

Aus dem Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE)

Institut der Ludwig-Maximilians-Universität München

Direktor: Univ. Prof. Dr. Ulrich Mansmann

***Entwicklung und Testung eines datenbasierten  
Diagnosealgorithmus für vestibuläre Erkrankungen im  
Bereich der hausarztzentrierten Versorgung***

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

vorgelegt von

Michael Grözinger, M.Sc., Dipl.-Ing. (FH)

aus

München

Jahr

2021

---

Mit Genehmigung der Medizinischen Fakultät  
der Universität München

Berichterstatter: Univ. Prof. Dr. Eva Grill, MPH

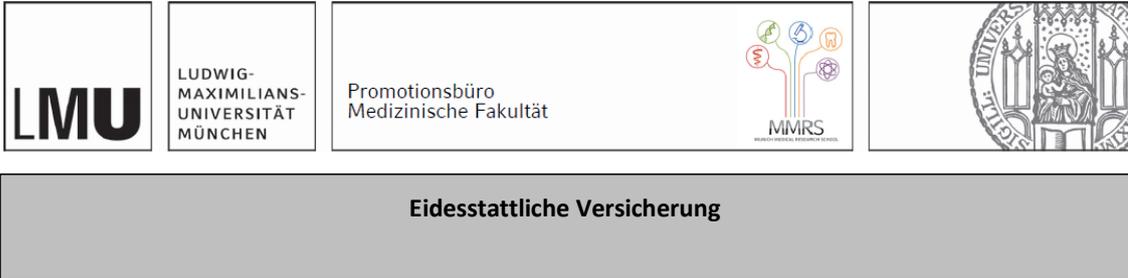
Mitberichterstatter: PD Dr. Larissa Schwarzkopf  
Prof. Dr. Christian Heumann

Mitbetreuung durch den  
promovierten Mitarbeiter: Dr. rer. biol. hum. Ralf Strobl

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 11.11.2021

## Affidavit



### Eidesstattliche Versicherung

Grözinger, Michael

\_\_\_\_\_  
Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

### **Entwicklung und Testung eines datenbasierten Diagnosealgorithmus für vestibuläre Erkrankungen im Bereich der hausarztzentrierten Versorgung**

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 11.11. 2021

Michael Grözinger

# Inhaltsverzeichnis

<b>Affidavit</b> .....	<b>3</b>
<b>Inhaltsverzeichnis</b> .....	<b>4</b>
<b>Abkürzungsverzeichnis</b> .....	<b>5</b>
<b>1. Einleitung</b> .....	<b>6</b>
1.1 Grundlegende Fragestellungen und Forschungsziele .....	7
1.1.1 Forschungsziele .....	7
1.2 Methoden .....	7
1.2.1 Datenquelle .....	7
1.2.2 Diagnosen und Kovariablen.....	8
1.2.3 Vorbereitung der Daten für das Machine Learning .....	8
1.2.4 Statistische Methoden.....	9
1.3 Ergebnisse .....	11
1.3.1 Vorhersagegüte schwer unterscheidbarer Diagnosen .....	11
1.3.2 Ermittlung relevanter Kovariablen und deren Relevanz .....	12
1.4 Diskussion und Ausblick .....	12
1.4.1 Diskussion.....	12
1.4.2 Ausblick.....	14
1.4.3 Wissenschaftlicher Beitrag der Publikationen.....	15
<b>2. Beitrag zu den Veröffentlichungen</b> .....	<b>16</b>
2.1 Beitrag zu Publikation I .....	16
2.2 Beitrag zu Publikation II .....	17
<b>3. Zusammenfassung</b> .....	<b>18</b>
<b>4. Abstract (English)</b> .....	<b>19</b>
<b>5. Information zu den Publikationen</b> .....	<b>20</b>
5.1 Publikationen als Grundlage dieser Dissertation .....	20
5.2 Weitere Publikationen des Doktoranden .....	20
<b>6. Abbildungsverzeichnis</b> .....	<b>21</b>
<b>7. Literaturverzeichnis</b> .....	<b>22</b>
<b>8. Danksagung</b> .....	<b>25</b>

## Abkürzungsverzeichnis

BDT	Boosted Decision Trees
BPLS	Benigner Paroxysmaler Lagerungsschwindel
CART	Classification and Regression Trees
DNN	Deep Neural Networks
DSGZ	Deutsches Schwindel- und Gleichgewichtszentrum
MICE	Multiple Imputation by Chained Equations
ML	Machine Learning
MM	Morbus Meniere
RF	Random Forest
VM	Vestibuläre Migräne

# 1. Einleitung

Mit einer Lebenszeitprävalenz von etwa 30% und steigender Inzidenz mit dem Alter ist Schwindel neben Kopfschmerz eines der häufigsten Leitsymptome in der ärztlichen Praxis und stellt für Patient:innen eine schwere Beeinträchtigung des täglichen Lebens dar [1].

Die oft unklare Symptomlage und fehlende Erfahrung bei der Behandlung von Schwindelerkrankungen führen im hausärztlichen Bereich häufig zu falscher Einordnung. Daraus können sich erfolglose Behandlungsversuche oder späte bzw. unpassende Überweisungen ergeben [2]. Über- oder Unterversorgung sind die häufigsten kurzfristigen Folgen. Langfristig kann eine nicht adäquate Therapie zu einer Chronifizierung führen. Die Konsequenzen dieses Prozesses sind schließlich erhebliche medizinische und gesellschaftliche Kosten [3].

Die Diagnose der verschiedenen Ursachen von Schwindel wird dadurch erschwert, dass sich häufig Symptome verschiedener Schwindelerkrankungen überlagern [4] oder auch von den Patient:innen selbst nicht klar formuliert werden können. Zu den vestibulären Diagnosen, die auch von Fachleuten nur mit erhöhtem Aufwand differenziert werden können, zählen Morbus Menière (MM) und die vestibuläre Migräne (VM). Beide Erkrankungen sind durch spontan wiederkehrende Schwindelattacken gekennzeichnet, die Minuten bis Stunden andauern können und das ohne erkennbare Auslöser [5]. Patient:innen berichten bei vestibulärer Migräne von Symptomen, die auch mit nicht-vestibulärer Migräne assoziiert sind, wie Kopfschmerzen, Phonophobie und Photophobie. Bei Morbus Menière dauern die Attacken ebenfalls Minuten bis Stunden. Diese treten in Verbindung mit beispielsweise Hörverlust, Tinnitus oder Übelkeit auf. Patient:innen mit Morbus Menière berichten aber auch von sporadischen Kopfschmerzen zwischen den Attacken [6]. Erschwerend kommt hinzu, dass bei Patient:innen mit Morbus Menière deutlich häufiger auch nicht-vestibuläre Migräne auftritt [4] und es daher schwierig sein kann, die beiden Diagnosen korrekt zu differenzieren.

Aufgrund dieser Randbedingungen ist es von Interesse, hausarztzentrierten Versorgung Ärzt:innen Möglichkeiten an die Hand zu geben, anhand derer sie Schwindelerkrankungen zuverlässiger einordnen können. Grundlage für die Entwicklung der dazu notwendigen diagnostischen Algorithmen bilden in der Regel klinische Datensätze, mit denen aufbauend auf den Charakteristika von Patient:innen mit bekannter Diagnose Entscheidungsregeln entwickelt werden. Einfache lineare Regressionsmodelle können zur Identifizierung solcher Regeln eingesetzt werden [7, 8]. Diese überschätzen die Vorhersagegüte jedoch in der Regel oder können mit der hohen Anzahl an potenziell relevanten, korrelierenden Charakteristika der Patient:innen nicht ausreichend umgehen.

Als Basis für die Erstellung von Werkzeugen zur Unterstützung von Diagnosen hat sich in den letzten Jahren zunehmend das Verfahren des Machine Learning aus dem Bereich der künstlichen Intelligenz durchgesetzt [9].

