hinterwand hingegen konnten 44 EKGs kumulativ richtig befundet werden, 61 wurden hier hingegen falsch negativ befundet. Es ergab im Chi-Quadrat-Test mit Yates' Korrekturreinen einen P-Wert von $<0,0018$. Im Fischer-Exact-Test konnte mit $<0,0014$ ebenfalls ein signifikanter Unterschied der Gruppen festgestellt werden.

KI-Erkennungsalgorithmus des STEMI (Grad-CAM)

Wir haben eine Grad-CAM verwendet, um die Stellen in den EKGs zu visualisieren, auf die sich der letzte Layer unserer CNNs bei der Erkennung konzentrierte. In einem EKG von MI-Fällen konzentrierte sich der CNN stark auf die ST-T-Hebung. In Nicht-MI-Fällen konzentrierte sich der CNN weitgehend auf den Bereich zwischen QRS-Komplexen und ST-T-Segment.

Erstellte Heat-Map: Im EKG markierte Bereiche der durch das CNN im letzten Layer am stärksten gewichteten Bereiche zur Entscheidungsfindung.



Reproduktion der Abbildung mit Genehmigung des Rechteinhabers [26]

Abb. 12: Erstellte Heatmap, die auf dem Ausgangs-EKG darstellt, welche Bereiche im letzten Layer des Convolutional Neural Network (CNN) besonders stark zur Entscheidungsfindung „Myokardinfarkt (MI) vs. Non-Myokardinfarkt (Non-MI)“ beitragen. **(a)** Drei Fälle mit MI, in welchen Ärzte und Künstliche Intelligenz (KI) eine richtige Diagnose trafen. **(b)** Zwei Fälle ohne MI, in welchen Ärzte und Künstliche Intelligenz (KI) eine richtige Diagnose trafen. **(c)** Vier Fälle mit MI, in denen nur die KI eine richtige

Diagnose traf. Die KI schien sich auf die ST-T-Segmente der präkardialen Ableitungen zu fokussieren und nur teilweise auf die Extremitätenableitungen. Diese MI-Fälle wiesen keine typischen ST-Segment-Veränderungen auf (Hebungen/Senkungen oder große Q-Wellen). Dies führt zu erschwerter Diagnosefindung[26].

In dem zusätzlichen Experiment mit 72 EKGs konnten alle zehn Ärzte in 8 Fällen (11,1 %) keine korrekte Diagnose stellen. Diese 8 EKGs stammten entweder von akuten Koronarsyndromen (non-ST-segment-elevation MI, n=6) oder von alten inferioren MI (n=2). Ein Fall mit altem inferiorem MI hatte einen kompletten Rechtsschenkelblock.

Die CNN-Modelle erreichten für diese 8 EKG-Aufzeichnungen eine durchschnittliche Genauigkeit von $87,5 \pm 11,8$ %. In diesen Fällen konzentrierte sich das CNN auch stark auf die ST-T-Segmente und QRS-Komplexe.

Diskussion

In dieser Arbeit wurde die Hypothese untersucht, ob eine KI zumindest gleich gut oder besser als in der EKG-Diagnostik erfahrene Ärzte einen akuten ST-Hebungsinfarkt erkennen kann. Zudem wurde die sekundäre Hypothese untersucht, ob die Lokalisation des Myokardinfarktes eine Auswirkung auf die Genauigkeit der Befundung der KI im Vergleich zu der Gruppe der Ärzte aufweist.

