Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Hochschule für Angewandte Wissenschaften Hamburg
Fakultät Life Sciences

Explainable Machine Learning in der Medizintechnik - Aufbau eines CDSS zur Unterstützung der Erkennung von kardiovaskulären Krankheiten

Bachelorarbeit, zur Erlangung des
Bachelor of Science (B. Sc.)

im Studiengang Medizintechnik
vorgelegt von

Christian Knuth

Erstkorrektur: Prof. Dr. Marina Tropmann-Frick

Zweitkorrektur: Prof. Dr. Petra Margaritoff

HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG
Hamburg University of Applied Sciences

Thema: Explainable Machine Learning in der Medizintechnik - Aufbau eines CDSS zur Unterstützung der Erkennung von kardiovaskulären Krankheiten

Die Bachelorarbeit wurde zusammen mit Excel Arbeitsmappen und einem Python Skript abgegeben. Für die Textverarbeitung wurde Latex genutzt.

Hamburg, den 26.06.2022

Themen: *Explainable Machine Learning, Medical technology*

Stichworte: Shapley additive explanations, clinical decision support system (CDSS), coronary heart disease

Inhaltsverzeichnis

Abkürzungen

CAD: Coronary Artery Disease

CHD: Coronary Heart Disease

CVD: Cardiovascular Disease

CDSS: Clinical Decision Support System

ML: Machine Learning

FP: False positive

FN: False negative

Zusammenfassung

In letzter Zeit ist das Thema Machine Learning (ML) immer bekannter geworden und wird bereits in vielen Disziplinen aus der Industrie und Wissenschaft angewendet. Unter Machine Learning wird dabei das automatisierte Lernen aus Datenbeständen mittels mathematischer und statistischer Methoden verstanden. Machine Learning wird auch in der Medizin und der Medizintechnik eingesetzt. So können z.B. mittels ML-Modellen Krankheiten von Patienten auf Basis gesammelter Daten vorhergesagt werden. Dadurch besteht die Perspektive, dass medizinische Diagnosen langfristig durch ML verbessert werden könnten.

Bei der Anwendung von ML in diesem Bereich gibt es aber auch genügend Faktoren, die dieses Vorgehen erschweren. Dies sind v.a. anwendungsbedingte Schwierigkeiten bei der Auswertung medizinischer Daten und anwendungsunabhängige Herausforderungen, die sich aus statistischer Sicht bei der Datenauswertung ergeben. In diesem Kontext ist wichtig, dass viele Modelle des ML sich durch Experten des jeweiligen Fachgebiets zunächst nicht direkt erklären bzw. nachvollziehen lassen. Die Machine Learning Modelle werden dann als sogenannte „Black Box“-Modelle bezeichnet. Die Anwendung dieser Modelle könnte dazu führen, dass ein Modell für einen Anwender, z.B. in der Medizin einem Arzt bei der Diagnose, nur teilweise nachvollziehbar ist. Ebenso kann es dazu führen, dass fehlerhafte Entscheidungen des ML-Algorithmus nur schwer erkannt werden können. Während in einigen Anwendungen aus der Medizin und Medizintechnik die Nachvollziehbarkeit nur eine geringe Rolle spielt, so ist in vielen der medizinischen Anwendungen von ML die Nachvollziehbarkeit vorrangig. Die Methoden des ML müssen deshalb an die Anforderung erklärbar zu sein angepasst werden. Dazu werden Methoden des Explainable Machine Learning angewendet. In dieser Arbeit wird versucht, dies am Beispiel der Erkennung von kardiovaskulären Krankheiten bzw. den koronaren Herzerkrankungen anhand des „Heart Disease“-Datensatzes prinzipiell zu implementieren. Die Anwendung wäre neben anderen möglichen Anwendungsgebieten v.a. klinisch als ein (fiktives) Clinical Decision Support System (CDSS) einzusetzen. Ein solches System könnte prinzipiell als Unterstützung für Ärzte dabei helfen, kardiovaskuläre Krankheiten von Patienten zu erkennen. Das aufgebaute ML-Modell soll dabei nicht nur die Möglichkeit bieten, Ergebnisse interpretieren zu können, sondern die Ergebnisse des ML-Algorithmus sollten idealerweise im Detail nachvollziehbar sein. Bei der Untersuchung des „Heart Disease“-Datensatzes werden bis zu 33 Merkmale betrachtet. Da der Datensatz in dieser Situation weniger häufig ausgewertet wurde, wird versucht in dieser Situation weitere Fragestellungen zu dem Datensatz zu beantworten.

Einleitung

In letzter Zeit sind die Themen Machine Learning (ML) und Big data analytics immer bekannter geworden. Unter Machine Learning werden das automatisierte Lernen und die Wissensextraktion aus Datenbeständen verstanden. Dazu werden mathematische und statistische Methoden angewendet. Die „Machines“ sind die Computer, deren gegebenes Ziel es ist, automatisiert Zusammenhänge aus den jeweils vorgegebenen Daten zu erlernen. Zu diesem Zweck wird auch die mathematische Optimierung genutzt. ML-Methoden ermöglichen es, komplexe Fragestellungen zu lösen; dabei sind die Vorteile bei der Verwendung von ML vielfältig. Fragestellungen mit unsicheren Lösungen werden oft sehr komplex, besonders da heutzutage die Anforderung besteht, detaillierte und genaue Ergebnisse zu finden. Die Anwendung rein konsistenter Lösungsmethoden durch die Expertise von Fachexperte, wird dann zunehmend ineffizient. Oft ist in dem jeweils betrachteten Fachgebiet auch der vorhandene Wissensbestand zu groß und komplex geworden, um zuvor gewonnene Erkenntnisse in bestimmten Situationen effizient und zielgerichtet nutzen zu können. Teilweise ist es auch z.B. aus Zeitmangel nicht möglich, genügend Expertise zu einer Fragestellung zu sammeln, oder ein Fachexperte ist gerade nicht erreichbar. In all diesen Fällen kann es sich anbieten zusätzlich Machine Learning als Hilfe einzusetzen.

Machine Learning wird bereits auch in der Medizin und der Medizintechnik eingesetzt. So können z.B. mittels Machine Learning Krankheiten von Patienten auf Basis gesammelter Daten vorhergesagt werden. Dabei können die Ergebnisse der ML-Algorithmen teilweise die Erwartungen übertreffen und bislang unbekannte medizinische Zusammenhänge aufzeigen. Die Ergebnisse des ML können sogar in der medizinischen Forschung neue Erkenntnisse für Experten aufzeigen, die anderenfalls nicht entdeckt worden wären, weil es z. B. zu aufwendig gewesen wäre. Desweiteren kann Machine Learning auch in klinischen Anwendungen genutzt werden, indem ML in medizintechnische Geräte implementiert wird. In der klinischen Anwendung kann ML dazu dienen, vorher gewonnenes Wissen in ML-Algorithmen zu integrieren, um damit jederzeit für Aufgaben eingesetzt werden zu können auch dann, wenn gerade kein Fachexperte in der Nähe ist oder dieser für eine vollständige Aufarbeitung der Aufgabe zu wenig Zeit hat. Damit kann medizinisches Wissen auch besser in klinischen Anwendungen praktisch genutzt werden. Es kann ebenso genutzt werden, um Routineaufgaben abzunehmen, was Zeitmangel in klinischen Fragestellungen ausgleichen könnte. Es bietet sich aber auch die Anwendung von Machine Learning als Unterstützung für einen Fachexperten an. Dabei besteht der Vorteil, dass die Diagnose genauer sein kann, wenn der Fachexperte zusätzlich ein ML-System verwendet. Zudem bietet ein erklärbares ML-System die Möglichkeit, objektiv Ergebnisse abgleichen zu können. So kann der Fachexperte die Diagnose mit einer Vorhersage des ML-Systems abgleichen, was als Korrekturmechanismus dienen kann. Damit besteht aber die Möglichkeit Fehldiagnosen zu korrigieren und auch dann richtig zu diagnostizieren, selbst wenn der Patient nicht die typischen Symptome aufweist.

Je nach Anwendung von Machine Learning können sich die notwendigen Eigenschaften der ML-Systeme unterscheiden. Für bestimmte Anwendungen ist es nicht notwendig, dass die Ergebnisse erklärbar sind oder eine vollständige Erklärung der Ergebnisse ist auch gar nicht möglich. In vielen Anwendungen von ML in der Medizintechnik erscheinen Lösungen oder auch Hypothesen, die nicht erklärbar sind und nicht ausreichend gesichert wurden, jedoch als nicht akzeptabel und das Risiko bei der Anwendung demzufolge als hoch. Besonders bei der computerunterstützten Diagnose wie bei einem klinischen Unterstützungssystem ist die Forderung nach der Verständlichkeit und Plausibilität natürlicherweise vorrangig, sodass aus den Daten valide Schlussfolgerungen oder geeignete Hypothesen gezogen werden können [1].

Als zentrales Konzept, um dies mittels ML erreichen zu können und damit eine sichere Anwendung im medizinischen Umfeld zu ermöglichen, ist dabei der Ansatz des Explainable Machine Learning zu nennen. Dieser Ansatz ermöglicht es vielen Schwierigkeiten zu entgegnen, die sich bei unverständlichen ML-Modellen ergeben können, die als „Black Box“-Modelle bezeichnet werden. Beim explainable ML wird der Fokus darauf gelegt, dass nicht nur die Ergebnisse der ML-Modelle interpretierbar sind sondern auch im Detail nachvollziehbar. Eine der interessierenden Fragestellungen ist dabei zunächst, warum „ Black Box“- Modelle für viele Anwendungen in der Medizin und Medizintechnik nicht ausreichend sind. Häufig sind anwendungsbedingte Schwierigkeiten und Anforderungen im Bereich der Medizin und Medizintechnik ein Grund dafür. Desweiteren sind es insbesondere Schwierigkeiten aus statistischer Sicht bei der Auswertung von medizinischen Daten, weshalb die Anwendung von „Black Box“-Modellen vielfach nicht geeignet ist. Die Auswertung hochdimensionaler Daten sorgt für Ungenauigkeiten. Insgesamt ist deshalb auch die statistische Evaluation und die Berücksichtigung statistischer Zusammenhänge im ML sehr wichtig [2]. Dies ist bei „Black Box“-Modellen meist nicht ausreichend möglich. Dies kann insofern relevant sein, wenn hohe Anforderungen an die Genauigkeit gestellt werden und gleichzeitig zunehmend im ML die Anforderung besteht, Aufgabenstellungen zu lösen, wie die Vorhersage der Ergebnisse bei der Anwendung auf größere Patientenmengen. Erst wenn die Funktionsweise der Algorithmen verstanden werden kann und die Ergebnisse evaluiert werden können, ist es jedoch möglich zu sagen, ob eine Anwendung von ML für diese Ziele überhaupt ausreichend geeignet ist.

Die Zielsetzung in dieser Arbeit ist beispielhaft die Implementierung eines (fiktiven) CDSS zur Vorhersage von kardiovaskulären Krankheiten bzw. koronaren Herzerkrankungen mittels Methoden des Explainable Machine Learning anhand des „Heart Disease“-Datensatzes [3]. Aus der Bearbeitung dieser Zielsetzung soll geklärt werden, ob es mit einfachen Mitteln möglich ist, ein solches System aufzubauen, das sich sinnvoll z.B. durch Ärzte nutzen ließe. Dazu werden unterschiedliche ML-Modelle verglichen. Gelingt es auch mit einfachen Methoden ein CDSS aufzubauen, so ist dies ein Anhaltspunkt dafür, dass die Verwendung von ML für CDSS zur präventiven Erkennung von koronaren Herzerkrankungen sinnvoll wäre, wenn es mit professionellem Aufwand umgesetzt werden würde. Dabei soll auf den Unterschied zum Vorgehen bei dem klassischen klinischen Entscheidungsfindungsprozess bei der Diagnose durch Ärzte eingegangen werden und welche Unterschiede sich ergeben bei dem zusätzlichen Einsatz von ML-Systemen als klinische Unterstützungssysteme in der Kardiologie. In diesem Fall würde ein Arzt zusätzlich durch ein ML-System, bei der Entscheidungsfindung bzw. Diagnose unterstützt werden. Ebenfalls ist es möglich, dass Ärzte bei den ML-Algorithmen mitarbeiten bzw. durch Entscheidungen das ML-System beeinflussen können. Ansätze, die dem Anwender Einfluss auf das ML-System ermöglichen, werden auch als „human in the loop“ [4] bezeichnet. Die-

ses Konzept ist auch für Explainability sehr wichtig. Es stellt sich deshalb die Frage, weshalb die kollaborative Entscheidungsfindung bzw. die Zusammenarbeit von Arzt und ML-System im Allgemeinen vorteilhaft ist unter der Bedingung, dass Explainable Machine Learning angewendet wird.

Der „Heart Disease“-Datensatz wurde mittels ML schon oft erfolgreich untersucht. Die Ergebnisse sind aber häufig nur aus der Sicht des ML beschrieben und es wird nicht die medizintechnische Anwendung berücksichtigt und wie ein ML-System konkret nutzbar wäre. Zudem wird in den meisten Publikationen nur ein Teil der Daten des „Heart Disease“-Datensatzes und nur ein Teil der Merkmale genutzt. In dieser Arbeit soll der komplette „Heart Disease“-Datensatz ausgewertet werden. Damit die Ergebnisse erklärbar sind, werden u. a. Shapley additive explanations verwendet. Dabei geht es primär darum, wie sich bei Nutzung von Explainable Machine Learning aus den Ergebnissen Schlussfolgerungen und Hypothesen herleiten lassen könnten. Um wirklich Schlussfolgerungen zu ziehen, sind die Mechanismen von den diversen Herzkrankheiten auch zu komplex und sollten im Detail auch immer durch Kardiologen bewertet werden. Desweiteren wird in dieser Situation der Einfluss der Stichprobengröße auf die Genauigkeit von Klassifikationsalgorithmen untersucht und verschiedene Wissensstände des Fachexperten simuliert.

Theoretischer Hintergrund

2.1 Beschreibung des Datensatzes

Der Datensatz, der untersucht wird, ist der „Heart Disease“-Datensatz, der verfügbar ist über die UC Irvine Machine Learning Repository Server von dem „Center for Machine Learning and Intelligent Systems“ der University of California, Irvine [3]. Für den Datensatz wurden von 1981 bis 1987 Daten erhoben, die es ermöglichen sollten, in einem gewissen Rahmen kardiovaskuläre Krankheiten mithilfe wahrscheinlichkeitstheoretischer Methoden besser vorherzusagen bzw. diagnostizieren zu können. Die Studien wurden durch Kardiologen an den unterschiedlichen Standorten, an denen Daten gesammelt wurden, geplant und der Datensatz wurde dann zunächst von den Erstellern im Rahmen von Studien für Publikationen genutzt [5, 6]. Die Daten wurden in Kooperation mit dem Institut erhoben, dort v.a. für die weitere Nutzung aufbereitet und dann auch über den UC Irvine Machine Learning Repository Server veröffentlicht. Dabei wurden die Patientendaten anonymisiert. Obwohl der Datensatz als „Heart Disease“-Datensatz bezeichnet wird, wird im Folgenden aber davon ausgegangen, dass mit dem Datensatz koronare Herzkrankheiten (CHD) und damit direkt nur ein Teil der Herzkrankheiten untersucht werden können. Hingegen können nicht direkt sämtliche Herzkrankheiten berücksichtigt werden. Diese Annahme ist dadurch gerechtfertigt, dass es bei den Publikationen der Ersteller um CHD ging [5]. Desweiteren kann der Datensatz aber auch mit gewissen Annahmen für kardiovaskuläre Krankheiten im Allgemeinen bzw. zur Bewertung der Herzgesundheit genutzt werden.

Bei der Datenerhebung wurden unterschiedlichste Merkmale über die Patienten gesammelt, wie grundlegende Patientendaten bis hin zu Daten, die aus bildgebenden Verfahren wie der Fluoroskopie Bildgebung erhoben wurden. Welche Merkmale das genau sind, wird im Weiteren erläutert (s. dazu auch Anhang 1). Dabei wurden die Daten in vier unterschiedlichen Kliniken erhoben. Aus der Dokumentation zum Datensatz kann geschlossen werden, dass die Daten in den Vereinigten Staaten durch Veterans Health Administration (V.A. Medical Center), an den Standorten Cleveland durch die Cleveland Clinic Foundation und Kalifornien, Long Beach, [7] erhoben wurden. Desweiteren wurden Daten im „Hungarian Institute of Cardiology“ [8] sowie den Universitätskrankenhäusern in Zürich und Basel erhoben. Die Daten aus Zürich und Basel wurden dabei in einem Rohdatensatz zusammengefasst. Die Datenerhebung an den unterschiedlichen Standorten erscheint v.a. aufgrund der großen Disparitäten, die regional z.B. durch gewisse Gewohnheiten bei CHD existieren, geeignet. Dadurch kann die Verzerrung der Daten, die regional auftreten können, begrenzt werden bzw. es gibt Vergleichspunkte, falls solche Verzerrungen auftreten. Damit ist es prinzipiell auch möglich, mit dem Datensatz regionale Disparitäten, die bei CHD auftreten, untersuchen zu können. Dieser Ansatz wurde vermutlich auch durch die Ersteller der Datensätze verfolgt, denen für ihre Studien die Stichproben aus Long Beach, Ungarn und der Schweiz als Testgruppen bzw. Kontrollgruppen dienten [5]. Insgesamt sind es damit vier unterschiedliche Stichproben bzw. vier unterschiedliche Datensätze. Die Datensätze werden im Folgenden als „Cleveland“-Datensatz, „VA Medical“-Datensatz, „Hungarian“-Datensatz und „Switzerland“-Datensatz bezeichnet. Es waren nach dem Einlesen der Rohdaten insgesamt 899 Untersuchungsobjekte bzw. Patienten aufgeteilt auf die vier Datensätze (s. Tabelle).

Der Cleveland Datensatz χ_1 und Hungarian Datensatz χ_2 werden im Folgenden genutzt. Der VA Medical und Switzerland Datensatz werden nur supplementär zur Erstellung weiterer Datensätze verwendet, da diese einen hohen Anteil an fehlenden Daten haben.

	Anzahl	Nutzung
Cleveland	282	χ_1
VA Medical	200	supp
Hungarian	294	χ_2
Switzerland	123	supp
Summe	899	

Tab. 1: Überblick über die eingelesenen Daten

Dabei ist zu erwähnen, dass in vielen Publikationen und Arbeiten (s. 2.2 Literaturrecherche), wenn der „Heart Disease"-Datensatz untersucht wurde, oft gemeint ist, dass nur ein Datensatz verwendet wurde, nämlich der „Cleveland"-Datensatz. Es wurden in vielen der Publikationen nur Daten ausgewertet, die in den Vereinigten Staaten gesammelt wurden. Desweiteren wurde oft nur eine Teilmenge der erhobenen Merkmale (d = 13) untersucht, während die rohen Datensätze sogar bis zu 80 Merkmale nach dem Einlesen hatten. Einige der Merkmale wurden jedoch nicht genutzt und deshalb werden oft 75 Merkmale genannt, wenn der Datensatz vollständig ausgewertet wurde. Insgesamt bietet der „Heart Disease"-Datensatz also deutlich mehr Informationen als vielfach ausgewertet wurden. Es kann vermutet werden, dass oft 13 Dimensionen untersucht wurden, da in den Publikationen der Ärzte für ihre Studien nur 13 der Merkmale relevant waren [5]. Werden jedoch die Daten in einem anderen Kontext untersucht, so können durchaus mehr Merkmale sinnvoll verwendet werden. Diese Merkmale sollen es ermöglichen, auf die Zielvariable zu schließen, die binär aussagt, ob für den jeweiligen Patienten eine koronare arterielle Herzkrankheit (CAD) als single vessel disease (SVD) oder multi vessel disease (MVD) diagnostiziert wurde. Wie bereits bei den Merkmalen liefert auch die Zielvariable allerdings weitere Informationen nämlich bezogen darauf, wieviele und welche Vesseln als blockiert diagnostiziert wurden. Zu der Skala gibt es direkt durch die Dokumentation des Datensatzes keine weitere Information, aber aus den Publikationen kann darauf geschlossen werden, dass eine Angiographie durchgeführt wurde und dann alle Vesseln, die blockiert waren, notiert waren [5]. Die Skala gibt die gesamte Anzahl der blockierten Vesseln an.

Insgesamt ist der „Heart Disease"-Datensatz immer noch einer der populärsten medizinischen frei verfügbaren Datensätze. Durch die Dokumentation, die Publikationen und Artikel aus dem Internet gibt es vergleichsweise viele Informationen und die Bedeutung der Merkmale kann relativ gut rekonstruiert werden. Dabei ist allerdings zu erwähnen, dass die Dokumentation nicht vollständig ist. Trotzdem lassen sich z.B. aus den Abkürzungen der Variablen nach eigener Recherche relativ gut die Bedeutung der Variablen und die dabei durchgeführten Untersuchungen herausfinden. Besonders die Publikationen sind hilfreich. Zu welchen Variablen Annahmen über deren Bedeutung gemacht werden, wird im Folgenden natürlich dokumentiert. Der „Heart Disease"-Datensatz ist desweiteren ein Referenzdatensatz, anhand dessen die Performance von ML-Algorithmen gut nachvollzogen werden kann, da er schon sehr oft mittels ML untersucht wurde. Der Datensatz zeichnet sich bei den medizinischen Datensätzen v.a. dadurch aus, dass er sehr breit angelegt ist und versucht,

zu vielen Aspekten von kardiovaskulären Krankheiten bzw. zu sämtlichen Diagnostiken Daten zu erheben. Dies ist für kardiovaskuläre Krankheiten auch sinnvoll, da sie nicht als einzelne Krankheiten betrachtet werden können sondern komplexe Wirkungsmechanismen haben. Unterschiedlichste Aspekte können also auf kardiovaskuläre Krankheiten hindeuten.

Es gibt weitere Datensätze, die ähnlich sind aber entweder weniger bekannt oder weniger ausführlich. Dies sind der „Z-Alizadeh Sani"-Datensatz, der ebenfalls im UCI ML Repository zu finden ist [9], und der „Framingham Heart Disease"-Datensatz [10]. Der „Framingham Heart Disease"-Datensatz untersucht zwar mehr Patienten aber dafür nur 5 Merkmale. Viele andere medizinische Datensätze sind auf bestimmte Diagnostiken limitiert wie z.B. nur auf die Untersuchungen beim EKG (wie beim „PTB-XL"-Datensatz für Elektrokardiographie). Aufgrund der breiten Untersuchungen beim „Heart Disease"-Datensatz mittels unterschiedlicher Diagnostiken ist er geeignet, um bestimmte Lernalgorithmen des ML wie Multiview Learning oder Multi Feature Set Learning zu bewerten. Die Datensätze sind natürlich wertvoller, je breiter die Auswahl der Merkmale ist, gerade wenn es um eine umfassende Bewertung von kardiovaskulären Krankheiten geht. Dies ist besonders der Fall, wenn es, wie in der präventiven Kardiologie, um eine vollständige Bewertung der Herzgesundheit geht, und ist damit auch für das CDSS gut geeignet. Trotzdem enthält der Datensatz auch detailliertere Informationen, wie sie z.B. für einen Kardiologen interessant wären, und hat damit eine gewisse Tiefe. In dem verwendeten Datensatz wurden trotzdem nicht sämtliche relevanten medizinischen Parameter erhoben, wenn bestimmte kardiovaskuläre Krankheiten im Detail untersucht werden sollen. Besonders aus EKG-Ergebnissen, die hohe diagnostische Bedeutung haben, hätten sich noch mehr Informationen ziehen lassen, wenn z.B. die Zeitreihen der EKG-Untersuchungen angegeben wären. Auch andere Diagnosemöglichkeiten wurden nicht ausgenutzt, wie die Auswertung von Herzgeräuschen wie sie in der „PhysioNet/CinC Challenge 2016" untersucht wurden. Insgesamt ist ein Teil des Interesses für den „Heart Disease"-Datensatz vermutlich darauf zurückzuführen, dass als einzige Information der Zielvariablen gegeben ist, ob ein Patient eine CAD hat aber nicht, was dazu geführt hat. Desweiteren ist es auch interessant herauszufinden, ob es aus einer Angabe der Zielvariablen möglich ist, anhand der Ergebnisse Wirkungsmechanismen für einzelne Herzkrankheiten herzuleiten.

2.2 Literaturrecherche

Zu dem betrachteten Datensatz gibt es bereits viele Untersuchungen und Publikationen. Dabei beziehen sich die meisten jedoch auf die Situation, in denen 13 oder weniger Merkmale verwendet wurden. Es gibt zahlreiche Publikationen, bei denen in dieser Situation als Vergleich verschiedene ML-Algorithmen auf ihre Eignung untersucht wurden [11]. Desweiteren wurden in dieser Situation verschiedene spezielle ML-Algorithmen verwendet [12]. Dabei gab es neben der Klassifikation auch Ansätze mittels Clustering. Dazu gehören auch die Artikel der Ersteller der Datensätze. Insgesamt war die Untersuchung mit ML dabei sehr erfolgreich [11]. Die Situation mit 13 betrachteten Merkmalen ist jedoch nicht direkt vergleichbar mit der Situation, in denen mehr Merkmale berücksichtigt werden (hier sind es 33). Ein Review von ML-Algorithmen, die bei diesen Datensatz für mehr als 13 Merkmale angewendet wurden, konnte bei der Literaturrecherche nicht gefunden werden. Es gibt jedoch Publikationen, bei denen einzelne ML-Algorithmen bei mehr als 13 Merkmalen untersucht wurden [13]. Bei dem genannten Artikel handelt es sich um 29 Merkmale und 303 Untersuchungsobjekte des Cleveland-Datensatzes. Es wurde dabei ein neuronales Netzwerk angewendet und eine Accuracy von 83,67 % erreicht. Bei der Literaturrecherche wurden keine Publikationen gefunden, bei denen andere Datensätze als der Cleveland-Datensatz verwendet wurde oder versucht wurde, die Daten aus allen Standorten integrativ zu nutzen. Damit ist unklar, wie sich größere Stichproben auswirken würden. Desweiteren wurde der Datensatz mit den 13 Merkmalen auch bereits mittels verschiedener Methoden des Explainable Machine Learnings untersucht [14].

Der Literatur ist zu entnehmen, dass der Datensatz also bereits mittels ML erfolgreich untersucht wurde. Jedoch tritt die Situation, in der mehr als 13 Merkmale verwendet wurden, weniger häufig auf. Es bietet sich deshalb für diese Arbeit an, für 33 Merkmale verschiedene ML-Algorithmen zu vergleichen. Die Ausgangslage ist dabei realistischer, da nicht immer bekannt sein muss, welche Merkmale relevant sind, und oft viele Merkmale berücksichtigt werden müssen, um überhaupt detaillierte Informationen dafür zu finden, dass die Anwendung von ML sinnvoll ist. Denn in der Situation, in der das ML-System nur Lösungen bietet, die ein Fachexperte ohnehin findet, ist es weitgehend nutzlos. Dabei würde als Referenz die hohe Accuracy von 83,67 % dienen [13]. Dabei soll allerdings auch berücksichtigt werden, ob die gefundenen Lösungen ausreichend erklärbar sind. Es stellt sich deshalb die Frage, ob diese Accuracy auch mit einfacheren Modellen als neuronalen Netzwerken erreicht werden kann. Um dies zu erreichen, sollen alle Daten aus allen Standorten integrativ genutzt werden und damit die Frage beantwortet werden, wie sich eine größere Stichprobe auswirkt. Zudem bietet es sich an, unterschiedliche Wissensstände des Fachexperten zu simulieren und die Ergebnisse zu vergleichen. In der Literaturrecherche wurde kein Artikel für einen solchen Vergleich gefunden.

2.3 Medizinischer Hintergrund

2.3.1 Einordnung koronare Herzerkrankungen

Kardiovaskuläre Krankheiten (CVD) ist der Überbegriff für sämtliche Krankheiten des Herzens oder des Herz-Kreislaufsystems [15]. Desweiteren gibt es den Begriff Herzkrankheiten, die sich von den kardiovaskulären Krankheiten unterscheiden, da sie nur das Herz betreffen. Damit ist jede Herzkrankheit eine kardiovaskuläre Krankheit aber nicht jede kardiovaskuläre Krankheit eine Herzkrankheit [16]. Außerdem gibt es periphere arterielle Erkrankungen, die im erweiterten Sinne das Kreislaufsystem beeinflussen, die ebenfalls zu den CVD gezählt werden. Bei den Herzerkrankungen sind v.a. die koronaren Herzerkrankungen relevant, da sie am häufigsten auftreten und unterschiedlichste weitere Erkrankungen hervorrufen können. Deshalb wird auch oft verallgemeinernd von Herzkrankheiten gesprochen auch dann, wenn eigentlich nur koronare Herzerkankungen gemeint sind. Selbst wenn über kardiovaskuläre Krankheiten geredet wird, ist oft Arteriosklerose der Auslöser und damit ist es eine koronare Herzerkrankung [15]. Es ist jedoch wichtig den Unterschied zwischen kardiovaskulären Krankheiten und Herzkrankheiten zu berücksichtige. Desweiteren muss man auch zwischen koronaren Herzerkrankungen und koronaren arteriellen Erkrankungen unterscheiden. Mit koronaren arteriellen Erkrankungen (CAD) ist im weitesten Sinne Arteriosklerose gemeint. Koronare Herzerkrankungen sind sämtliche Störungen des Kreislaufs durch teilweise oder weitgehend zugesetzte Blutgefäße, die auf den koronaren arteriellen Erkrankungen basieren. Das Zusetzen von Vesseln (Gefäßen) durch Plaque bei der Arteriosklerose kann zu Herzattacken führen und, wenn sich Gefäße im Gehirn zusetzen, zum Schlaganfall. Neben den koronaren Herzerkrankungen gibt es die Herzinsuffizienz, dabei reicht der Kreislauf nicht mehr für die Aktivität des Körpers aus. Auch die Herzinsuffizienz kann auf koronare Herzerkrankungen zurückzuführen sein [17]. Es gibt weitere Herzkrankheiten wie Herzrhythmusstörungen, Herzklappenfehler und Bluthochdruck, von denen einige weniger kritisch sind als koronare Herzerkrankungen, da sie therapiert werden können. Einige treten davon auch nur temporär auf wie die Perikarditis oder die Myocarditis. Zusammen beeinflussen diese Erkrankungen die allgemeine Herzgesundheit. Weitere Herzkrankheiten werden in diesem Abschnitt nicht genauer betrachtet, da der Datensatz v.a. nur Aussagen über koronare Herzerkrankungen oder die allgemeine Herzgesundheit ermöglicht.

2.3.2 Koronare arterielle Erkrankung und Folgen

Die koronaren arteriellen Erkrankungen können zu unterschiedlichen Erscheinungen bzw. Erkrankungen führen. Dazu gehören Herzattacken (auch als myokardialer Infarkt bezeichnet), Herzarrhythmien, der plötzliche Herztod und viele weitere Krankheiten [18]. Es ist jedoch zu berücksichtigen, dass eine koronare arterielle Erkrankung nicht mit diesen aus der Erkrankung folgenden Erscheinungen bzw. Komplikationen gleichgesetzt werden kann. Eine koronare arterielle Erkrankung kann zwar zu den genannten Komplikationen führen, aber der direkte Schluss, dass eine koronare Herzerkrankung in jedem Fall zu den Komplikationen führt, ist medizinisch nicht korrekt. Dies ist darauf zurückzuführen, dass die medizinischen Zusammenhänge, die zu diesen Komplikationen führen, medizinisch deutlich komplexer sind. Deshalb sollten diese Komplikationen oder Folgeerkrankungen immer kardiologisch beurteilt werden. Aus diesem Grund kann das CDSS auch nicht eine medizinische Beurteilung ersetzen, es kann nur CAD vorhersagen. Außerdem gibt es sehr viel mehr Komplikationen, auf die auch wegen der medizinischen Komplexität nicht eingegangen werden kann. Wie es zu den koronaren arteriellen Erkrankungen kommen kann, wird falls nötig genauer bei der Auswertung des Algorithmus berücksichtigt. Aus diesem Grunde wäre es auch nicht korrekt von den Ergebnissen des CDSS direkt auf eine aus CAD folgende Komplikation zu schließen, was teils bei einigen Auswertungen des „Heart Disease“-Datensatzes der Fall war. Es wurde dabei nicht berücksichtigt, dass ursprünglich nur koronare Herzerkrankungen untersucht wurden [5]. Damit ist

es zunächst die Aufgabe des CDSS, koronare Herzerkrankungen vorherzusagen. Diese können jedoch Indikatoren für die daraus folgenden Komplikationen sein.

2.3.3 Einfluss Ernährung

Die Ernährung hat einen großen Einfluss auf Herzkrankheiten im Allgemeinen und im Besonderen auf die koronaren Herzkrankheiten. Eine falsche Ernährung kann Herzkrankheiten hervorrufen. Die Ernährung ist oft geprägt von vielen tierischen Fettsäuren, die vor allem in verarbeiteten Fleischprodukten, fettem Käse, Wurst und Süsswaren zu finden sind. Das heißt, der Körper erhält zu viele gesättigte Fettsäuren und zu wenige ungesättigte. Gleichzeitiger Bewegungsmangel kann zu Übergewicht und Fettleibigkeit führen. Dies begünstigt koronare Herzerkrankungen. Dabei ist v.a. Cholesterin relevant. Durch das Cholesterin können sich in den Gefäßwänden Ablagerungen bilden und damit koronare Herzerkrankungen auslösen. Cholesterin ist dabei direkt an der Bildung von Plaques in den Gefäßen beteiligt. Besonders sind dabei Fette relevant; hochwertige Fette haben guten Einfluss und minderwertige Fette schlechteren Einfluss. Aber auch Kohlenhydrate können zu hohen Cholesterinspiegeln führen. Es ist also eine ausgewogene Ernährung wichtig. Besonders in Indien sind die hohen Zahlen an koronaren Herzkrankheiten auf die falsche Ernährung zurückzuführen. Auch Gewohnheiten können einen Einfluss haben, so ist es z.B. negativ, Fett mehrfach zu verwenden (wie z.B. in Indien). Dauerhafter erhöhter Konsum von rotem Fleisch ist negativ (wie z.B. in Indien) statt als Alternative weißes Fleisch zu wählen [19]. In den Datensätzen wird dieses Problem in der Variablen „X_6: chol“ berücksichtigt. Das hier gemessene Cholesterin im Blut umfasst sowohl das „gute“ HDL-Cholesterin als auch das „schlechte“ LDL-Cholesterin. Als günstig gilt ein gesamter Cholesterinwert von 200 mg/dl [20].

2.3.4 Einfluss genetischer Faktoren

Genetische Einflussfaktoren gibt es, die Auswirkungen sind aber verglichen mit anderen Einflüssen nur moderat. Es kann z.B. ein erhöhtes Risiko für koronare Herzkrankheiten durch Gene weitervererbt werden. Deshalb ist es eine Routine von Kardiologen nach Fällen von koronaren Herzkrankheiten in der Verwandtschaft zu fragen. In dem Datensatz ist diese Untersuchung durchgeführt und als hist_cad notiert worden. Desweiteren können sich prädisponierende Bedingungen wie Diabetes als genetischer Einflussfaktor auswirken. Auch Mutationen können entweder positiven oder negativen Einfluss haben. Tatsächlich können auch de novo Mutationen kardiovaskuläre Krankheiten bevorzugt oder weniger oft auslösen.