Auch für Schwindelerkrankungen wurden diese Ansätze bereits eingesetzt. Die ersten Ansätze mit den anfangs verfügbaren, wenig komplexen Machine Learning Modellen waren jedoch nicht vielversprechend genug, um weiter verfolgt zu werden. Einzelne Krankheitsentitäten konnten, sofern sie vorgegeben wurden (One-vs-All-Ansatz) [10, 11], erkannt werden. Wenn es darum ging, ohne Vorhypothese zu einer diagnostischen Entscheidung zu kommen, versagten die Modelle in der Regel. Außerdem konnten die Modelle nicht transparent darstellen, welche Charakteristika der Patient:innen hohe Vorhersagekraft haben und welche nicht [12]. Die Modelle und Lernverfahren des Machine Learning haben sich seither jedoch grundlegend weiterentwickelt, insbesondere im Bereich des Deep Learning [13].

## 1.1 Grundlegende Fragestellungen und Forschungsziele

Die Fragestellung dieser Arbeit war es daher, zu untersuchen, ob Verfahren des Machine Learning anhand symptomorientierter Charakteristika der Patient:innen in der Lage sind, vestibuläre Erkrankungen zu unterscheiden. Die zweite Forschungsfrage untersuchte, welche vestibulären Symptome besonders relevant für die Vorhersagegüte sein könnten.

### 1.1.1 Forschungsziele

Ausgehend von diesen Fragestellungen waren die Ziele dieser Arbeit:

- Beurteilung der Vorhersagegüte bei der Klassifizierung ausgewählter statistischer Lernverfahren von schwer unterscheidbaren Diagnosen am Beispiel von vestibulärer Migräne und Morbus Menière (Publikation I)
- Ermittlung der für die Diagnostik von vestibulären Krankheiten relevanten Symptome und klinischen Zeichen (Publikation I und II)
- Identifizierung von einfach zu ermittelnden Symptomen für eine Klassifizierung zwischen den sieben häufig diagnostizierten vestibulären Krankheiten (Publikation II)

## 1.2 Methoden

### 1.2.1 Datenquelle

In beiden Studien wurden Daten aus der DizzyReg Datenbank des Deutschen Schwindel- und Gleichgewichtszentrums (DSGZ) verwendet [14]. Das DSGZ ist ein spezialisiertes tertiäres Behandlungszentrum der Universitätsklinik der Ludwig-Maximilians-Universität in München. Es wurde 2009 gegründet und behandelt etwa 3000 Patient:innen pro Jahr. Als Register hat DizzyReg das Ziel, alle relevanten Informationen der Patient:innen aus elektronischen Akten und Arztbriefen des DSGZ bereitzustellen [14, 15]. Nach einer neuro-otologischen Untersuchung erhalten die Patient:innen eine validierte standardisierte Diagnose nach den derzeitigen internationalen

Standards [16, 17]. Aufgrund der langjährigen Erfahrung der Expert:innen in der Klinik stellen die Diagnosen für diese Arbeit den Gold-Standard dar.

### 1.2.2 Diagnosen und Kovariablen

Für die Klassifizierung von Schwindelkrankheiten mit Machine Learning Modellen stehen aus DizzyReg insgesamt 582 Kovariablen bereit, die eingesetzt werden können [14]. Mit einer großen Anzahl an Kovariablen benötigen Machine Learning Modelle beim Training eine entsprechend exponentiell größere Anzahl an Untersuchungen, was auch als „Curse of Dimensionality“ bezeichnet wird [18]. Aufgrund der niedrigen Anzahl an untersuchten Patient:innen in DizzyReg, war es notwendig, die verwendeten Kovariablen auf die relevanten zu reduzieren. Dazu gibt es verschiedene Verfahren. Sie können mit Hilfe statistischer Eigenschaften, zum Beispiel der Korrelation einzelner Kovariablen mit den Diagnosen, bestimmt werden; dies wird in der Literatur zu Machine Learning auch als Filterung bezeichnet [19]. Damit kommen nur die Kovariablen beim Training eines Machine Learning Modells zum Einsatz, die die für das Training notwendige Information beinhalten.

In **Publikation I** war die Aufgabenstellung auf die Diagnosen vestibuläre Migräne und Morbus Menière fokussiert. Am DSGZ sind diese Diagnosen der Expert:innen an den Kriterien der Bárány Gesellschaft für Neuro-Otologie von 2012 [17] und 2015 [20] orientiert, was für die weitere Arbeit in diesem Gebiet für mehr Vergleichbarkeit sorgt. Die in dieser Studie verwendeten Kovariablen wurden mit Hilfe ihrer Korrelation mit den beiden Zielvariablen MM und VM ausgewählt.

In **Publikation II** wurden die im DSGZ am häufigsten vergebenen sieben Diagnosen untersucht [14]. Um in dieser Studie die relevanten Symptome zu identifizieren, die in der hausarztzentrierten Versorgung einfach ermittelt werden können, wurde eine stufenweise Vorgehensweise gewählt. Zunächst erfolgte die Abfrage bei Expert:innen aus der Praxis nach ihren Kriterien und Fragestellungen bei der Diagnose von Schwindelkrankheiten. Daran schloss sich eine Kategorisierung an und jede Kategorie bzw. Unterkategorie bekam die Kovariable, welche die Kategorie inhaltlich am besten abbildet, aus DizzyReg zugeordnet.

### 1.2.3 Vorbereitung der Daten für das Machine Learning

Für die Klassifizierung mit Machine Learning Verfahren sind verschiedene vorbereitende Schritte notwendig. Unter anderem erfordern Verfahren wie Deep Neural Networks, dass der Datensatz keine fehlenden Werte hat. Da es sich bei Machine Learning um statistische Verfahren handelt [18], sollte das Auffüllen dieser Werte (Imputation) mit Verfahren geschehen, welche die Wahrscheinlichkeitsdichteverteilung der Trainingsdaten möglichst wenig verändern. In dieser Arbeit kam für den Datensatz zur Unterscheidung von MM und VM das Verfahren des Multiple Imputation by Chained Equations (MICE) [21] zum Einsatz, das diese Anforderung erfüllt. MICE basiert auf der Ermittlung der Werte durch Markov Chain Monte Carlo Methoden und schätzt die Wahrscheinlichkeitsdichteverteilung der Kovariablen mit fehlenden Werten ab [22]. Das ergibt mehrere

imputierte Datensätze, die für diese Studien eingesetzt wurden. Bei den Daten für die Klassifizierung mit sieben Diagnosen wurden die fehlenden Werte durch neutrale Werte aufgefüllt, was beispielsweise die gängige Praxis reflektiert, dass bestimmte Untersuchung nur durchgeführt werden, wenn Sie notwendig sind, das heißt die Werte sind vermutlich neutral.

### 1.2.4 Statistische Methoden

Bei den Verfahren des Machine Learning liegt die Annahme zugrunde, dass ein Ergebnis  $Y$  in Beziehung zu  $p$  verschiedenen unabhängigen Variablen  $X = (X_1, X_2, X_3, \dots, X_p)$  steht. Für die optimale Klassifizierung gilt es eine Funktion  $f(X)$  zu identifizieren und deren Parameter optimal an die Trainingsdaten anzupassen [18]. Um die Qualität der Klassifizierung der verschiedenen Modelle vergleichen zu können, setzten wir die Accuracy, wie in (1) definiert, und das F-Measure [23], wie in (4) beschrieben, als Kriterien ein. Das F-Measure kombiniert die Sensitivität (3), welche angibt, wie gut Patient:innen mit einer bestimmten Krankheit richtig erkannt werden mit der Präzision (2), die aufzeigt, mit welcher Genauigkeit der Klassifizierungsalgorithmus positive Diagnosen erkennt.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Präzision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivität = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-Measure} = 2 * \frac{Präzision * Sensitivität}{Präzision + Sensitivität} \quad (4)$$

Dabei sind TP (True Positives) die Anzahl der Personen mit und TN (True Negatives) die Anzahl der Personen ohne die Krankheit, die richtig erkannt wurden. Entsprechend sind FP (False Positives) die Anzahl der Personen mit und FN (False Negatives) die Anzahl der Personen ohne die Diagnose Schwindel, die falsch klassifiziert wurden.