Die wichtigsten Ergebnisse der hier vorliegenden Arbeit waren, dass

- eine künstliche Intelligenz (mit 6-schichtigem neuronalen Netzwerk) eine bessere Erkennungsleistung bei ST-Hebungsinfarkt -EKG-Bildern erreicht als eine Gruppe in der EKG-Diagnostik erfahrenerer Ärzte,
- die Reduktion der Anzahl der EKG-Ableitungen keinen signifikanten Einfluss auf die Erkennungsleistung des CNN hatte,
- die Bildqualität, selbst wenn sie auf ein Viertel komprimiert wurde, die Erkennungsleistung des CNN nicht signifikant beeinträchtigte, was auf die Übertragbarkeit dieser Architektur hinweist, und
- dass das CNN sich bei der Erkennung von Herzinfarkten in EKGs auf die ST-T-Hebungen konzentrierte, vergleichbar zu menschlichen Versuchspersonen.
- dass Ärzte und eine KI einen Vorderwandinfarkt sicherer erkennen als einen Nicht-Vorderwandinfarkt

Soweit wir wissen, ist dies die erste Arbeit, die die Fähigkeiten von Deep Learning (das anhand eines kleinen Datensatzes von EKGs trainiert wurde) und Ärzten direkt vergleicht. Eine automatisierte, einfach nutzbare und effiziente künstliche Intelligenz zum Einlesen und Diagnostizieren von EKGs sollte für die Notfallmedizin unerlässlich sein. Unsere primäre Absicht ist es, die Machbarkeit von CNN-Architekturen zu bewerten, die mit kommerziell verfügbaren und kostengünstigen Rechenkapazitäten erreicht werden können.

In einer früheren Studie wurde berichtet, dass Deep Learning eine hohe Erkennungsleistung mit der gleichen PTB-EKG Datenbank aufwies [29]. In dieser Studie von Strodthoff, N. wurde ein 6-Layer-CNN mit digitalen EKG-Rohdaten eingesetzt. Die erzielte Sensitivität und Spezifität waren mit etwa 0,90 höher als bei unserem CNN mit Bildanalyse (Tabelle 3). Die Vergleichsgruppe „Ärzte“ wurde von Strodthoff, N. in der zuvor genannten Arbeit jedoch nicht auf EKGs aus der PTB EKG-Datenbank getestet, sodass die Vergleichbarkeit eingeschränkt ist.

Um die EKG-Bildanalyse zu verbessern, sind weitere Untersuchungen vonnöten. In unserer Studie haben wir 10 Ärzte einbezogen, die alle zum Zeitpunkt der Datenerfassung klinisch tätig waren, 5 davon als Fachärzte. Unsere Ergebnisse unterstreichen daher die hohe Leistungsfähigkeit der KI. Auch mit einer sehr kleinen Datenbank und limitierter Rechenleistung konnte eine KI erreicht werden, die gleichauf mit praktizierenden Ärzten ein EKG treffsicherer oder zumindest gleich gut befunden kann.

Außerdem haben wir die EKGs als Bilder (PNG-Format) mit mehreren Ableitungen und nicht als digitale Rohdaten verwendet. Dies ist genau die Art und Weise, wie Kardiologen EKGs in der klinischen Praxis sehen und diagnostizieren.

Hannun et al. haben berichtet, dass die KI, die CNN verwendet, mindestens so effizient ist wie Kardiologen bei der Erkennung von Herzrhythmusstörungen. Sie verwendeten die digitalen EKG-Rohdaten als Eingabe. In der vorliegenden Studie haben wir die Leistungsfähigkeit einer "Bildererkennung" zwischen menschlichen Internisten/Kardiologen und KI verglichen. Jun et al. haben vor kurzem die effiziente Befundung von Herzrhythmusstörungen durch CNN mit Graustufenbild-EKGs erfolgreich demonstriert.[30] Sie verwendeten die EKG-Bilder jedes einzelnen Schlags eines EKG-Streifens, d. h. einer einzelnen Ableitung, um verschiedene Arrhythmien

anhand von Holter-Monitoring-EKG-Daten zu erkennen. Unsere derzeitige KI erstellt eine Diagnose auf der Grundlage des gesamten 12-Kanal-EKGs und nicht nur auf einem einzelnen Schlag, um akute ST-Hebungsinfarkte zu detektieren. Obwohl die Aufgaben der KI unterschiedlich waren, deuten diese Ergebnisse auf die Machbarkeit hin, mit CNN effizient EKG-Bildbefundung betreiben zu können.