2.3.5 Einfluss Rauchen

Rauchen hat einen negativen Einfluss auf koronare Herzkrankheiten. Durch das Rauchen kann es verstärkt zu Arteriosklerose kommen [21]. Dabei ist aber zu bedenken, dass dieser Einfluss bezogen auf die Folgen der koronaren Herzerkrankung nur in geringerem Ausmaß einen langfristigen Einfluss hat. So konnte nachgewiesen werden, dass das Aufhören mit dem Rauchen zur Reduktion des Risikos sehr effektiv ist, bereits nach mehreren Jahren konnte kein signifikanter Einfluss des Rauchens mehr festgestellt werden [22]. Dies zeigt desweiteren, dass koronare Herzkrankheiten und die Folgen aus koronaren Herzkrankheiten nicht gleichgesetzt werden können. Der Einfluss des Rauchens ist im Datensatz mit den Merkmalen cigs und cig_time berücksichtigt.

2.3.6 Einfluss Alkohol

Alkohol ist einer der Risikofaktoren für Herzkrankheiten, die mit Gewohnheiten zu tun haben. Alkohol kann zu Kardiomyopathie und auch zu koronaren Herzerkrankungen (ischämischen Herz-

erkrankungen) führen und damit auch zu Herzattacken. Die Ergebnisse der Studien sind dabei auf den ersten Blick sehr uneindeutig [23]. Häufig wird erwähnt und auch durch Studien bestätigt, dass selbst moderater Alkoholkonsum an mehreren Tagen in der Woche das Risiko für koronare Herzerkrankungen senken könnte. Gleichzeitig gibt es auch viele Anzeichen, dass Alkoholkonsum das Risiko stark ansteigen lässt. Das größte epidemiologische Anzeichen sind die übermäßig hohen Inzidenzen an ischämischen Herzerkrankungen in Russland [24]. Aufgrund der stark erhöhten Fallzahlen in Russland wurden viele Studien durchgeführt, die zu dem Ergebnis kommen, dass neben anderen möglichen Faktoren v.a. der starke Alkoholkonsum (dabei in Russland sicherlich Wodka als hauptsächliche Spirituose) der Grund wäre. Sicher ist wohl nur, dass leichter Alkoholkonsum das Risiko senkt, insgesamt ein höherer Alkoholkonsum aber für ein höheres Risiko an ischämischen Herzerkrankungen sorgen kann. In der genannten Studie werden die Faktoren kontrovers diskutiert und festgestellt, dass ein erheblicher Anteil auch durch Konfundierungseffekte zustande kommen könnte. Welche weiteren Gründe zu den unterschiedlichen Studienergebnissen führen, müsste noch weiter untersucht werden. Alkohol kann aber nach den russischen Studien als alleiniger Faktor ausreichen. Der Alkoholkonsum wurde durch den Datensatz jedoch nicht berücksichtigt, so dass bei einigen Patienten eine Herzkrankheit durch den Algorithmus möglicherweise nicht vorausgesagt werden kann.

2.3.7 Einfluss Diabetes

Diabetes ist einer der Risikofaktoren für verschiedene kardiovaskuläre Krankheiten. Dazu gehören v.a. die koronaren Herzkrankheiten und ein erhöhtes Risiko für Schlaganfälle. Dabei ist es allgemein wahrscheinlicher, dass es zu koronaren Herzkrankheiten kommt, je länger ein Patient Diabetes hat. Bezogen auf koronare Herzkrankheiten wurde bei Patienten mit Diabetes mellitus ein 2- bis 4-fach höheres Risiko festgestellt [25]. Deshalb haben Patienten mit Diabetes häufig auch eine koronare Herzerkrankung. Dabei ist zu berücksichtigen, dass bezogen auf sämtliche Ursachen für koronare Herzkrankheiten Diabetes nur eine relativ geringe Rolle spielt. Dies ist darauf zurückzuführen, dass die Anzahl an Patienten mit koronarer Herzkrankheit, die Diabetes haben, relativ gering ist. So wurden bei allen Datensätzen nur ca. 18 % der Patienten mit Diabetes für die Studien ausgewählt. Dies ist repräsentativ für den Anteil an Patienten mit Diabetes. So haben in Deutschland 18 % der Männer und 14 % der Frauen Diabetes. Der Anteil an Patienten mit koronarer Herzkrankheit und Diabetes gleichzeitig war mit 11 % nur gering. Jedoch bezogen auf die Patienten mit Diabetes hatten 65 % auch eine koronare Herzkrankheit (dabei ist allerdings bei dem Datensatz zu beachten, dass die koronare Herzkrankheit bei vielen dieser Patienten auch auf andere Faktoren zurückgeführt werden könnte). Es ist zu berücksichtigen, dass sich Diabetes Typ 1 und Diabetes Typ 2 bezogen auf die Wirkungsmechanismen unterscheiden. Besonders bei Typ 2 spielt wie zuvor erwähnt eine fehlerhafte Ernährung häufig eine Rolle. Diese Unterscheidung kann jedoch nicht anhand der Datensätze vorgenommen werden, da nicht in Diabetes Typ 1 und Typ 2 unterschieden wurde. Desweiteren tritt bei Diabetes häufig Bluthochdruck auf, welches einer der hauptsächlichen Gründe für eine koronare Herzkrankheit sein kann. Ebenso kann Diabetes in Zusammenhang mit zu hohem Cholesterin für diese Krankheit sorgen. Festgestellt wurde Diabetes durch das Merkmal „X_5: dm, fbs“, wobei entweder basierend auf einer vorherigen Diagnose oder einem Test auf Diabetes vor der kardiologischen Untersuchung festgestellt wurde, ob die Patienten Diabetes hatten.

2.4 Machine Learning und verwendete Algorithmen

2.4.1 Beispiel für eine Klassifikation

In diesem Abschnitt wird dargestellt, wie die Klassifikation durchgeführt wird. Klassifikationen können als Schätzer für die Assoziation von Untersuchungsobjekten zu Klassen dienen. Bei der Klassifikation wird ein Datensatz χ in die Datenmatrix X und die Zielwerte Y aufgeteilt [26]. Dazu werden zunächst Daten in einer Studie gesammelt oder es werden bereits vorhandene Datensätze genutzt. Müssen sie gesammelt werden, wie in einer Studie, so muss ein geeignetes Stichprobendesign gewählt werden. Danach könnte es sich anbieten den Datensatz aufzuteilen. Dies hängt u.a. von dem jeweiligen Lernansatz ab. Klassisch wird der gesamte Datensatz untersucht, es findet dann keine Aufteilung statt. Dies wird als Full-Space Machine Learning bezeichnet. Bei einer Datenanalyse strukturierter Daten wird dann eine Datenmatrix X mit Features (Feature Ansatz) erstellt. Für unstrukturierte Daten (wie Daten zu Texten oder medizinischen Bildern) wäre ein möglicher Ansatz, sie ebenfalls in den Feature Ansatz zu überführen. So wird dieselbe Darstellung für sämtliche Untersuchungen möglich.

Ausgangssituation auf Basis der Daten:

X: Datenmatrix

Y: Zielwerte

N: Anzahl der Untersuchungsobjekte

m: Index der Untersuchungsobjekte

d: Anzahl der Merkmale

$\vec{x}_1 = (x_{1,1},\ldots,x_{1,d}),\ldots,\ \vec{x}_N = (x_{N,1},\ldots,x_{N,d})$

$\chi = (\vec{x}_1,\ y_1),\ldots,(\vec{x}_N,\ y_N)$

Für eine binäre Klassifikation mit einer Zielvariablen:

$y_m = 0$ für Klasse 0

$y_m = 1$ für Klasse 1

$\chi \subseteq\ M_x \times\ M_y = \{\vec{x}_1,\ldots,\vec{x}_N\} \times \{0,1\}$

$$X_{(N \times d)} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,d} \end{pmatrix}$$

Jede Zeile dieser Datenmatrix, die einem einzelnen Objekt (den Beobachtungen) entspricht, lässt sich algebraisch betrachtet als Datenpunkt, typischerweise als ein Punkt in einem d-dimensionalen reellen Vektorraum $\mathbb{R}^d$ auffassen. Geometrisch betrachtet entspricht jeder Beobachtung dann ein Punkt in einem Koordinatensystem. Der dargestellte Raum wird als Merkmalsraum (Featureraum) bezeichnet. Desweiteren sollte der Raum gewisse Eigenschaften erfüllen, er sollte topologisch sein und auf ihm muss eine Metrik definiert sein (metrischer Raum).

$$f\colon M_x \rightarrow M_y$$

Dann werden Informationen aus dem Datensatz extrahiert. Dabei wird die Abbildung f bzw. die allgemeine Vorschriften gesucht, wie man die gegebenen bekannten Zielwerte 0 und 1 erhält [26]. Bei mehreren Zielvariablen ergibt sich eine Matrix mit den Zielwerten und kein Vektor. Teilweise wird dies auch als Optimierung einer Zielfunktion bezeichnet. Hierfür werden mathematische oder statistische Verfahren der Optimierung verwendet, da es nicht immer möglich ist, genau auf die Zielwerte zu schließen. Dies entspricht einer Partitionierung des Datenraums. Die Abbildung bzw. die Vorschrift, um auf die Zielwerte zu schließen, repräsentiert das ML-Modell. Das Prinzip ist, dass möglichst ähnliche Datenpunkte in derselben Partition (demselben Gebiet) des Merkmalsraums liegen. Welches Konzept für Ähnlichkeit verwendet wird, kann sich darin unterscheiden, welches ML-Modell angewendet wird. Aus dem Modell lassen sich die Klassenzuordnungen für alle der Datenpunkte deduzieren. Es soll im Prinzip sowohl die Ähnlichkeit von Datenpunkten als auch der gegebenen Zielwerte vorliegen. Da beim „Heart Disease“-Datensatz Zielwerte zu allen Objekten, also allen Patienten, gegeben sind, ist es supervised. Da binäre Zielwerte gegeben sind, ist es eine binäre Klassifikation.

Mit Hilfe der Abbildung f wird für eine neue Beobachtung $\vec{x}_l$ der Zielwert y_l vorhergesagt:

$$\vec{x}_l \xrightarrow{f} y_l$$

Es folgt die Evaluation, bei der die Performance des ML-Algorithmus betrachtet wird, und untersucht, ob sich das gefundene Modell für die jeweiligen Ziele eignet. Dazu wird typischerweise Kreuzvaliderung genutzt. Besteht das Ziel nur in der Modellierung, ist die Klassifikation an dieser Stelle beendet. Oft bestehen aber darüber hinausgehende Ziele, wie die Vorhersage oder Prognose. Dann wird das Modell auf neue Objekte angewendet. Aus dem Modell können neue Vorhersagen getroffen werden. Die Zuordnung eines neuen Punktes wird als Allokation bezeichnet.

2.4.2 Darstellung weiterer verwendeter Lernansätze

Bei den verwendeten Algorithmen können nicht nur Fullspace-Machine-Learning-Methoden verwendet werden, sondern es gibt zahlreiche weitere Lernansätze. Für einen der folgenden Machine-Learning-Algorithmen ist z.B. das Multi View Learning relevant [27, 28]. Dabei werden die Merkmale, die untersucht werden, aufgeteilt in unterschiedliche Teilmengen. Die Anzahl der Merkmale, die jeweils zusammen untersucht werden, wird damit effektiv reduziert. Dabei stellen unterschiedliche „views“ die verschiedenen Mengen von Merkmalen als ein Blickwinkel bzw. Aspekt aus den Daten dar. Das wird auch als Multi Feature Set Learning bezeichnet. Dies hat verschiedene Vorteile. Es ist keine Aufteilung nötig, wenn bereits verschiedene Datensätze gegeben sind.

v: Index der Datenviews

Aufteilung der Datenmatrix:

$X^{(v=1)}, X^{(2)}, \ldots, X^{(V)}$

Im Gegensatz zu Fullspace-Machine-Learning-Methoden wird dann z.B. nicht nur ein ML-Modell aufgebaut sondern eines für jeden Datenview. Bezogen auf die durchzuführende Klassifikation generiert jedes der ML-Modelle ein Klassifikationsschema, das einzeln ausgewertet werden könnte.

$f_1 : M_{x,v=1} \rightarrow M_y$
$f_2 : M_{x,v=2} \rightarrow M_y$
...
$f_V : M_{x,v=V} \rightarrow M_y$

Optimalerweise sollte der Algorithmus aber einen Konsenz für die verschiedenen Klassifikationsschemata finden. Dazu werden die Ergebnisse mehrerer einzelner ML-Modelle zu einer Abbildung f zusammengeführt oder rekombiniert. Es resultiert die Klassifikation. Dies wird in unterschiedlichen Kontexten auch als ensemble learning oder classifier fusion bezeichnet.

$f : d(f_1, f_2, \ldots, f_V)$

Diese Modelle sind für Entscheidungssysteme relevant und wurden dort am häufigsten verwendet. Das Ziel ist eine Integration von Informationen aus unterschiedlichen Datenquellen und wird auch als data integration bezeichnet.

2.4.3 kNN Algorithmus

Der kNN Algorithmus (k-Nächste Nachbarn Klassifikation) ist ein Klassifikationsalgorithmus. Das Vorgehen des kNN Algorithmus kann einfach beschrieben werden. Die Klassifikation beruht auf den Klassen bzw. Zielwerten und darauf, ob die jeweils zu klassifizierenden Beobachtungen sich ähnlich sind. Dazu wird ein d-dimensionaler Hyperball um jeden Datenpunkt definiert, wenn als Metrik der euklidische Abstand genutzt wird. Der Radius kann sich für jeden Datenpunkt unterscheiden und hängt von dem Parameter k ab (Anzahl der k nächsten Nachbarpunkte). Die jeweils ähnlichen Beobachtungen sind nun Teilmenge des Hyperballs bzw. liegen innerhalb. Jede Beobachtung wird demnach klassifiziert, welche Klassen die jeweils ähnlichen Beobachtungen haben. Es wird die Klasse zugewiesen, die eine größere Anzahl an ähnlichen Beobachtungen haben [29].

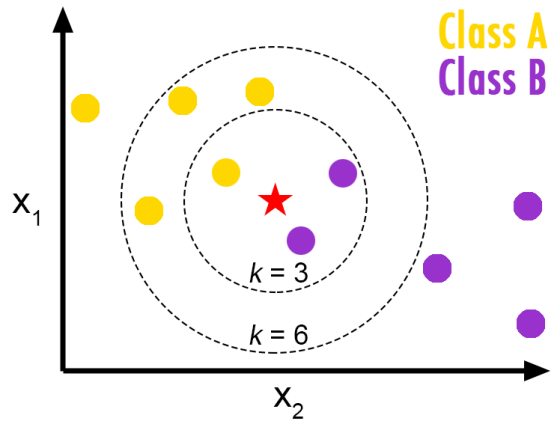


Abb. 1: Prinzip der kNN Klassifikation, Quelle: https://helloacm.com/a-short-introduction-to-k-nearest-neighbors-algorithm

2.4.4 Entscheidungsbäume

Entscheidungsbäume gehören zu den nichtmetrischen Methoden des Machine Learning, da sie keine Distanzmetrik der Datenpunkte zur Klassifikation benötigen. Sie werden zur Klassifikation angewendet, indem binäre Bäume bzw. gerichtete Graphen aufgebaut werden (s. Abb. 2). Die Graphen bestehen aus Knoten, die durch gerichtete Pfade verbunden sind, die auch als Verzweigungen bezeichnet werden. Dabei steht jeder Knoten für eine Teilgruppe der Daten. Bei Entscheidungsbäumen werden die Untersuchungseinheiten an den Knoten (nodes) bzw. entlang der Pfade hierarchisch aufgeteilt in verschiedene Gruppen, die dann als Zweige des Entscheidungsbaums bezeichnet werden. Von jeder der intermediären Knoten geht ein Teilbaum aus, durch den die Daten weiter aufgeteilt werden. Für die Aufteilung wird jeweils eines der Merkmale aus den Daten gesucht, nach dem sich die Daten in einer Weise bestmöglich aufteilen lassen. Als Kriterium dafür dient ein Maß der Inhomogenität, das sich unterscheiden kann je nachdem welcher Algorithmus konkret verwendet wird. Übliche Maße sind die Gini Inhomogenität oder die hier angewendete Entropie. So wird versucht die Inhomogenität in den jeweiligen Gruppen bestmöglich zu reduzieren. Die Entropie lässt sich damit nach der physikalischen Bedeutung des Wortes aus der Thermodynamik verstehen. Je stärker unterschiedliche Komponenten vermischt sind, hier die verschiedenen Klassen der Untersuchungseinheiten innerhalb einer Gruppe an Untersuchungseinheiten, umso größer ist die Entropie. Das

Ziel des Entscheidungsbaumes ist es nun, diese Entropie möglichst zu reduzieren. Dazu wird der Merkmalsraum durch achsenparallele Entscheidungsoberflächen in verschiedene Gruppen an Daten aufgeteilt.

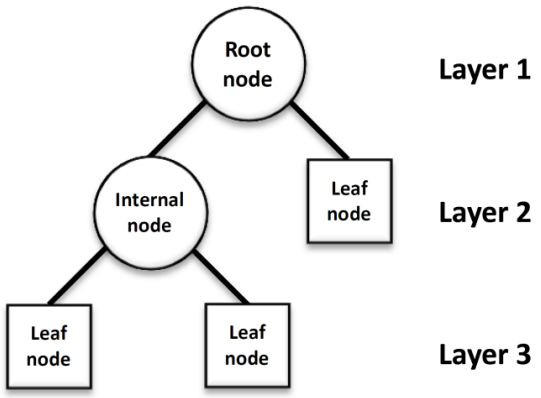


Abb. 2: Aufbau eines Entscheidungsbaum, Quelle: https://ichi.pro/de/entscheidungsbaume-in-5-minuten-161351977790824

Die genauere Vorgehensweise wird im Folgenden dargestellt und bezieht sich auf die weiter unten angegebenen Formeln und Abbildung 3. Zunächst wird am Start auf Ebene 0 für die gesamten Daten bzw. alle Untersuchungseinheiten die Entropie $H_{\varepsilon=0}(X,\vec{y})$ berechnet (s. 1). Dann wird das Merkmal gesucht, bei dem die Aufteilung des Merkmalsraumes die Entropie bestmöglich reduziert (s. 2). Dies ist in dem Beispiel das Merkmal 3. Dies entspricht einer Aufteilung der Daten durch einen Schwellenwert $\theta_{X-3,opt}$ des jeweiligen Merkmals, da die Aufteilungen achsenparallel sind. Diese Aufteilung ist die jeweilige Entscheidungsregel. Dadurch gibt es eine binäre Aufteilung der Daten in zwei Teilgruppen (in Gruppe 1 und Gruppe 2, s. Abb. 3). Für jede der Gruppen wird dieses Verfahren nun auf den folgenden Ebenen wiederholt. Das Verfahren ist dabei für jeden der Teilbäume iterativ. Die Teilmengen werden also mit jeder Ebene des Entscheidungsbaumes entsprechend neuer Entscheidungsregeln immer weiter aufgeteilt. Diese Entscheidungsregeln suchen dabei mittels Optimierung für jede Teilgruppe ein Merkmal und dessen Schwellenwert, für den sich die Daten wieder bestmöglich aufteilen lassen. Für jede Teilgruppe kann sich dieses Merkmal unterscheiden. Dadurch wird eine hierarchische Klassifikation erstellt, bei dem die Entropie der Teilgruppen immer kleiner wird. Entscheidungsbäume klassifizieren damit durch Entscheidungsfunktionen auf mehreren Ebenen. Mit jeder Ebene ε wird die Anzahl der Teilgruppen größer und ist bei Ebene n 2^n, wenn alle für eine Aufteilung nötigen Bedingungen erfüllt sind. Dies soll dazu dienen, dass die Teilgruppen an den jeweils letzten Verzweigungen der Bäume (die als Wurzeln bezeichnet werden) möglichst viele Untersuchungseinheiten mit derselben Klasse enthalten. Die Klassenwahrscheinlichkeit ist also möglichst hoch und damit sind die Teilgruppen möglichst aussagekräftig in verschiedene Klassen unterteilt. Werden für eine Teilgruppe die Bedingungen nicht mehr erfüllt, wird für diesen Zweig nicht weiter aufgeteilt.

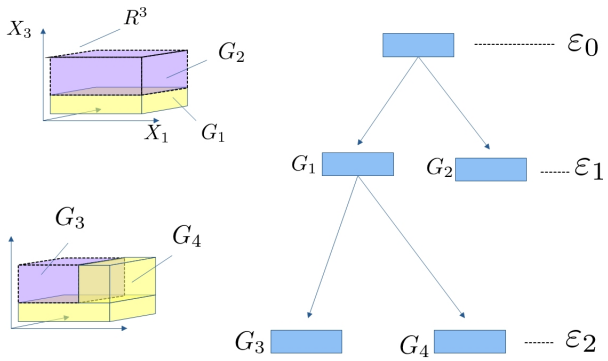


Abb. 3: Aufteilung des Merkmalsraum, Quelle: eigene Darstellung

Abkürzungen für alle folgenden Formeln:

ε: Ebene des Entscheidungsbaums

p: Klassenwahrscheinlichkeit

j: Index der Klassen 0 oder 1

X-1, X-2, X-3: Merkmal X_1,X_2, X_3 (um einen doppelten Index zu vermeiden)

Beschreibung des Algorithmus:

Root/Start, $\varepsilon = 0$:

$p_{j,\varepsilon=0} = \frac{n_{j,\varepsilon=0}}{n_{\varepsilon=0}}$

$$H_{\varepsilon=0}(X,\vec{y}) = \sum_{j=0}^{1} p_{j,\varepsilon=0} \cdot log_2(p_{j,\varepsilon=0}^{-1}) = p_{0,\varepsilon=0}\ log_2(p_{0,\varepsilon=0}^{-1}) + p_{1,\varepsilon=0}\ log_2(p_{1,\varepsilon=0}^{-1}) \qquad (1)$$

A.1: berechne die Entropie der eingegebenen Daten $H_{\varepsilon=0}(X,\vec{y})$ aus den Klassenwahrscheinlichkeiten

Aufteilung am Start, $\varepsilon = 0$:

$$\Delta H_{\varepsilon=0,X-1} \overset{B=w}{=} H_{\varepsilon=0} - H_{\varepsilon=1,X-1}(\theta_{X-1,opt}) = H_{\varepsilon=0} - \sum_{j=0}^{1} p_{j,\varepsilon=1} \cdot \log_2\left(p_{j,\varepsilon=1}^{-1}\right)$$

Mögliche Bedingung B:

$\Delta H > \Delta H_{\min}$

berechne die maximale Reduktion der Entropie für alle weiteren Merkmale, hier $\Delta H_{\varepsilon=0,X-2}$ und $\Delta H_{\varepsilon=0,X-3}$

$$\Delta H_{\varepsilon=0} = \sup\{\Delta H_{\varepsilon=0,X-1}, \Delta H_{\varepsilon=0,X-2}, \ldots, \Delta H_{\varepsilon=0,X-I}\} = \Delta H_{\varepsilon=0,X-3} \qquad (2)$$

Entscheidungsfunktion für Gruppe 1 (G 1):

$d_{\varepsilon=0}(\theta_{X-3,opt}) \geqslant \theta_{X-3,opt}$

Entscheidungsregel für Gruppe 2 (G 2):

$d_{\varepsilon=0}(\theta_{X-3,opt}) < \theta_{X-3,opt}$

A.2: berechne $\Delta H_{\varepsilon=0}$ als maximale Reduktion der Entropie aus den möglichen Reduktionen aus allen Merkmalen

A.3: bestimme das Merkmal, das mit $\Delta H_{\varepsilon=0}$ zur maximalen Reduktion der Entropie führt

A.4: teile den Merkmalsraum durch die Entscheidungsfunktion, wie hier $d_{\varepsilon=0}(\theta_{X-3,opt})$ entsprechend dieses Merkmals auf (mit einem optimalen Schwellenwert)

A.5: erhöhe die Ebene um 1, weise auf der Ebene 1 den Knoten die jeweiligen Untersuchungseinheiten bzw. Gruppen an Daten zu

Aufteilung an Ebene 1:

Für Knoten 1:

$$\Delta H_{\varepsilon=1,X-1}(G_1, \vec{y}) \overset{B \equiv w}{=} H_{\varepsilon=1} - H_{\varepsilon=2,X-1}(\theta_{X-1,opt}) = H_{\varepsilon=1} - \sum_{j=0}^{1} p_{j,\varepsilon=2} \cdot \log_2\left(p_{j,\varepsilon=2}^{-1}\right)$$

$$\Delta H_{\varepsilon=1} = \sup\{\Delta H_{\varepsilon=1,X-1}, \Delta H_{\varepsilon=1,X-2}, \Delta H_{\varepsilon=1,X-3}\} = \Delta H_{\varepsilon=0,X-1}$$

für Gruppe 3:

$$d_{\varepsilon=1}(\theta_{X-1,opt}) \geqslant \theta_{X-1,opt}$$

wiederhole das Vorgehen für Knoten 2:

A.7: berechne für die jeweils betrachtete Ebene für alle Knoten die maximale Reduktion der Entropie

A.8: bestimme das Merkmal, das mit $\Delta H_{\varepsilon=1}$ zur maximalen Reduktion der Entropie führt für alle Knoten

A.9: teile den Merkmalsraum auf für die Entscheidungsfunktionen für alle Knoten

A.10: erhöhe die Ebene um 1, weise auf der nächsten Ebene den Knoten die jeweiligen Untersuchungseinheiten bzw. Gruppen an Daten zu

Aufteilung an weiteren Ebenen:

wiederhole A.7-10 für jede weitere Ebene

Das Ergebnis des Algorithmus ist eine Abbildung aller möglichen Punkte des Merkmalsraums auf die Zielwerte 0 oder 1. Bei weiteren Aufteilungen könnte das Ergebnis wie in Abbildung 4 dargestellt aussehen. Die Verallgemeinerung des Vorgehens in mehr Dimensionen als im Beispiel teilt den Merkmalsraum nicht in 3 dimensionale Hypercubes auf sondern in d-dimensionale Hypercubes auf.

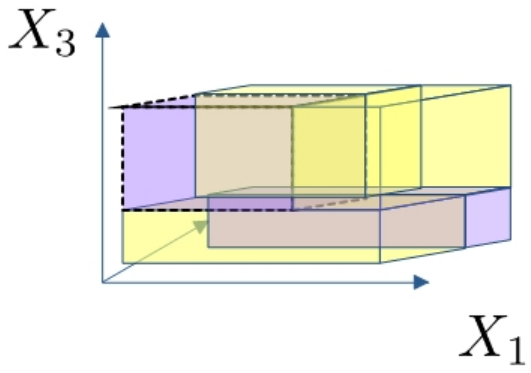


Abb. 4: mögliche Partition des Merkmalsraums in 3 dimensionale Hypercubes als Abbildung der Datenmatrix auf die Zielwerte, Quelle: eigene Darstellung

Durch das hierarchische Verfahren hängen Entscheidungsbäume stark mit bedingten Wahrscheinlichkeiten und der Aussagenlogik zusammen. Denn jede Aufteilung des Merkmalsraums erfolgt unter der Bedingung aller vorherigen Aufteilungen in unterschiedliche Gruppen und den jeweiligen bedingten Wahrscheinlichkeiten. Dabei ist zu berücksichtigen, dass Entscheidungsbäume typischerweise Algorithmen sind, die „greedy“ sind. Dies bedeutet, dass der Algorithmus nur berücksichtigt, welche Aufteilung der Daten momentan bestmöglich die Inhomogenität reduziert unter der Bedingung von allen vorherigen Aufteilungen der Daten (end cut preference). Um diesen möglichen Nachteil auszugleichen, gibt es verschiedene Erweiterungen der Algorithmen wie den Extra Trees Algorithmus, bei dem aus zufälligen Aufteilungen ausgewählt wird. Dadurch können besser unterschiedliche Bedingungen berücksichtigt werden.

2.4.5 Algorithmus 1

In diesem Abschnitt soll ein neu aufgebauter Algorithmus beschrieben werden. Er kann eingeordnet werden zu Multiclassifier-Systemen (MCS). Verglichen mit Algorithmen, die zu den MCS gehören, gibt es aber bei diesem Algorithmus Erweiterungen, damit die Vorgehensweise verständlich ist und zu Explainable Machine Learning dazugerechnet werden kann. Zunächst wird der eingelesene Datensatz von Ausreißern bereinigt. Dann wird der Datensatz χ in die Datenmatrix und die Zielwerte aufgeteilt (s. Abb. 5). Die Datenmatrix enthält alle Merkmale, die als Prädiktoren für die Klassifikation genutzt werden. Es wird hier von einer binären Klassifikation ausgegangen. Die Zielwerte ergeben einen Vektor bzw. eine Matrix (bei Vorhersage mehrerer Zielvariablen) mit den Angaben zu den Klassen, die vorhergesagt werden sollen.

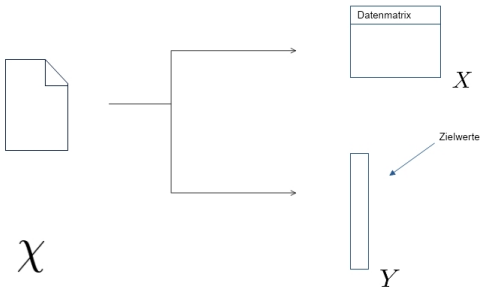


Abb. 5: Aufteilung des Datensatzes, Quelle: eigene Darstellung

Die Datenmatrix X wird dann für verschiedene Klassifikationsmodelle $v = 1,\ldots,V$ in unterschiedliche Teilmengen von Merkmalen aufgeteilt (s. Abb. 6). Es resultieren Datenmatrizen für jede Teilmenge $X^{(v=1)},\ldots,X^{V}$ der Multiview-Klassifikation. Jede Teilmenge beinhaltet alle Untersuchungseinheiten bzw. Beobachtungen. Jede Matrix wird dann als Input für ein Klassifikationsmodell verwendet.

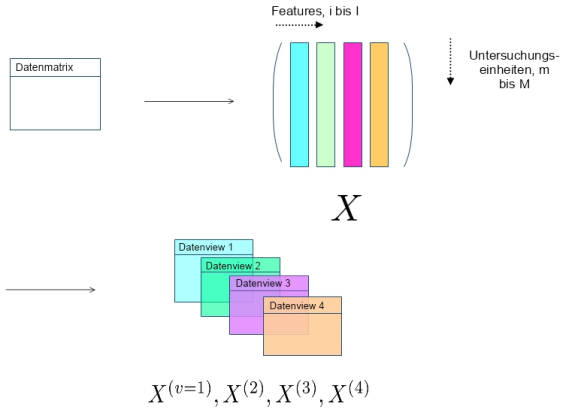


Abb. 6: Aufteilung für verschiedene Datenviews, Quelle: eigene Darstellung

Für die verschiedenen Klassifikationsaufgaben werden jeweils ein oder mehrere Entscheidungsbäume genutzt. Alternativ wäre es auch möglich, für jede Klassifikationsaufgabe einen unterschiedlichen Klassifikationsalgorithmus zur binären Klassifikation zu wählen. Für jede der Klassifikationsaufgaben wird dann ein ML-Modell aufgebaut. Das Ergebnis sind ML-Modelle, hier jeweils ein oder mehrere Entscheidungsbäume zu jeden Datenview, die einzeln ausgewertet werden können (s. Abb. 7).

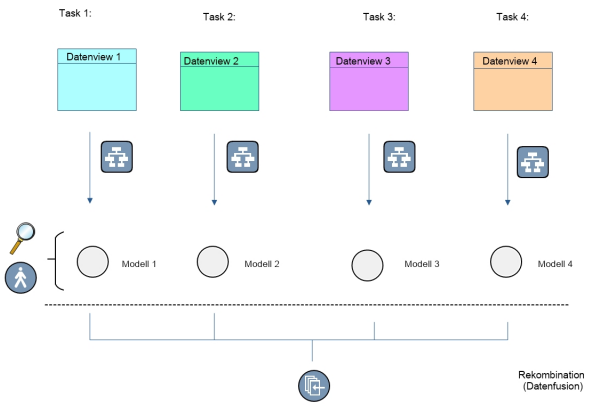


Abb. 7: Modellarchitektur, Quelle: eigene Darstellung

Normalerweise werden bei der Auswertung nur die Klassifikationen genutzt, hier sollen die Wahrscheinlichkeiten für eine Klassifikation verwendet werden. Für jede Beobachtung bzw. jede der Untersuchungseinheiten lässt sich aus den Entscheidungsbäumen die geschätzte Wahrscheinlichkeit für die Zugehörigkeit zu dieser Klasse ablesen. Diese ergibt sich aus dem Verhältnis der Anzahl an Untersuchungseinheiten der Klasse 0 bzw. Klasse 1 zur Gesamtzahl der Einheiten in diesem Zweig. Die Klassenwahrscheinlichkeiten geben damit die Unsicherheit der Klassifikation an. Das Ergebnis ist die Matrix $\widetilde{P}_{(V\times M)}$, die bei einer binären Klassifikation für die Klasse 1 für jede Untersuchungseinheit bzw. Beobachtung die Klassenwahrscheinlichkeit angibt.

$\widetilde{P}_{(V\times M)}$: Matrix der Klassenwahrscheinlichkeiten für die Klasse 1 für alle Untersuchungseinheiten und Modelle

$n_{y=1}$: Anzahl der Untersuchungseinheiten des jeweiligen Zweiges im Entscheidungsbaum, die zur Klasse 1 zählen

n: Gesamtanzahl der Untersuchungseinheiten des Zweiges

$$\widetilde{P}_{(V\times M)} = \begin{pmatrix} \widetilde{P}_{v=1,m=1} & \cdots & \widetilde{P}_{V,1} \\ \vdots & \ddots & \vdots \\ \widetilde{P}_{1,M} & \cdots & \widetilde{P}_{V,M} \end{pmatrix}$$

Für alle Untersuchungseinheiten und Datenviews gilt:

$\widetilde{P}_{v,m} = \widetilde{P_v}(C_1/\vec{x}_m) = \frac{n_{y=1}}{n}$, bei einem Modell pro Datenview

Bei mehreren Modellen pro Datenview: Berechnung des Mittelwerts aus den Klassenwahrscheinlichkeiten aller Modelle des jeweiligen Datenviews

In der Folge werden die verschiedenen Ergebnisse basierend auf dieser Matrix rekombiniert (Datenfusion bzw. Datenintegration). Zur Rekombination der Ergebnisse wird Formel (1) verwendet. Dazu werden, wenn mehrere Entscheidungsbäume zu einem Datenview aufgebaut wurden, zunächst die Klassenwahrscheinlichkeiten aller Modelle zu einem Datenview gemittelt. Das Resultat ist zu jedem Datenview die mittels des arithmetischen Mittelwerts berechnete Klassenwahrscheinlichkeit. Alternativ wäre es möglich zur Regularisierung der Ergebnisse z.B. den James Stein Schätzer zu verwenden. Für das Entscheidungsmodell werden nicht die Klassifikationen $C^{(v=1)},...,C^{V}$ verwendet sondern es ist ein lineares Modell der abgeleiteten Klassenwahrscheinlichkeiten aus den verschiedenen Modellen bzw. Datenviews. Dies entspricht einer Akkumulation der Klassenwahrscheinlichkeiten aus den unterschiedlichen Modellen (Datenviews) für jede Untersuchungseinheit. Dies ist sinnvoll, da Entscheidungsbäume verglichen zu wahrscheinlichkeitstheoretischen Modellen oft keine guten Schätzungen der Klassenwahrscheinlichkeiten ergeben. Die Akkumulation sorgt für einen Fehlerausgleich, der umso effektiver ist, je größer V ist also je mehr Klassenwahrscheinlichkeiten akkumuliert werden. Ein solcher Fehlerausgleich wird in vielen statistischen Methoden verwendet. Dabei kann der Faktor, zu dem jeder Datenview zu dem Mittelwert beiträgt, für jede Untersuchungseinheit gesondert gewählt werden oder für alle Untersuchungseinheiten gleich. Jeder Untersuchungseinheit wird damit ein score zugewiesen, der aussagt, welcher der Klassen die Untersuchungseinheit eher zugewiesen werden kann. Es ist eine Quasiwahrscheinlichkeit als gemischte Wahrscheinlichkeit der Klassenwahrscheinlichkeiten der unterschiedlichen Modelle. Statt Formel (1) könnte ein Regressionsmodell genutzt werden. Dazu würde es sich anbieten, dass auch gemischte Regressionsmodelle verwendet werden. Dies hätte den Vorteil, dass für unterschiedliche Teilgruppen der Daten unterschiedliche Faktoren für das Modell gewählt werden könnten.