Es stehen im Machine Learning eine Vielzahl an Modellen und Lernverfahren zur Verfügung [13, 18]. Für die Differenzierung von MM und VM fiel die Wahl auf die Modelle der Deep Neural Networks (DNN) [13] und der Boosted Decision Trees (BDT) [24]. Insbesondere DNN haben sich bereits in anderen Bereichen der Medizin als leistungsfähig erwiesen [9] und BDT ermöglichen zusätzlich die Ermittlung der Wichtigkeit der Variablen. Des weiteren wurden Classification and Regression Trees (CART) [25] und darauf aufbauende Ensembles wie Random Forests (RF) [26] verwendet, da sie für die Klassifizierung und die Identifikation der relevanten Kovariablen die geforderte Einfachheit und Transparenz bieten.

In **Publikation I** war das Ziel, vestibuläre Migräne und Morbus Menière möglichst genau zu differenzieren. In den verwendeten Daten aus DizzyReg kommen die Diagnosen mit sehr unterschiedlicher Häufigkeit vor, was dazu führen kann, dass der Lernalgorithmus im Training be-

stimmte Klassen bevorzugt [27]. Ein Viertel der Fälle sind der Diagnose des funktionellen Schwindels zugordnet (24,1%), sowie 15,5% der vestibulären Migräne und 9,9% der Diagnose Morbus Menière. Deshalb wurde wie in [10] ein One-vs-All Ansatz [28] verfolgt. Dabei wird eine bestimmte Diagnose, zum Beispiel Morbus Menière, gegenüber allen anderen Diagnosen, die in eine zweite Klasse zusammengefasst wurden, differenziert. Für jede der beiden Diagnosen VM und MM erfolgte das Training eines separaten Modells.

Das erste für die Differenzierung von MM und VM verwendete Verfahren, Deep Neural Networks [13] lehnt sich an die Struktur des menschlichen Gehirns an. Ein Beispiel ist in Abbildung 1 dargestellt. Die Anzahl der Schichten im Modell (hier Hidden Layers) und die Anzahl der Knoten pro Schicht, ermöglichen es, beliebig komplexe Netzwerke zu erstellen, die dann die Funktion  $f(X)$  möglichst genau annähern. Um die beste Netzwerkkonfiguration des DNN zu finden, wurden beide Parameter unabhängig variiert und das F-Measure zur Beurteilung verwendet. Wir unterschieden dabei „Shallow Networks“ mit maximal zwei Hidden Layers und „Deep Networks“ mit maximal acht Hidden Layers.

Während des Trainingsvorgangs des DNN werden die Parameter des Modells wiederholt in kleinen Schritten angepasst. Es kommt dabei vor, dass der Lernalgorithmus das Training in einer lokalen, sub-optimalen Lösung beendet. Um das Ergebnis des Trainings aufgrund solcher Fälle noch weiter zu verbessern, wurde durch wiederholtes Trainieren des DNN mit variierten Trainingsdaten versucht, die Klassifizierungsergebnisse zu verbessern. Dieses Vorgehen nannten wir Pre-Training.

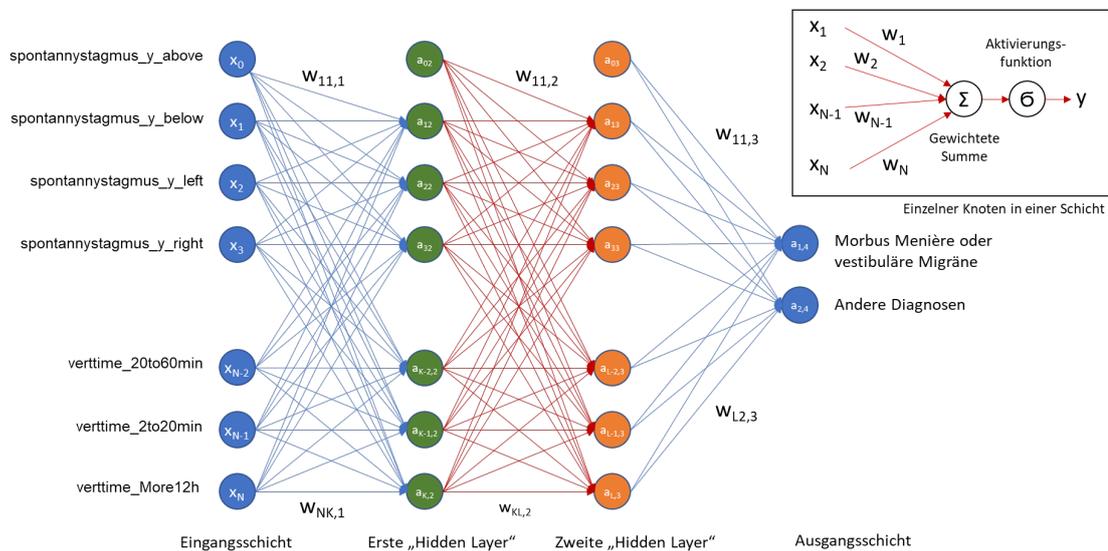


Abbildung 1 Schematische Darstellung eines Deep Neural Network mit 2 Hidden Layer und K Knoten in der ersten Hidden Layer sowie L Knoten in der zweiten Hidden Layer. Nach dem One-vs-All Ansatz wird die gesuchte Diagnose (z.B.: VM) gegen alle anderen Diagnosen differenziert. In einem einzelnen Knoten werden die Eingangswerte aufsummiert und mit einer Aktivierungsfunktion bewertet. (eigene Darstellung)

Das zweite Verfahren, das im Rahmen von Publikation I verwendet wurde, sind die Boosted Decision Trees (BDT) [29]. Sie bestehen aus einem Ensemble aus einfachen Entscheidungsbäumen, die sukzessive trainiert werden. Die endgültige Klassifizierung wird als Mehrheitsentscheidung unter den einzelnen Bäumen getroffen. In unseren Untersuchungen variierten wir die Komplexität der BDT über die Anzahl der trainierten Entscheidungsbäume.

In der Studie zu **Publikation II** wurde die Klassifizierung mit möglichst großer Transparenz und einfach zu ermittelnden Kovariablen angestrebt. Classification and Regression Trees bieten hier den Vorteil, dass sie eine visuelle Darstellung in Form eines Entscheidungsbaumes bereitstellen, die die menschliche Entscheidungsfindung nachbildet [25]. CART verwenden jedoch nach dem Training für die Klassifizierung nicht alle Variablen des Trainingsdatensatzes. Es kann deshalb nur für diese Kovariablen die relative Wichtigkeit ermittelt werden. Um das für alle Kovariablen im Trainingsdatensatz zu erreichen, setzten wir Random Forest Modelle ein, die zufällig Untermenüen der Kovariablen auswählen und dazu jeweils einen CART trainieren. Das Klassifizierungsergebnis entsteht ebenfalls durch einen Mehrheitsentscheid unter den erstellten Entscheidungsbäumen.