Bei den Modellen mit unterschiedlicher Layerarchitektur (5, 6 und 7 Layer) konnten wir keinen signifikanten Unterschied in der Effizienz des Deep Learnings feststellen, aber die erforderlichen Parameter waren bei dem 6-Schicht-CNN-Modell am besten (Tab. 5). Daher haben wir das 6-Layer CNN-Modell für das Deep Learning übernommen. Unser CNN erkannte MI-EKGs besser als Kardiologen, selbst wenn sie mit einer kleinen Datenbank (<300 EKGs) trainiert wurden. Unsere Daten zeigen, dass die Verwendung einer KI mit CNN zur Erkennung pathologischer EKGs auch ohne große Datenmengen möglich ist und mindestens so genau arbeiten kann wie Kardiologen. Die für die vorliegende Studie verwendete Grafikkarte (GPU) war mit einer Kapazität von 8 GB im regulären Handel erhältlich. Dies ermöglicht einen einfachen Zugang zu dieser Technologie.

Vergleich der <i>Convolutional Neural Networks</i> (CNN) mit unterschiedlicher Layeranzahl [26]							
	Parameter	Genauigkeit	P-Wert Genauigkeit	F1	P-Wert F1	„Loss“	P Wert (Loss)
CNN 5 Layer	17,711,745	0.818±0.054	0.24	0.825±0.053	0.28	1.82±0.33	0.15
CNN 6 Layer	7,787,201	0.814±0.044		0.823±0.044		1.56±0.31	
CNN 7 Layer	7,824,129	0.784±0.046		0.793±0.051		1.59±0.30	

Tab. 5 : Vergleich der *Convolutional Neural Networks* (CNN) mit unterschiedlicher Layeranzahl. Beschrieben ist der Vergleich der Parameter der einzelnen CNN mit je 5, 6 und 7 Layern. Ziel des Vergleichs war das effektivste Convolutional Neural Network (CNN) zu identifizieren. Die Entscheidung wurde getroffen anhand der Genauigkeit der KI, dem F1-Wert und dem Verlustwert („Loss“). Gewünscht ist eine möglichst hohe Genauigkeit und F1-Wert, bei möglichst geringem „Loss“. Das ausgewählte CNN mit 6 Layern traf auf die Anforderungen am besten zu. F1-Maß: (harmonisches Mittel aus Sensitivität und positiv prädiktivem Wert) [26]

Unter den EKGs mit verschiedenen Ableitungskonfigurationen führte nur die Konfiguration mit ausschließlich Extremitäten-EKGs zu einer geringeren Erkennungsleistung des CNN im Vergleich zu den anderen Ableitungskonfigurationen. In unseren Ergebnissen war die Verringerung der Erkennungsleistung (mit Ausnahme der EKGs mit Ableitungen an den Extremitäten) nicht signifikant. Die relativ abnehmende Erkennungsleistung in Abhängigkeit von der Anzahl der Ableitungen erscheint jedoch logisch, da die EKG-Anomalien je nach Lokalisation unter MI in bestimmten Ableitungen auftreten – gegebenenfalls wären diese Ableitungen in diesen Konfigurationen nicht enthalten. Auch wenn unser CNN in dieser Studie nur mit Extremitäten-EKGs trainiert wurde, zeigte es dennoch eine vergleichbare Erkennungsleistung wie die Kardiologen bei den 12-Kanal-EKGs. Dies könnte auf die Möglichkeiten hinweisen, diese Technologie in Situationen mit begrenzten medizinischen Ressourcen oder bei der Verwendung von tragbaren Geräten einzusetzen.

Wir haben EKGs im PNG-Format als Eingabedaten verwendet, einem gängigen Format zum Speichern von Bildern. Dies ist in klinischen Situationen praktisch, da EKGs in der Regel in einem 2-dimensionalen Bildformat dargestellt, diagnostiziert und gespeichert werden. Es gab keine signifikanten Unterschiede in den Erkennungsfähigkeiten unseres CNN zwischen den Bildqualitäten der Eingabe-EKGs und selbst EKGs mit schlechteren Qualitäten führten zu einer vergleichbaren Erkennungsfähigkeit wie bei Kardiologen.