$\vec{\epsilon}_v$: Vektor mit den Faktoren für die verschiedenen Datenviews

$\widetilde{P}_m$: Zeilenvektor aus der Matrix $\widetilde{P}_{(V\times M)}$

$$score_m = \vec{\epsilon}_v \cdot \widetilde{P}_m = \epsilon_1 \cdot \tilde{P}_{1,m} + \epsilon_2 \cdot \tilde{P}_{2,m} + \epsilon_3 \cdot \tilde{P}_{3,m} + \epsilon_4 \cdot \tilde{P}_{4,m} \quad (1)$$

Beispiel für eine Formel mit Faktoren:

$$score_m = 30\ \% \cdot \widetilde{P}_{1,m} + 20\ \% \cdot \widetilde{P}_{2,m} + 10\ \% \cdot \widetilde{P}_{3,m} + 40\ \% \cdot \widetilde{P}_{4,m}$$

Es können in der Folge die scores aller Untersuchungseinheiten verglichen werden (s. Abb. 8). Zusammen mit den bekannten Klassen der Untersuchungseinheiten ergeben sich Verteilungen der scores für die verschiedenen Klassen. Um die Klassenzugehörigkeit durch den Algorithmus zu bestimmen, wird ein Threshold Modell verwendet. Dazu werden alle scores und die zugehörigen Untersuchungseinheiten, die größer sind als der Threshold Parameter θ der Klasse 1 zugewiesen.

θ: Threshold Parameter

$$\widetilde{C}_{ges,m} = \begin{cases} 0, \text{ für } score_m < \theta \\ 1, \text{ für } score_m \geqslant \theta \end{cases}$$

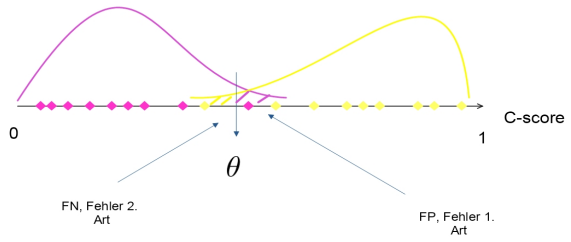


Abb. 8: Threshold Modell, Quelle: eigene Darstellung

Der Threshold Parameter kann dabei unterschiedlich bestimmt werden. Er kann z.B. durch Optimierung der Classification Accuracy festgelegt werden. Der Parameter wird dann so gewählt, dass die Classification Accuracy möglichst hoch ist. Ebenso könnte, basierend auf den Ergebnissen, der Parameter so gewählt werden, dass der Fehler 1. Art und daraus folgend der Fehler 2. Art bestimmte Werte hat. Alternativ wäre es möglich, ein Kostenmodell zu definieren und dieses zu optimieren [30]. Dabei können für falschpositive Klassifkationen (FP) und falschnegative Klassifkationen (FN) unterschiedliche Kosten definiert werden. Zur Evaluation der Ergebnisse wäre es möglich, statistische Tests durchzuführen (bei Normalverteilung t-Tests), um zu untersuchen, ob sich die scores für die jeweiligen Klassen zu einem gegebenen Signifikanzniveau signifikant unterscheiden. Dieser könnte auch für Teilgruppen der Daten durchgeführt werden und damit überprüft werden, für welche Gruppen das Klassifikationsmodell die Klassen signifikant unterscheiden kann.

Um den Algorithmus besser erklärbar zu machen, werden Shapley Werte mittels Shapley additive explanations eingesetzt (s. 2.5). Dabei wird jeweils der Tree Shap Algorithmus für die Modelle eines Datenviews verwendet. Dabei werden die Erklärungsanteile des Tree Shap Algorithmus nicht für die Klassifikationen als Output bestimmt sondern die Klassenwahrscheinlichkeiten. Das Ergebnis sind Shapley Scatterplots, die zusammen mit den Entscheidungsbäumen die Erklärbarkeit erhöhen. Der Vorteil davon ist, dass man kontinuierliche Maße erhält, die aussagen, wie die Merkmale zu den Klassifikationen beitragen. Die Shapley Scatterplots werden mittels nichtparametrischer Regression als kontinuierliche Funktionen bestimmt. Desweiteren können mittels Bootstrapping gebootstrappte Konfidenzintervalle bestimmt werden, die Unsicherheiten zu den Effektabschätzungen aus den Shapley Werten der unterschiedlichen Variablen beitragen. So ließe sich beispielhaft an Abb. 9 ablesen, dass bei einem Alter von 40 Jahren bezogen auf den jeweiligen Datenview ein durchschnittlicher

Effekt von einer um 15 % verringerten Wahrscheinlichkeit gegeben ist an CAD zu erkranken mit einer Unsicherheit der Klassenwahrscheinlichkeit von 5 % als Variation des Effekts für unterschiedliche Patienten.

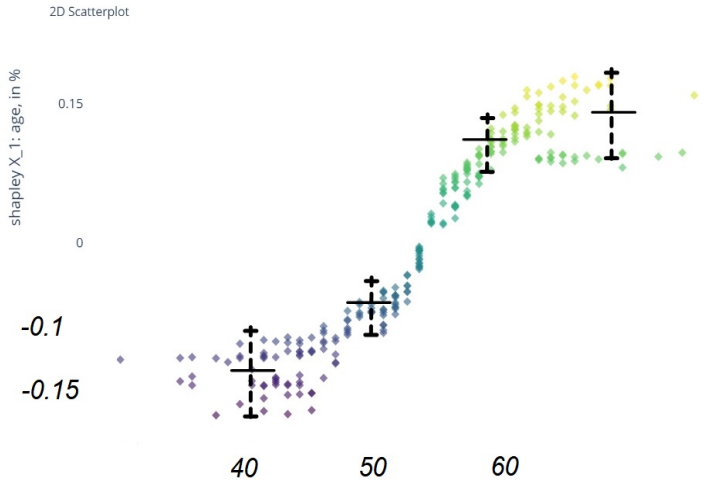


Abb. 9: Shapley Scatterplot mit Effektabschätzungen, Quelle: eigene Darstellung

2.5 Explainable Machine Learning

Unter Explainable Machine Learning werden Methoden verstanden, die dafür sorgen sollen, dass die Ergebnisse der ML-Modelle erklärbar und nachvollziehbar sind. Dabei werden die Begriffe Erklären und Interpretieren unterschieden. Während mit Interpretieren das Interpretieren der Lösungen von ML-Modellen gemeint ist, wird mit Erklären verstanden, dass bestenfalls vollständig nachvollziehbar ist, wie ein Algorithmus im Detail zu einem Ergebnis kommt. Während Interpretieren damit auch bei „Black Box"-Modellen möglich ist, ist es nicht möglich diese Modelle zu erklären. Es werden dabei unter „Black Box"-Modellen Modelle verstanden, die keine weiteren Outputs liefern, die zu einer Erklärung herangezogen werden könnten, als den Modelloutput (das Ergebnis). Bei den Explainable Machine Learning Methoden wird strukturell die Erklärbarkeit erhöht und damit die „Black Box"-Modelle so angepasst, dass auf sämtliche Ergebnisse geschlossen werden kann. Das ML-System wird dann so ausgelegt, dass es verständlich ist. Es ist aber ebenso möglich, die „Black Box"-Modelle nicht weiter anzupassen und nur aus dem Modelloutput weitere Informationen, die die Erklärbarkeit erhöhen, zu extrahieren. Die Anwendung von Explainable Machine Learning kann vielen Zielen dienen. Dabei ist die Verständlichkeit von ML-Modellen zentral. Die Verständlichkeit kann indirekt z.B. auch ethischen Aspekten (AI ethics) und Fairness dienen. Diese werden im Weiteren nicht genauer betrachtet.

2.5.1 Umsetzung von Explainable Machine Learning

Eine einfache Form Explainable Machine Learning anzuwenden ist es, für das ML-Modell ein Modell zu wählen, dass inhärent verständlich ist. Dies bedeutet, dass Modelle gewählt werden, aus denen sich direkt Entscheidungsregeln extrahieren lassen. Dabei ist es relevant, ob sich aus dem Modell sämtliche Entscheidungsregeln extrahieren lassen, da in diesem Fall in sämtlichen Situationen darauf geschlossen werden kann, wie sich ein ML-System auch bei neuen Vorhersagen verhält. In Bezug auf die Entscheidungsregeln ist v.a. relevant, ob diese auch von Fachexperten verstanden werden können in der Situation, in der die Erklärbarkeit auch regulär für Fachexperten erhöht werden soll. Viele der ML-Modelle haben komplexe mathematische Entscheidungsfunktionen, die prinzipiell nicht für Fachexperten verständlich wären. Ein Beispiel für Modelle mit verständlichen Entscheidungsregeln sind Entscheidungsbäume.

Desweiteren wurde zur Verbesserung der Verständlichkeit auch Dimensionsreduktion eingesetzt [31]. Mittels Dimensionsreduktion können auch bei hoher Anzahl an Dimensionen in Ansätzen interpretierbare Lösungen gefunden werden. Zu beachten ist jedoch, dass die Bedeutung der dann entstandenen transformierten Merkmale nicht direkt klar sein muss. Die Anwendung durch einen Fachexperten, der sich nicht mit Machine Learning auskennt, könnte deshalb nicht geeignet sein. Außerdem geht Dimensionsreduktion mit einem Informationsverlust einher.

Ein weiterer Ansatz sind hybride Modelle, die versuchen unterschiedliche ML-Modelle zu kombinieren, um die Vorteile der jeweiligen Modelle in Bezug auf Explainability zu erhalten. Ein Beispiel sind dafür Adaptive Neural Trees (ANT), die versuchen Neuronale Netzwerke und Entscheidungsbäume zu kombinieren [32]. Desweiteren ist es möglich, die Entscheidungsregeln eines Modelles durch ein anderes zu erlernen, das besser erklärbar ist z.B. mittels verschiedener spezieller Methoden des Explainable Machine Learning. Ein Beispiel dafür ist der „neuralization"- Trick, der verwendet werden kann, um Modelle in neuronale Netzwerke umzuwandeln und diese mittels Layerwise Relevance Propagation erklären zu können [33]. So konnten z.B. Clustermodelle mittels Layerwise Relevance

Propagation erklärt werden.

Eine weiterere Möglichkeit ist es strukturell Modelle anzupassen, sodass diese erklärbar und besser verständlich sind. Die ML-Modelle werden dabei so ausgelegt, dass diese verständlich sind. Ein Ansatz ist es dabei, die Struktur der ML-Modelle modular zu wählen, sodass bestenfalls auf jeder Ebene des ML-Systems Outputs generiert werden und die Funktionsweise jeder Komponente verständlich ist. Die Anpassung aus struktureller Sicht dient meist dazu Komplexität zu reduzieren. Ein Beispiel ist es die Zusammenhänge explizit zu modellieren mittels generalisierter additiver Modelle [34].

2.5.2 Methoden des Explainable Machine Learning

Die am häufigsten verwendeten Methoden sind Shapley Werte mittels Shapley additive explanations und Layerwise Relevance Propagation (LRP). Diese Methoden sind anwendbar auch ohne ein ML-Modell strukturell anzupassen und können deshalb auch bei dem Modelloutput von „Black Box"-Modellen verwendet werden und ohne die vorher genannten Umsetzungen von Explainable Machine Learning. Dabei werden Shapley Werte häufiger für strukturierte Daten verwendet und LRP tendenziell eher für unstrukturierte Daten wie medizinische Bilder. Es ist jedoch auch möglich Shapley Werte für unstrukturierte Daten zu nutzen und auch LRP lässt sich für strukturierte Daten nutzen [35]. Dies ist möglich, da prinzipiell beide Methoden dasselbe Ergebnis liefern. Es soll in beiden Fällen die Erklärbarkeit erhöht werden, indem meist für eine Klassifikation angegeben wird, welche der Merkmale zu welchem Anteil der Klassifikation beigetragen haben. Ein positiver Anteil trägt z.B. bei einer binären Klassifikation eher zu Klasse 1 bei und ein negativer Anteil eher zu Klasse 0. So soll ein Vergleich von Input und Output des ML-Modells ermöglicht werden. Dies wird meist dadurch erreicht, dass der Input geringfügig verändert wird und der Einfluss auf den Output bestimmt wird. Da es für solche Perturbationen sehr viele Möglichkeiten gibt, ist dieser Vorgang für Explainable Machine Learning sehr rechenaufwändig. Für die Shapley Werte und LRP gibt es jeweils verschiedene Methoden, von denen einige modellagnostisch sind, also für jegliche Modelle genutzt werden können. Andere sind nur für bestimmte ML-Modelle geeignet. So ist LRP an sich nur für neuronale Netzwerke anwendbar. Es kann jedoch auch mittels des „neuralization"-Trick für andere ML-Modelle angewendet werden. Im Folgenden werden die Shapley Werte betrachtet, weil nur sie hier verwendet werden sollen.

2.5.3 Shapley additive explanations

Shapley Werte werden in verschiedenen Algorithmen genutzt, wie z. B. Kernel Shap und Tree Shap [36, 37]. Dies dient meist dazu, um lokal und global auf den Beitrag einzelner Merkmale für die Klassifikation zu schließen. Diese Beiträge nennt man Shapley Werte. Unter lokaler Explainability wird verstanden, dass man für ein Untersuchungsobjekt für jedes Merkmal angeben kann, welchen Beitrag dieses für die Klassifikation dieses Untersuchungsobjektes hatte. Merkmale mit einem positiven Beitrag sorgen für Klasse 1 und Merkmale mit negativem Merkmal sorgen für Klasse 0 bei einer binären Klassifikation. So wird ein Erklärungsbeitrag dafür geliefert, was zu der Klassifikation geführt hat. Die Shapley Werte einzelner Untersuchungseinheiten können aggregiert werden und so kann auch global für alle Untersuchungseinheiten darauf geschlossen werden, welche Merkmale jeweils zu der Klassifikation geführt haben. Es lassen sich Shapley Werte jedoch auch für viele weitere Anwendungen und Ziele nutzen, von denen viele der Explainability dienen können. Dies ist darauf zurückzuführen, dass Shapley Werte ein allgemein anwendbarer Ansatz aus der Spieltheorie sind. So lässt sich mittels Shapley Werten fast jeder Beitrag eines Input auf einen Output erklären.

Shapley Werte wurden in der Spieltheorie genutzt und gehen auf Lloyd Shapley zurück. Sie stellen in einem Koalitionsspiel den Durchschnitt von marginalen Beiträgen von unterschiedlichen Spielern auf einen Output dar. Es könnten z.B. die Spieler verschiedene Unternehmen sein, die zusammen durch Kooperation einen Gewinn erwirtschaften. In der Folge sollen die Gewinne fair auf die verschiedenen Unternehmen aufgeteilt werden. Es ist jeweils bekannt, welchen Gewinn die einzelnen Unternehmen haben bzw. jeder der möglichen Koalitionen mit anderen Spielern. Für sämtliche der Permutationen an Spielern wird dann der Gewinn (Output) der Spieler jeweils ohne den betrachteten Spieler bestimmt und jeweils mit dem Spieler. Die Differenz dieser Werte ist der marginale Beitrag, den der betrachtete Spieler für die jeweilige Konstellation der Kooperation liefert. Der Durchschnitt dieser marginalen Beiträge liefert den Shapley Wert des Spielers und damit die faire Auszahlung, die dieser erhalten sollte. Es wird dabei fair der Beitrag von dem betrachteten Spieler auf den Output bestimmt. Diese Berechnung kann für alle Spieler durchgeführt werden. Die Summe aller Shapley Werte (von allen Spielern) ergibt dann den Zielwert bzw. den Output. Die Shapley Werte nutzen deshalb additive Modelle.

$n!$: Anzahl der möglichen Koalitionen

i: Index der Koalitionen

$\phi_m = \frac{1}{n!} \sum_{i=1}^{I} \varphi_{m,i} = \frac{1}{n!} \sum_{i=1}^{n!} v(C) - v(C \S_m)$

$y = \sum_{m=1}^{M} \phi_m$

Dieser Ansatz lässt sich auf verschiedene Situationen des Explainable Machine Learning beziehen. Soll z.B. ein Output eines ML-Modells wie eine Klassifikation erklärt werden, so wären die Spieler die verschiedenen Merkmale und der Zielwert die Klassifikation. Ebenfalls wäre es möglich die Beiträge von Merkmalswerten an Klassenwahrscheinlichkeiten zu bestimmen. Dadurch würde für jede Untersuchungseinheit für jeden Merkmalswert ein Shapley Wert zugeordnet werden, der den Beitrag dieses Merkmalswertes aus allen Koalitionen mit anderen Merkmalen darstellt. Zur Bestimmung der Shapley Werte würde das ML-Modell einmal mit dem betrachteten Merkmal angewendet werden und einmal ohne das betrachtete Merkmal. Der Output aus beiden Situationen würde dann verglichen werden. Da die Anwendung der ML-Modelle für alle Koalitionen an Merkmalen jedoch sehr rechenaufwendig werden kann, werden die Shapley Werte nur geschätzt. Deshalb werden für unterschiedliche ML-Modelle verschiedene Verfahren zur Schätzung genutzt. Während für Kernel Shap ein Regressionsmodell dabei verwendet wird, so wird bei Tree Shap die Hierarchie des Entscheidungsbaumes dafür genutzt. Während Shapley Werte verglichen zu anderen Verfahren des Explainable Machine Learning aus theoretischer Sicht gewisse Optimalitätskriterien erfüllen, so hängt praktisch die Genauigkeit der Ergebnisse davon ab, ob die Schätzungen der Shapley Werte ausreichend genau sind. Diese Shapley Werte können verschieden dargestellt werden.

Material und Methoden

3.1 Verwendete Programme

Um für die Algorithmen das Programm zu erstellen, wurde Excel und Anaconda mittels Jupyter Notebooks verwendet. Die jeweils in Jupyter Notebooks verwendete Programmiersprache war Python. Es musste zunächst im Skript „utility“ programmiert werden, wie die Daten in Jupyter Notebooks importiert werden sollten. Der dargestellte Vorgang war notwendig, weil die Daten nur als Text ohne bestimmtes Format von dem UCI ML Repository vorlagen. Dann wurden sie im CSV-Format abgespeichert und in Excel importiert. In Excel wurden sie in der Arbeitsmappe „Datenvorbereitung“ vorbereitet, was durch die dort vorhandene graphische Oberfläche gut funktioniert hat, und danach wieder in Jupyter Notebooks importiert. In Jupyter Notebooks konnten dann im Skript „Klassifikation“ die Daten visualisiert und die Algorithmen programmiert werden. Weitere Softwarekomponenten waren Bibliotheken aus Python. Durch xlwings konnten die Daten direkt aus Excel importiert werden, ohne sie vorher in Excel zu speichern. Das ist ein Beispiel dafür, dass Python gut mit Excel zusammenarbeitet. Außerdem wurde die Bibliothek plotly zur interaktiven Visualisierung genutzt. Für die Algorithmen wurde sklearn verwendet und für das Explainable Machine Learning die Bibliothek shap. Desweiteren wurden verschiedene Addins selber geschrieben, in Excel als VBA Makros und in Python als Funktionen. Diese dienten v.a. der Datenverarbeitung, der Darstellung und der Visualisierung.

3.2 Rekonstruktion der Bedeutung von Merkmalen

Zunächst wurde die Bedeutung der Merkmale der Datensätze rekonstruiert. Dies war nötig, da die Dokumentation zu der Bedeutung der Merkmale im UCI Machine Learning Repository nicht vollständig und direkt verständlich war. Auch aus weiteren Quellen aus dem Internet konnte nicht sofort auf die Bedeutung der Merkmale geschlossen werden. Teilweise gibt es auch unterschiedlichste Interpretationen über die Bedeutung der Merkmale in unterschiedlichen Quellen, die sich teils auch widersprachen oder die Angaben waren ungenau. Dies ist darauf zurückzuführen, dass jeweils auf andere Artikel und Quellen verwiesen wird, aber nicht auf die Originalartikel der Ersteller des Datensatzes [5]. Erst nach Auswertung der originalen Artikel und weiterer Recherche konnte die Bedeutung der Merkmale weitgehend rekonstruiert werden. Die Rekonstruktion der Bedeutung von Features ist sehr wichtig für explainable ML, da erst dann Ergebnisse des Machine Learning ausreichend interpretiert werden können. Meist ist die Bedeutung von Merkmalen bezogen auf das Machine Learning irrelevant. Für das Explainable Machine Learning ist es jedoch notwendig. Besonders relevant war die korrekte Bedeutung der Zielvariablen, erst damit war klar, dass mit diesen Daten primär koronare Herzkrankheiten vorhergesagt werden können. Die Bedeutung zu kennen, kann auch Feature Engineering vereinfachen. Die Liste der Merkmale und deren Bedeutung ist in Anhang 1 und in der Excel-Arbeitsmappe zum Datensatz zu finden.

3.3 Struktur des Datensatzes und Validität

Die Struktur des Datensatzes wurde untersucht. Dies diente dazu, den Datensatz besser zu verstehen und festzustellen, ob er logisch konsistent ist und damit intern valide für die generelle Untersuchung der Anwendung von ML für CDSS. Dies war relevant, da aus den Notizen zum Datensatz vermutet werden könnte, dass im Datensatz Fehler sind. Dies konnte nicht letztendlich überprüft werden, da eine Email an das UCI ML Repository zu diesem Thema nicht beantwortet wurde. Aus mehreren Gründen werden die Daten aber trotzdem als intern valide eingeschätzt. Zum einen gibt es keinen Aufschluss darüber, dass Fehler in sämtlichen Datensätzen vorhanden sind oder diese nicht bereits ausgeglichen wurden. Zum anderen wurden keine Datenwerte gefunden, die logisch inkonsistent wären und alle Kombinationen an Werten unterschiedlicher Merkmale erscheinen grundsätzlich logisch. Es wird deshalb davon ausgegangen, dass nur geringe Fehler vorhanden sind. Außerdem haben medizinische Datensätze aber allgemein geringere Datenqualität, wenn die Daten im regulären Umfeld wie bei einem CDSS gesammelt werden. Insofern könnten geringe Fehler auch als realistisch angesehen werden. Desweiteren kann es in dieser Arbeit nur um die allgemeine Anwendbarkeit von Machine Learning für CDSS gehen und nicht primär um valide Schlussfolgerungen aus den Daten. Dies bedeutet nicht unbedingt, dass die verwendeten Daten extern valide sind, wie es z.B. für medizinische Forschung der Fall sein müsste. Weil in der medizinischen Forschung die Anforderungen höher sind, müsste die externe Validität der Daten überprüft werden.

3.4 Feature Selection

Bei der Feature Selection wurden irrelevante oder unvollständige Merkmale mit weitgehend fehlenden Daten reduziert. Dabei wurde die Anzahl der Merkmale von 75 nach dem Einlesen der Daten auf 33 Prädiktoren und 11 Zielvariablen reduziert. Bei den irrelevanten Merkmalen handelt es sich um solche, die von den Erstellern des Datensatzes selbst so bezeichnet wurden. Im Einzelnen kann man das aus der Excel-Arbeitsmappe entnehmen. Die Liste der verwendeten Merkmale ist auch im Anhang zu finden. Mathematische Verfahren zum Ausschluss von Merkmalen wurden nicht verwendet. Auch sachlogisch wurden keine Merkmale weggelassen, wie es z. B. ein Mediziner handhaben würde, sondern es wurde sich darauf verlassen, dass die Ersteller des Datensatzes die relevanten Merkmale ausgewählt haben. Aus Sicht des Machine Learnings war es sinnvoll, mit relativ vielen Merkmalen zu arbeiten. Es ist zwar schwierig, so viele Merkmale zu verwenden, aber auch notwendig, damit das Machine Learning sehr detaillierte Informationen finden kann. Wenn das ML nur Informationen liefert, die ein Mediziner ohnehin weiß, dann erübrigt sich eine Anwendung. Es wurde keine Dimensionsreduktion genutzt, da sie zu Informationsverlust führen würde.

3.5 Fehlende Daten und Datenimputation

Aus den Datensätzen wurden die Untersuchungseinheiten entfernt, die weitgehend für viele der Merkmale fehlende Daten hatten. Desweiteren wurden die Untersuchungseinheiten entfernt, bei denen einzelne Datenwerte fehlten in ansonsten vollständigen Merkmalen. Nach dem Entfernen der fehlenden Daten wurde eine neue PatientenID vergeben. Die Tabelle zeigt einen Überblick, wieviele Untersuchungseinheiten von jedem der Datensätze entfernt wurden. Welche dies waren, ist in Anhang 2 dokumentiert.

	entfernt	Anzahl	nach Entfernen
Cleveland	12	282	270
VA Medical	71	200	129
Hungarian	47	294	247
Switzerland	31	123	92
Summe	161	899	738

Tab. 2: Überblick über die Daten

Bei der Datenimputation wurden fehlende Daten in den Fällen durch 0 ersetzen, bei denen es sinnvoll war. Die fehlenden Daten waren mit „-9“ gekennzeichnet. Auch inhaltlich unlogische Datenwerte wurden entfernt (z.B. 0 für Cholesterin). Es wurden ebenfalls sachlogisch Daten imputiert. Die Formeln, mittels derer in Excel Daten imputiert worden sind, sind in der Arbeitsmappe nachvollziehbar. Datenspalten, für die Werte imputiert wurden, sind gesondert in der Arbeitsmappe ohne imputierte Werte aufgelistet, sodass prinzipiell im Nachhinein auf inkorrekte Datenimputation geschlossen werden könnte. Der folgende Überblick zeigt, wie Daten imputiert worden sind.

X_5: dm, fbs, für dm und fbs wurden alle fehlenden Daten mit 0 ersetzt, dies entspricht Diabetes nicht vorhanden.

X_7: hist_cad, die fehlenden Daten wurden durch 0 ersetzt, dies entspricht keine Historie von CAD vorhanden.

X_23: slope, die fehlenden Daten wurden durch 1 ersetzt.

X_31: xhypo, fehlende Daten wurden durch 0 ersetzt.

Für Zielvariablen: fehlende Daten der Angiographie wurden durch 0 ersetzt.

3.6 Feature Transformationen

An einzelnen Stellen, wo es sinnvoll war, wurden verschiedene Merkmale verändert oder in neue Merkmale zusammengefasst. Die Formeln, mittels derer in Excel Daten transformiert wurden, sind in der Arbeitsmappe nachvollziehbar. Es folgt ein Überblick, welche Merkmale dies waren.

X_5: für dm, fbs wurden die Merkmale dm und fbs zusammengefasst, da die Bedeutung weitgehend identisch und Diabetes im Datensatz tendenziell unterrepräsentiert ist.

restecg_0, restecg_1, restecg_2 sind die hot encoded Merkmale aus restecg, dafür wurde restecg aus dem Datensatz entfernt.

rdlve wurde im Datensatz Cleveland durch 10 geteilt, um Konsistenz zu den anderen Datensätzen zu erreichen.

3.7 Ausreißererkennung

Die kontinuierlichen Merkmale wurden mittels des GESD-Tests (Generalized Extreme Studentized Deviate Test) auf Ausreißer untersucht. Dazu wurden jeweils aus allen vier Datensätzen die Datenwerte eines Merkmals als Datenreihe zusammen verwendet. Es wurde ein Alpha-Fehler von 5 % gewählt. Dazu wurde die Excel Arbeitsmappe „Ausreißererkennung“ genutzt. Für jeden der möglichen Ausreißer wurde sachlogisch festgestellt, ob es sinnvoll ist, den jeweiligen Ausreißer zu entfernen. Vorher wurde nach Normwerten für die jeweiligen Merkmale recherchiert. Insgesamt sollen die Ausreißer nur moderat entfernt werden. Denn es wäre nicht effizient zu viele Ausreißer zu entfernen, da in gewisser Weise für Patienten mit CAD bzw. kardiovaskulären Krankheiten ungewöhnliche Werte zu erwarten sind. Es würden ansonsten die Werte entfernt werden, die gerade klassifiziert werden sollen. Im Folgenden wird ein Überblick über die Ausreißererkennung gegeben. Für den Ausreißertest wird dies beispielhaft anhand des Merkmals Cholesterin dargestellt. Die Ausreißererkennung für weitere Merkmale ist aus der Excel Arbeitsmappe zu entnehmen.

Cholesterin: Alle Cholesterinwerte größer als 417 wurden durch den Test als Ausreißer erkannt (s. Tabelle 3). Insgesamt wurde der Wert 458 nicht als Ausreißer entfernt, da es sich sachlogisch auch um einen erhöhten Cholesterinwert handeln könnte.

Wert	603	564	529	491	458	417	412	409	85
G	6,312	5,829	5,350	4,755	4,203	3,451	3,388	3,364	3,275
G-crit	3,904	3,904	3,903	3,902	3,902	3,902	3,901	3,901	3,901
sig	yes	yes	yes	yes	yes	no	no	no	no

Tab. 3: Ausgabe des GESD-Tests für Cholesterin

thalrest: Es wurde der Wert 134 entfernt, da der Wert zu hoch erscheint (s. Bedeutung des Merkmals).

	entfernt	Anzahl	nach Entfernen
Cleveland	2	270	268
VA Medical	1	129	128
Hungarian	4	247	243
Switzerland	0	92	92
Summe	7	738	731

Tab. 4: Überblick über die Daten

3.8 Erstellen neuer Datensätze

Aus den vier vorhandenen Datensätzen wurden neue Datensätze erstellt. Während der Cleveland Datensatz und Hungarian Datensatz mittels der ML-Algorithmen im Weiteren untersucht werden sollten, dienten der VA Medical Datensatz und Switzerland Datensatz nur der effektiven Erhöhung der Stichprobengrößen für weitere Datensätze. Es wurden der „doctor“, „doctor 2“, „doctor 3“ und „doctor 4“ Datensatz neu erstellt. Desweiteren wurden aus Teilmengen an Merkmalen der Datensätze die Werte aller Datensätze zusammengeführt, sodass die Stichprobengröße für den Datensatz der jeweiligen Teilmenge von Merkmalen erhöht wird. Die neu erstellten Datensätze dienten der folgenden Untersuchung mittels ML-Algorithmen und unterschiedlichen Vergleichen dabei. Im Folgenden wird dargestellt, welche Datensätze untersucht werden und aus welchen Daten sich diese zusammensetzen. Die genauere Angabe, um welche Merkmale es sich jeweils handelt, findet man in Anhang 3.

Cleveland Datensatz: Der Datensatz enthält die Daten aus Cleveland, verglichen zum klassischen Cleveland Datensatz, der mit 13 untersuchten Merkmalen zu finden ist, hat dieser Datensatz 31 Merkmale. Der Datensatz kann auch als „cardiology“-Datensatz bezeichnet werden, da er auch alle kardiologisch erhobenen Daten enthält.

Hungarian Datensatz: Der Datensatz enthält die Daten aus Ungarn, verglichen zum Cleveland Datensatz fehlen z.B. die Angaben zum Rauchen.

doctor Datensatz: Der Datensatz besteht aus den Daten aus Cleveland und des VA Medical Datensatzes.

doctor 2 Datensatz: Der Datensatz besteht aus den Daten des Cleveland Datensatzes. Dies soll dazu dienen, verglichen mit dem Cleveland Datensatz eine vergleichbare Stichprobengröße zu haben. Ansonsten ist der Datensatz bzgl. der Merkmale identisch zum doctor Datensatz.

doctor 3 Datensatz: Der Datensatz besteht aus den Daten des Cleveland Datensatzes, des VA Medical Datensatzes und des Hungarian Datensatzes. Dies dient dazu, die Stichprobengröße effektiv zu erhöhen. Verglichen zum doctor Datensatz fehlen Angaben zum Rauchen.

doctor 4 Datensatz: Der Datensatz besteht aus den Daten des Cleveland Datensatzes. Dies soll dazu dienen, verglichen mit dem Cleveland Datensatz vergleichbare Stichprobengröße zu haben. Der Datensatz hat eine reduzierte Anzahl an Merkmalen.

3.9 Anwenden der Algorithmen

Nach der Erstellung der verschiedenen Datensätze wurden im Folgenden die ML-Algorithmen zur Klassifikation angewendet. Dabei soll jeweils klassifiziert werden, ob eine koronare arterielle Herzkrankheit vorliegt oder nicht. Zunächst wurden Referenzmodelle angewendet. Dazu gehören die

Standardmodelle kNN und die Entscheidungsbäume, deren Ergebnisse dann in der Folge mit komplexeren Modellen wie dem neuronalen Netzwerk und dem aufgebauten ML-System verglichen werden konnten. Desweiteren wurden innerhalb dieser Untersuchungen die Ergebnisse der Datensätze an verschiedenen Standorten verglichen. Dabei wurden die Modelle jeweils auf die Daten aus Cleveland und Ungarn angewendet. Außerdem sollten die Ergebnisse für unterschiedliche Kenntnisstände des jeweiligen Experten bzgl. des ML-Systems verglichen werden. So wurde jeweils der Cleveland Datensatz untersucht, der die Nutzung des CDSS für einen Kardiologen widerspiegeln kann, ebenso die „doctor"-, „doctor 2"- und „doctor 3"-Datensätze, die für unterschiedliche Daten die Nutzung durch einen Allgemeinarzt widerspiegeln sollen und dann auch die Verwendung des „doctor 4"-Datensatzes, der eine Untersuchung von Merkmalen ermöglicht, die auch ohne ärztliche Untersuchung bekannt sein können. Es wurde jeweils evaluiert mittels der Classification Accuracy und der falschpositiven und falschnegativen Klassifikationen.

3.10 Evaluation der Ergebnisse

Für die Evaluation der Ergebnisse wurde k fold cross validation angewendet. Dabei wurde k=7 verwendet. Dabei wurden zu jeweils unterschiedlichen Teilen der verschiedenen Datensätze Kreuzvalidierungsdaten erstellt. Aus der Classification Accuracy für die unterschiedlichen Kreuzvalidierungsdaten wird der Mittelwert berechnet.

$\overline{CA} = \frac{1}{k} \sum_{i=1}^{k} CA_i$

Ergebnisse und Auswertung

Mittels der Auswertung sollen verschiedene Fragestellungen beantwortet werden. Es soll z.B. beurteilt werden, wie verständlich die gefundenen Lösungen aus praktischer Sicht für einen Experten des jeweiligen Fachgebietes sind. Damit wird die Explainability bzw. der Aspekt des Explainable Machine Learning beurteilt. Desweiteren soll dargestellt werden, wie die verwendeten ML-Systeme praktisch zur Diagnose genutzt werden könnten. Dadurch kann der Unterschied bei einer Diagnose zu dem Vorgehen bei der klassischen klinischen Entscheidungsfindung dargestellt werden. Es soll dabei betrachtet werden, weshalb die Entscheidungsfindung durch einen Experten und ein zusätzliches ML-System vorteilhaft sein kann. Desweiteren ist relevant, ob die Classification Accuracy der Algorithmen ausreicht, um die ML-Systeme sinnvoll zur Diagnose einsetzen zu können. Die Fragestellungen sollen anhand unterschiedlicher Vergleiche beantwortet werden, bei denen die verschiedenen Algorithmen auf unterschiedliche Datensätze zur Klassifikation angewendet wurden. Durch den Vergleich der Ergebnisse für verschiedene ML-Modelle kann daraus geschlossen werden, welche davon geeignet sind. Außerdem wurde der Einfluss der Stichprobengröße untersucht und unterschiedliche Wissensstände des Fachexperten simuliert mittels der verschiedenen Datensätze (s. Literaturrecherche). Ein weiteres Ergebnis waren die aus den Quellen des UCI Machine Learning Repository neu erstellten Datensätze, die für weitere Untersuchungen verwendet werden könnten.