Für die Ermittlung der Wichtigkeit der Variablen auf Basis der RF stehen mehrere Verfahren zur Verfügung. Eines, die Permutation Importance, basiert auf der zufälligen Vertauschung (Permutation) der Werte nur einer einzelnen Kovariablen und dem anschließenden erneuten Training eines RF Modelles [26]. Wenn die Genauigkeit der Klassifizierung mit den permutierten Daten variiert, dann spricht das für eine hohe Wichtigkeit der Kovariablen. Dieses Verfahren hat die Einschränkung, dass andere, nicht permutierte Variablen als eventuell Stellvertreter agieren können und den Effekt der Permutation ausgleichen. Deshalb schlägt die Forschung in diesem Bereich unter anderem vor, die Häufigkeit zu betrachten, wie oft eine Variable in den einzelnen CART als Wurzel des Baumes verwendet wurde [30]. In dieser Arbeit wurden beide Verfahren verwendet und berichtet.

## 1.3 Ergebnisse

### 1.3.1 Vorhersagegüte schwer unterscheidbarer Diagnosen

Insgesamt wurden in der Studie zu Publikation I Daten von 1357 Patient:innen verwendet (Durchschnittsalter: 53,9 Jahre, STD: 15,9 Jahre, 54,7% sind weiblich). Für 9,9% der Fälle diagnostizierten Expert:innen im DSGZ Morbus Menière und für 15,6% vestibuläre Migräne.

Zur Beurteilung der Vorhersagegüte bei den schwer unterscheidbaren Diagnosen MM und VM konnten gemittelt über alle imputierten Datensätze mit einem DNN für Morbus Menière ein F-Measure von  $55,5 \pm 2,7\%$  (Accuracy  $91,4 \pm 0,6\%$ ) ermittelt werden. Für VM erhielten wir ein F-Measure von  $36,8 \pm 0,8\%$  (Accuracy  $81,8 \pm 0,6\%$ ). Das Pre-Training eines DNN verbesserte die Ergebnisse, beispielsweise für Morbus Menière auf einen Wert für das F-Measure von  $90,0 \pm 4,6\%$  und

einer Accuracy von  $98,0 \pm 1,0\%$ . Für beide Diagnosen wurden mit einem Shallow Network mit zwei Hidden Layers das beste Ergebnis erzielt.

Der Einsatz von Boosted Decision Trees, trainiert auf Morbus Menière, ergab ein F-Measure von  $53,3 \pm 3,0\%$  (Accuracy  $93,3 \pm 0,7\%$ ). Bei VM erhielten wir mit BDT ein F-Measure von  $27,6 \pm 5,1\%$  und eine Accuracy  $84,5 \pm 0,5\%$ .

Aus dem Training des BDT Modells wurden für die relative Wichtigkeit der Kovariablen das Alter (MM  $9,13\%$  und VM  $11,93\%$ ) und kalorische Seitendifferenz (MM  $7,60\%$  und VM  $7,06\%$ ) als die wichtigsten Kovariablen identifiziert.

### 1.3.2 Ermittlung relevanter Kovariablen und deren Relevanz

Bei der Erfassung der in der Praxis wichtigen Fragestellungen für Publikation II beteiligten sich 21 Expert:innen aus 16 verschiedenen Ländern (19 Zentren und 23 Jahre Erfahrung). Ihre Befragung ergab insgesamt 152 verschiedene Statements. In dieser Studie wurden 1066 Patient:innen mit einem mittleren Alter von 51,1 Jahren (STD: 15,3 Jahre) betrachtet, wovon 56% weiblich waren. Funktioneller Schwindel war die häufigste Diagnose ( $32,5\%$ ) gefolgt von VM ( $20,2\%$ ) und MM ( $13,3\%$ ). Die Analyse der Statements der Expert:innen führte zu 9 Haupt- und 23 Unterkategorien. Diesen konnten 98 Variablen aus DizzyReg zugeordnet werden.

Aus den Experimenten mit den Classification and Regression Trees wurden folgende acht Kovariablen als relevant ermittelt: Dauer der Attacke, Drehschwindel, Hörprobleme, die Trigger Drehen im Bett und Sport bzw. schwere Hausarbeit, Alter, Probleme beim Gehen in der Dunkelheit und Erbrechen. Als übergeordnete Genauigkeit bei der Klassifizierung aller sieben Diagnosen ergab eine Genauigkeit von  $42,2\%$ .

Bei der Beurteilung der Wichtigkeit der Variablen mit den beiden eingesetzten Verfahren auf Basis der RF wurden Erbrechen, Alter, Hörprobleme, Drehen im Bett als Trigger und Dauer der Attacke von weniger als zwei Minuten als die wichtigsten Variablen ermittelt.

## 1.4 Diskussion und Ausblick

### 1.4.1 Diskussion

Wir konnten in dieser Arbeit zeigen, dass Machine Learning (ML) Verfahren mit guter Vorhersagegüte, klinisch schwer unterscheidbare vestibuläre Diagnosen differenzieren können. Dabei ist es uns gelungen, am Beispiel von Morbus Menière und der vestibulären Migräne relevante Symptome und klinische Zeichen zu identifizieren, die besonders für eine pragmatische Einordnung dieser beiden Krankheitsbilder geeignet sind. ML Verfahren können zudem eingesetzt werden, um nicht nur VM und MM zu unterscheiden, sondern auch andere häufige Schwindelkrankheiten sicher zu klassifizieren. Wir haben acht Schlüsselfragen, welche die sieben häufigsten Diagnosen

Klassifizieren können, identifiziert, nämlich Dauer der Attacke, Drehschwindel, Hörstörungen, typische Trigger wie Drehen im Bett und Sport bzw. schwere Hausarbeit, Alter, Orientierung im Dunkeln und Erbrechen.

Die von uns erzielte Genauigkeit von 91,4% für MM ist der in einer früheren Studie mit der k-Nearest-Neighbour-Klassifizierung erzielten mittleren Genauigkeit von 79,8% deutlich überlegen [11]. Eine andere Studie erreichte eine Genauigkeit von 76% bei der Klassifikation der unilateralen Vestibulopathie mit Hilfe von Support Vector Machines [31].

Unsere Ergebnisse zeigten, dass DNN Morbus Menière besser klassifizieren können als vestibuläre Migräne. Dies ist zu erwarten, da die Differenzierung beider Diagnosen auch in der klinischen Praxis schwierig ist. Mehr als die Hälfte der Patient:innen mit MM erfüllen zusätzlich die diagnostischen Kriterien von VM und umgekehrt [32].

Ein relevanter Nachteil von DNN ist, dass diese Modelle keine Aussage über den Beitrag einzelner Kovariablen zur Prädiktion geben. Wir setzten deshalb zusätzlich Boosted Decision Trees [29] ein, einerseits, um die Genauigkeit zu vergleichen, andererseits aber, weil aus Boosted Decision Trees die Relevanz der Kovariablen für die Klassifizierung ermittelbar ist. Als wichtigste Kovariablen konnten Parameter vestibulärer Testung (kalorische Seitendifferenz, Gain) für VM und MM identifiziert werden. Das deckt sich mit den Beobachtungen, dass Patient:innen mit VM beispielsweise milde zentrale okulomotorische Störungen in Form eines Blick-evozierten Lagerungsnystagmus aufweisen [33-36]. Unter den Kovariablen, die keine apparativen Untersuchungen voraussetzen, identifizierten wir Hörverlust, Erbrechen und die Dauer der Attacke als wichtigste Kovariablen für MM und Kopfschmerz für VM. Mit Hörverlust bei MM und Kopfschmerz bei VM als typische Symptome sind diese Ergebnisse nicht überraschend und lassen sich bei der Weiterentwicklung der Klassifizierungsalgorithmen einsetzen.

In einem dritten Schritt konnten wir die Meinung von Expert:innen aus der Praxis einsetzen, um aus den Kovariablen von DizzyReg einen Datensatz abzuleiten, der einfach zu ermittelnde Parameter enthält. Wir setzten Machine Learning Verfahren ein und erzielten plausible Ergebnisse. Es war nicht überraschend, dass Funktioneller Schwindel, der häufig diagnostiziert wird, mit der höchsten Präzision klassifiziert wurde. Ebenfalls konnten wir die Diagnose Benigner Paroxysmaler Lagerungsschwindel (BPLS) durch eine kurze Dauer der Schwindelattacke sowie die Bewegung des Körpers und des Kopfes in der horizontalen Ebene klassifizieren, was sich mit der Erfahrung aus der Praxis deckt.