Dies könnte auf mögliche Verwendungszwecke unseres CNN hinweisen, z. B. die Übermittlung der fotografierten EKGs an eine mit dem CNN ausgestattete Notrufzentrale für eine medizinische Diagnose oder sogar eine Anwendung in mobilen Geräten wie zum Beispiel Smartphones.

Wir visualisierten die Fokuspunkte des CNN auf den EKGs während des Erkennungsprozesses mit Heatmaps (Grad-CAM). Es zeigte sich, dass unser CNN sich auf die ST-T-Hebung im EKG konzentrierte, um einen Herzinfarkt zu erkennen, wie es auch Kardiologen tun. Bei MI-Fällen, die nur der CNN korrekt als MI erkannte, konzentrierte sich der CNN auch auf die ST-T Segmente. Für die Kardiologen war es in diesen Fällen jedoch schwierig, die MI anhand dieses Fokusbereichs im EKG zu beurteilen.

Wenn bei einem Herzinfarkt keine ST-T-Segment-Hebung/-Senkung oder Q-Wellen vorhanden sind wird die Erkennung eines akuten oder alten Herzinfarkts schwierig. Heatmaps könnten uns neue Kriterien für das Lesen von EKGs liefern, wenn wir die Heatmaps in EKGs in größeren Datenbanken sammeln. Andererseits konnten wir in der vorliegenden Studie nicht erklären, warum sich der CNN auf das ST-T-Segment eines bestimmten Schlags konzentriert. Dies sollte mit größeren Datenbanken weiter untersucht werden.

Das CNN in der vorliegenden Studie wurde mit einer kleinen Datenbank trainiert und konnte auf EKGs eine MI-Diagnose stellen. Obwohl die KI bereits in der Lage war, EKGs so korrekt wie Ärzte/Kardiologen zu befunden, könnte man mit größeren Trainingsdatenbanken noch präzisere Ergebnisse bei der Bildanalyse erzielen.

Bemerkenswert ist hierbei, dass die Sensitivität und der negativ prädiktive Wert des CNN trotzdem deutlich höher waren als bei Ärzten. Diese Ergebnisse legen nahe, eine solche KI zur Verringerung von Fehldiagnosen in der Klinik einzusetzen.

Die zehn Modelle des CNN in diesem Projekt zeigten durchweg vergleichbare bis bessere Diagnostikfähigkeiten als Ärzte/Kardiologen, allerdings war die Leistung leicht schwankend (Abb. 7). Wahrscheinlich ist dies auf die geringe Datenmenge zurückzuführen. Mit einer größeren EKG-Datenbank auch für zusätzliche Herzkrankheiten, ist es vorstellbar, eine KI zu entwickeln, die EKGs nutzt, um ein breiteres Spektrum an Herz-Kreislauf-Erkrankungen zu diagnostizieren.