Überblick über Datensätze:

Index 1: Cleveland Datensatz

Index 2: Hungarian Datensatz

Index 3: doctor Datensatz, neu erstellt

Index 4: doctor 2 Datensatz, neu erstellt

Index 5: doctor 3 Datensatz, neu erstellt

Index 6: doctor 4 Datensatz, neu erstellt

Eine Beschreibung zu den verschiedenen Datensätzen, wie welche der Merkmale und Untersuchungseinheiten jeweils verwendet wurden ist im Anhang zu finden. Die Ergebnisse beziehen sich jeweils auf die theoretisch erläuterten Algorithmen aus Abschnitt 2.4.3, 2.4.4 und 2.4.5. Eine Übersicht über die verwendeten Algorithmen ist in Anhang 4.

4.1 Modell 1, kNN Klassifikation

Zunächst wurde die kNN Klassifikation untersucht. Dazu wurden die Daten jeweils z-transformiert. Es wurde als Parameter jeweils k=8 genutzt und damit die 8 nächsten Nachbarn für jeden Datenpunkt ausgewertet. Als Distanzmetrik wurde die euklidische Distanz basierend auf den standardisierten Daten genutzt. Das Ergebnis aus dem kNN-Algorithmus ist nur die Klassifikation der Untersuchungseinheiten. Es lässt sich eine Liste mit den Untersuchungseinheiten herleiten, bei der für jede angegeben ist, zu welcher Klasse (1: CAD vorhergesagt, 0: keine CAD vorhergesagt) sie nach dem Algorithmus zugerechnet wird.

	id	Y_2: target	prediction
0	233.0	0.0	0.0
1	234.0	1.0	0.0
2	235.0	0.0	1.0
3	236.0	1.0	1.0
4	237.0	1.0	0.0
...	...	...	...
33	266.0	0.0	0.0
34	267.0	1.0	0.0
35	268.0	0.0	1.0
36	269.0	1.0	0.0
37	270.0	0.0	0.0

Abb. 10: Liste der Untersuchungseinheiten, aus den Kreuzvalidierungsdaten des Cleveland Datensatzes

Es lässt sich damit nicht erkennen, wie es zu der Klassifikation gekommen ist. Der kNN-Algorithmus ist damit aus Sicht des Explainable Machine Learning nicht ausreichend. Prinzipiell wäre es nur möglich, parallel zur Funktionsweise des Algorithmus anhand der Distanzmetrik zu jeder Untersuchungseinheit auf ähnliche Untersuchungseinheiten zu schließen. Ein Experte könnte dann möglicherweise basierend auf den Merkmalswerten ähnlicher Untersuchungseinheiten (ähnlicher Patienten) darauf schließen, weshalb eine der Klassen gewählt wurde. Der Algorithmus liefert jedoch keine weitere Erklärung, was dazu beigetragen hat, dass eine Untersuchungseinheit einer Klasse zugeordnet wird, außer dass diese ähnlich zu anderen Untersuchungseinheiten ist. Damit ist Modell 1 praktisch so nutzbar, indem ein Arzt zu einer Diagnose kommt (hier CAD ja/nein) und diese mit der Klassifikation verglichen wird. In dem Fall, in dem die beiden Klassen übereinstimmen, gibt es für den Arzt einen weiteren Grund seiner Diagnose zu vertrauen. In dem Fall, dass die Diagnose und die vorhergesagte Klasse nicht übereinstimmen, ist Modell 1, da es nicht ausreichend erklärbar ist, nur bedingt sinnvoll nutzbar. Dies ist der Fall, da die Auswertung der Klassifikation von Modell 1 nur durch die Merkmalswerte möglich ist, die wiederum der Arzt selber interpretieren müsste. Das Modell liefert damit keine Erklärungsbeitrag oder erklärbare alternative Hypothesen zur Diagnose des Arztes.

Der kNN-Algorithmus wurde angewendet, da er einer der einfachsten der standardmäßigen Algorithmen des Machine Learnings ist. Aus der Classification Accuracy (s. Tab. 5) für die Kreuzvalidierungsdaten mit ca. 66 % für die meisten Modelle lässt sich entnehmen, dass der Algorithmus für die Anwendung als CDSS nicht geeignet ist und auch nicht für die genannten speziellen Datensätze. Auch für die Modelldaten war die Klassifikation mit maximal 73 % nicht sehr effizient. Für die Anwendung von CDSS wird noch höhere Classification Accuracy benötigt. Dies zeigt, dass die Anwendung standardmäßiger Algorithmen des Machine Learnings damit häufig nicht ausreicht. Desweiteren ergeben sich bei den unterschiedlichen Daten auch kaum Unterschiede in Bezug auf die Classification Accuracy. Damit konnten mehr Merkmale bzw. größere Stichprobengrößen und damit mehr Informationen nicht zu einer besseren Klassifikation beitragen. Der Grund dafür, dass mehr Merkmale nicht zu besserer Klassifikation beitragen konnten, liegt daran, dass nach der Standardisierung der Daten der kNN-Algorithmus alle Merkmale in Bezug auf die Distanzmetrik als gleich relevant betrachtet. Eine Möglichkeit wäre gewesen, die Anzahl der Merkmale auf die wichtigsten Merkmale zu reduzieren. Dies wäre jedoch einem Informationsverlust gleichgekommen und zu einer zu wenig detaillierten Bewertung von Informationen, die ein Experte sowieso wissen würde.

	$\overline{CA}_m$	$\overline{CA}_{cv}$	n_{cv}	FP	FN
1, Cleveland, n=268	0,729	0,664	38	9	7
2, Hungarian, n=243	0,717	0,675	34	6	3
3, doctor, n=335	0,739	0,702	47	6	7
4, doctor 2, n=268	0,729	0,664	38	9	7
5, doctor 3, n=578	0,743	0,649	82	17	8
6, doctor 4, n=268	0,729	0,646	38	11	7
Durchschnitt	0,731				

Tab. 5: Ergebnisse mittels k-fold cross validation und k=7, m: Modelldaten, cv:Kreuzvalidierungsdaten

4.2 Modell 2, Entscheidungsbäume Klassifikation

Dann wurden Entscheidungsbäume untersucht. Da Entscheidungsbäume invariant gegenüber Feature Transformationen reagieren, wurden die Daten nicht z-transformiert. Dies hat den Vorteil, dass Merkmalswerte direkt aus den Entscheidungsregeln abgelesen werden können ohne transformierte Merkmalswerte als z-Werte interpretieren zu müssen. Dies sorgt für bessere Erklärbarkeit als bei Modell 1. Im Gegensatz zur kNN Klassifikation wurde nicht nur die Klassifikation ausgegeben sondern auch ein Entscheidungsbaum, aus dem sich alle Entscheidungsregeln verständlich ablesen lassen (s. Anhang 5). Damit sind Entscheidungsbäume aus Sicht des Explainable Machine Learning prinzipiell gut geeignet. Aus Sicht des Machine Learning sind die Entscheidungsbäume sogar vollständig nachvollziehbar, da sich alle Entscheidungsregeln leicht ablesen lassen. Damit kann auch für weitere Vorhersagen darauf geschlossen werden, wie sich das ML-System verhalten würde. Dabei ist zu berücksichtigen, dass die achsenparallelen Entscheidungsfunktionen als Schwellenwerte medizinisch gut interpretiert werden können. Die Interpretation durch Schwellenwerte einzelner Merkmale oder Kombination mehrerer Schwellenwerte entspricht der diagnostischen Entscheidungsfindung von Ärzten. Desweiteren ist die Form als Entscheidungsbaum gut geeignet für diese Interpretation, da diese sehr stark den durch Ärzte genutzten sogenannten medizinischen Algorithmen ähneln (s. Abb. 11). Damit sollten Fachexperten wie Ärzte in der Lage sein, die Aussage von Entscheidungsbäumen prinzipiell gut verstehen zu können.

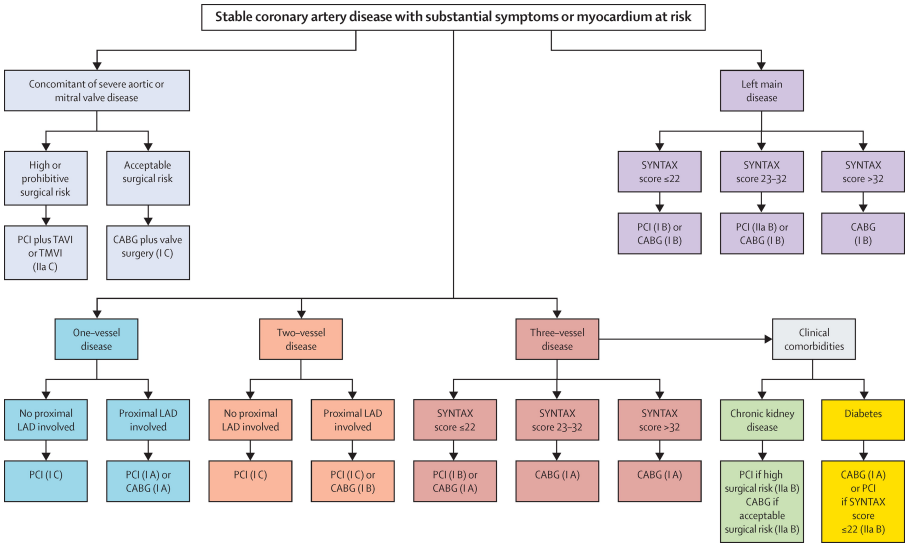


Abb. 11: Beispiel für einen medizinischen Algorithmus, Quelle: https://www.thelancet.com/cms/attachment/cdd5ce30- ebc3-4d54-a1e8-2f0e327ef13d, The lancet Volume 386,Number 9994, p702 , e2-e6

Es lässt sich für jede Untersuchungseinheit aus dem Entscheidungsbaum anhand logisch kombinierbarer Entscheidungsregeln nachvollziehen, wie es zu der Klassifikation gekommen ist. Das Prinzip ist dabei, dass für jeden Patienten der Entscheidungspfad entlang des Entscheidungsbaumes nachvollzogen werden kann. Die Vorhersage ergibt sich aus der Klassenbezeichnung an den Wurzelknoten.

Auf diese Weise ist es auch möglich, die Entscheidungspfade bzw. die Diagnosen für unterschiedliche Patienten zu vergleichen.

Es ist jedoch zu berücksichtigen, dass aus Sicht des jeweiligen Fachgebietes dies meist trotzdem nicht ausreichen könnte. Die Entscheidungsbäume sind zwar vollständig erklärbar, aber in dieser Form eher schwierig zu interpretieren. Zum einen ist die Informativität gering, da nicht alle Merkmale für den Entscheidungsbaum berücksichtigt wurden. Es wird damit nur eine Auswertung der wichtigsten Merkmale vorgenommen. Es können keine Aussagen zu Merkmalen abgeleitet werden, die nicht durch den Entscheidungsbaum berücksichtigt wurden. Zum anderen können die Entscheidungsregeln auf verschiedenen Ebenen des Entscheidungsbaumes nur schwer sinnvoll verknüpft werden, da die Merkmale grundlegend unterschiedliche Bedeutung haben. Es ist z.B. nicht unmittelbar klar, weshalb es sinnvoll ist, auf der einen Ebene Informationen über ECG und auf der nächsten Ebene direkt mit Angaben über Cholesterin zu verknüpfen. Der Vorteil von Entscheidungsbäumen, bedingte Verteilungen zu berücksichtigen und diese verknüpfen zu können, wurde dabei deshalb nicht gut ausgenutzt. Um die Erklärbarkeit zu verbessern, wurde TreeShap genutzt. Bei den Shapley Werten konnten dabei nur Merkmale berücksichtigt werden, die auch durch den Entscheidungsbaum berücksichtigt wurden. Es konnten für bestimmte Merkmale teils charakteristische Verläufe der Shapley Werte abgelesen werden. Ein Beispiel sieht man in Abbildung 12. Einige der Diagramme waren jedoch nur schwierig zu interpretieren. Zudem waren die Shapley Werte bei jeder Durchführung des Algorithmus stark unterschiedlich. Insgesamt waren die Ergebnisse der Anwendung von TreeShap auf Modell 2 damit nicht vollständig überzeugend. Im Gegensatz zu Modell 1 ist Modell 2 nicht nur praktisch nutzbar, wenn die Diagnose eines Fachexperten und die Klassifkation des ML-System übereinstimmen. Da Entscheidungsbäume und TreeShap weitere Erklärungsbeiträge liefern, kann der Fachexperte, wenn diese nicht übereinstimmen eine Hypothese aufstellen, weshalb das ML-System eine unterschiedliche Diagnose hat. Damit kann er entweder die Diagnose des ML-System verwerfen oder seine eigene Diagnose verwerfen und auch die korrekte Diagnose ableiten, wenn entweder der Fachexperte oder das ML-System nicht korrekt waren.

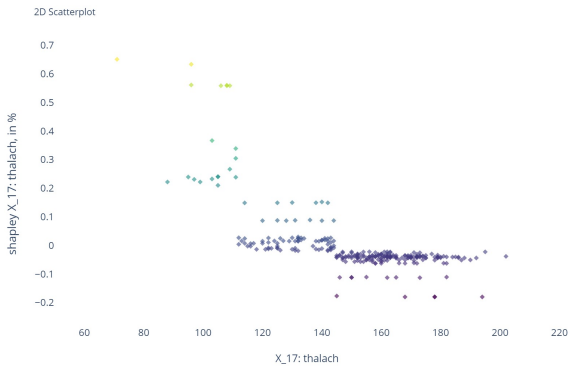


Abb. 12: Shapley Scatterplot

Die Entscheidungsbäume waren prinzipiell bereits besser für den Datensatz geeignet als der kNN Algorithmus. Aus der Classification Accuracy (s. Tab. 6) für die Kreuzvalidierungsdaten lässt sich

dies entnehmen. Es wurde eine Classification Accuracy von 65,3 % bis 74,1 % für die Kreuzvalidierungsdaten erreicht. Für die Modelldaten waren die Werte deutlich höher, was auf end cut preference bzw. overfitting hindeutet. Desweiteren haben sich gut interpretierbare Unterschiede für unterschiedliche Wissensstände des Fachexperten und unterschiedliche Stichprobengrößen ergeben.

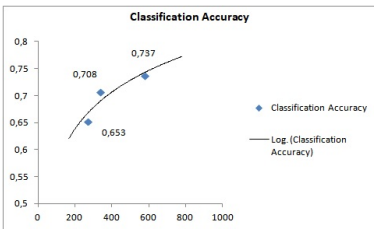


Abb. 13: Stichprobengröße und Classification Accuracy

Für das allgemeinärztliche Niveau war die Classification Accuracy mit 0,653 geringer als bei dem Wissenstand eines Kardiologen mit 0,709. Aus der Untersuchung der Stichprobengrößen ist zu entnehmen, dass die Classification Accuracy bei größerer Stichprobe höher ist (s. Abb.). Ein Grund könnte sein, dass Entscheidungsbäume für die heterogenen Daten gut geeignet waren, da nach der Aufteilung die weiteren Entscheidungen unabhängig von den vorherigen Entscheidungsregeln sind, was einer Strukturierung von den Daten entspricht. Die Daten können damit auch aus unterschiedlichen Datenquellen sein. Ist das Ziel die Inferenz, so wurde dies auch als selektive bzw. unabhängige Inferenz bezeichnet. Insgesamt waren die Entscheidungsbäume besser geeignet als Modell 1. Für die Anwendung von CDSS wird eine noch höhere Classification Accuracy benötigt. Außerdem waren die Ergebnisse teils auch mittel der Methoden des Explainable Machine Learning nicht optimal interpretierbar.

	$\overline{CA}_m$	$\overline{CA}_{cv}$	n_{cv}	FP	FN
1, Cleveland, n=268	0,897	0,709	38	9	3
2, Hungarian, n=243	0,883	0,741	34	8	3
3, doctor, n=335	0,844	0,708	47	8	4
4, doctor 2, n=268	0,859	0,653	38	8	6
5, doctor 3, n=578	0,843	0,737	82	8	7
6, doctor 4, n=268	0,848	0,687	38	6	8
Durchschnitt	0,862				

Tab. 6: Ergebnisse mittels k-fold cross validation und k=7, m: Modelldaten, cv: Kreuzvalidierungsdaten

4.3 Modell 3, Klassifikation

Bei Modell 3 wurden die Merkmale in die unten angegebenen Datenviews aufgeteilt und für jede Teilmenge 50 Entscheidungsbäume erstellt. Eine Auswahl aus den insgesamt 250 Entscheidungsbämen findet man in Anhang 6. Um die Ergebnisse mit dem vorherigen Modell vergleichen zu können, wurde der Cleveland-Datensatz gewählt. Auch der doctor 2 Datensatz (ohne Datenview Kardiologie) wurde verwendet.

1: Präposition, Gewohnheiten	2: ECG	3: Herz Kreislauf	4: Vorerkrankungen	5: Kardiologie (nur Cleveland)
X_1: age	X_28: oldpeak	X_22: thalach	X_5: dm, fbs	X_31: ca
X_2: sex	X_29: slope	X_23: thalrest	X_8: hist_cad	X_32: thal
X_3: cigs	X_30: rldv5e	X_24: tpeakbps	X_9: dig	
X_4: cig_time	X_22: thalach	X_25: tpeakbpd	X_10: prop	
X_5: dm, fbs	X_35: thaldur	X_26: trestbpd	X_11: nitr	
X_6: chol	X_14: restecg_0	X_33: trestbps	X_12: pro	
	X_15: restecg_1	X_34: htn	X_13: diuretic	
	X_16: restecg_2	X_37: xhypo	X_34: htn	
			X_13: met	

Tab. 7: Merkmale der Datenviews

Datenview 6: Clustering der Zugehörigkeit zu Wurzelknoten, Mittelwert der Klassenwahrscheinlichkeiten der Cluster

Da man die verwendeten Merkmale pro Datenview auswählen und die logisch verknüpften Merkmale benutzen kann, sind die Entscheidungsbäume besser interpretierbar. Die Interpretierbarkeit wurde auch dadurch verbessert, da die Anzahl der betrachteten Merkmale pro Entscheidungsbaum reduziert wurde. Allerdings muss man verglichen zu Modell 2, um die Klassifikation nachvollziehen zu können, mehr Entscheidungsbäume auswerten; für eine vollständige Bewertung muss man alle 50 verschiedenen Entscheidungsbäume ansehen. Aus dem Entscheidungsbaum zu Datenview 1 (s. Abb. 14) lässt sich beispielsweise ablesen, dass jüngere Frauen eine geringe Klassenwahrscheinlichkeit haben an CAD zu erkranken. In weiteren Entscheidungsbäumen des Datenviews ist abzulesen, dass diese Reduktion der Wahrscheinlichkeit bei Frauen besonders effektiv ist, wenn zudem der Cholesterinwert in einem optimalen Bereich ist (s. dazu auch den Shapley Scatterplot zu Cholesterin in Abb. A.16 in Anhang 8, bei dem der Wert bei 180 mg/dl bis 230 mg/dl optimal ist). Aus dem Entscheidungsbaum aus Abb. 14 ist erkennbar, dass eine solche Reduktion auch bei Männern im Alter von 43 bis 60 Jahren möglich ist (in geringerem Ausmaß). Der Einfluss der Merkmale kann auch aus den Shapley Scatterplots nachvollzogen werden. Verglichen zu Modell 2 ist relevant, dass Diabetes mit berücksichtigt wird. Selbst bei reduzierten Merkmalen wird Diabetes bei den Entscheidungsbäumen aus Modell 2 nicht berücksichtigt. Dies kann mehrere Gründe haben wie end cut

preference. Ein weiterer ist, dass Diabetes verglichen zu anderen Faktoren, wie z.B. dem Rauchen, im Datensatz unterrepräsentiert ist. Dass Diabetes in Modell 3 mit berücksichtigt wird ist darauf zurückzuführen, dass die verwendeten Entscheidungsbäume Randomisierung ausnutzen.

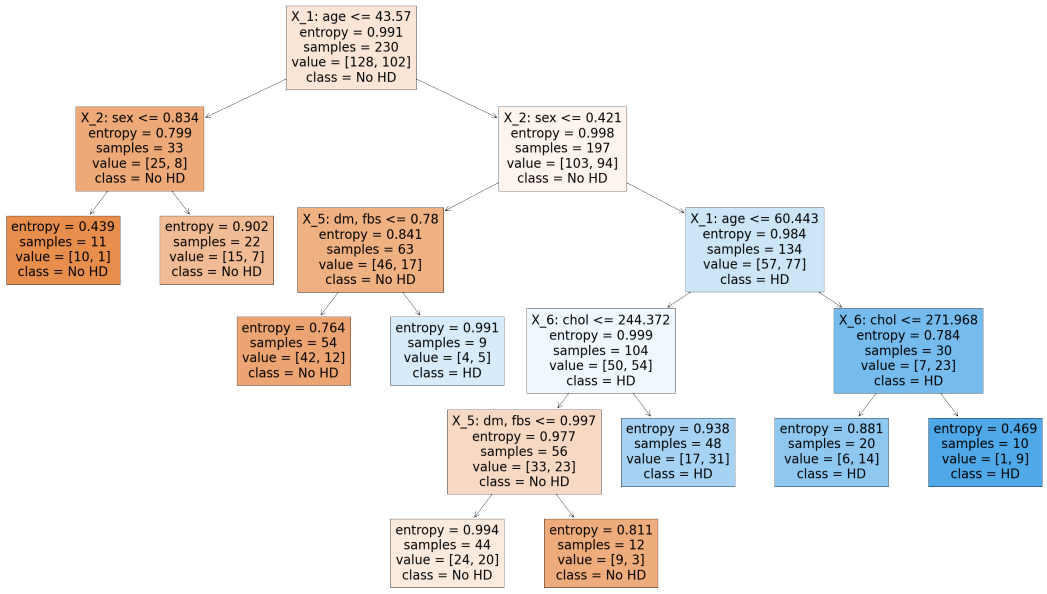


Abb. 14: Entscheidungsbaum aus Datenview 1

Für jeden Datenview wurde eine Klassenwahrscheinlichkeit errechnet und alle Klassenwahrscheinlichkeiten, wie in der Formel unten angegeben, zu einem Wert zusammengefasst. Unterschiedliche Klassenwahrscheinlichkeiten sind besser zu vergleichen, als nur eine reine Klassifikation. Zum Vergleich kann man die Klassenwahrscheinlichkeiten aus den Entscheidungsbäumen nutzen oder für den jeweiligen Datenview oder man betrachtet den gesamten Wert für die Wahrscheinlichkeit bzw. den score. Jede Untersuchungseinheit erhält einen speziellen Wert, sodass verschiedene Einheiten miteinander verglichen werden können und sich die Erklärbarkeit dadurch erhöht (s. dazu Abb. 15). Man hat auch die Möglichkeit für eine Untersuchungseinheit zu einem späteren Zeitpunkt diesen Wert ein weiteres Mal zu bestimmen, um eine Veränderung festzustellen. So wäre es möglich für einen Patienten einen Verlauf von Werten zu erstellen und beispielsweise einen Patienten an einen Kardiologen zu überweisen, wenn der Wert einen bestimmten Wert wie den Threshold Parameter überschreitet. Dabei würden die falschpositiven und falschnegativen Überweisungen reduziert werden.

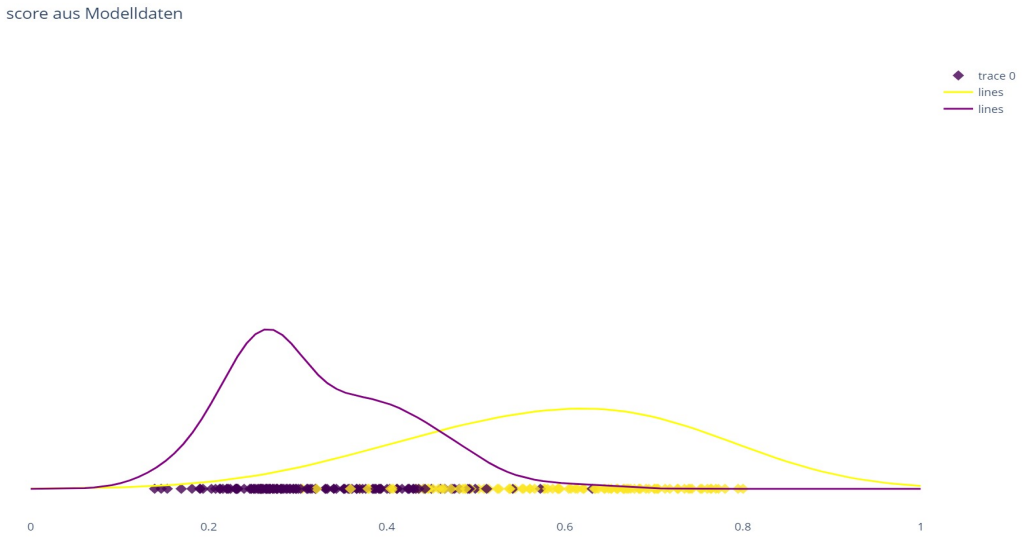


Abb. 15: scores aus den Modelldaten

Im Gegensatz zu Modell 2 sind auch die Shapley Werte besser zu interpretieren (s. dazu Anhang 7 und 8). Man sieht in den Shapley-Diagrammen eindeutige Verläufe, eine Verbesserung zu den Diagrammen in Modell 2. Da die Shapley Werte jetzt auf die Klassenwahrscheinlichkeit bezogen sind, kann der Betrag der Merkmalswerte als Wahrscheinlichkeit für eine koronare Herzkrankheit interpretiert werden. Im Vergleich zu Modell 2, bei dem nur Shapley Werte für die im Entscheidungsbaum verwendeten Merkmale berechnet wurden, hat man jetzt die Werte für fast alle Merkmale. Der Grund dafür ist, dass man mehr Entscheidungsbäume hat und man in die Teilmengen von Merkmalen aufgeteilt hat. Die Shapley Werte ergeben eine gute Möglichkeit zur Zusammenfassung der Ergebnisse und um die Erklärbarkeit der Entscheidungsbäume damit zu erhöhen.

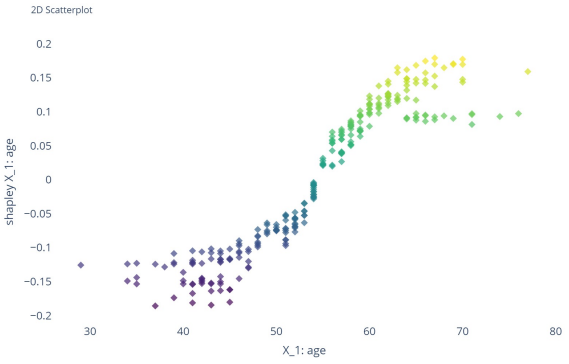


Abb. 16: Shapley Scatterplot zum Merkmal Alter

Ein Arzt kann jetzt die Shapley-Diagramme mit seinem eigenen Wissen oder mit Ergebnissen aus der medizinischen Forschung abgleichen. In der medizinischen Forschung verwendet man ähnliche Diagramme, die auch Verläufe mit Wahrscheinlichkeiten für unterschiedliche Merkmale enthalten. Die Ergebnisse stimmen z.B. für das Merkmal „age“ grundsätzlich mit denen aus der medizinischen Forschung überein (s. Abb. 17 und 18). So sind die Verläufe sehr ähnlich. Desweiteren ist das Alter, bei dem kein positiver oder negativer Effekt vgl. zum durchschnittlichen Patienten gefunden wurde, mit 50-60 Jahren sehr ähnlich (s. Odds Ratio gleich 1 in Abb. 18). Aus der Gruppe an Werten, die bei höherem Alter geringere Shapley Werte aufweisen, könnte vermutet werden, dass dies auf ethnische Differenzen zurückzuführen ist (s. dazu Abb. 18). Insgesamt können sehr detailliert Erklärungen durch die Shapley Werte gefunden werden.

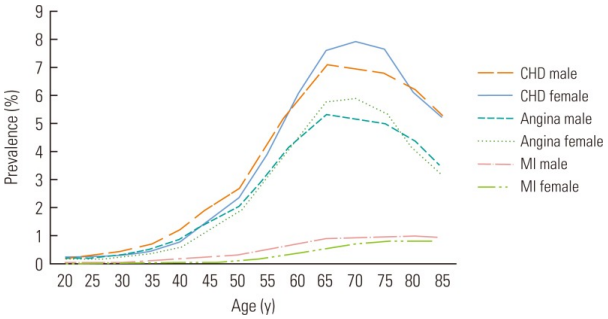


Abb. 17: Zusammenhang von Alter und koronarer Herzerkrankung einer Studie in Korea, Quelle: https://www.researchgate.net/profile/Hoo-Sun-Chang/publication/232612101/figure/fig1/AS:213424287031301@1427895723360/Prevalence-of-treated-coronary-heart-disease-by-age-and-sex-in-Korea-2005-CHD-coronary.png

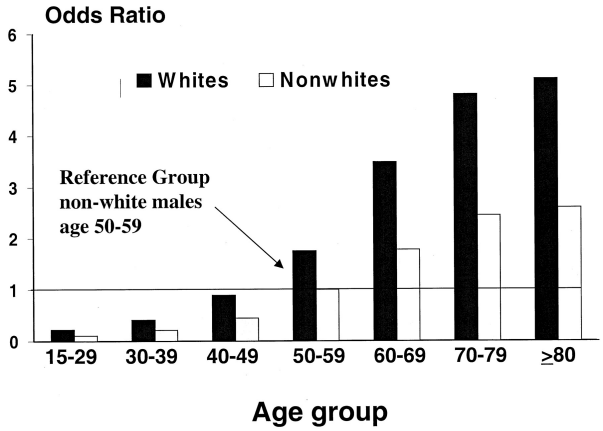


Abb. 18: Zusammenhang von Alter und koronarer Herzerkrankung einer Studie, Quelle: https://jasn.asnjournals.org/content/12/7/1516

Es lassen sich desweiteren die Shapley Werte nicht nur global als Zusammenfassung und erweiterten Erklärung der Entscheidungsbäume nutzen sondern auch lokal. So kann für jede Untersuchungseinheit zu jedem Datenview bzw. jeder Klassenwahrscheinlichkeit die Beiträge verschiedener Merkmalswerte festgestellt werden. Dies wird beispielhaft für Patient mit der Id 240 dargestellt (s. dazu Anhang 9). Er hat die angegebenen Klassenwahrscheinlichkeiten für verschiedene Datenviews. Diese setzen sich aus den Shapley Werten für unterschiedliche Merkmale zusammen, da als Zielwert für die Shapley Werte die Klassenwahrscheinlichkeiten gewählt wurden. Der Aspekt Präposition und Herz Kreislauf suggeriert, dass der Patient CAD haben könnte. Die Aspekte ECG, Vorerkrankungen und Kardiologie sprechen jedoch gegen diese Diagnose (s. dazu auch die Shapley plots und die Shapley Scatterplots in Anhang 8). Wie aus Abb. 19 zu erkennen ist, hat beispielsweise das Alter für den Datenview einen Einfluss von +10% vgl. zum durchschnittlichen Patienten der übereinstimmt mit den Shapley Werten, die für das Alter 62 Jahre auch aus dem Shapley Scatterplot in Abb. A.11 abzulesen ist. Desweiteren stimmt der gesamte Einfluss des Alters von 0,075 auf dem score ungefähr mit den Werten überein, die aus der Studie aus Abb. 17 bei einem Alter von 62 Jahren aus der Prävalenz abgelesen werden können. Insgesamt hat er einen score von 0.3647, der kleiner ist als der Threshold Parameter, und wird daher als nicht CAD klassifiziert. Der Patient hat nach der Zielvariable tatsächlich kein CAD.

$$score_{240} = \vec{\epsilon_v} \cdot \tilde{\vec{P}}_{240} = \epsilon_1 \cdot \tilde{P}_{1,240} + \epsilon_2 \cdot \tilde{P}_{2,240} + \epsilon_3 \cdot \tilde{P}_{3,240} + \epsilon_4 \cdot \tilde{P}_{4,240} + \epsilon_5 \cdot \tilde{P}_{5,240} + \epsilon_6 \cdot \tilde{P}_{6,240} = 0.25 \cdot 49\ \% + 0.23 \cdot 41\ \% + 0.19 \cdot 48\ \% + 0.08 \cdot 36\ \% + 0.18 \cdot 12\ \% + 0.07 \cdot 9\ \% = 0.3647$$

$$\tilde{C}_{ges,240} = \begin{cases} 0, \text{ für } score_{240} < 0.47 \\ 1, \text{ für } score_{240} \geqslant 0.47 \end{cases}$$

Einfluss vom Alter

$$\epsilon_1 |(\varphi_{x_1=62} - \varphi_{x_1=20})| = 0{,}075$$

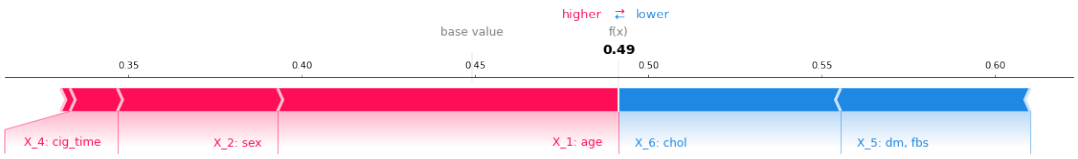


Abb. 19: Shapley plot zu Datenview 1

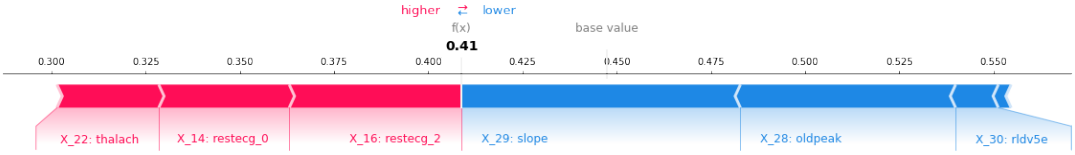


Abb. 20: Shapley plot zu Datenview 2

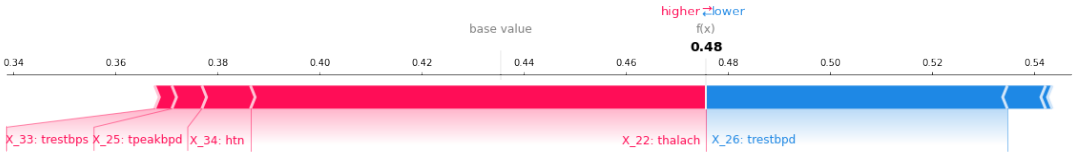


Abb. 21: Shapley plot zu Datenview 3

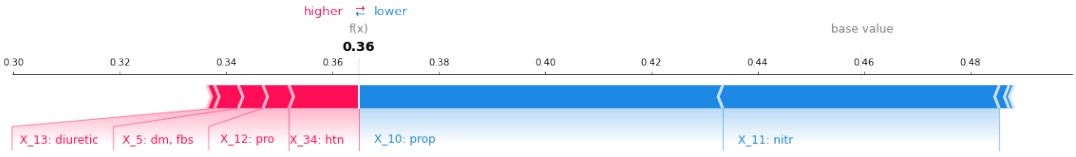


Abb. 22: Shapley plot zu Datenview 4

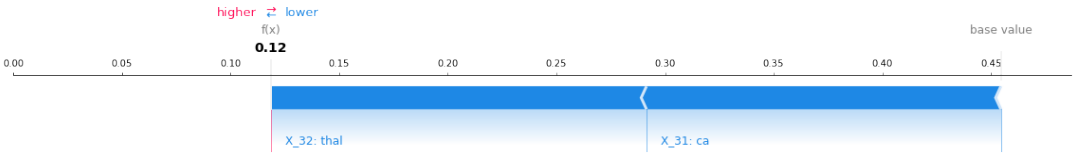


Abb. 23: Shapley plot zu Datenview 5

Insgesamt ist die Erklärbarkeit von Modell 3 besser als bei Modell 2. Damit ist das aufgebaute ML-System auch praktisch gut nutzbar. Ein Fachexperte kann z.B. detailliert erkennen, warum eine Diagnose des Systems wie angegeben ausfällt. Zudem wird auch eine bessere Classification Accuracy erreicht. Die Classification Accuracy erscheint, da die Ergebnisse erklärbar sind, als ausreichend

für ein CDSS. Die Classification Accuracy war desweiteren im Gegensatz zu Modell 2 bei den Daten auf allgemeinärztlichen Niveau (doctor 2 Datensatz) mit 0,807 nur unwesentlich geringer als für den Wissensstand eines Kardiologen (Cleveland Datensatz).