Ein spezifisches Ziel dieser Arbeit war die Klassifizierung mit Transparenz zur Entscheidung, einfache Nutzung in der Praxis und Einfachheit des Ergebnisses. In der Vergangenheit wurden bereits Machine Learning Modelle zur Klassifikation vestibulärer Erkrankungen eingesetzt [37, 38], jedoch nicht mit der geforderten Transparenz. Wir setzten Classification and Regression Trees und die darauf aufbauenden Random Forest Modelle ein, die intuitiv verständliche Ergebnisse bereitstellen und eine leicht lesbare Baumdarstellung der Entscheidung bieten.

Diese Arbeit hat drei Limitationen, erstens verwendeten wir Daten aus einem Patientenregister eines tertiären Referenzzentrums für Gleichgewichtsstörungen, das nicht repräsentativ für Patient:innen ist, die sich mit Schwindel und Benommenheit in der Primär- oder Sekundärversorgung vorstellen. Patient:innen, die das DSGZ aufsuchen, sind in der Regel eine Auswahl von schweren oder chronischen Fällen mit einer langen Krankheitsgeschichte oder erfolgloser Therapie [14]. Dies könnte die relativ geringe Genauigkeit unserer Ergebnisse erklären. Zweitens haben wir wegen der niedrigen Prävalenz der Diagnosen in DizzyReg den One-vs-All Ansatz verwendet. Es ist aber von praktischer Relevanz, alle Diagnosen gleichzeitig zu klassifizieren. Nur dann wäre die geforderte Transparenz zur Entscheidungsfindung erreichbar. Aus diesem Grund haben wir in der zweiten Studie CART verwendet und die sieben häufigsten Diagnosen eingeschlossen. CART sind jedoch, und das ist die dritte Limitation, im Gegensatz zu anderen Klassifizierungsmodellen, sehr anfällig bei kleinen Variationen der Trainingsdaten, was dann zu einem völlig anderen Entscheidungsbaum führen kann. Dies wird jedoch kompensiert durch die hohe visuelle Qualität des resultierenden Entscheidungsbaums mit intuitiv verständlichen Ergebnissen und durch die bestechende Einfachheit der Klassifizierungsschritte.

In unserer zweiten Studie war es ein wesentlicher Schritt, dass wir die Trainingsdaten auf die Erfahrung von Expert:innen stützten, die wir aus Interviews und einer Kategorisierung ermittelten. Das ist ein gutes Beispiel dafür, das statistische Lernverfahren immer durch den Sachverstand von Expert:innen ergänzt werden sollten.

Es lässt sich zusammenfassen, dass DNN sehr gute Klassifizierungsergebnisse lieferten, aber aufgrund der inhärenten Struktur des Modells und der Verwendung des One-vs-All Vorgehens nicht die geforderte Transparenz schaffen. Mit diesen guten Ergebnissen in der Klassifizierung und der Möglichkeit zur Schaffung von mehr Transparenz bei DNN ist es vielversprechend, sich in der weiteren Arbeit auf diese Themen zu fokussieren. Die bei der Klassifizierung der häufigsten sieben Diagnosen verwendeten Algorithmen basieren auf einem transparenten und einfach anzuwendenden Ansatz. Sie können für eine initiale Triage der Patient:innen verwendet werden, aber es muss sich eine klinische Untersuchung der vestibulären und okularen Funktionen anschließen, um die Genauigkeit der Diagnose zu verbessern.

Um herauszuarbeiten, ob die verwendeten Modelle als Basis für einfach zu verwendende Algorithmen in der hausarztzentrierten Versorgung dienen, sind weitere Studien im hausärztlichen Bereich notwendig.

## 1.4.2 Ausblick

Unter dem Begriff der „Explainable Artificial Intelligence“ wird an Verfahren geforscht [39], um so bei Machine Learning Modellen mehr Transparenz zu schaffen. Verfahren wie die „Layer-wise Relevance Propagation“ [40] für DNN wurden allerdings auf Basis homogener Daten im Kontext der Bilderkennung entwickelt. Es muss noch überprüft werden, ob sie mit heterogenen Daten wie denen aus DizzyReg vergleichbare Ergebnisse liefern können.

Machine Learning bietet eine Vielzahl von Modellen [13, 41, 42]. In dieser Arbeit wurde zunächst auf eine Untermenge der Modelle fokussiert, die es ermöglichen, die gesetzten Forschungsziele zu erreichen. Aus den „No Free Lunch Theoremen“ [43] lässt sich ableiten, dass es für jede Aufgabenstellung notwendig ist, möglichst viele verschiedene Verfahren zu evaluieren. Für die weitere Forschung zur Klassifizierung mit Daten aus DizzyReg bedeutet das, in einem erweiterten Umfang noch andere Machine Learning Modelle zu testen, die auf anderen Prinzipien basieren. Diese sollten durch erweiterte Methoden der „Explainable Artificial Intelligence“ ergänzt werden, die gegebenenfalls auf die Daten von Patient:innen mit Schwindelsymptomen angepasst werden.

Die Daten aus DizzyReg sind im Umfang begrenzt, was die Anzahl der Untersuchungen betrifft. Um mit Machine Learning bessere Klassifizierungsqualität zu erreichen, muss die Menge der Daten bezüglich der Anzahl an Untersuchungen erweitert werden. Das kann über mehrere Schwindelzentren hinweg oder im Kontext von DizzyReg geschehen. Für die Daten in DizzyReg wird daran im Projekt Poise [44] gearbeitet, vor allem werden dabei auch Ärzt:innen aus der hausarztzentrierten Versorgung eingebunden. Diese Daten stehen jedoch erst in den nächsten 3-4 Jahren zur Verfügung.

### 1.4.3 Wissenschaftlicher Beitrag der Publikationen

Der wissenschaftliche Beitrag der Arbeit für **Publikation I** lässt sich wie folgt zusammenfassen:

- Es wurde gezeigt, dass Deep Neural Networks für die Differenzierung von Morbus Menière und vestibuläre Migräne sehr gut geeignet sind, jedoch mit wenig Transparenz. Boosted Decision Trees erreichten vergleichbare Ergebnisse für die Genauigkeit, aber erheblich schlechtere beim F-Measure.
- Die von Boosted Decision Trees ermittelte Wichtigkeit der Variablen kann mit der klinischen Erfahrung nachvollzogen werden.

Der Beitrag der Arbeit im Kontext von **Publikation II** stellt sich folgendermaßen dar:

- Über die wichtigen Fragestellungen von Expert:innen in Bezug auf die medizinische Historie von Patient:innen konnten die für sie wichtigen Fragestellungen ermittelt, kategorisiert und den Kovariablen aus DizzyReg zugeordnet werden.
- Durch die Verwendung von CART und RF konnte eine kleine Menge an Variablen identifiziert werden, die eine gute und transparente Klassifizierung ermöglichen und sich mit der klinischen Praxis decken.

## 2. Beitrag zu den Veröffentlichungen

Im Rahmen dieser Forschungsarbeit wurden verschiedene Machine Learning Verfahren angewandt, um deren Fähigkeit, Diagnosen für Schwindelerkrankungen treffen können, zu beurteilen. Die beiden folgenden Kapitel fassen den Inhalt der beiden Publikationen zusammen und erläutern den Beitrag des Doktoranden.

### 2.1 Beitrag zu Publikation I

In der ersten Studie wurde untersucht, wie zuverlässig zwei verschiedene Machine Learning Modelle aus dem Deep Learning vestibuläre Migräne und Morbus Menière klassifizieren können und in welcher Struktur der Netzwerke optimale Ergebnisse erzielt werden.