Eine Limitierung der hier vorliegenden Arbeit ist, dass die verwendete Rechenleistung für Deep Learning begrenzt war. Im aktuellen Fall wurden die Trainingsläufe auf einer Grafikkarte mit acht Gigabyte eigenem Speicher durchgeführt. Die Durchführbarkeit und Zuverlässigkeit sollte daher in zukünftigen Forschungen bestätigt werden. Wahrscheinlich wird es weitere Fortschritte in der AI-Hardware-Technologie geben. Aufgrund des explorativen Charakters dieser Studie wurde keine Berechnung der Stichprobengröße durchgeführt. Wegen der Unausgewogenheit der beiden Studienarme und der daraus folgenden Datenaugmentation haben wir keine Kreuzvalidierung durchgeführt. Stattdessen haben wir 10 verschiedene Trainings-/Validierungs-/Test-EKG-Sets erstellt und die Ergebnisse der Analyse gemittelt. Obwohl in der veröffentlichten Datenbank (PTB Diagnostic ECG Database) die zugewiesene Diagnose von MI oder Nicht-MI auf der Grundlage der anschließenden klinischen Beurteilung einschließlich Herzkatheteruntersuchung klassifiziert wurde, waren weitere detaillierte klinische Informationen nicht verfügbar. Obwohl in der vorliegenden Studie Grad-CAM zur Visualisierung des CNN-Fokus verwendet wurde, werden die Visualisierungen auf der Grundlage der Gradienten der letzten Layer-Aktivierung hochgerechnet. Daher sollte die Auflösung des Fokus sorgfältig bewertet werden. Außerdem konnte gezeigt werden, dass Grad-CAM Objekte in einem Bild nicht richtig lokalisieren kann, wenn das Bild mehrere Vorkommen der gleichen Klassifikation enthält. Dieser Nachteil könnte durch die Weiterentwicklung von Algorithmen zur Visualisierung von Foci neuronaler Netze überwunden werden[13]. Schließlich war es nicht das Ziel dieser Forschung, die Unterlegenheit von menschlichen Ärzten/Kardiologen zu beweisen. Wir betonen, dass eine Diagnose nicht nur auf einer reinen EKG-Interpretation, sondern auch auf einer klinischen Beurteilung beruht, die zumindest zum jetzigen Zeitpunkt für KI-Systeme nicht machbar ist. Diese Technologie könnte jedoch als erstes Triage- und Bewertungsinstrument zur Unterstützung der endgültigen klinischen Diagnose eingesetzt werden.

Die schlechter diagnostizierten Nicht-Vorderwandinfarkte könnten gegebenenfalls sicherer mit erweiterter EKG-Diagnostik erfasst werden. Eine weitere Studie mit Erfassen von V7-V9 beispielsweise wäre wünschenswert.

Schlussfolgerung und Ausblick

Künstliche Intelligenz mit einer einfachen CNN-Architektur, die aus einer kleinen EKG-Datenbank trainiert wird, kann bei der Erkennung von akuten ST-Hebungsinfarkten einen Vorteil gegenüber Kardiologen bieten – zumindest ist sie nicht unterlegen. Eine Unterstützung von Berufsanfängern oder fachfremden Ärzten könnte die Sicherheit von Patienten verbessern.

Im ersten Projektteil konnten wir darlegen, dass die Entfernung der Goldberger-Ableitungen die Erkennungsleistung nicht wesentlich beeinträchtigt. Eine Analyse von EKG's in Überwachungsmonitoren auf einer Intensivstation könnte beispielsweise kontinuierlich EKGs analysieren und so ST-Hebungsinfarkte zeitnah entdecken. Weiter könnte im Rettungsdienst trotz nur weniger Ableitungen eine hohe Erkennungsrate von Herzinfarkten erreicht werden.

Selbst die Bildkompression auf bis zu einem Viertel der Ausgangsauflösung führte nicht zu einer relevanten Reduktion der Erkennungsleistung der KI, was die Anforderung an Hardware (Kamera, Computer) deutlich reduziert.

Weitere prospektive randomisierte Arbeiten sind erforderlich, um den möglichen Vorteil, der Verwendung einer KI zur Beurteilung von EKGs als Triage- oder Risikostratifizierungsinstrument im klinischen Kontext weiter beurteilen zu können.

Verfügbarkeit der Daten

Die verwendeten Daten (PTB ECG Datenbank) sind alle im Physionet verfügbar (frei veröffentlicht). Die Autoren haben keine Einschränkungen für die Offenlegung der verwendeten EKGs oder Informationen über 10 Gruppen von randomisierten EKG-Sets.