	$\overline{CA}_m$	$\overline{CA}_{cv}$	n_{cv}	FP	FN
1, Cleveland, n=268	0,866	0,813	38	6	3
4, doctor 2, n=268	0,889	0,807	38	7	3
Durchschnitt	0,862				

Tab. 8: Ergebnisse mittels k-fold cross validation und k=7, mittels Cleveland Daten

i	CA_{cv}
1	0,821
2	0,794
3	0,868
4	0,842
5	0,789
6	0,816
7	0,763

Tab. 9: Ergebnisse aus der Kreuzvalidierung, für Cleveland

Beispiele für mögliche weitere Interpretationen:

Es sind die Shapley Scatterplots zu allen Merkmalen in Anhang 8 dokumentiert. Aus den Shapley Scatterplots lassen sich viele weitere Interpretationen ziehen.

Einfluss Rauchen: Die durch Rauchen festgestellten Effekte sind nur gering. Bei längerer Zeit erhöht sich ab 25 Jahren Rauchen und ab 20 Zigaretten pro Tag die Wahrscheinlichkeit für eine Krankheit (s. dazu Abb. A.13 und A.14). Dies entspricht der theoretischen Annahme. Jedoch sind die Ergebnisse relativ uneindeutig. Es ist auch die Wahrscheinlichkeit relativ hoch, wenn der Patient nicht raucht. Dies könnte auf mögliche Verzerrungen und Fehler durch den Einfluss Rauchen hindeuten. Trotzdem sind die Ergebnisse sinnvoll, da durch Explainable ML gerade solche Verzerrungen erkannt werden sollen. Tatsächlich treten diese auch auf, da in der Cleveland Stichprobe ein Großteil der Patienten raucht. Deshalb könnte Rauchen durch den Algorithmus als normal eingeordnet werden, was auch die geringen und teils nicht eindeutigen Effekte erklärt.

zu dm, fbs: Im Gegensatz zu Modell 2 konnte ein Effekt von Diabetes nachgewiesen werden. Bei Diabetes ist die Wahrscheinlichkeit für CAD höher (s. dazu Abb. A.15). Dies ist jedoch nicht für alle Patienten der Fall. Einige Patienten mit Diabetes haben geringere Wahrscheinlichkeit für CAD.

Dies könnte darauf zurückzuführen sein, dass diese Personen generell weniger neigen an CAD zu erkranken.

zu chol: Der Cholesterinwert hat einen moderaten Effekt. Es ist generell bei geringerem Cholesterinwert die Wahrscheinlichkeit für CAD geringer (s. dazu Abb. A.16). Dies entspricht der Theorie. Für bestimmte Patientengruppen kann sehr effektiv die Wahrscheinlichkeit verringert werden, wenn diese einen optimalen Cholesterinwert von 180 mg/dl bis 230 mg/dl haben. Für höhere Cholesterinwerte ist die Schwankung hoch und es könnte damit stark von dem jeweiligen Patienten abhängen, ob ein hoher Cholestinwert die Wahrscheinlichkeit für CAD ansteigen lässt.

zu oldpeak: Es lässt sich aus dem ECG eindeutig aus einer höheren ST-Segment Depression auf eine höhere Wahrscheinlichkeit für CAD schließen (s. dazu Abb. A.17). Der Verlauf ist dabei linear bis zu einer ST- Segment Depression von 3.

zu slope: Es lässt sich aus dem ECG eindeutig aus einer negativen Steigung oder einem flachen Verlauf des ST-Segments auf eine erhöhte Wahrscheinlichkeit für CAD schließen (s. dazu Abb. A.18).

Fazit

In dieser Arbeit sollten verschiedene ML-Modelle auf die Eignung für ein CDSS verglichen werden. Dabei wurde sowohl untersucht, ob die Classification Accuracy ausreichend ist, als auch, ob die Modelle dabei erklärbar sind und damit sinnvoll für ein CDSS als Unterstützung für einen Arzt anwendbar wären. Dabei wurde der „Heart Disease“-Datensatz jedoch nicht wie häufig mit 13 Merkmalen verwendet sondern mit 33 Merkmalen. Die Situation war damit realistischer und prinzipiell können damit mehr Informationen gewonnen werden. Die Situation mit 13 Merkmalen zu untersuchen, mag aus Sicht des Machine Learnings sinnvoll sein. Aus Sicht eines Fachexperten könnte sich jedoch herausstellen, dass die Untersuchung mit geringerer Anzahl an Merkmalen häufig nicht ausreichend ist, da sich aus den Ergebnissen zu wenige Informationen schließen lassen oder nur Informationen, die ein Arzt ohnehin wüsste. Es wurden für die Situation mit 33 Merkmalen die Klassifikation mittels kNN, mittels Entscheidungsbäumen und mit einem neu aufgebauten Multiclassifier System untersucht. Das Ziel war dabei der Vergleich zu der Classification Accuracy von 83,67 %, die mittels eines neuronalen Netzwerks [13] erreicht wurden. Während die Klassifikation mittels kNN weder in der Accuracy noch in der Erklärbarkeit ausreichend war, so waren die Entscheidungsbäume bereits besser geeignet, besonders, da sie vergleichsweise gut erklärbar waren. Am besten geeignet war das Multiclassifier System, das mit 81 % Accuracy aber nicht die Accuracy des neuronalen Netzwerks erreichen konnte. Wurden jedoch die Daten aus allen Standorten integrativ genutzt, so wurde eine höhere Accuracy erreicht. Dies zeigt wie wichtig es ist, ausreichend große Stichprobengrößen zu berücksichtigen, um repräsentative und detaillierte Ergebnisse erhalten zu können. Die Verwendung von 300 Untersuchungseinheiten ist dabei nicht ausreichend. Insgesamt war die Untersuchung von CAD damit wie in der Literatur erfolgreich. Durch Simulation verschiedener Wissensstände konnte festgestellt werden, dass bereits die Untersuchungen auf allgemeinärztlichem Niveau für eine solche Classification Accuracy ausreichen. Es würde sich deshalb eine Anwendung eines CDSS zur Unterstützung der Erkennung von CAD anbieten. Würden mehr Merkmale erhoben werden, könnte sogar höhere Classification Accuracy erreicht werden. Dabei ist zu berücksichtigen, dass sicherlich nicht in jedem Fall so hohe Accuracy wie bei der Untersuchung von CAD erreicht werden kann. Desweiteren müssten die gefundenen Ergebnisse in professioneller Umgebung durch einen Fachexperten evaluiert werden.

In Bezug auf die Explainability wurden verschiedene Methoden untersucht. Zunächst wurde ein ML-Modell verwendet, das inhärent erklärbar ist und versucht strukturell die Erklärbarkeit zu erhöhen. Besonders die inhärent verständlichen Entscheidungsbäume haben dazu beigetragen, dass die Modelle erklärbar waren. Ebenso hat der strukturelle Aufbau bei Modell 3 dazu beigetragen. So konnte ein ML-System aufgebaut werden, das durchgängig nachvollziehbar war und zahlreiche Outputs liefert, die die Erklärbarkeit erhöhen. Dabei wurden zusätzlich Shapley Werte genutzt, die dazu beigetragen haben. Besonders bei der Interpretation der Ergebnisse waren die Shapley Werte hilfreich. Die Interpretationen konnten häufig generell durch die Theorie bestätigt werden. Während die Shapley Werte in bestimmten Anwendungen, wie z.B. den medizinischen Bildern, zur Erklärung ausreichen können, so waren sie für die Anwendung des CDSS jedoch nicht ausreichend.

Erst in Kombination mit den Entscheidungsregeln der Entscheidungsbäume konnten sie effizient zur Erklärung beitragen. Damit wäre es nicht ausreichend „Black Box“-Modelle mittels Shapley Werten erklärbar zu machen, sondern es sind jeweils auch verständliche Entscheidungsregeln nötig. Während viele ML-Modelle Entscheidungsregeln nutzen, so sind viele jedoch gerade für einen Fachexperten nicht gut verständlich. In dieser Situation die Erklärbarkeit zu verbessern, indem generelle Verfahren des Explainable Machine Learnings verwendet werden, wie Shapley Werte, war für die Anforderungen an die Erklärbarkeit des CDSS nicht ausreichend. Bei den Shapley Werten war besonders auffällig, dass sie gezielt eingesetzt werden sollten. Während sie bei Modell 2 deutlich weniger gut geeignet waren, so waren sie bei Modell 3 bereits deutlich besser geeignet, da bei diesem Modell mehr Entscheidungsbäume aufgebaut wurden und Randomisierung ausgenutzt wurde. Inwieweit diese korrekt sind, müsste natürlich durch einen Fachexperten bestätigt werden.

In Bezug darauf, welche Modelle für ein CDSS geeignet wären, konnte festgestellt werden, dass auch bei der Untersuchung mittels einfacher und bereits lange bekannter Modelle gute Ergebnisse erreicht werden können. Dies ist relevant, da eine mögliche Begründung ML-Systeme z.B. für ein CDSS nicht einzusetzen wäre, dass noch höhere Accuracy benötigt wird, die nur durch neuartige und komplexe ML-Modelle erreicht werden könnten. Dies kann in Bezug auf andere Anwendungen auch nötig sein. Bei der Untersuchung der CAD konnte der Vorteil von komplexeren und dabei unverständlicheren Modellen wie z.B. neuronalen Netzwerken insgesamt nicht bestätigt werden. Die Accuracy von komplexeren Modellen war nur moderat höher als von dem aufgebauten System und würde nicht rechtfertigen, dass die Ergebnisse weniger verständlich sind. Stattdessen sind viele der erklärbaren Modelle, wie Entscheidungsbäume und generalisierte additive Modelle, bereits lange bekannt und konnten bereits gute Ergebnisse bei der Vorhersage von CAD erreichen. Damit kann in einigen Anwendungen nicht herangezogen werden, dass komplexere Modelle generell nötig wären, um ML überhaupt erst sinnvoll anwenden zu können. Dabei ist auch zu berücksichtigen, dass, je besser der Fachexperte das ML-System versteht, eine umso geringere Accuracy nötig ist für einen sinnvollen Einsatz. Dass trotz der prinzipiellen Möglichkeit ML-Systeme zu nutzen, diese nicht regulär verwendet werden, erscheint deshalb erstaunlich, da das aufgebaute ML-Modell in einer professionelleren Ausführung in der Lage wäre präventiv CAD zu erkennen. Dass die frühzeitige Erkennung von CAD sehr wichtig ist, zeigen die hohen Fallzahlen in Russland und Indien. Dabei könnte das CDSS einem Kardiologen helfen. Jedoch wäre es hilfreicher in Situationen die häufiger auftreten, in denen ein Arzt geringeres kardiologisches Fachwissen hat. Dies ist der Fall, wenn der Patient durch einen Allgemeinarzt behandelt wird. Es könnte z.B. dabei helfen Patienten rechtzeitig zum Kardiologen zu überweisen, um die hohen Fallzahlen aus Folgeerkrankungen der CAD zu vermeiden. Dabei ist die Erkennung von CAD auch bei möglichen Ungenauigkeiten des ML-Systems risikofrei, da nicht das System selbst eine Entscheidung trifft. Deshalb sollte es bei dieser Anwendung keine regulatorischen oder ethische Einwände geben. Insgesamt erscheint damit v.a. die Notwendigkeit für Explainability und bislang die fehlende Nutzung von Explainable Machine Learning als ein Grund, weshalb solche Systeme nicht professionell umgesetzt worden sind, denn erst bei Nutzung verständlicher ML-Systeme steigt die Akzeptanz der Fachexperten.

Literaturverzeichnis

[1] Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3(17).

[2] Witten D, editor. The Big Data Blog, Part II: Daniela Witten [blog on the Internet]. Washington: American Association for the Advancement of Science; 2014 [cited 2022 Mai 26]. Available from: `https://www.aaas.org/news/big-data-blog-part-ii-daniela-witten`.

[3] Heart Disease Data Set [dataset]. UCI ML Repository, 1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. , 2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D. , 3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D. , 4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D., Donor: David W. Aha; 1988. Available from: `https://archive.ics.uci.edu/ml/datasets/Heart+Disease`.

[4] Human in the Loop [website]. Agile Im; [cited 2022 Mar 30]. Available from: `https://www.agile-im.de/2019/08/14/human-in-the-loop/`.

[5] Detrano R, Gianrossi R, Mulvihill D, Lehmann K, Dubach P, Colombo A, et al. Exercise-Induced ST Segment Depression in the Diagnosis of Multivessel Coronary Disease: A Meta Analysis. JACC. 1989;14(6):1501-8.

[6] Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, et al. International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease. The American Journal of Cardiology. 1989;64:304-10.

[7] Veterans Health Administration [website]; [cited 2022 Mar 30]. Available from: `https://www.va.gov/health/`.

[8] Hungarian Institute of Cardiology [website]; [cited 2022 Mar 30]. Available from: `https://www.gokvi.hu`.

[9] Z-Alizadeh Sani Data Set [dataset]. UCI ML Repository, Dr Zahra Alizadeh Sani; 2017. Available from: `https://archive.ics.uci.edu/ml/datasets/Z-Alizadeh%20Sani`.

[10] Framingham Dataset [dataset]. Framingham Heart Study; 2022. Available from: `https://www.framinghamheartstudy.org`.

[11] Misra R, Gupta P, Jain P. Prediction of Heart Disease Using Machine Learning Algorithms. IJIRT. 2021;8(2):643-6.

[12] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Computational Intelligence and Neuroscience. 2021;2021.

[13] Miao KH, Miao JH. Coronary Heart Disease Diagnosis using Deep Neural Networks. International Journal of Advanced Computer Science and Applications. 2021;9(10).

[14] Dave D, Naik H, Singhal S, Patel P. Explainable AI meets Healthcare: A Study on Heart Disease Dataset. International Journal of Advanced Computer Science and Applications. 2020.

[15] Kardiovaskuläre Erkrankung [website]. Köln: DocCheck; [cited 2022 Mar 30]. Available from: https://flexikon.doccheck.com/de/Kardiovaskul%C3%A4re_Erkrankung.

[16] Know the Differences Cardiovascular Disease, Heart Disease, Coronary Heart Disease [website]; [cited 2022 Mar 30]. Available from: https://www.nhlbi.nih.gov/sites/default/files/media/docs/Fact_Sheet_Know_Diff_Design.508_pdf.pdf.

[17] Heart failure [website]. Mayo Clinic; [cited 2022 Mar 30]. Available from: https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142.

[18] Coronary Artery Disease (CAD) Complications [website]; [cited 2022 Mar 30]. Available from: https://www.healthline.com/health/coronary-artery-disease/complications.

[19] Cardiac Disease Among South Asians: A Silent Epidemic [website]. Indian Heart Association; [cited 2022 Mar 30]. Available from: http://indianheartassociation.org/why-indians-why-south-asians/overview.

[20] Cholesterinwerte Tabelle [website]; [cited 2022 Mar 30]. Available from: https://www.grossesblutbild.de/cholesterinwerte-tabelle.

[21] Smoking and Coronary Artery Disease [website]. University of Michigan; [cited 2022 Mar 30]. Available from: https://www.uofmhealth.org/health-library/hw79682.

[22] Quitting smoking drops heart attack risk to levels of never smokers [website]. European Society of Cardiology; [cited 2022 Mar 30]. Available from: https://www.escardio.org/The-ESC/Press-Office/Press-releases/Quitting-smoking-drops-heart-attack-risk-to-levels-of-never-smokers.

[23] Emberson JR, Bennett DA. Effect of Alcohol on Risk of Coronary Heart Disease and Stroke: Causality, Bias, or a Bit of Both? Vascular Health and Risk Management. 2006;2(3):239–49.

[24] Murphy A, Johnson CO, Roth GA, Forouzanfar MH, Naghavi M, Ng M, et al. Ischaemic heart disease in the former Soviet Union 1990-2015 according to the Global Burden of Disease 2015 Study. Heart. 2018;104(1):58-66.

[25] Kannel WB MD. Diabetes and cardiovascular disease, The Framingham study. JAMA. 1979;241(19):2035-8.

[26] Morik, K . Vorlesung Maschinelles Lernen, Klassifikation und Regression: Lineare Modelle. Technische Universität Dortmund; 2008. https://www-ai.cs.tu-dortmund.de/LEHRE/VORLESUNGEN/MLRN/WS0809/2MLVbiasVariance.de.pdf.

[27] Li Y, Wu FX, Ngom A. A review on machine learning principles for multi-view biological data integration. Briefings in Bioinformatics. 2018;19(2):325–40.

[28] Rooney, N . Stacking for supervised learning. NIKEL, University of Ulster; 2007. http://www.ukkdd2007.org/slides/UKKDD-2007-Niall-talk.pdf.

[29] Morik, K . Maschinelles Lernen. Technische Universität Dortmund; 2013. https://www-ai.cs.tu-dortmund.de/LEHRE/VORLESUNGEN/MLRN/WS1314.

[30] Margineantu, D and Dietterich, T . Improved Class Probability Estimates from Decision Tree Models; 2002. https://web.engr.oregonstate.edu/~tgd/publications/tr-msri-2002.pdf.

[31] Fuhrman JD, Gorre N, Hu Q, Li H, Naqa I, Giger ML. A review of explainable and interpretable AI with applications in COVID-19 imaging. Medical Physics. 2022;49(1):1-14.

[32] Tanno R, Arulkumaran K, Alexander DC, Criminisi A, Nori A, editors. Adaptive Neural Trees; 2019.

[33] Samek, W . Meta-Explanations, Interpretable Clustering and Other Recent Developments. Fraunhofer HHI, Machine Learning Group; 2019. https://xai.kaist.ac.kr/static/img/event/ICCV_2019_VXAI_Samek_Talk.pdf.

[34] Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In: KDD'15, August 10-13, 2015, Sydney, NSW, Australia. ACM; 2015. Available from: https://www.microsoft.com.

[35] Ullah H, Rios A, Gala V, McKeever S. Explaining Deep Learning Models for Structured Data using Layer-Wise Relevance Propagation. CoRR. 2020;abs/2011.13429. Available from: https://arxiv.org/abs/2011.13429.

[36] Lundberg S, Lee SI. A unified approach to interpreting model predictions. CoRR. 2017;abs/1705.07874. Available from: textit{http://arxiv.org/abs/1705.07874}.

[37] Lundberg S, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. Explainable AI for Trees: From Local Explanations to Global Understanding. CoRR. 2019;abs/1905.04610. Available from: http://arxiv.org/abs/1905.04610.

Eidesstaatliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Bachelorarbeit mit dem o. a. Thema ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Anhang 1: Liste der Merkmale des Datensatzes

X_1 : age; Alter des Patienten

X_2 : sex; Geschlecht des Patienten, 1: männlich und 0: weiblich

X_3 : cigs; durchschnittliche Anzahl an Zigaretten pro Tag geraucht im Zeitraum seit X_4 (cig_time) Jahren

X_4 : cig_time; seit wie vielen Jahren ist Patient Raucher

X_5 : dm, fbs; es wurde bei dem Patienten bereits Diabetes mellitus diagnostiziert oder mittels Blutzuckermessung nach dem Fasten ein Wert über 120 mg/dl diagnostiziert, ja/nein

X_6 : chol; Blutparameter: Cholesterol im Blut in mg/dl, im Blutserum

X_7 : hist_cad; es gibt basierend auf Fällen in der Verwandtschaft die Vermutung, dass koronare Herzerkrankungen bzw. CAD möglich sind, ja/nein

X_8 : dig; Medikation während exercise ECG (bzw. allgemein Medikation): Digitalis verabreicht, 1: ja und 0: nein

X_9 : prop; Medikation während exercise ECG (bzw. allgemein Medikation): Beta blocker verabreicht, 1: ja und 0: nein

X_{10} : nitr; Medikation während exercise ECG (bzw. allgemein Medikation): Nitrate verabreicht, 1: ja und 0: nein

X_{11} : pro; Medikation während exercise ECG (bzw. allgemein Medikation): Kalziumkanalblocker verabreicht, 1: ja und 0: nein

X_{12} : diuretic; Medikation während exercise ECG (bzw. allgemein Medikation): Diuretika verabreicht, 1: ja und 0: nein

X_{13} : met; Metabolic Equivalents (MET) erreicht während des exercise-tests, kann als Maß der Leistungsfähigkeit angesehen werden

X_{14} : restecg_0; ECG-Ergebnisse ohne sportliche Betätigung, ECG Zeitreihe klassifiziert als normal condition 1: ja und 0: nein

X_{15} : restecg_1; ECG-Ergebnisse ohne sportliche Betätigung, abnormality in the ST-T wave (T-wave inversions or ST depression $>$ 0,05 mV or both)

X_{16} : restecg_2; ECG-Ergebnisse ohne sportliche Betätigung, possibility or certainty of LV (left ventricular) Hypertrophie per Estes' criteria 1: ja und 0: nein

X_{17} : thalach; maximale Herzrate in 1/min, während exercise ECG

X_{18} : thalrest; Herzrate in 1/min, im Ruhezustand bei Einlieferung

X_{19} : tpeakbps; systolischer Blutdruck in mmHg, Maximalwert (vermutlich während peak exercise, während exercise test)

X_{20} : tpeakbpd; diastolischer Blutdruck in mmHg, Maximalwert (vermutlich während peak exercise, während exercise test)

X_{21} : trestbpd; diastolischer Blutdruck in mmHg, in Ruhe

X_{22} : oldpeak; ECG-Merkmal: exercise-induced ST-segment depression, verglichen zum Ruhe-ECG Senkung um den gegebenen Wert

X_{23} : slope; ECG-Merkmal: Steigung der ECG-Kurve of the peak exercise element, ST segment 1: upsloping, 2: flat, 3: downsloping

X_{24} : rldv5e; exercise-induced R-wave change

X_{25} : ca; Anzahl an major Vessel, die Calcium enthalten, durch Fluoroskopie-Bildgebung festgestellt

X_{26} : thal; exercise thallium scintigraphy 3 = none, normal, 6 = fixed defect, 7 = reversable defect

X_{27} : trestbps; Blutdruck in mmHg bei Einlieferung in Klinik, ohne Belastung systolisch

X_{28} : htn; bei dem Patienten wurde Bluthochdruck (hpyertension, htn) diagnostiziert

X_{29} : thaldur; Dauer des Exercise-Tests in Minuten

X_{30} : thaltime; Zeit von Beginn des exercise ECG bis es zu bestimmtem Ereignis (vermutlich ST depression) kommt, in Minuten

X_{31} : xhypo; hypotension, während exercise test change

X_{32} : rldv5; R-wave change

X_{33} : cathef; Ejektionsfraktion des Herzens, Blutvolumen, das aus dem Herzen gepumpt wird vgl. zu dem, was ins Herz gepumpt wird, in Prozent, über Herzkatheter, obwohl es auch anders gemessen werden kann

Y_3 : lmt; angiographic variable: lmt als blockiert diagnostiziert, ja/nein

Y_4 : ladprox; angiographic variable: lad proximal als blockiert diagnostiziert, ja/nein

Y_5 : laddist; angiographic variable: lad distal als blockiert diagnostiziert, ja/nein

Y_6 : cxmain; angiographic variable: cxmain als blockiert diagnostiziert, ja/nein

Y_7 : om1; angiographic variable: om1 als blockiert diagnostiziert, ja/nein

Y_8 : rcaprox; angiographic variable: rca proximal als blockiert diagnostiziert, ja/nein

Y_9 : rcadist; angiographic variable: rca distal als blockiert diagnostiziert, ja/nein

Y_{10} : om2; angiographic variable: om2 als blockiert diagnostiziert, ja/nein

Y_{11} : ramus; angiographic variable: ramus als blockiert diagnostiziert, ja/nein

Y_1 : target_sev; angiographic variable: Indikator aus der Angiographie, wieviele Vessel als blockiert diagnostiziert wurden, 1,2,3,4: gibt die Anzahl an und 0: nein

Y_2 : target; Indikator, ob bei dem Patienten eine SVD oder MVD vorhanden ist, abgeleitet „target_sev>0“, 1: ja und 0: nein

hist_cad: vermutlich basierend auf Angaben des Patienten bei einem Interview

chol: entspricht dem Gesamtcholesterinwert als Summe der Werte für LDL Cholesterin und HDL Cholesterin

zu Zielvariablen: als CAD diagnostiziert, wenn Veränderung des Durchmessers von mehr als 50 %

Anhang 2: Liste der bei der Datenvorvereitung entfernten Untersuchungseinheiten, basierend auf der PatientenId der Ersteller des Datensatzes

Cleveland: 21, 23, 87, 121, 159, 166, 182, 192, 195, 207, 250, 266

VA Medical: 13, 23, 26, 29, 30, 33, 36, 38, 42, 44, 48, 51, 57, 58, 60, 62, 68, 69, 72, 77, 87, 89, 93, 94, 95, 96, 109, 110, 111, 112, 113, 114, 115, 119, 120, 124, 127, 129, 132, 133, 134, 135, 136, 137, 140, 141, 142, 144, 145, 147, 149, 155, 158, 159, 160, 161, 163, 164, 167, 168, 169, 174, 176, 181, 183, 185, 188, 193, 196, 198, 199

Hungarian: 1, 4, 28, 30, 31, 38, 40, 60, 61, 75, 80, 85, 90, 103, 123, 132, 137, 138, 150, 151, 154, 156, 157, 163, 165, 172, 178, 183, 186, 188, 194, 197, 214, 220, 225, 227, 233, 235, 238, 264, 270, 275, 278, 287, 288, 289, 291

Switzerland: 1, 3, 4, 13, 15, 20, 30, 35, 41, 43, 53, 60, 65, 66, 73, 79, 81, 84, 85, 87, 89, 92, 93, 100, 103, 107, 109, 112, 114, 121

Anhang 3: Liste der Merkmale der Datensätze

Cleveland: X_1 bis X_31

Hungarian: X_1, X_2, X_5, X_6, X_8 bis X_24, X_27 bis X_32

fehlende Merkmale Hungarian (Rauchen und kardiologisch): cigs, cig_time, hist_cad, ca, thal

doctor: X_1 bis X_24, X_27 bis X_31

fehlende Merkmale doctor (kardiologisch): ca, thal

doctor 2: X_1 bis X_24, X_27 bis X_31

fehlende Merkmale doctor 2 (kardiologisch): ca, thal

doctor 3: X_1, X_2, X_5, X_6, X_8 bis X_24, X_27 bis X_31

fehlende Merkmale doctor 3 (Rauchen und kardiologisch): cigs, cig_time, hist_cad, ca, thal

doctor 4: X_1 bis X_13, X_17 bis X_21, X_27, X_28, X_31

fehlende Merkmale doctor 4 (ECG und kardiologisch): restecg_0, restecg_1, restecg_2, oldpeak, slope, rldv5e, ca, thal, thaldur, thaltime

Anhang 4: Verwendete Algorithmen

Modell 1:

Implementierung: sklearn.neighbors.KNeighborsClassifier

Parameter: n_neighbors, metric

n_neighbors=8

metric=„euclidean“

Modell 2, für alle Datensätze:

Implementierung: sklearn.tree.DecisionTreeClassifier

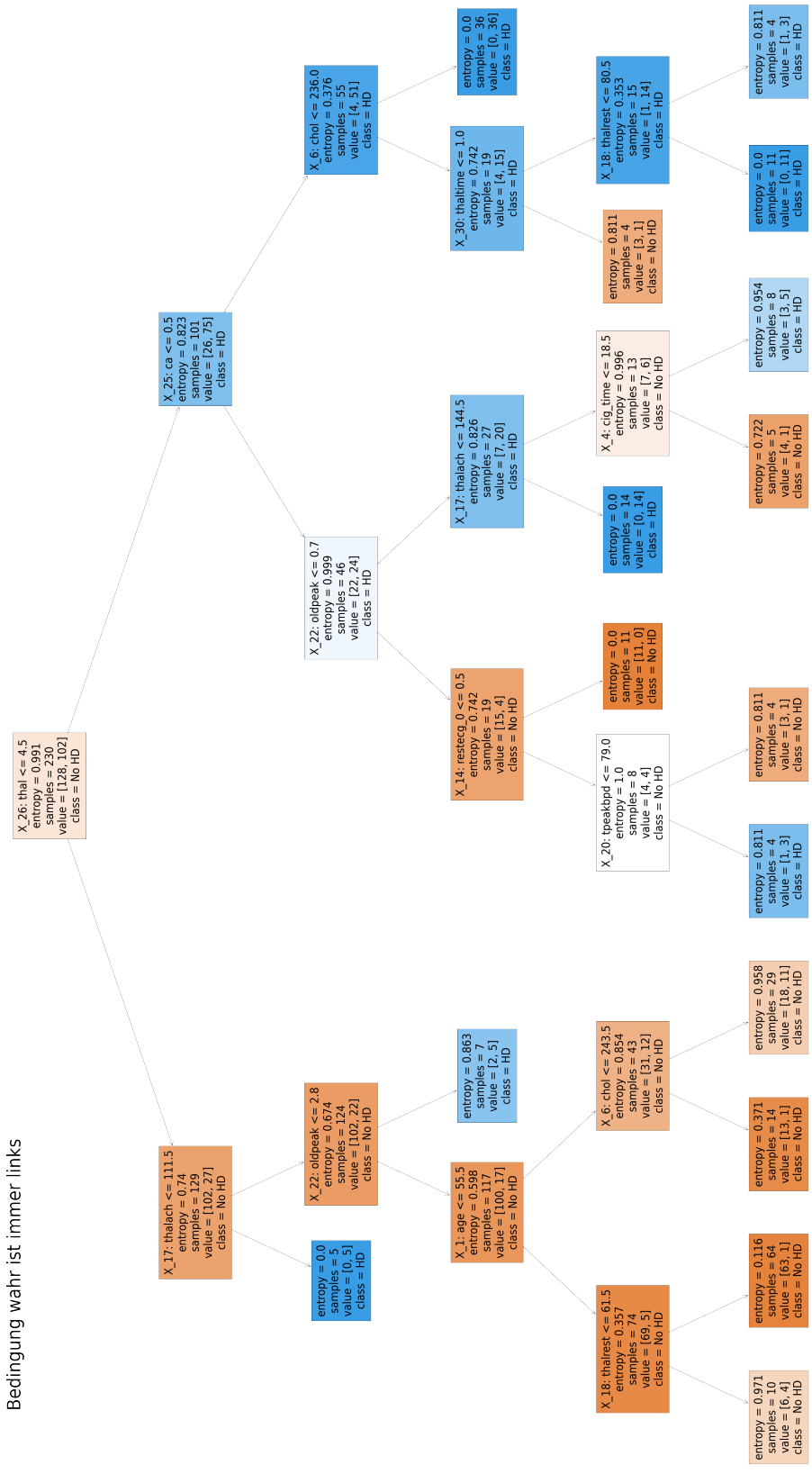
Parameter: criterion, max_depth, splitter, min_samples_split, min_samples_leaf, max_leaf_nodes, min_impurity_decrease

criterion=„entropy“, max_depth=5, splitter=„best“, min_samples_split=5, min_samples_leaf=4, max_leaf_nodes=20, min_impurity_decrease=0.003

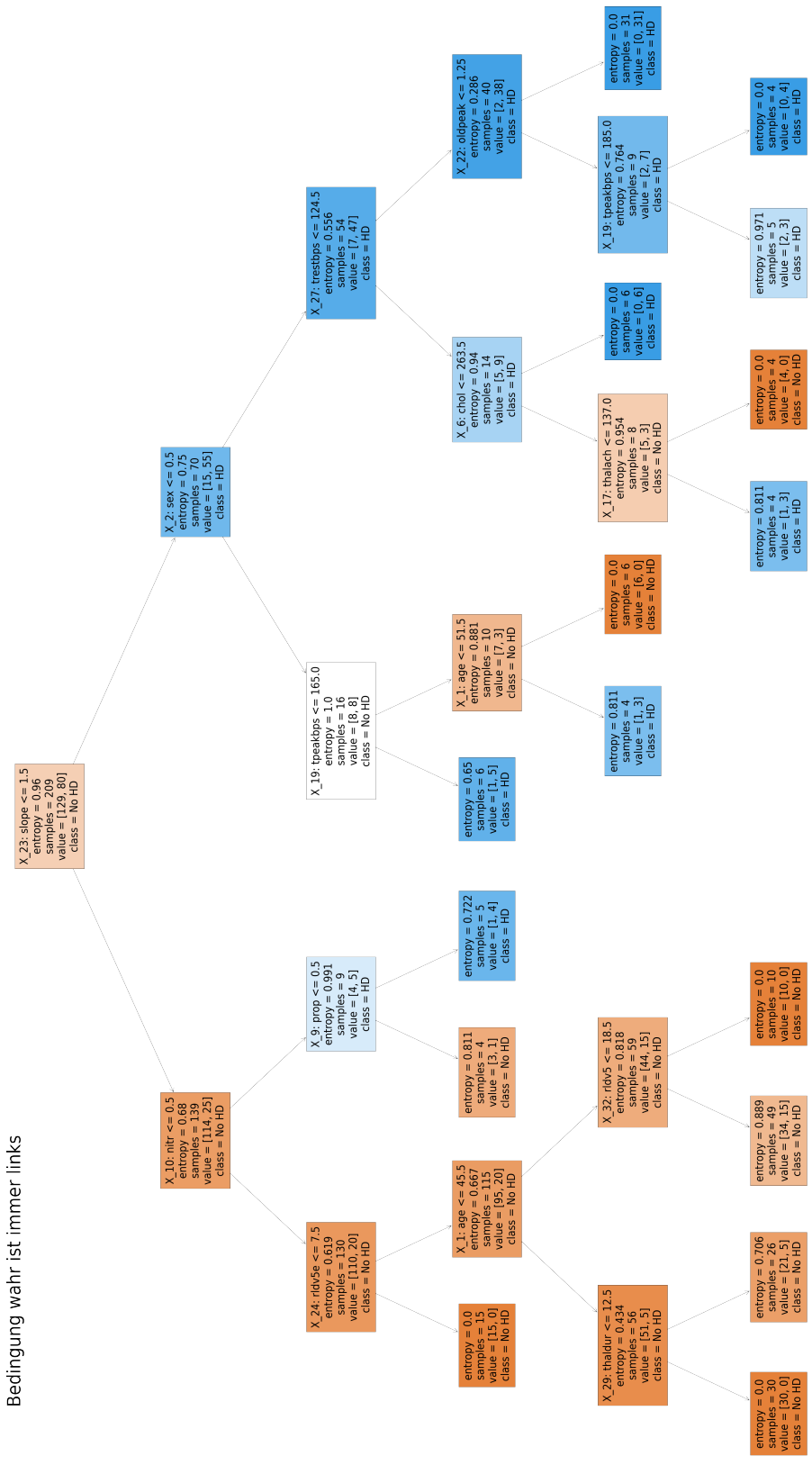
Modell 3, für alle Datensätze:

Implementierung: sklearn.ensemble.ExtraTreesClassifier

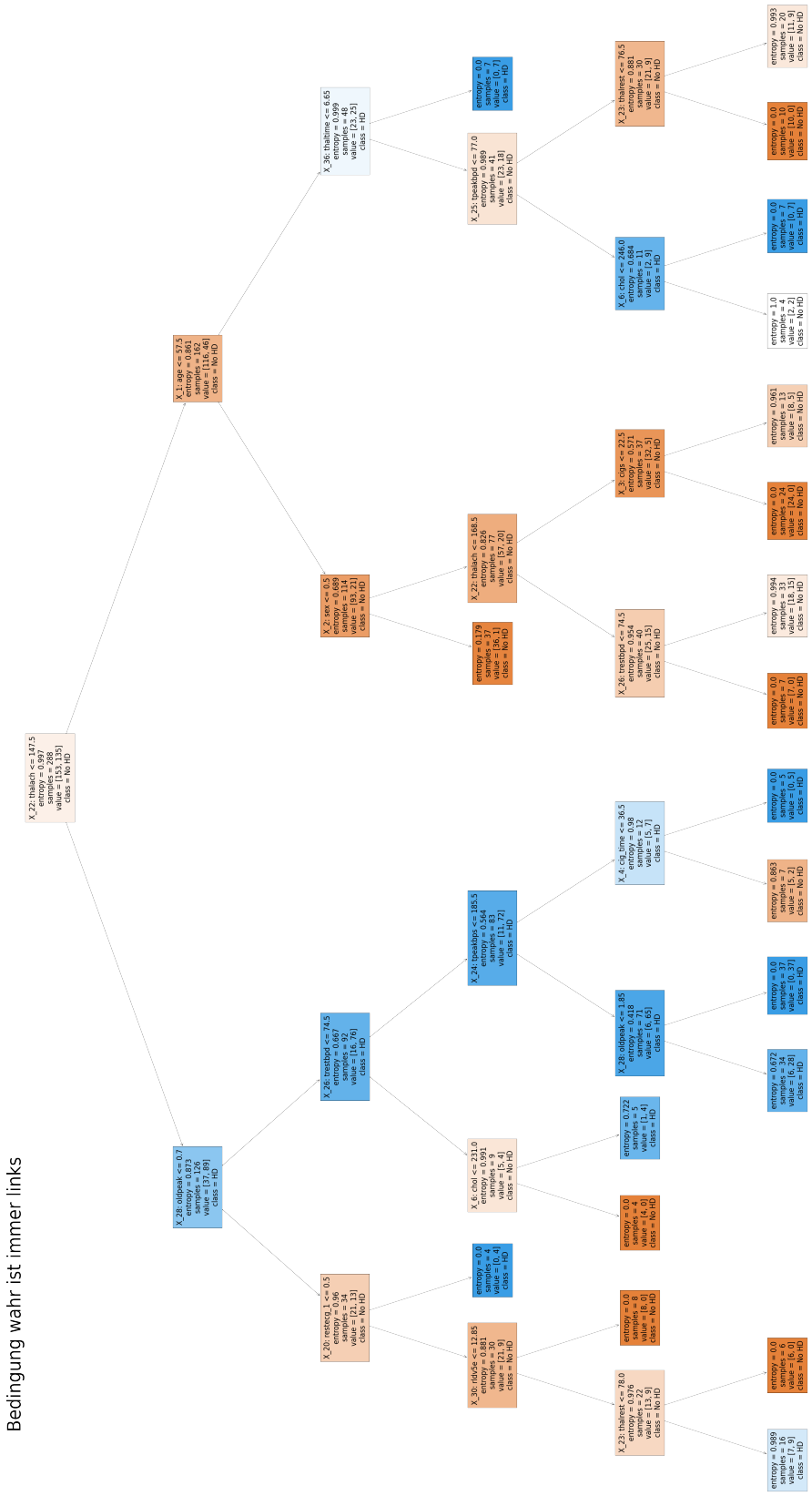
Anhang 5, Modell 2: Darstellung Entscheidungsbaum aus Cleveland
$CA_m = 0{,}883$ und $CA_{cv} = 0{,}684$



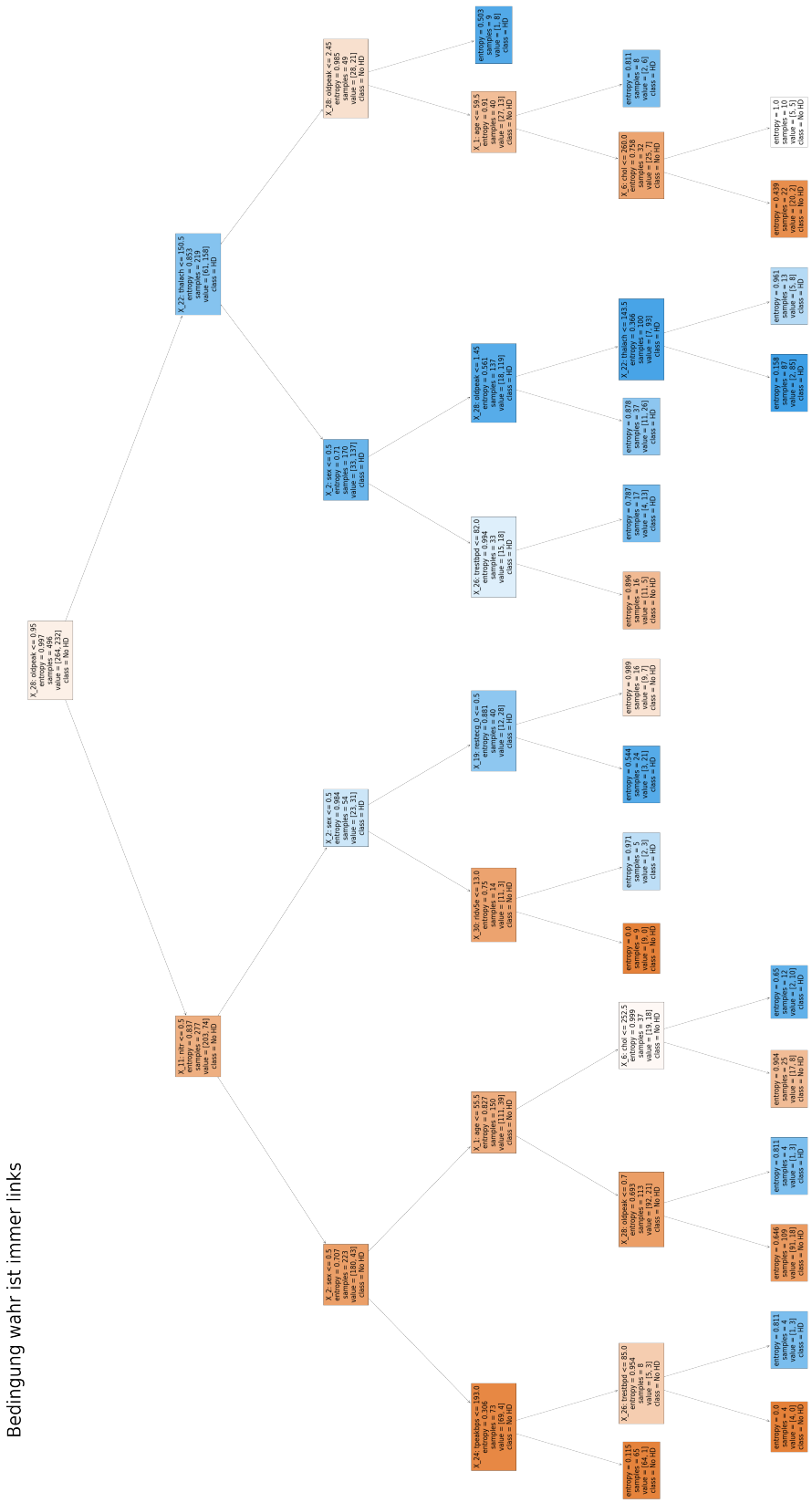
Anhang 5, Modell 2: Darstellung Entscheidungsbaum aus Hungarian
$CA_m = 0{,}871$ und $CA_{cv} = 0{,}941$



Anhang 5, Modell 2: Darstellung Entscheidungsbaum aus doctor
$CA_m = 0{,}833$ und $CA_{cv} = 0{,}745$

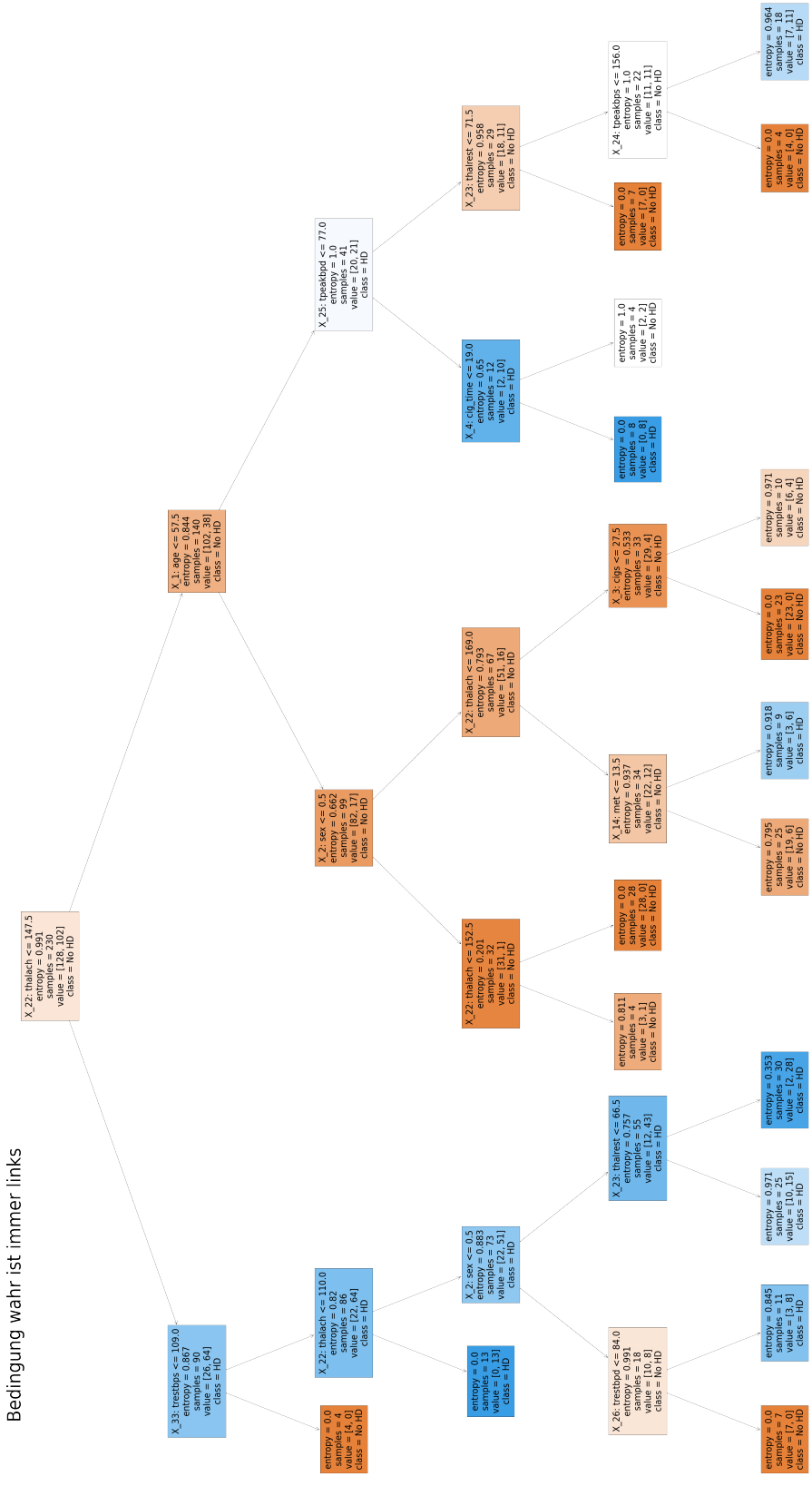


Anhang 5, Modell 2: Darstellung Entscheidungsbaum aus doctor 3
$CA_m = 0{,}839$ und $CA_{cv} = 0{,}817$

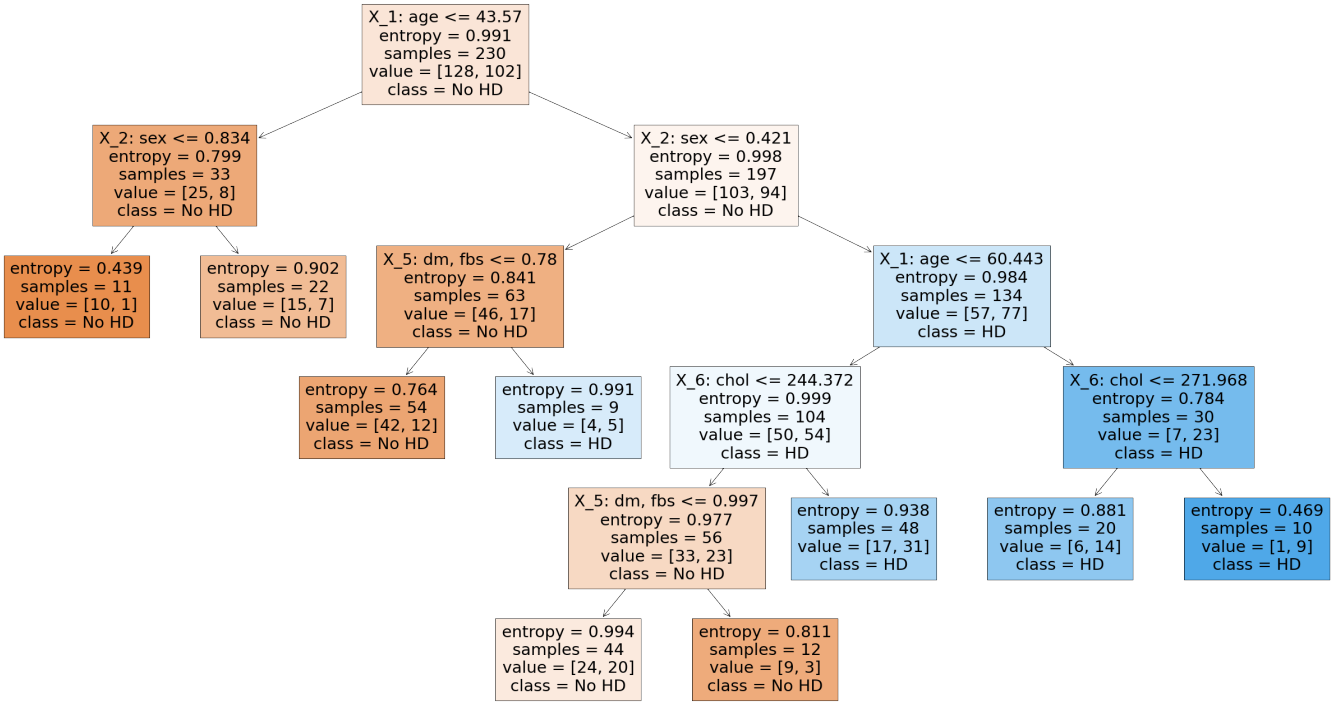


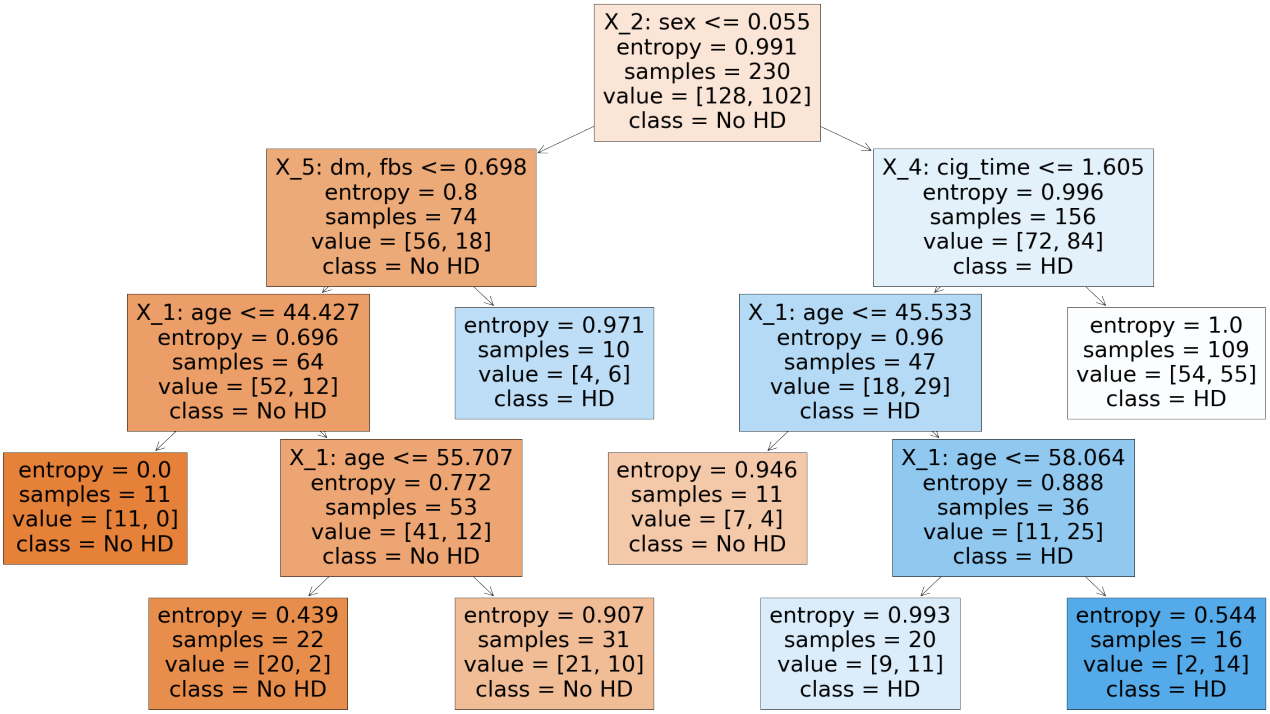
Bedingung wahr ist immer links

Anhang 5, Modell 2: Darstellung Entscheidungsbaum aus doctor 4
$CA_m = 0{,}835$ und $CA_{cv} = 0{,}632$

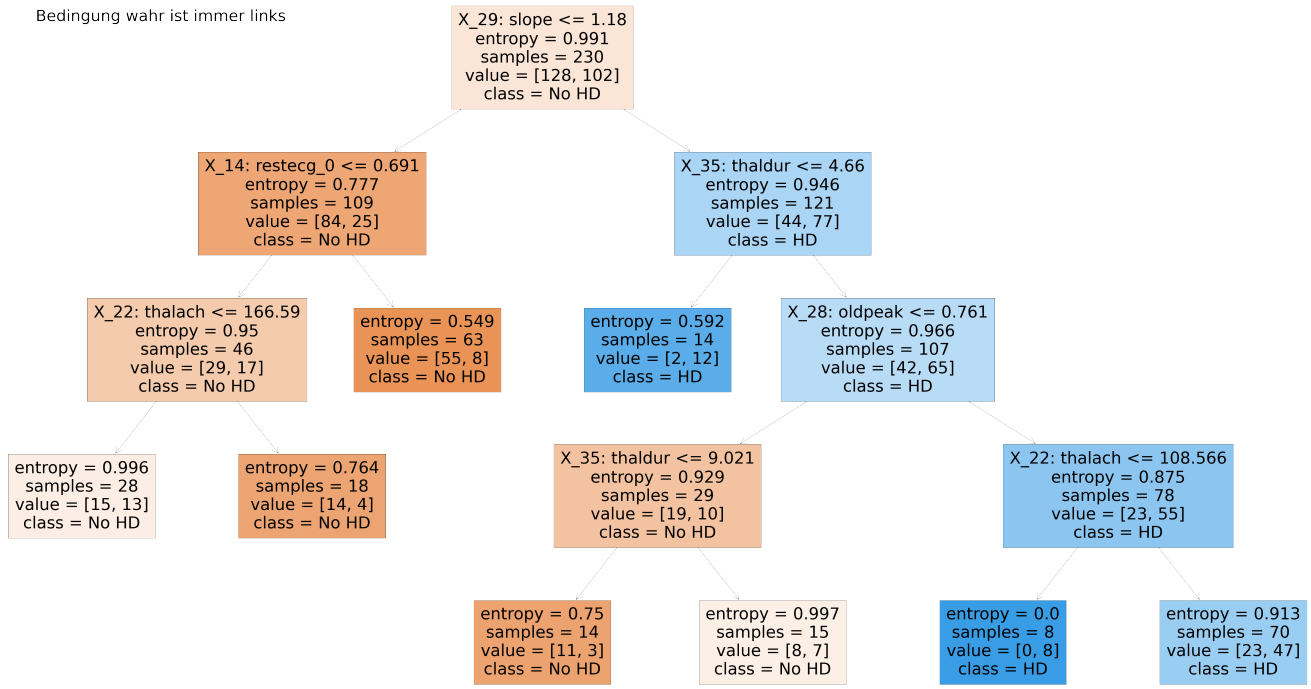


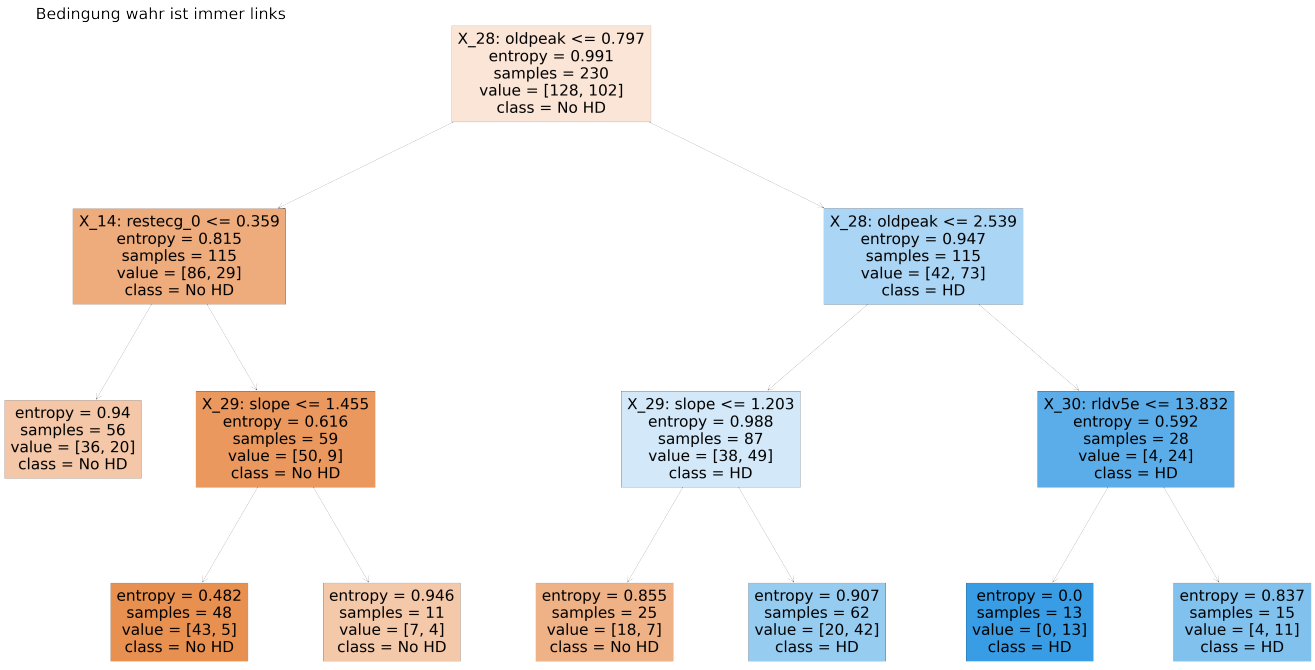
Anhang 6, Modell 3: Darstellung Entscheidungsbäume Cleveland, Datenview 1