Dazu wurden aus den Daten der DizzyReg Datenbank des DSGZ mit Hilfe statistischer Verfahren die Variablen ausgewählt, die eine hohe Korrelation mit beiden Diagnosen aufzeigen.

In Vorbereitung auf das Training der Deep Learning Modelle war es notwendig, die fehlenden Werte in den Daten aufzufüllen. Ein geeignetes Verfahren musste sicherstellen, dass die Wahrscheinlichkeitsdichteverteilung der Daten nicht stark verändert wird. Daher kam Multiple Imputation by Chained Equations [21, 22, 45] zum Einsatz. Nach dem Auffüllen der fehlenden Werte, wurde die Qualität der Imputation validiert.

Das erste Modell, Deep Neural Network, hat viele Freiheitsgrade in seiner Struktur und dem Ablauf des Lernverfahrens. Um den Umfang der Studie zu begrenzen, variierten wir die Anzahl der Knoten in den „Hidden Layers“ und die Anzahl der „Hidden Layers“. Diese vergrößerten wir sukzessive, um die Komplexität der Modelle und damit die Fähigkeit, Wissen im Modell abzubilden, zu erhöhen [18].

Zusätzlich wurde ein anderes Modell, Boosted Decision Trees, eingesetzt, um deren Ergebnisse mit denen der Deep Neural Networks zu vergleichen und zudem die Wichtigkeit der Variablen zu beurteilen.

Zur Beurteilung der Güte der Klassifizierung wurden verschiedene Parameter, wie Accuracy und F-Measure verwendet. Zusätzlich ergab sich die Fragestellung, inwieweit sich die Güte eines bereits trainierten Deep Learning Netzwerks verbessert, wenn es rekursiv mit veränderten Stichproben für die Trainingsdaten wiederholt trainiert wird.

Auf Basis der bereitgestellten Daten aus DizzyReg, hat der Doktorand folgende Aufgaben durchgeführt:

- Statistische Analyse der Daten und die Beurteilung der Struktur der fehlenden Werte, Auffüllen der fehlenden Werte mit einem Verfahren der Multiple Imputation, Überprüfung und Beurteilung der Qualität der ergänzten Daten
- Planung der Durchführung der Experimente mit dem Ziel der Analyse der Abhängigkeit der Qualität der Vorhersagegüte von der Komplexität des Modells
- Implementierung des Deep Neural Network und der Boosted Decision Trees auf Basis einer verfügbaren Python Bibliothek (Tensorflow), Durchführung und Auswertung der Experimente

- Analyse und Auswertung der Ergebnisse, Strukturierung und Erstellung des Forschungsberichtes
- Untersuchung der Möglichkeit des wiederholten Lernens mit einem Deep Neural Network

## 2.2 Beitrag zu Publikation II

In der zweiten Arbeit war es das Ziel, einen Diagnosealgorithmus zu implementieren und zu testen, der es ermöglicht, einfach zu ermittelnde Symptome für eine Klassifizierung zwischen häufig diagnostizierten Schwindelerkrankungen zu identifizieren, und damit die Anzahl der für eine Diagnose notwendigen Variablen zu minimieren.

Der verwendete Datensatz leitet sich aus Fragestellungen ab, die sich Expert:innen bei der Diagnose von Schwindelerkrankungen stellen. Diese Fragestellungen waren das Ergebnis eines Online-Fragebogens und dem Einsatz der Methode der strukturierten Inhaltsanalyse [46-48]. Die Fragestellungen wurden in Kategorien und Unterkategorien eingeteilt. Jeder Unterkategorie wurde eine passende Kovariable aus DizzyReg zugeordnet.

Die in der ersten Studie verwendeten Verfahren der Deep Neural Networks [13] haben eine wichtige Einschränkung, sie können nicht die geforderte Transparenz der Klassifizierungsentscheidung bieten. Deshalb kamen in dieser Studie Modelle auf Basis von Entscheidungsbäumen zum Einsatz. Das Modell der Classification and Regression Trees [25] erstellt aus den Trainingsdaten Entscheidungsbäume, die visualisiert werden können und damit „lesbar“ sind. Allerdings verwenden CART nur eine Untermenge der Kovariablen im Trainingsdatensatz für die Klassifizierung und bieten deshalb nicht die Möglichkeit, die Relevanz für alle Variablen zu ermitteln. Dieser Nachteil wurde mit dem Modell der „Random Forests“ [26] und mit Verfahren zur Ermittlung der relativen Wichtigkeit der Kovariablen ausgeglichen [49, 50]. Es wurden zwei weitere Verfahren verwendet.

Der Doktorand hat in der Studie, nachdem der Datensatz aufbereitet und verfügbar war, folgende Aufgaben durchgeführt:

- Implementierung und Testung der CART und RF Modelle, inklusive der Vorbereitung der Daten für das Training
- Auswahl und Recherche möglicher Visualisierungslösungen für CART, sowie der Implementierung der Visualisierung
- Recherche und Auswahl der möglichen Verfahren zur Ermittlung der Relevanz der Variablen mit einem Random Forest
- Implementierung und Testung einer Klasse an Verfahren (z.B.: Permutation Importance) [49-51]

### 3. Zusammenfassung

Mit einer Lebenszeitprävalenz von etwa 30% und steigender Inzidenz mit dem Alter ist Schwindel eines der häufigsten Leitsymptome und stellt für Patient:innen eine schwere Beeinträchtigung des täglichen Lebens dar [1]. Die oft unklare Symptomlage und fehlende Erfahrung im hausärztlichen Bereich führt häufig zu falscher Einordnung, somit zu erfolglosen Behandlungsversuchen [2]. Die Diagnose wird dadurch erschwert, dass sich häufig Symptome verschiedener Schwindelerkrankungen überlagern [4]. Somit ist es von Interesse, Ärzt:innen Möglichkeiten an die Hand zu geben, dass sie Schwindelerkrankungen in einem ersten Schritt zuverlässiger einordnen können. Die ersten Ansätze für die Klassifizierung von Schwindelerkrankungen mit den anfangs verfügbaren einfachen Modellen waren jedoch nicht vielversprechend genug, um weiter verfolgt zu werden [9].

Die Fragestellung dieser Arbeit war es daher, zu untersuchen, ob diese Verfahren in ihren weiterentwickelten Formen anhand symptomorientierter Charakteristika der Patient:innen dazu in der Lage sind, vestibuläre Erkrankungen zu differenzieren. Es war zudem wichtig, welche vestibulären Symptome besonders relevant für die Vorhersagegüte sein könnten.

In beiden Studien wurden Daten aus der DizzyReg Patientendatenbank des Deutschen Schwindel- und Gleichgewichtszentrum verwendet [14]. Aus der Vielzahl der Machine Learning Modelle [13, 18] fiel die Wahl für die Differenzierung von Morbus Menière und vestibulärer Migräne, aufgrund ihrer Leistung in anderen medizinischen Anwendungen [9], auf die Modelle Deep Neural Networks [13] und der Boosted Decision Trees [29]. Für die Klassifizierung mit möglichst großer Transparenz zur Entscheidung und die Ermittlung der Relevanz der Variablen wählten wir Classification and Regression Trees und Random Forests aus. CART bieten den Vorteil, dass sie eine visuelle Darstellung bereitstellen, die die menschliche Entscheidungsfindung nachbildet [25].

Gemittelt über alle fünf imputierten Datensätze ergab sich mit einem DNN für MM ein F-Measure von 55,5% (Accuracy 91,4%). Zum Vergleich erreichten wir mit DNN für VM ein F-Measure von 36,8% (Accuracy 81,8%). Boosted Decision Trees, trainiert auf VM, ergaben lediglich ein F-Measure von 27,6% und eine Accuracy von 84,5%. Aus den Experimenten mit den CART wurden acht Variablen als relevant ermittelt. Als übergeordnete Genauigkeit bei der Klassifizierung aller sieben Diagnosen ergab eine Genauigkeit von 42,2%.