Quellenverzeichnis

1. Pardee, H.E., *Concerning the electrodes used in electrocardiography*. American Journal of Physiology-Legacy Content, 1917. **44**(1): p. 80-83.
2. Einthoven, W., *Über die deutung des elektrokardiogramms*. Pflüger's Archiv für die Gesamte Physiologie des Menschen und der Tiere, 1912. **149**(1-3): p. 65-86.
3. Einthoven, W., G. Fahr, and A. De Waart, *Über die Richtung und die manifeste Grösse der Potentialschwankungen im menschlichen Herzen und über den Einfluss der Herzlage auf die Form des Elektrokardiogramms*. Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere, 1913. **150**(6-8): p. 275-315.
4. Wilson, F.N., et al., *Recommendations for standardization of electrocardiographic and vectorcardiographic leads*. Circulation, 1954. **10**(4): p. 564-573.
5. Goto, S., et al., *Artificial intelligence to predict needs for urgent revascularization from 12-lead electrocardiography in emergency patients*. PloS one, 2019. **14**(1): p. e0210103.
6. Kelm, M., et al., *Kommentar zu den Leitlinien 2017 der Europäischen Gesellschaft für Kardiologie (ESC) zur Therapie des akuten Herzinfarktes bei Patienten mit ST-Streckenhebung*. Der Kardiologe, 2018. **12**(2): p. 145-149.
7. Chowdhary, K. and K. Chowdhary, *Natural language processing*. Fundamentals of artificial intelligence, 2020: p. 603-649.
8. Searle, J.R., *Minds, brains, and programs*. Behavioral and Brain Sciences, 1980. **3**(3): p. 417-424.
9. McIlroy-Young, R., et al., *Aligning Superhuman AI with Human Behavior: Chess as a Model System*, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, Association for Computing Machinery: Virtual Event, CA, USA. p. 1677–1687.
10. Kaelbling, L.P., M.L. Littman, and A.W. Moore, *Reinforcement learning: A survey*. Journal of artificial intelligence research, 1996. **4**: p. 237-285.
11. Kashou, A.H., A.M. May, and P.A. Noseworthy, *Artificial Intelligence-Enabled ECG: a Modern Lens on an Old Technology*. Curr Cardiol Rep, 2020. **22**(8): p. 57.
12. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *ImageNet classification with deep convolutional neural networks*. Commun. ACM, 2017. **60**(6): p. 84–90.
13. Chattopadhyay, A., et al., *Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks*. arXiv. arXiv preprint arXiv:1710.11063, 2017.
14. Buda, M., A. Maki, and M.A. Mazurowski, *A systematic study of the class imbalance problem in convolutional neural networks*. Neural Netw, 2018. **106**: p. 249-259.
15. Mak, R.H., et al., *Use of Crowd Innovation to Develop an Artificial Intelligence-Based Solution for Radiation Therapy Targeting*. JAMA Oncol, 2019. **5**(5): p. 654-661.
16. Kusunose, K., et al., *A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images*. JACC Cardiovasc Imaging, 2020. **13**(2 Pt 1): p. 374-381.
17. Hannun, A.Y., *Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network*. 2019.
18. Jung, C. and A. Elsässer, *Update ESC-Leitlinie 2017–Akuter Myokardinfarkt (STEMI)*. DMW-Deutsche Medizinische Wochenschrift, 2018. **143**(11): p. 797-801.
19. Siontis, K.C., et al., *Artificial intelligence-enhanced electrocardiography in cardiovascular disease management*. Nature Reviews Cardiology, 2021. **18**(7): p. 465-478.
20. Busselot, R., D. Kreiseler, and A. Schnabel, *Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet*. 1995.

21. Goldberger, A.L., et al., *PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals*. *Circulation*, 2000. **101**(23): p. E215-20.
22. Moody, G.B. *LightWAVE*. [Program] 2013; LightWAVE is a lightweight waveform and annotation viewer and editor. Use it to view any of the recordings of physiologic signals and time series in PhysioBank, together with their annotations (event markers).]. Available from: <https://physionet.org/lightwave/>.
23. Salehi, A. and M. Balasubramanian, *DDCNet-Multires: Effective Receptive Field Guided Multiresolution CNN for Dense Prediction*. *Neural Processing Letters*, 2022: p. 1-25.
24. Werner, M., *Neuronale Netze mit Faltungsschichten*, in *Digitale Bildverarbeitung: Grundkurs mit neuronalen Netzen und MATLAB®-Praktikum*. 2021, Springer Fachmedien Wiesbaden: Wiesbaden. p. 409-465.
25. Graham, B., *Fractional max-pooling*. arXiv preprint arXiv:1412.6071, 2014.
26. Makimoto, H., et al., *Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction*. *Sci Rep*, 2020. **10**(1): p. 8445.
27. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. *Nature*, 2015. **521**(7553): p. 436-44.
28. Selvaraju, R.R., *Grad-CAM: visual explanations from deep networks via gradient-based localization*.
29. Strodthoff, N. and C. Strodthoff, *Detecting and interpreting myocardial infarction using fully convolutional neural networks*.
30. Jun, T.J., *ECG arrhythm*. 2018.