Anhang 6, Modell 3: Darstellung Entscheidungsbäume Cleveland, Datenview 2





Anhang 7, Modell 2: Shapley Scatterplots Cleveland

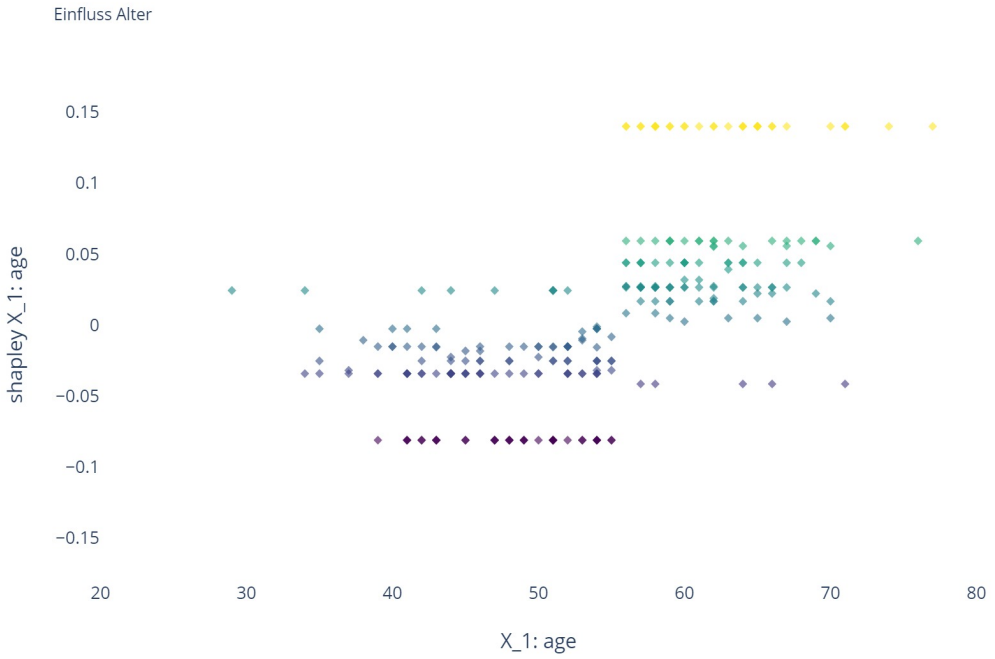


Abb. A.1: Einfluss Alter

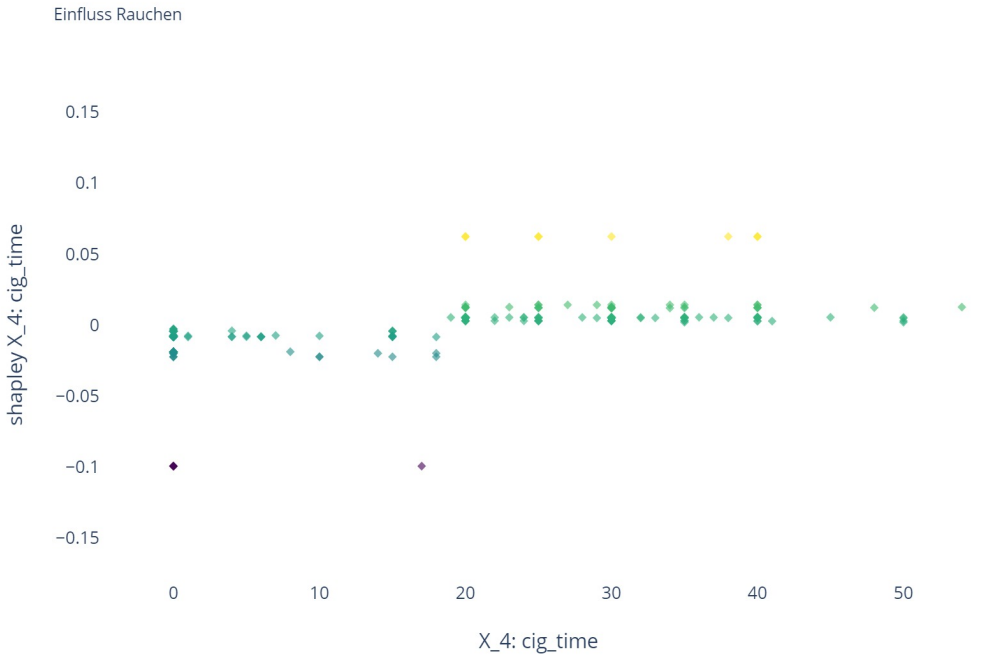


Abb. A.2: Einfluss Rauchen

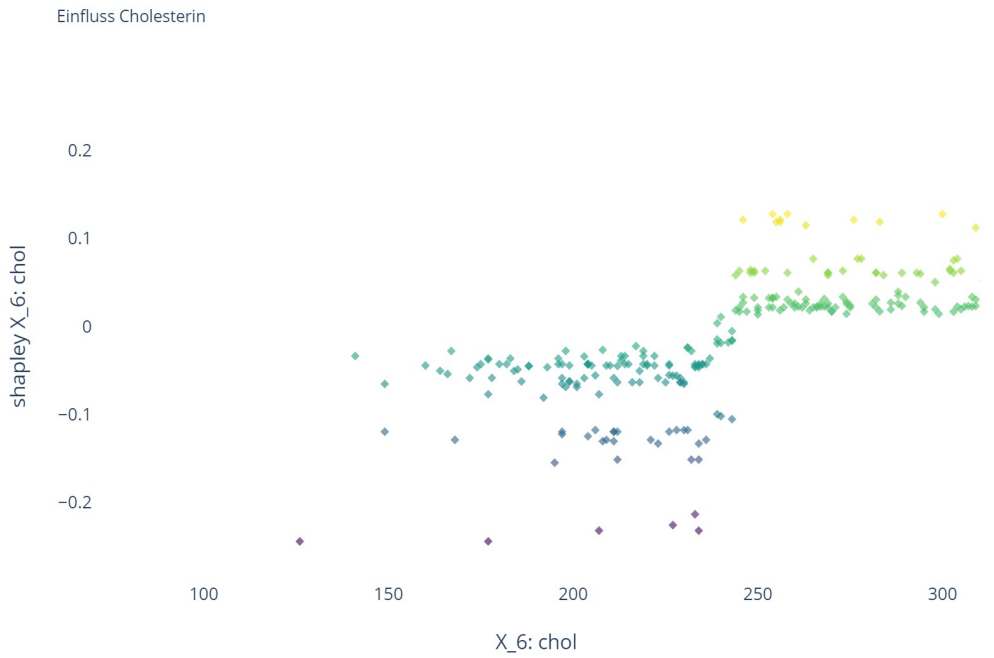


Abb. A.3: Einfluss Cholesterin in mg/dl

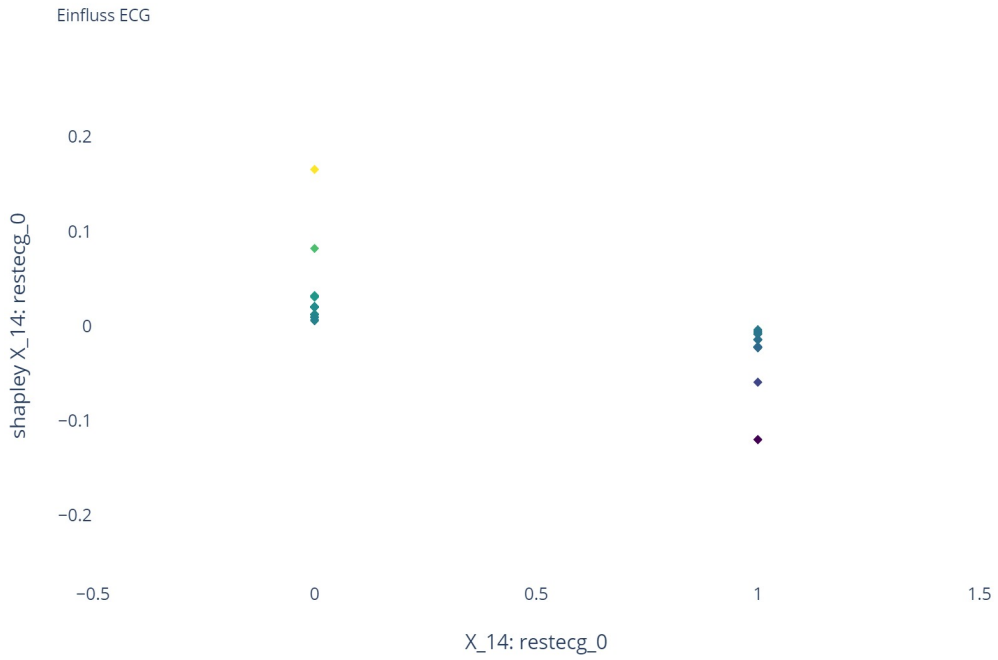


Abb. A.4: Einfluss ECG, normal condition 1: ja und 0: nein

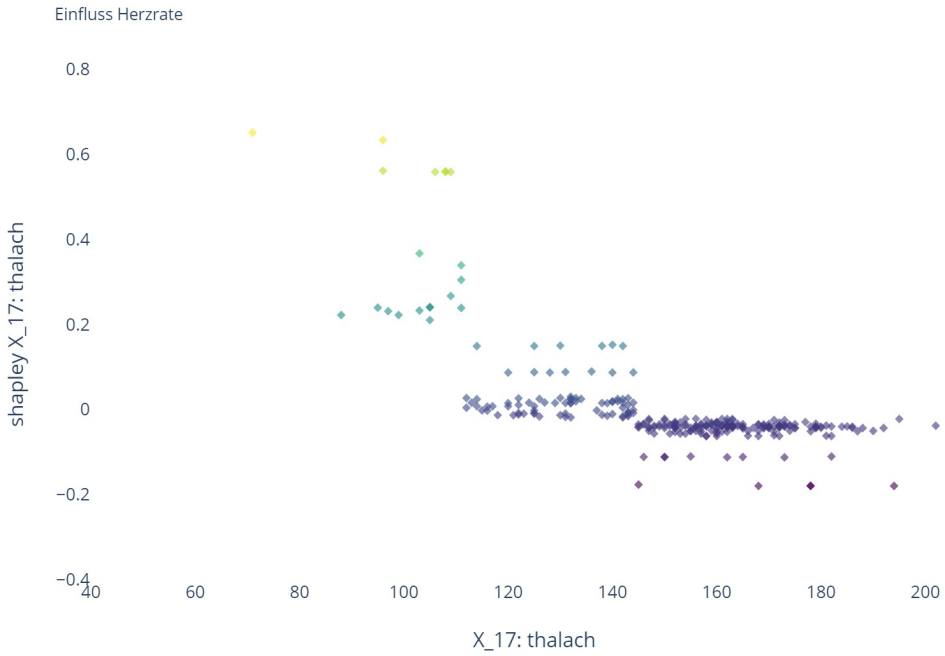


Abb. A.5: Einfluss Herzrate in 1/min

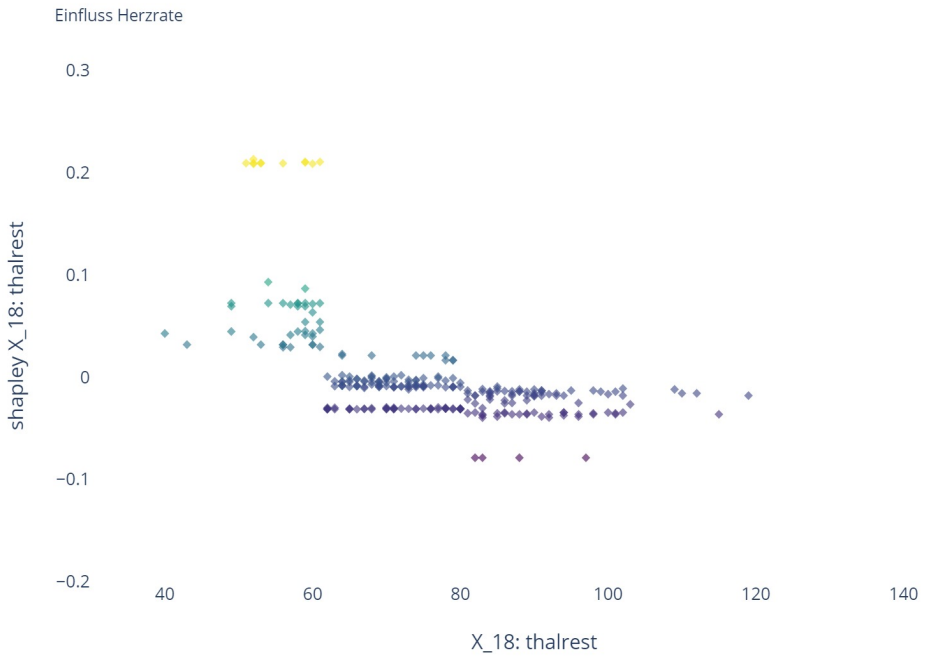


Abb. A.6: Einfluss Herzrate in Ruhe in 1/min

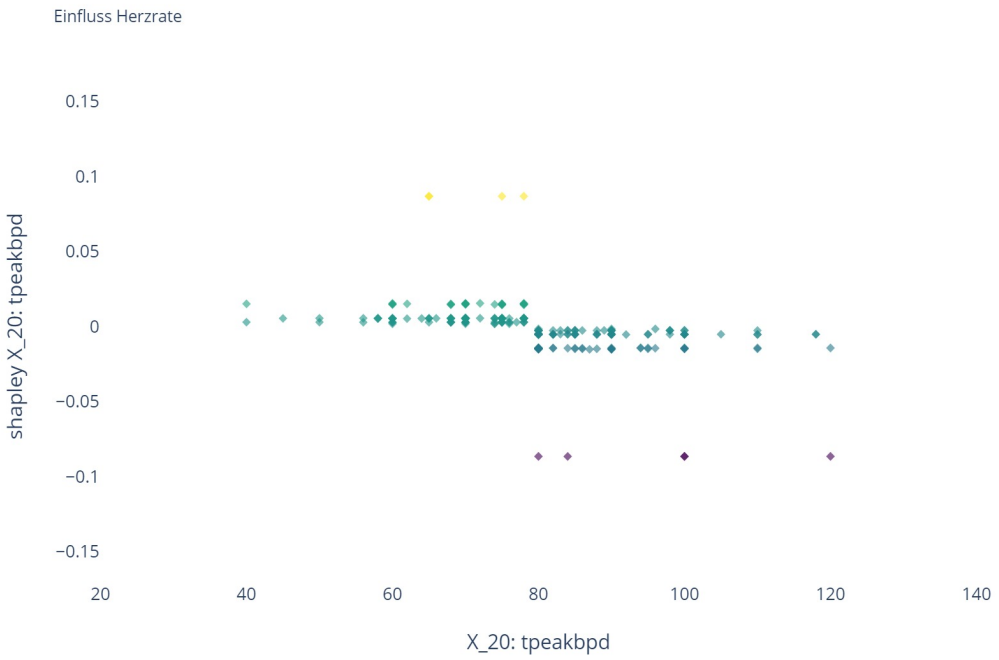


Abb. A.7: Einfluss Kreislauf in mmHg

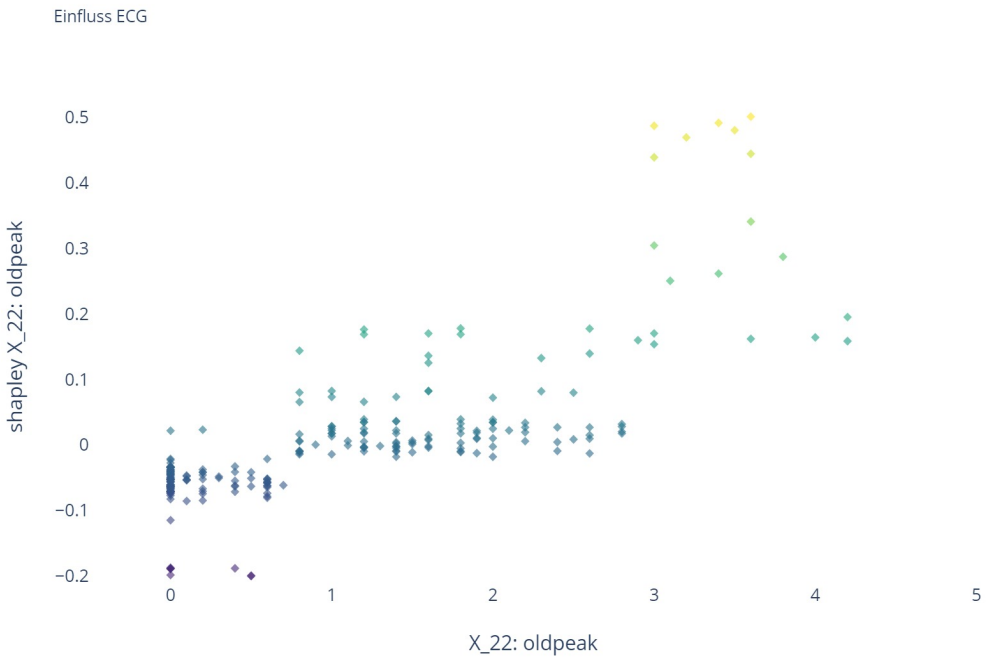


Abb. A.8: Einfluss ECG

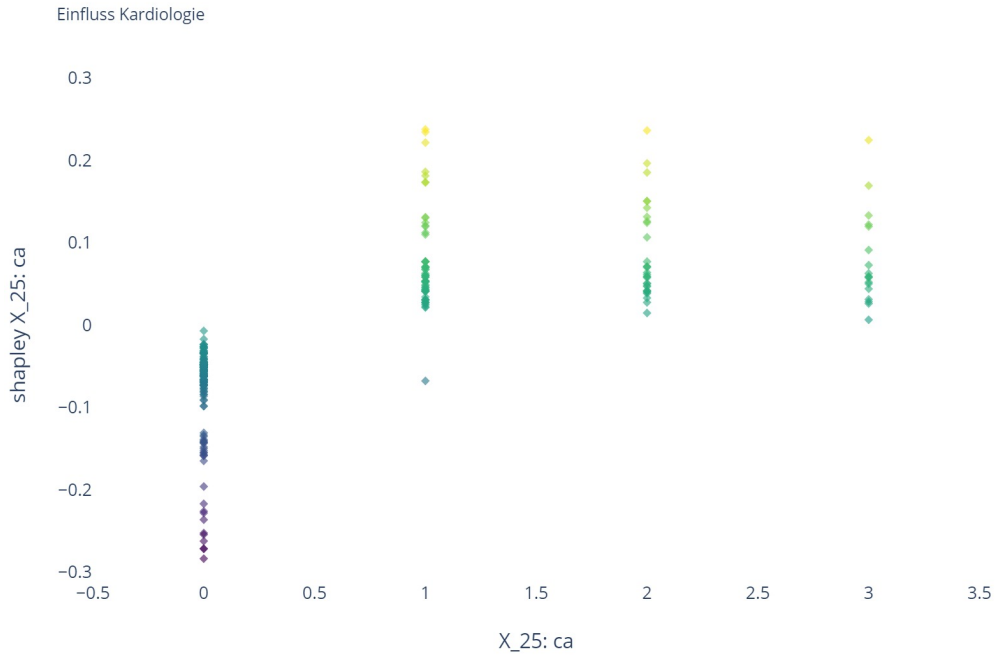


Abb. A.9: Einfluss Kardiologie

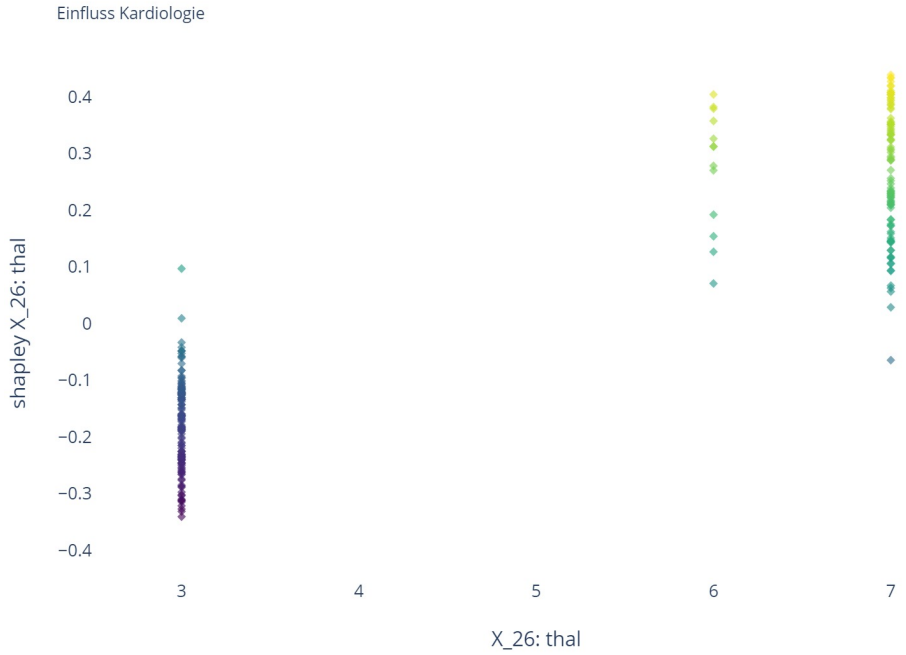


Abb. A.10: Einfluss Kardiologie, exercise thallium scintigraphy 3 = none, normal, 6 = fixed defect, 7 = reversable defect

Anhang 8, Modell 3: Shapley Scatterplots Cleveland, $CA_m = 0{,}863$ und $CA_{cv} = 0{,}825$

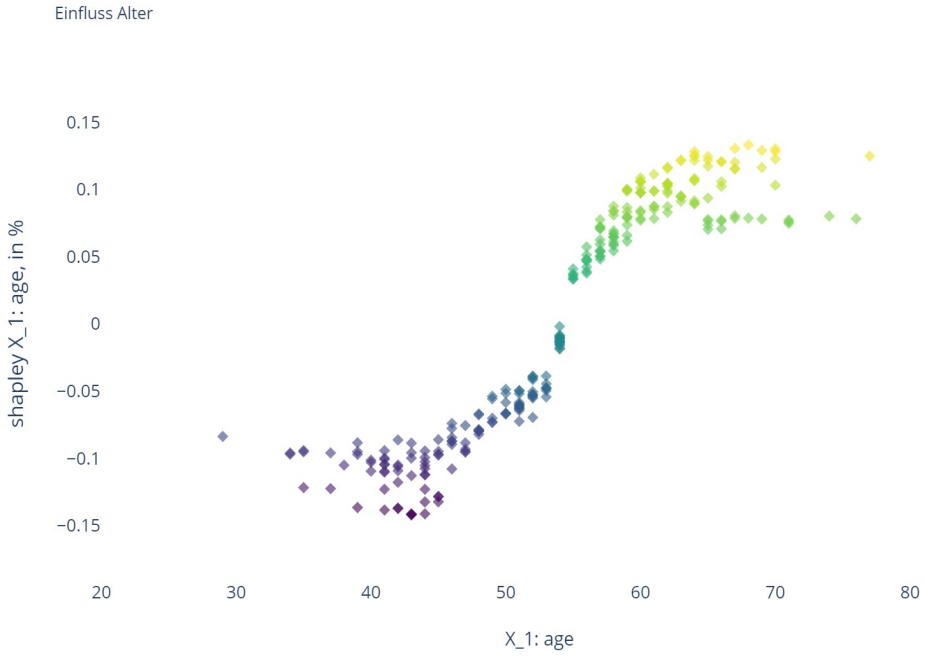


Abb. A.11: Einfluss Alter

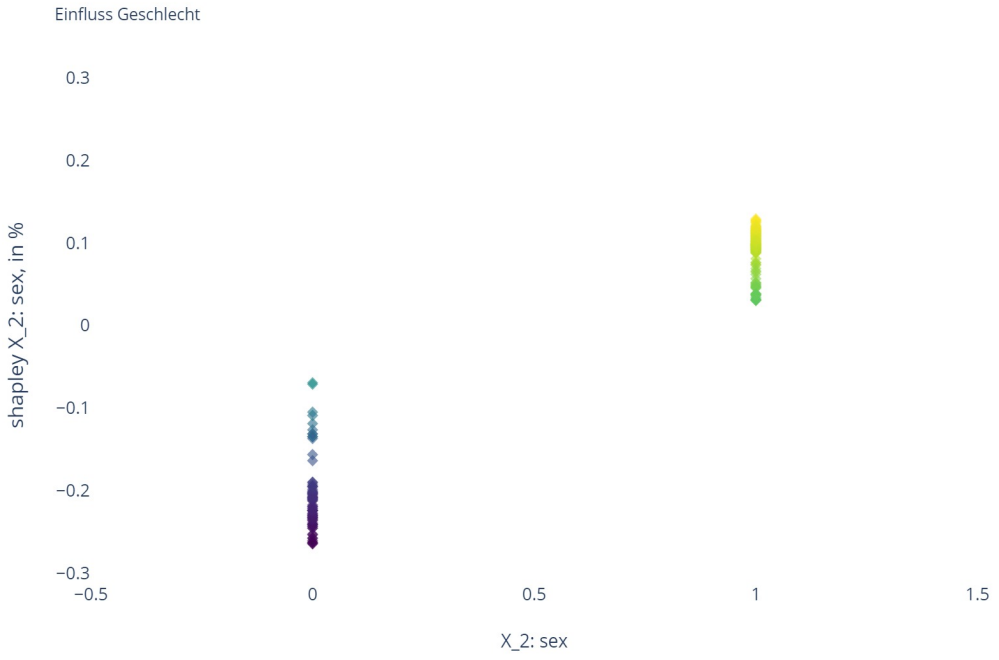


Abb. A.12: Einfluss Geschlecht, 1: männlich und 0: weiblich

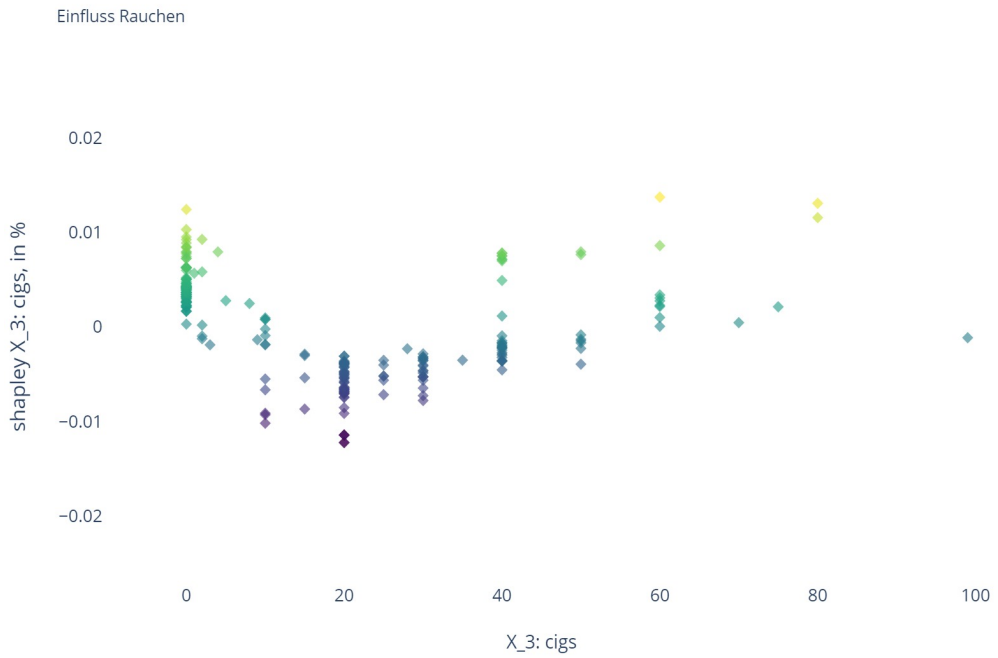


Abb. A.13: Anzahl Zigaretten geraucht pro Tag

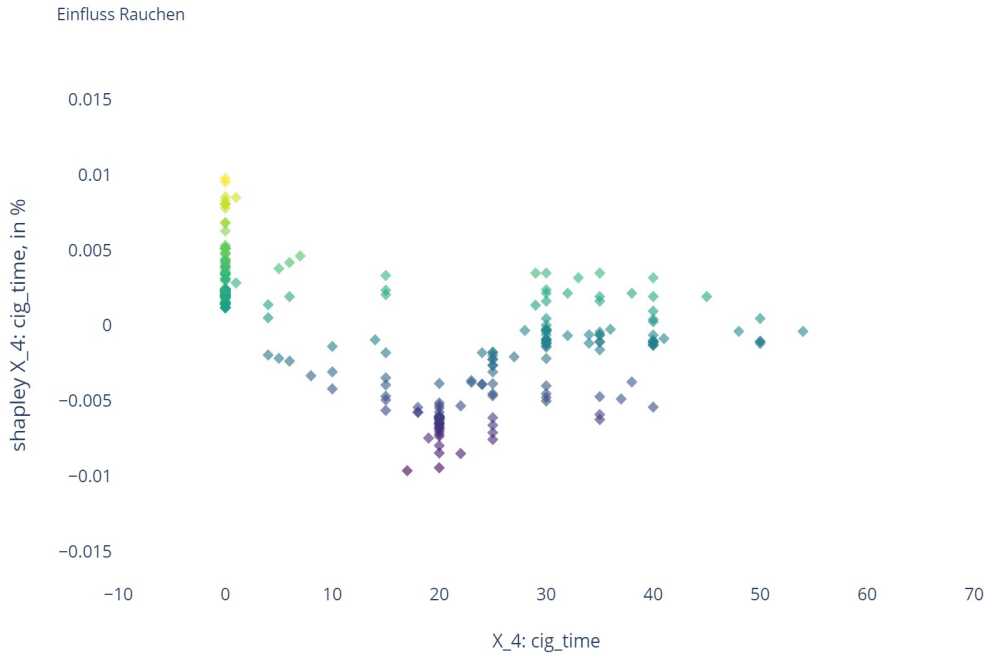


Abb. A.14: Einfluss Dauer des Rauchen

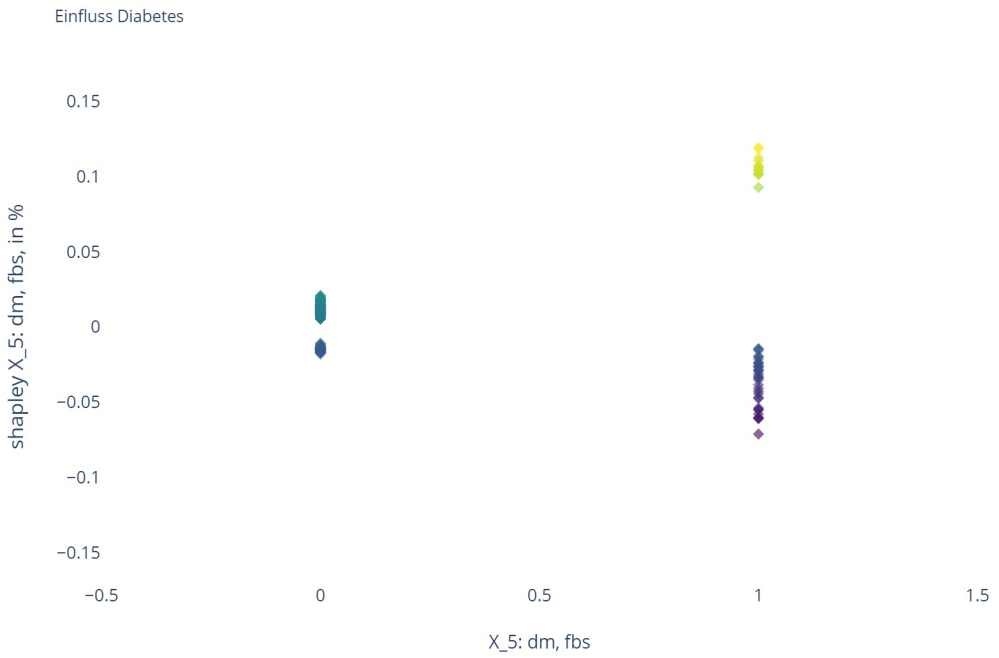


Abb. A.15: Diabetes diagnostiziert, 1: ja und 0: nein

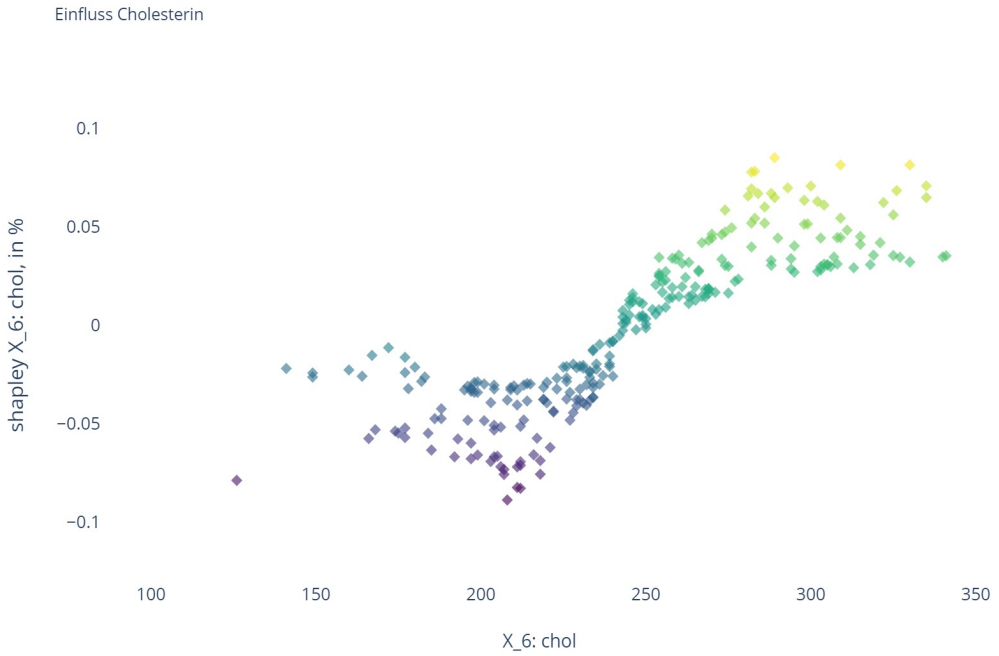


Abb. A.16: Einfluss Cholesterin in mg/dl

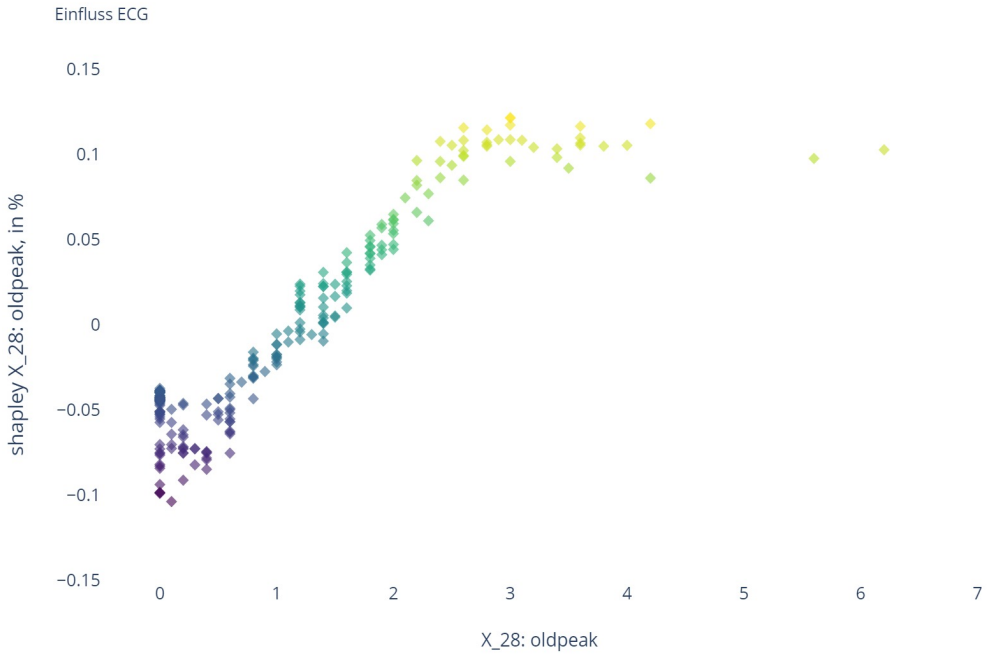


Abb. A.17: Einfluss ECG

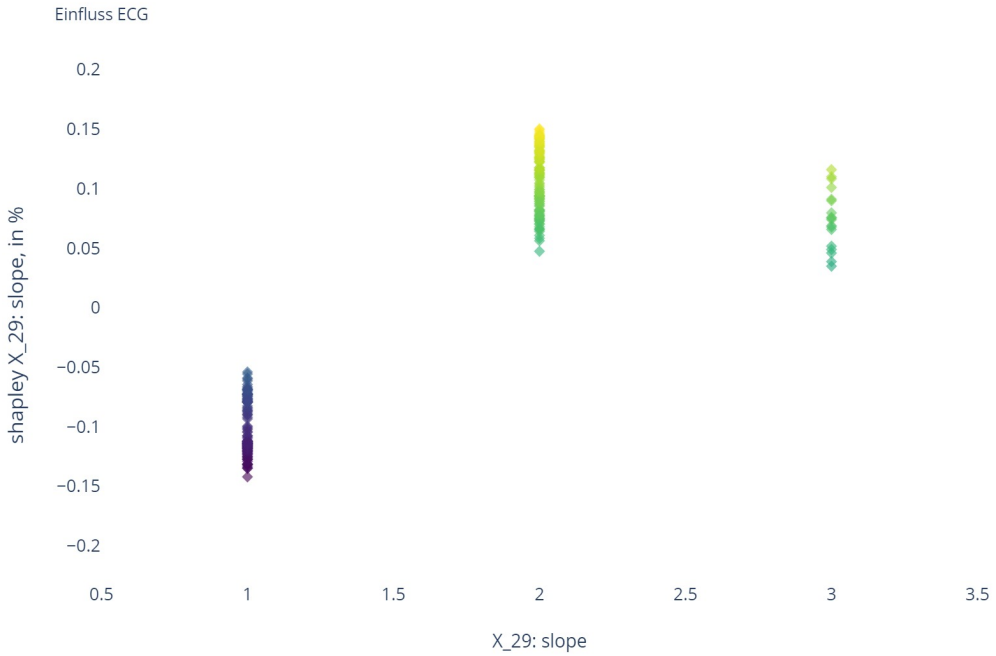


Abb. A.18: Einfluss Steigung ST segment 1: upsloping, 2: flat, 3: downsloping

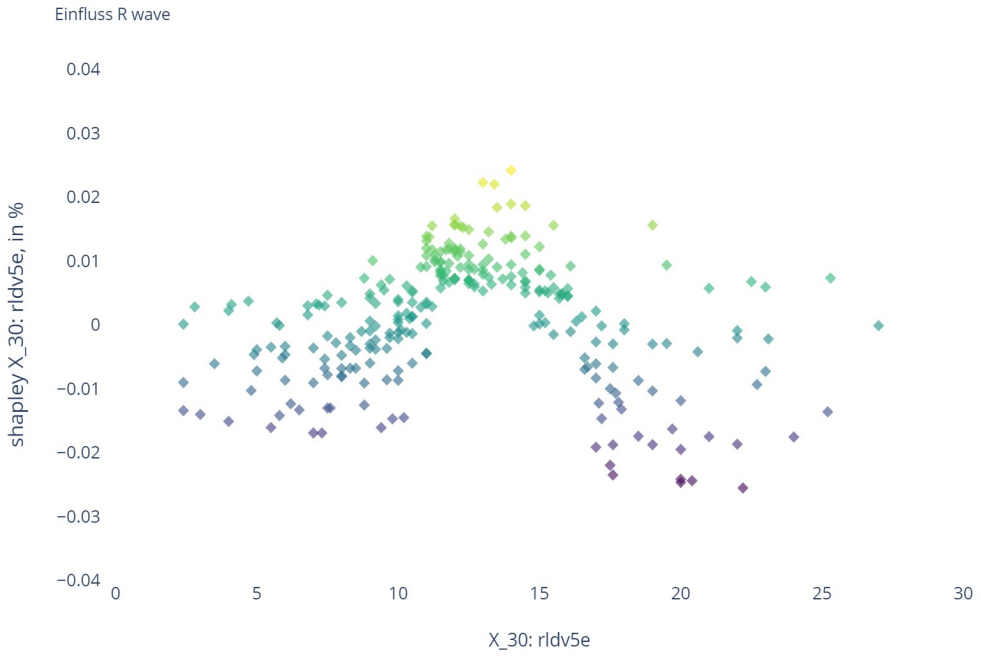


Abb. A.19: Einfluss R wave

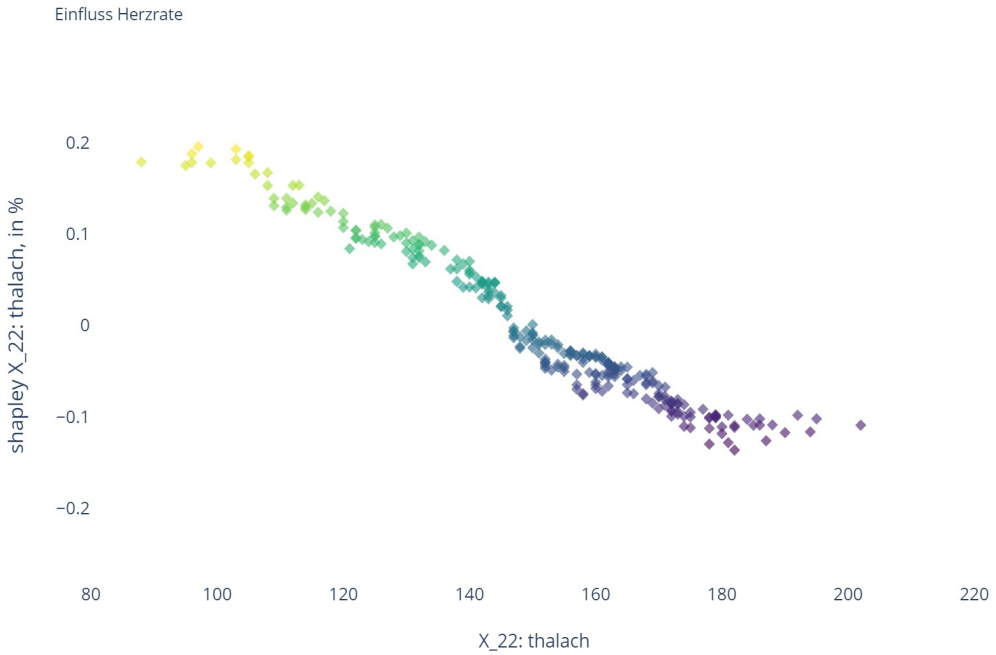


Abb. A.20: Einfluss Herzrate in 1/min

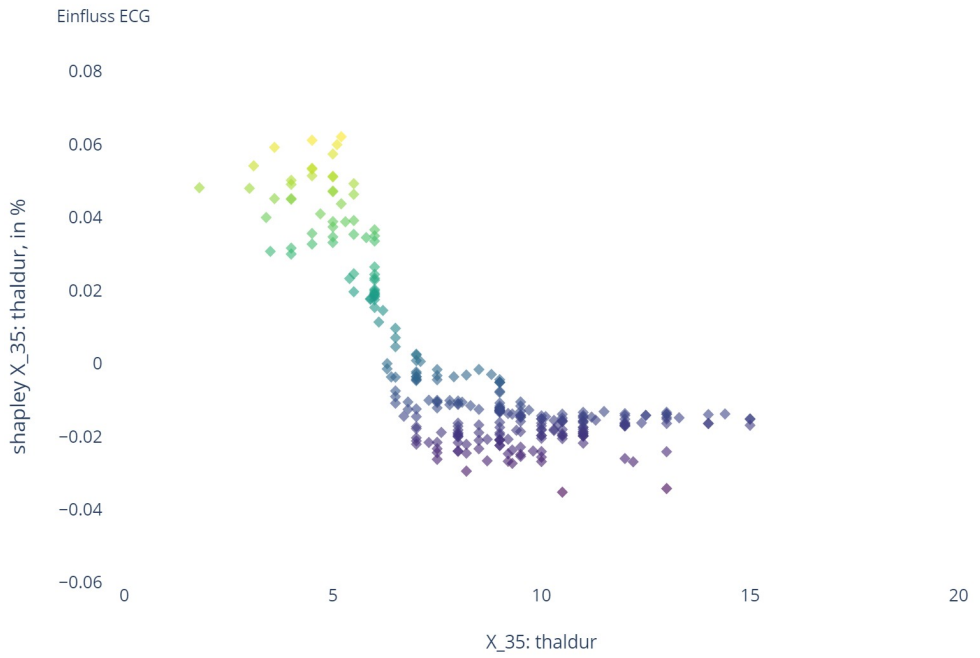


Abb. A.21: Einfluss ECG

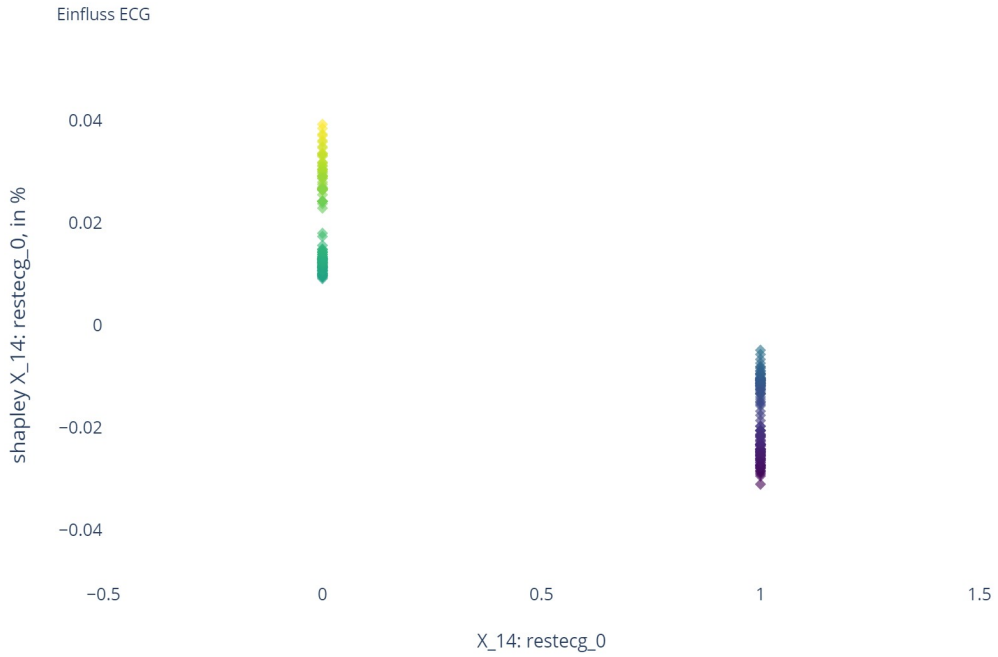


Abb. A.22: Einfluss ECG, normal condition 1: ja und 0: nein

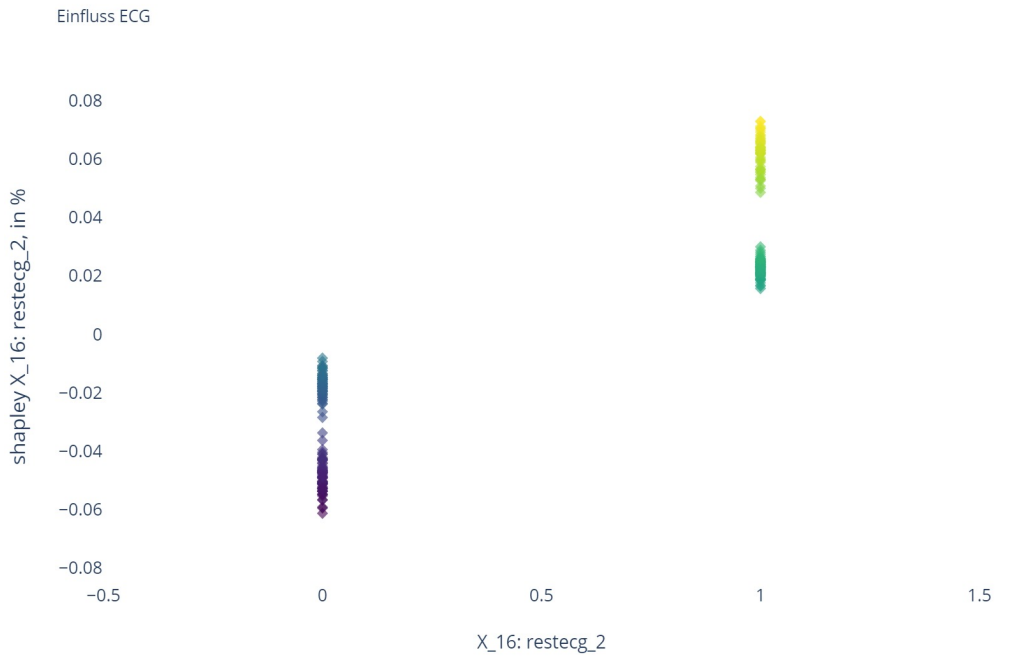


Abb. A.23: possibility or certainty of LV (left ventricular) Hypertrophie per Estes' criteria 1: ja und 0: nein