DNN lieferten sehr gute Klassifizierungsergebnisse, aber aufgrund der inhärenten Struktur des Modells nicht die geforderte Transparenz. Die bei der gleichzeitigen Klassifizierung der häufigsten Diagnosen verwendeten Algorithmen basieren auf einem transparenten Ansatz. Sie können für eine initiale Triage der Patient:innen verwendet werden, aber es muss sich eine klinische Untersuchung der vestibulären und okularen Funktionen anschließen, um die Genauigkeit der Diagnose zu verbessern. Für eine abschließende Beurteilung ist noch weitere Forschung notwendig.

## 4. Abstract (English)

With a lifetime prevalence of about 30% and an increasing incidence with age, dizziness is one of the most common symptoms and represents a severe impairment of daily life for patients [1]. The often-unclear symptoms and lack of experience with general practitioners often lead to incorrect classification and thus to unsuccessful attempts at treatment [2]. The diagnosis is complicated by the fact that symptoms of different vertigo diseases often overlap [4]. Therefore, it is of interest to provide general practitioners with possibilities to classify dizziness more reliably in a first step. However, the initial approaches to classify vertigo disorders with the simple models available at the beginning were not promising enough to be pursued further [9].

The question of this work was therefore to investigate whether these methods in their more developed forms can differentiate vestibular disorders based on symptom-oriented patient characteristics. It is also important to determine which vestibular symptoms might be particularly relevant to predictive accuracy.

In both studies, data from the DizzyReg patient database of the German Schwindel- und Gleichgewichtszentrum were used [14]. From the multitude of machine learning models [13, 18], the choice for differentiating Meniere's disease and vestibular migraine fell on the Deep Neural Networks and Boosted Decision Trees [29] models, due to their performance in other medical applications [9]. We selected Classification and Regression Trees for classification with the greatest possible transparency and for determining the relevance of variables. They offer the advantage of providing a visual representation that mimics human decision making [25].

Averaged over all five imputed datasets, a DNN for Menière's Disease yielded an F-measure of 55.5% (accuracy 91.4%). In comparison, with DNN for Vestibular Migraine we achieved an F-Measure of 36.8% (Accuracy 81.8%). Boosted decision trees trained on Vestibular Migraine yielded only an F-Measure of 27.6% and an accuracy of 84.5%. Eight variables were determined to be relevant from the experiments with the CART. The overall accuracy when classifying all seven diagnosis yielded a result of 42.2%.

DNN achieved very good classification results but did not provide the required transparency due to the inherent structure of the model. The algorithms used in the classification of the most common diagnoses are based on a transparent approach. They can be used for initial triage of patients but must be followed by clinical examination of vestibular and ocular functions to improve the accuracy of the diagnosis. More research is needed to make a final assessment.

## 5. Information zu den Publikationen

### 5.1 Publikationen als Grundlage dieser Dissertation

**Publikation I:** “Development and validation of a classification algorithm to diagnose and differentiate spontaneous episodic vertigo syndromes: results from the DizzyReg patient registry”

Grözinger, Michael; Huppert, Doreen; Strobl, Ralf; Grill, Eva (2020) Development and validation of a classification algorithm to diagnose and differentiate spontaneous episodic vertigo syndromes: results from the DizzyReg patient registry

In: Journal of Neurology (2020)

Impact factor 2019/2020: 3.956

<https://link.springer.com/article/10.1007/s00415-020-10061-9>

**Publikation II:** “A set of eight key questions helps to classify common vestibular disorders – results from the DizzyReg patient registry”

Strobl, Ralf, Grözinger, Michael; Zwergal, Andreas; Huppert, Doreen; Filippopoulos, Philipp; Grill, Eva (2021) A set of eight key questions helps to classify common vestibular disorders – results from the DizzyReg patient registry

In: Frontiers of Neurology (2021)

Impact factor 2019/2020: 0.733

<https://www.frontiersin.org/articles/10.3389/fneur.2021.670944/full>

### 5.2 Weitere Publikationen des Doktoranden

Grill, Eva; Grözinger, Michael; Feil, Katharina; Strupp, Michael (2016) [Developing and Implementing Diagnostic Prediction Models for Vestibular Diseases in Primary Care](#)

In: Studies in Health Technology and Informatics, 2016. 228: p. 735-9

Grill, Eva; Grözinger, Michael; Feil, Katharina; Strupp, Michael (2016) [Developing and Implementing Diagnostic Prediction Models for Vestibular Diseases in Primary Care](#). In: Grill, E., M.

Muller, and U. Mansmann, [Health-exploring complexity: an interdisciplinary systems approach HEC 2016: 28 August-2 September 2016, Munich, Germany](#). Eur J Epidemiol, 2016. **31 Suppl**

1: p. 1-239. [52] (siehe Seite S70, Abstract #135)

## 6. Abbildungsverzeichnis

Abbildung 1	Schematische Darstellung eines Deep Neural Network mit 2 Hidden Layer und K Knoten in der ersten Hidden Layer sowie L Knoten in der zweiten Hidden Layer. ....	10
-------------	--	----

## 7. Literaturverzeichnis

1. Brandt, T., M. Dieterich, and M. Strupp, *Vertigo - Leitsymptom Schwindel*. 2012: Springer Berlin Heidelberg.
2. Grill, E., M. Penger, and E. Kentala, *Health care utilization, prognosis and outcomes of vestibular disease in primary care settings: systematic review*. J Neurol, 2016. **263 Suppl 1**: p. S36-44.
3. Kovacs, E., X. Wang, and E. Grill, *Economic burden of vertigo: a systematic review*. Health Econ Rev, 2019. **9**(1): p. 37.
4. Neuhauser, H. and T. Lempert, *Vertigo and dizziness related to migraine: a diagnostic challenge*. Cephalalgia, 2004. **24**(2): p. 83-91.
5. Brandt, T., M. Strupp, and M. Dieterich, *Five keys for diagnosing most vertigo, dizziness, and imbalance syndromes: an expert opinion*. J Neurol, 2014. **261**(1): p. 229-31.
6. Lopez-Escamez, J.A., et al., *Accompanying Symptoms Overlap during Attacks in Meniere's Disease and Vestibular Migraine*. Front Neurol, 2014. **5**: p. 265.
7. Bayer, O., J.C. Warninghoff, and A. Straube, *Diagnostic indices for vertiginous diseases*. BMC Neurol, 2010. **10**: p. 98.
8. Friedland, D.R., et al., *Development of a Statistical Model for the Prediction of Common Vestibular Diagnoses*. JAMA Otolaryngol Head Neck Surg, 2016. **142**(4): p. 351-6.
9. Krittanawong, C., et al., *Deep learning for cardiovascular medicine: a practical primer*. Eur Heart J, 2019. **40**(25): p. 2058-2073.
10. Juhola, M., *On machine learning classification of otoneurological data*. Stud Health Technol Inform, 2008. **136**: p. 211-6.
11. Varpa, K., et al., *Applying one-vs-one and one-vs-all classifiers in k-nearest neighbour method and support vector machines to an otoneurological multi-class problem*. Stud Health Technol Inform, 2011. **169**: p. 579-83.
12. Exarchos, T.P., et al., *Mining balance disorders' data for the development of diagnostic decision support systems*. Comput Biol Med, 2016. **77**: p. 240-8.
13. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. Adaptive computation and machine learning. 2016, Cambridge, Massachusetts: The MIT Press. xxii, 775 pages.
14. Grill, E., et al., *DizzyReg: the prospective patient registry of the German Center for Vertigo and Balance Disorders*. J Neurol, 2017. **264**(Suppl 1): p. 34-36.
15. Grill, E., et al., *Multicenter data banking in management of dizzy patients: first results from the DizzyNet registry project*. J Neurol, 2018. **265**(Suppl 1): p. 3-8.
16. Lopez-Escamez, J.A., et al., *[Diagnostic criteria for Meniere's disease. Consensus document of the Barany Society, the Japan Society for Equilibrium Research, the European Academy of Otology and Neurotology (EAONO), the American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS) and the Korean Balance Society]*. Acta Otorrinolaringol Esp, 2016. **67**(1): p. 1-7.
17. Lempert, T., et al., *Vestibular migraine: diagnostic criteria*. J Vestib Res, 2012. **22**(4): p. 167-72.
18. Hastie, T., R. Tibshirani, and J.H. Friedman, *The elements of statistical learning : data mining, inference, and prediction*. 2nd ed. Springer series in statistics,. 2009, New York, NY: Springer. xxii, 745 p.
19. Sucar, L.E., *Bayesian Networks: Learning*, in *Probabilistic Graphical Models: Principles and Applications*, L.E. Sucar, Editor. 2015, Springer London: London. p. 137-159.
20. Lopez-Escamez, J.A., et al., *Diagnostic criteria for Meniere's disease*. J Vestib Res, 2015. **25**(1): p. 1-7.