Anhang

Beispieldatensatz der PTB-Datenbank (Ein Patient) [20]

s0010_re 15 1000 38400

s0010_re.dat 16 2000 16 0 -489 -8337 0 i

s0010_re.dat 16 2000 16 0 -458 -16369 0 ii

s0010_re.dat 16 2000 16 0 31 6829 0 iii

s0010_re.dat 16 2000 16 0 474 4582 0 avr

s0010_re.dat 16 2000 16 0 -260 11687 0 avl

s0010_re.dat 16 2000 16 0 -214 -16657 0 avf

s0010_re.dat 16 2000 16 0 -88 -12469 0 v1

s0010_re.dat 16 2000 16 0 -241 5636 0 v2

s0010_re.dat 16 2000 16 0 -112 -14299 0 v3

s0010_re.dat 16 2000 16 0 212 -17916 0 v4

s0010_re.dat 16 2000 16 0 393 -6668 0 v5

s0010_re.dat 16 2000 16 0 390 -17545 0 v6

s0010_re.xyz 16 2000 16 0 -3 -13009 0 vx

s0010_re.xyz 16 2000 16 0 120 7109 0 vy

s0010_re.xyz 16 2000 16 0 -18 -1992 0 vz

age: 81

sex: female

ECG date: 01/10/1990

Diagnose:

Reason for admission: Myocardial infarction

Acute infarction (localization): infero-latera

Former infarction (localization): no

Additional diagnoses: Diabetes mellitus

Smoker: no

Number of coronary vessels involved: 1

Infarction date (acute): 29-Sep-90

Previous infarction (1) date: n/a

Previous infarction (2) date: n/a

Hemodynamics:

Catheterization date: 16-Oct-90

Ventriculography: Akinesia inferior wall

Chest X-ray: Heart size upper limit of norm

Peripheral blood Pressure (syst/diast): 140/80 mmHg

Pulmonary artery pressure (at rest) (syst/diast): n/a

Pulmonary artery pressure (at rest) (mean): n/a

Pulmonary capillary wedge pressure (at rest): n/a

Cardiac output (at rest): n/a

Cardiac index (at rest): n/a

Stroke volume index (at rest): n/a

Pulmonary artery pressure (load) (syst/diast): n/a

Pulmonary artery pressure (load) (mean): n/a

Pulmonary capillary wedge pressure (load): n/a

Cardiac output (load): n/a

Cardiac index (load): n/a

Stroke volume index (load): n/a

Aorta (at rest) (syst/diast): 160/64 cmH2O

Aorta (at rest) mean: 106 cmH2O

Left ventricular enddiastolic pressure: 11 cmH2O

Left coronary artery stenoses (RIVA): RIVA 70% proximal to ramus diagonalis_2

Left coronary artery stenoses (RCX): No stenoses

Right coronary artery stenoses (RCA): No stenoses

Echocardiography: n/a

Therapy:

Infarction date: 29-Sep-90

Catheterization date: 16-Oct-90

Admission date: 29-Sep-90

Medication pre admission: Isosorbit-Dinitrate Digoxin Glibenclamide

Start lysis therapy (hh.mm): 19:45

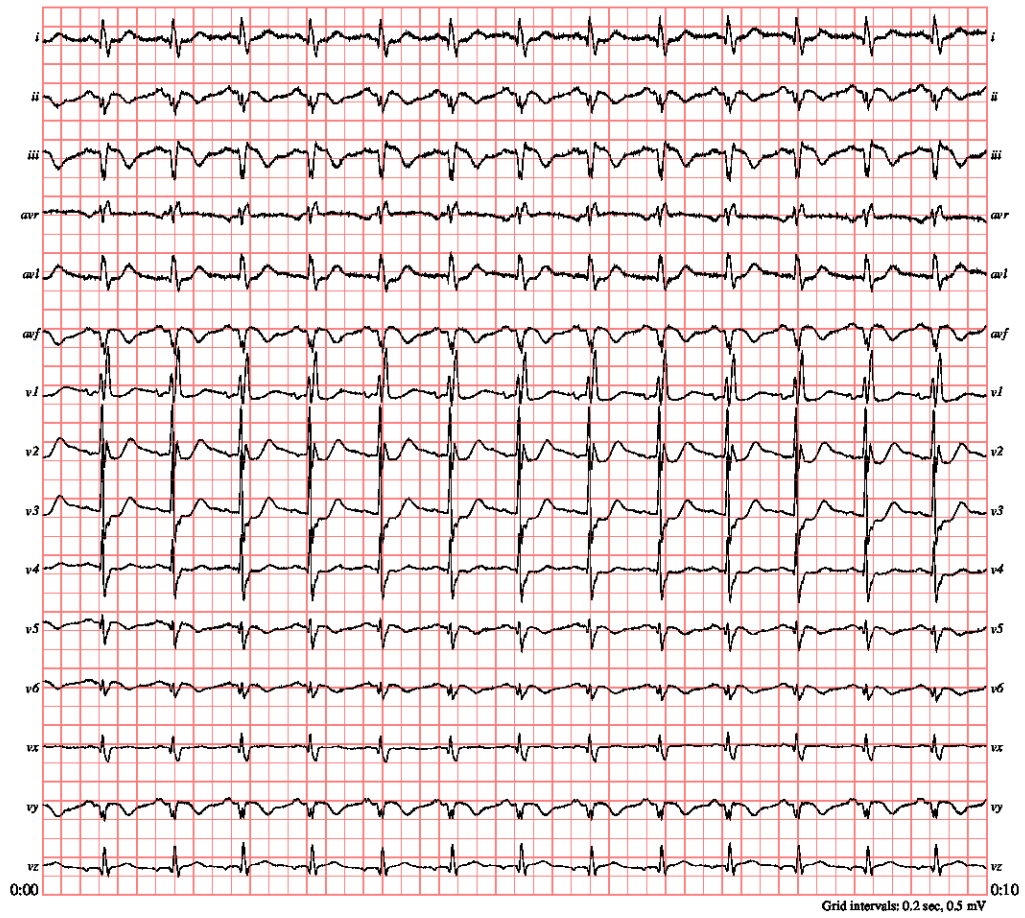
Lytic agent: Gamma-TPA

Dosage (lytic agent): 30 mg

Additional medication: Heparin Isosorbit-Mononitrate ASA Diazepam

In hospital medication: ASA Isosorbit-Mononitrate Ca-antagonist
Amiloride+Chlorothiazide Glibenclamide Insulin

Medication after discharge: ASA Isosorbit-Mononitrate Amiloride+Chlorothiazide
Glibenclamide



Danksagung

An dieser Stelle möchte ich allen allen beteiligten Personen danken, die mich bei der Anfertigung und Bearbeitung meiner Dissertation unterstützt haben. Mein besonderer Dank gilt Professor Makimoto, der sich die Zeit genommen hat mit mir meine ersten wissenschaftlichen Projekte zu beginnen und diese Doktorarbeit zu begleiten und mich maßgeblich auf diesem Wege unterstützt hat. Weiter möchte ich Professor Aubin für die Cobetreuung danken. Dr. David Glöckner und Dr. Daniel Oehler danke ich - sie leisteten große Unterstützungsarbeit mit praktischen Tips für die Umsetzung des Schreibprozesses. Nicht zuletzt danke ich meiner Familie und Freunden, die mich mit motivierten und mich ertrugen, wenn die Statistik mal nicht so wollte, wie ich.