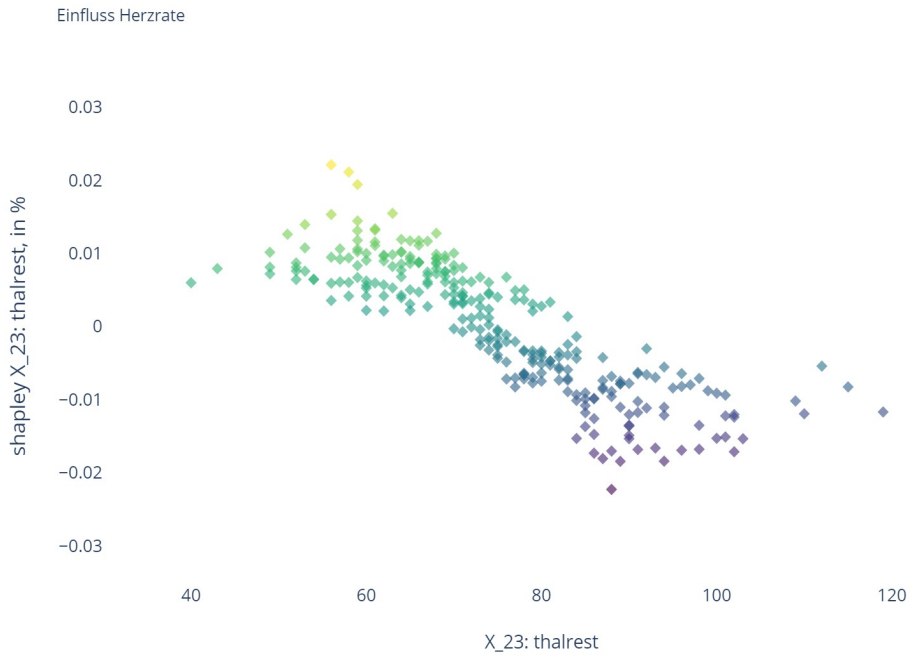


Abb. A.24: Einfluss Herzrate in Ruhe in 1/min

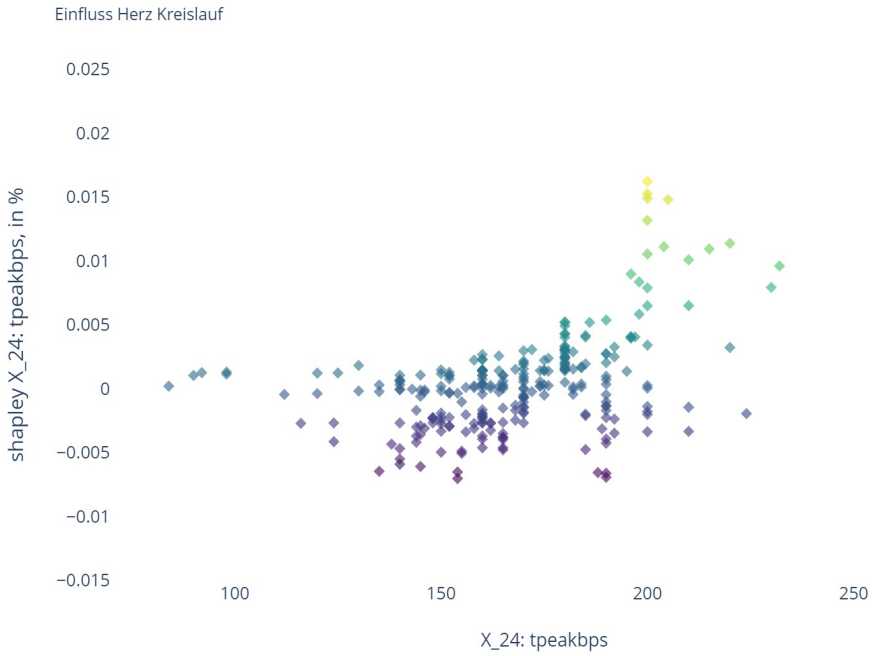


Abb. A.25: Einfluss Kreislauf in mmHg

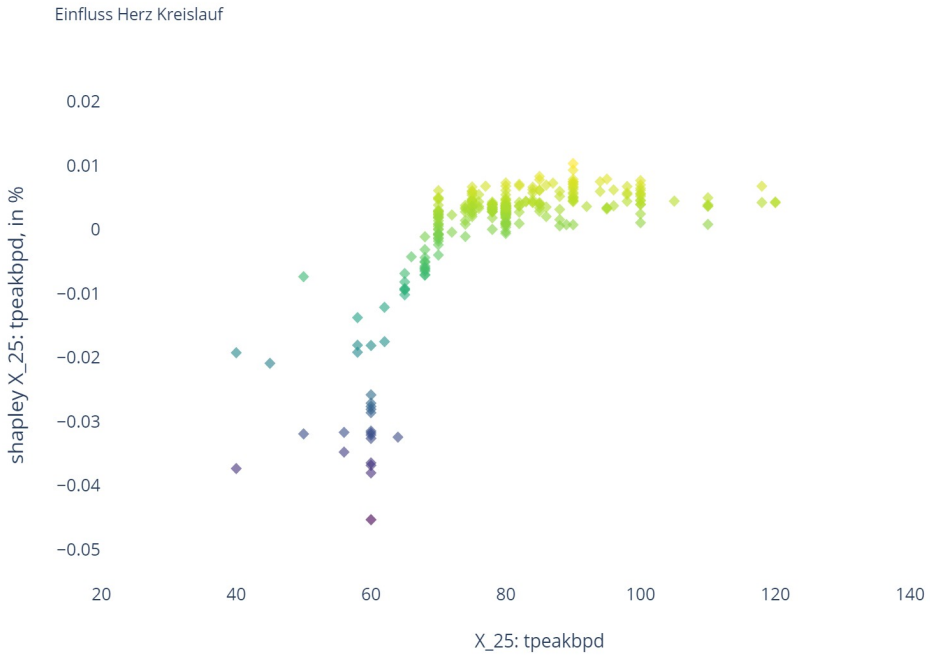


Abb. A.26: Einfluss Kreislauf in mmHg

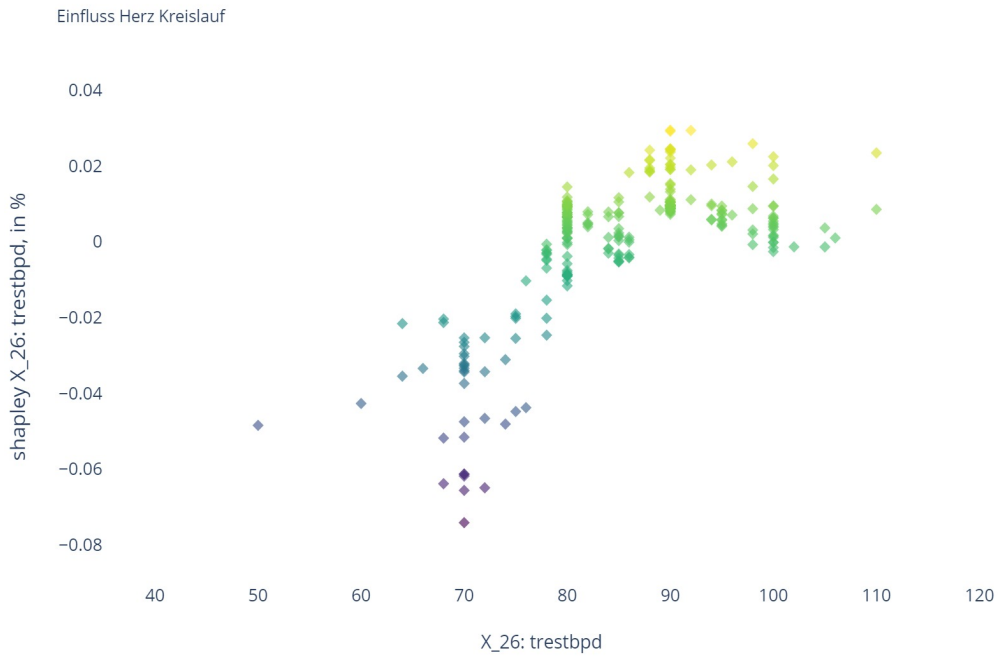


Abb. A.27: Einfluss Kreislauf in mmHg

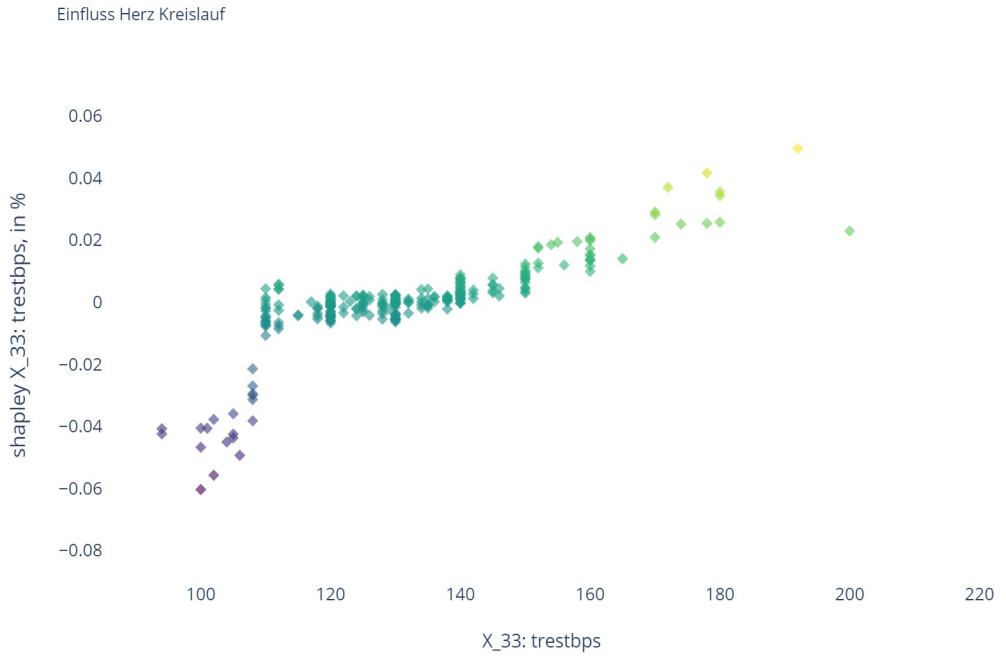


Abb. A.28: Einfluss Kreislauf in mmHg

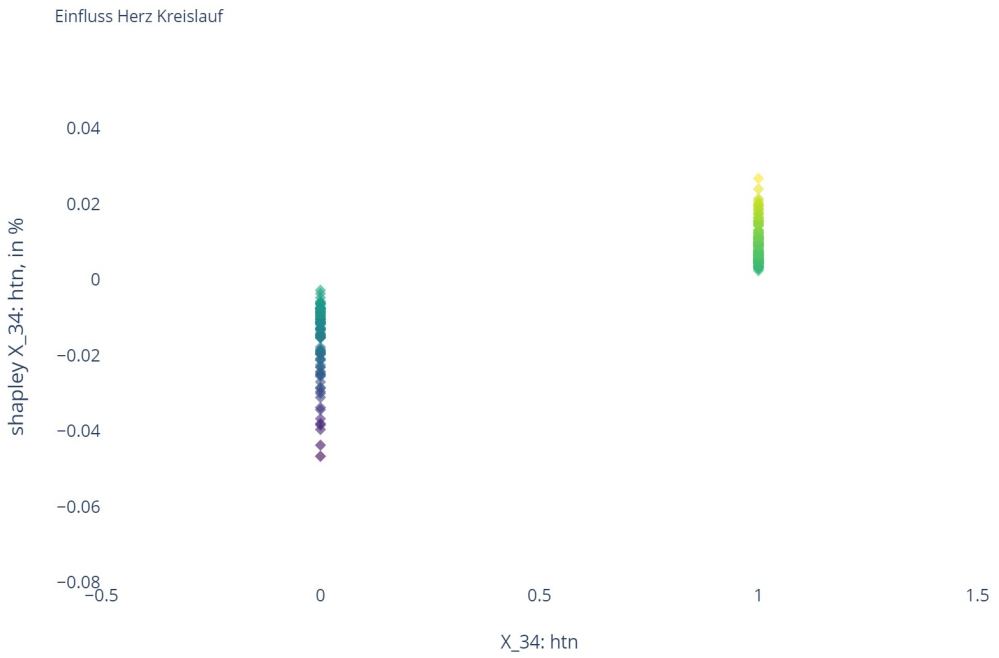


Abb. A.29: bei dem Patienten wurde Bluthochdruck (hypertension, htn) diagnostiziert, 1: ja und 0: nein

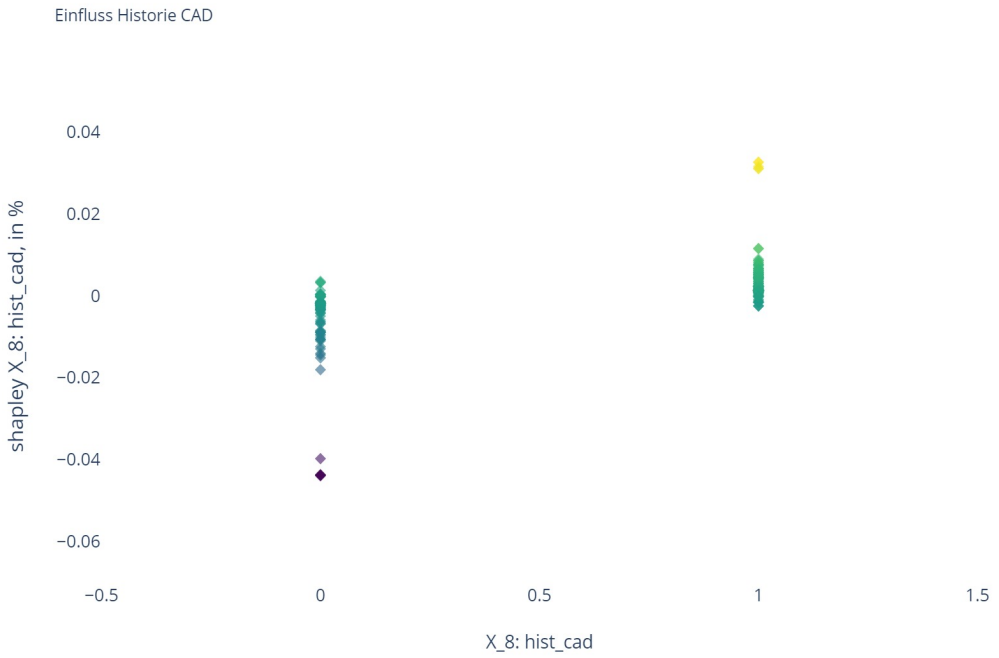


Abb. A.30: Historie von CAD in Verwandschaft, 1: ja und 0: nein

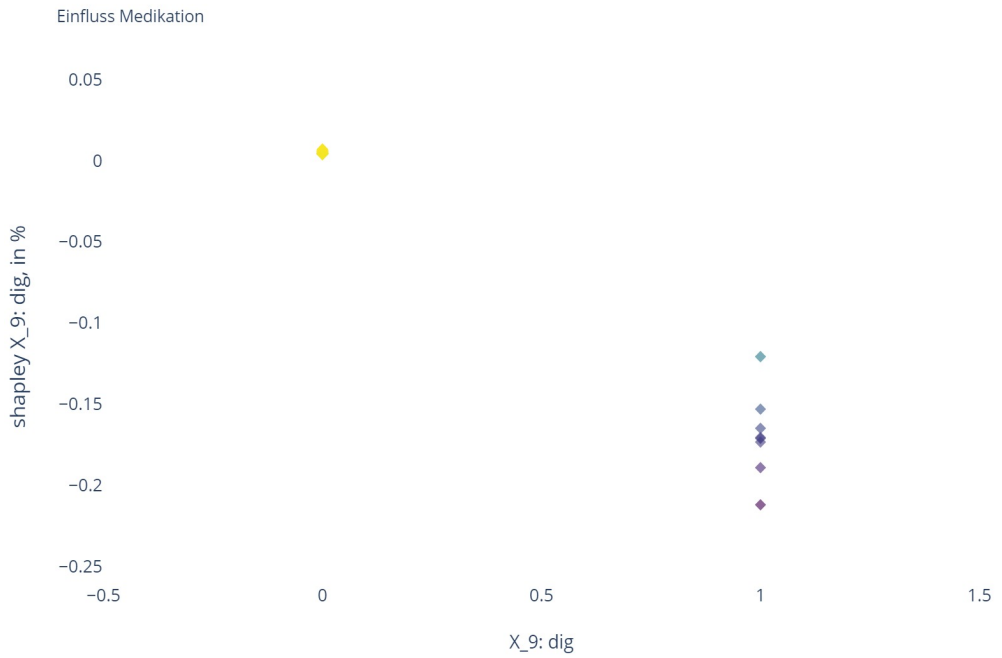


Abb. A.31: Digitalis verabreicht, 1: ja und 0: nein

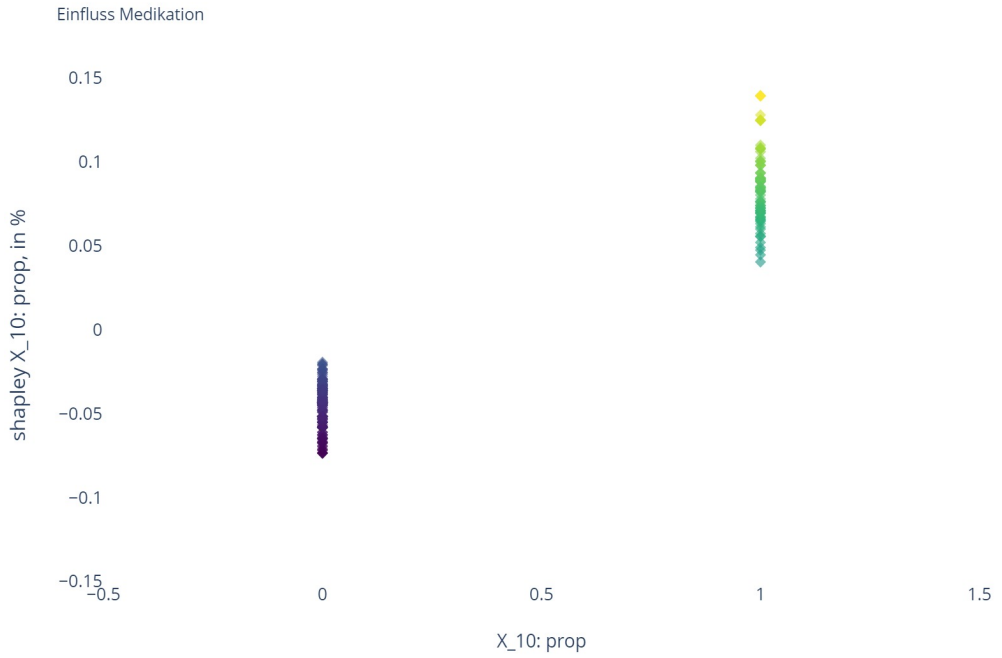


Abb. A.32: Beta blocker verabreicht, 1: ja und 0: nein

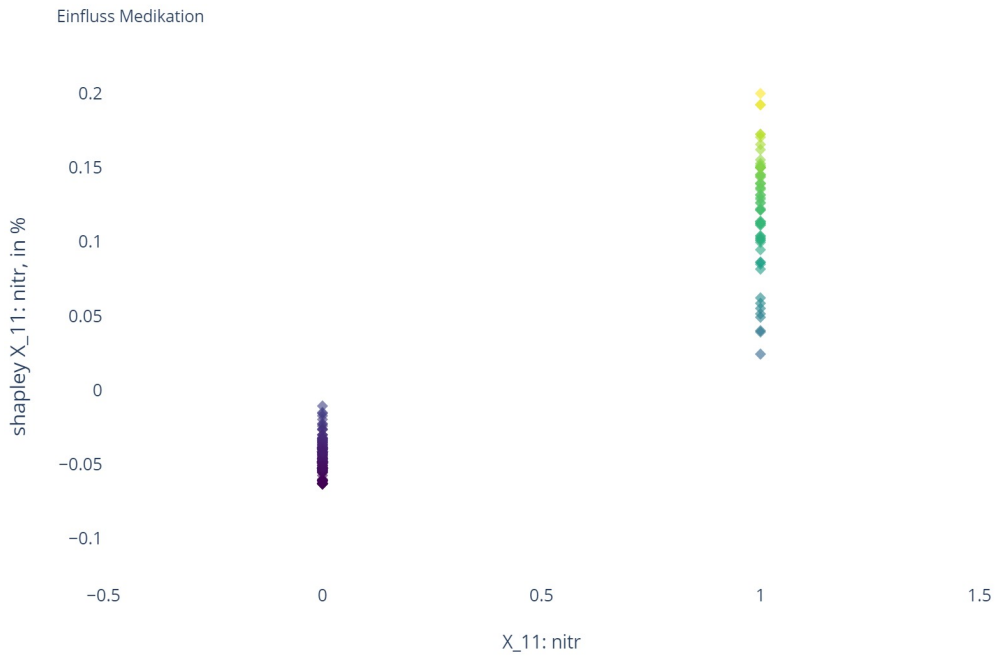


Abb. A.33: Nitrate verabreicht, 1: ja und 0: nein

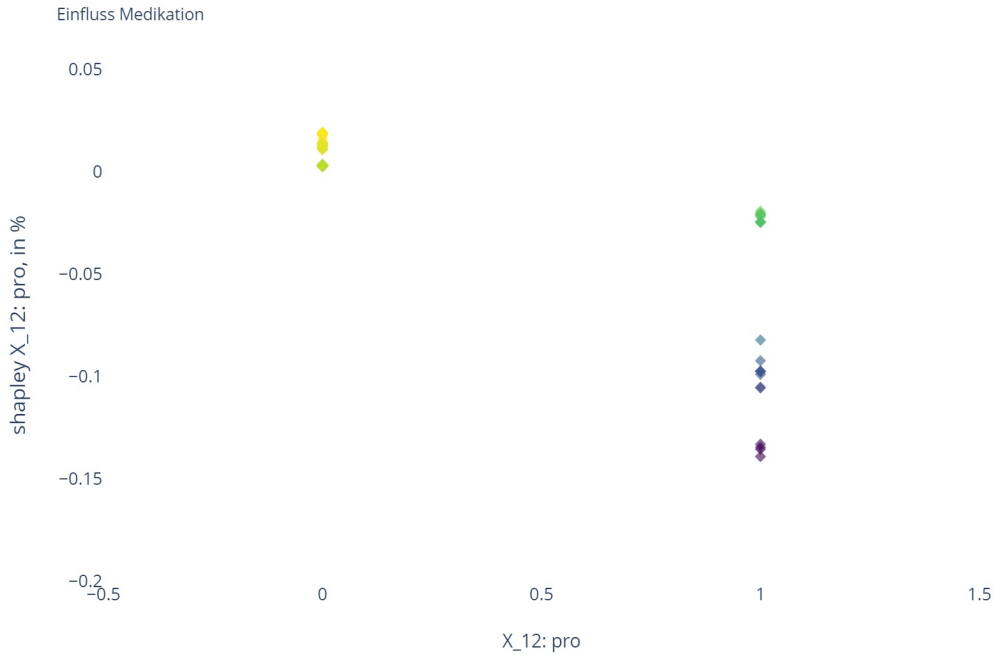


Abb. A.34: Kalziumkanalblocker verabreicht, 1: ja und 0: nein

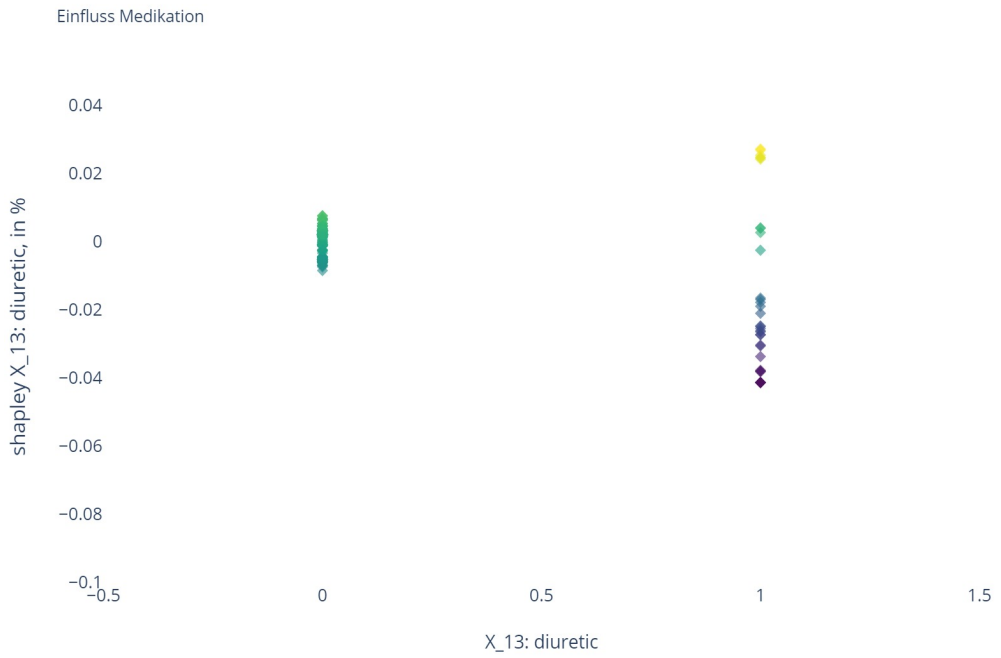


Abb. A.35: Diuretika verabreicht, 1: ja und 0: nein

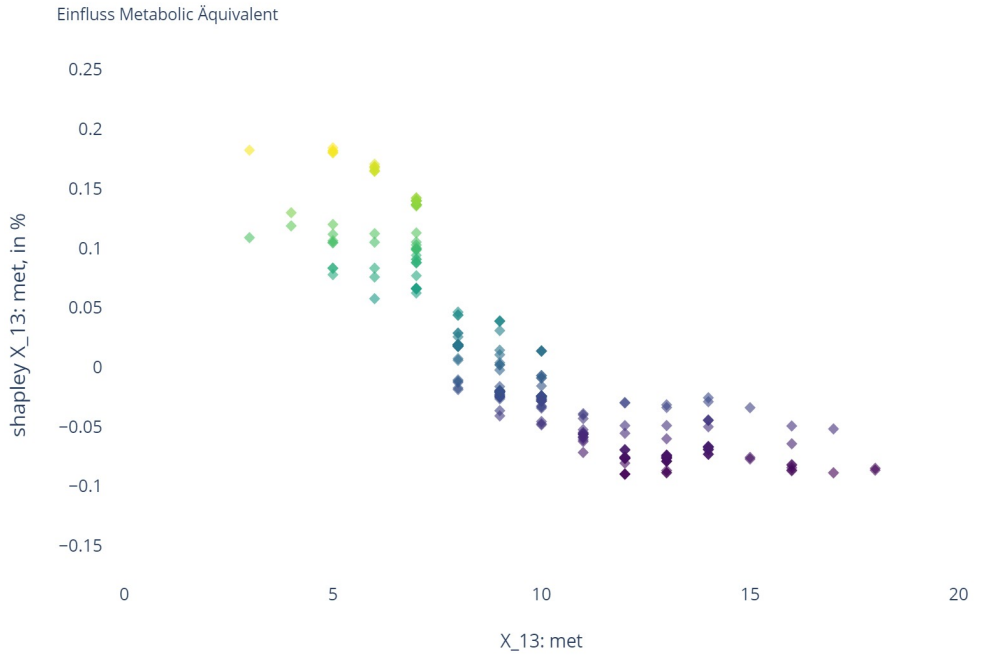


Abb. A.36: Metabolic Äquivalent, in MET

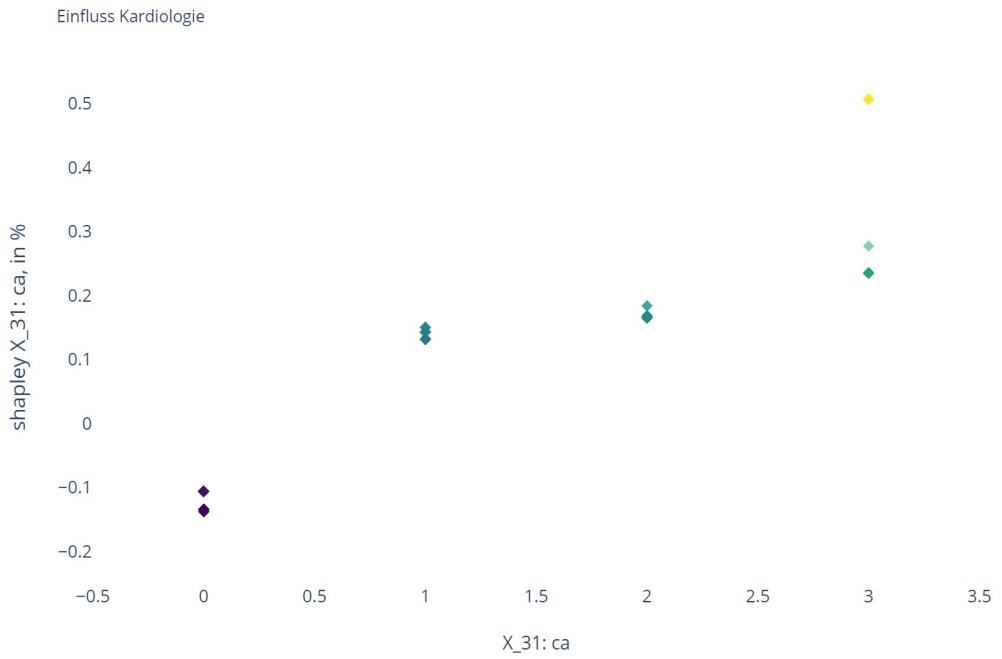


Abb. A.37: Einfluss Kardiologie

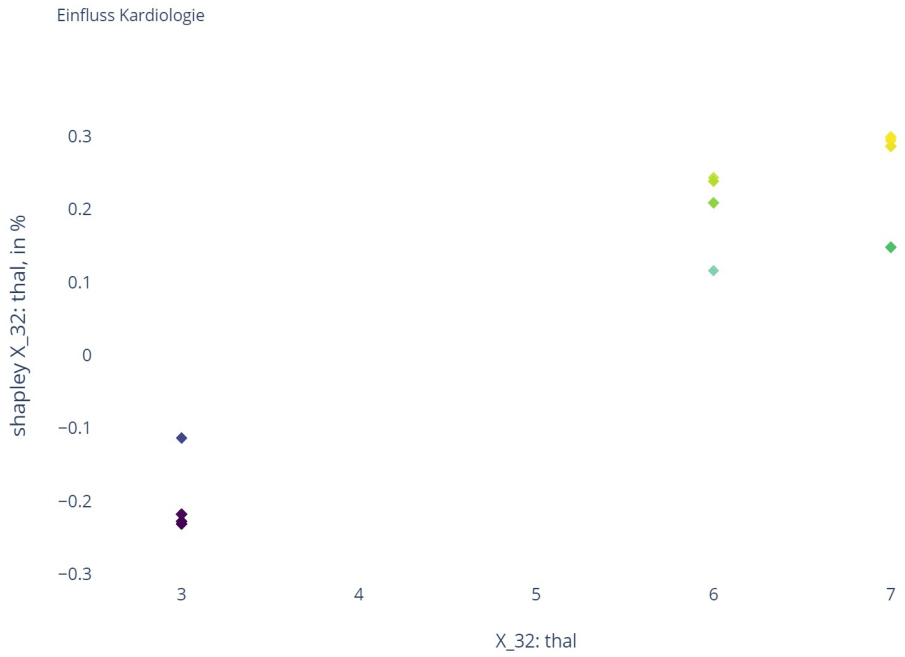


Abb. A.38: Einfluss Kardiologie, exercise thallium scintigraphy 3 = none, normal, 6 = fixed defect, 7 = reversable defect

Anhang 9: Beispieldaten für einen Patienten

Datenview 1

X_1: age	X_2: sex	X_3: cigs	X_4: cig_time	X_5: dm, fbs	X_6: chol
62	1	0	0	1	208

Datenview 2

X_28: oldpeak	X_29: slope	X_30: rldv5e	X_22: thalach	X_35: thaldur	X_14: restecg_0	X_15: restecg_1	X_16: restecg_2
0	1	9,6	140	7	0	0	1

Datenview 3

X_22: thalach	X_23: thalrest	X_24: tpeakbps	X_25: tpeakbpd	X_26: trestbpd	X_33: trestbps	X_34: htn	X_37: xhypo
140	91	160	70	72	128	1	0

Datenview 4

X_5: dm, fbs	X_8: hist_cad	X_9: dig	X_10: prop	X_11: nitr	X_12: pro	X_13: diuretic	X_34: htn	X_13: met
1	0	0	0	0	0	0	1	8

Datenview 5

X_31: ca	X_32: thal
0	3

Zielvariablen

Y_3: lmt	Y_4: ladprox	Y_5: laddist	Y_6: cxmain	Y_7: om1	Y_8: rcaprox	Y_9: rcadist	Y_1: target_sev	Y_2: target
0	0	0	0	0	0	0	0	0