21. Buuren, S.v., *Flexible imputation of missing data*. Second edition. ed. Chapman and Hall/CRC interdisciplinary statistics series. 2018, Boca Raton: CRC Press, Taylor & Francis Group. xxvii, 415 pages.
22. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Wiley series in probability and mathematical statistics Applied probability and statistics,. 1987, New York ;: Wiley. xxix, 258 p.
23. Rousseau, R., *The F-measure for research priority*. Journal of Data and Information Science, 2018. **3**(1): p. 1-18.
24. James, G., et al., *An introduction to statistical learning : with applications in R*. Springer texts in statistics,. 2013, New York: Springer. xvi, 426 pages.
25. Breiman, L., *Classification and regression trees*. The Wadsworth statistics/probability series. 1984, Belmont, Calif.: Wadsworth International Group. x, 358 p.
26. Breiman, L., *Random Forests*. Machine Learning, 2001. **45**(1): p. 5-32.
27. Varpa, K., K. Iltanen, and M. Juhola, *Machine learning method for knowledge discovery experimented with otoneurological data*. Comput Methods Programs Biomed, 2008. **91**(2): p. 154-64.
28. Rifkin, R. and A. Klautau, *In Defense of One-Vs-All Classification*. J. Mach. Learn. Res., 2004. **5**: p. 101–141.
29. Friedman, J.H., *Recent Advances in Predictive (Machine) Learning*. Journal of Classification, 2006. **23**(2): p. 175-197.
30. Jiang, A.P.P.B.Y., *randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance*. 2020.
31. Priesol, A.J., et al., *Clinical vestibular testing assessed with machine-learning algorithms*. JAMA Otolaryngol Head Neck Surg, 2015. **141**(4): p. 364-72.
32. Neff, B.A., et al., *Auditory and vestibular symptoms and chronic subjective dizziness in patients with Meniere's disease, vestibular migraine, and Meniere's disease with concomitant vestibular migraine*. Otol Neurotol, 2012. **33**(7): p. 1235-44.
33. Dieterich, M. and T. Brandt, *Episodic vertigo related to migraine (90 cases): vestibular migraine?* J Neurol, 1999. **246**(10): p. 883-92.
34. Neugebauer, H., et al., *Long-term changes of central ocular motor signs in patients with vestibular migraine*. Eur Neurol, 2013. **69**(2): p. 102-7.
35. Radtke, A., et al., *Vestibular migraine--validity of clinical diagnostic criteria*. Cephalalgia, 2011. **31**(8): p. 906-13.
36. Radtke, A., et al., *Vestibular migraine: long-term follow-up of clinical symptoms and vestibulo-cochlear findings*. Neurology, 2012. **79**(15): p. 1607-14.
37. Groezinger M, et al., *Development and validation of a classification algorithm to diagnose and differentiate spontaneous episodic vertigo syndromes: results from the DizzyReg patient registry*. J Neurol, 2020.
38. Grill, E., et al., *Developing and Implementing Diagnostic Prediction Models for Vestibular Diseases in Primary Care*. Stud Health Technol Inform, 2016. **228**: p. 735-9.
39. Samek, W., et al., *Explainable AI : interpreting, explaining and visualizing deep learning, in Lecture notes in artificial intelligence*. 2019, Springer,: Cham. p. 1 online resource (xi, 439 pages).
40. Bach, S., et al., *On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation*. PLoS One, 2015. **10**(7): p. e0130140.
41. Hinton, G.E., S. Osindero, and Y.W. Teh, *A fast learning algorithm for deep belief nets*. Neural Comput, 2006. **18**(7): p. 1527-54.
42. Hinton, G.E., *Training products of experts by minimizing contrastive divergence*. Neural Comput, 2002. **14**(8): p. 1771-800.

43. Wolpert, D.H. and W.G. Macready, *No free lunch theorems for optimization*. IEEE Transactions on Evolutionary Computation, 1997. **1**(1): p. 67-82.
44. Filippopoulos, F.M., et al., *Computerized clinical decision system and mobile application with expert support to optimize management of vertigo in primary care: study protocol for a pragmatic cluster-randomized controlled trial*. J Neurol, 2020. **267**(Suppl 1): p. 45-50.
45. van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate Imputation by Chained Equations in R*. Journal of Statistical Software; Vol 1, Issue 3 (2011), 2011.
46. Mayring, P., *Qualitative Content Analysis*, in *A Companion to Qualitative Research* E. Kardorff, Steinke, I., Flick, U., Editor. 2004, SAGE Publications: London. p. 266-269.
47. Flick, U., *An Introduction to Qualitative Research*. 4 ed. 2009, London: SAGE Publications.
48. Schmidt, C., *The Analysis of Semi-structured Interviews*, in *A Companion to Qualitative Research* E. Kardorff, Steinke, I., Flick, U., Editor. 2004, SAGE Publications: London. p. 253-258.
49. Strobl, C., et al., *Bias in random forest variable importance measures: Illustrations, sources and a solution*. BMC Bioinformatics, 2007. **8**(1): p. 25.
50. Louppe, G., et al., *Understanding variable importances in forests of randomized trees*. 2013: p. 431--439.
51. Strobl, C., et al., *Conditional variable importance for random forests*. BMC Bioinformatics, 2008. **9**: p. 307.
52. Grill, E., M. Muller, and U. Mansmann, *Health-exploring complexity: an interdisciplinary systems approach HEC 2016: 28 August-2 September 2016, Munich, Germany*. Eur J Epidemiol, 2016. **31 Suppl 1**: p. 1-239.

## 8. Danksagung

Mein erster und besonderer Dank gilt Frau Prof. Dr. Eva Grill für die Möglichkeit, nach langer Zeit in der beruflichen Praxis, wieder in die akademische Welt zurückzukehren und diese Arbeit anzufertigen zu können. Ich danke für ihre Geduld, ihre kontinuierliche Unterstützung mit inspirierenden und konstruktiven Diskussionen.

Dr. Ralf Strobl gebührt Dank für viele hilfreiche Impulse und die ausgezeichnete Unterstützung in allen Bereichen der verwendeten Daten und deren Statistik. Bei der Zusammenarbeit zur zweiten Publikation hat er mit ausgezeichnetem Fachwissen meinen Beitrag unterstützt und in die Arbeit eingebunden.

Frau Prof. Dr. Doreen Huppert und PD Dr. Andreas Zwergal unterstützten diese Arbeit mit ihrem ausgezeichneten Wissen und ihrem großen Erfahrungsschatz aus der Arbeit im Schwindelzentrum des Klinikums Großhadern. Ich bedanke mich bei Ihnen für diesen großartigen und wertvollen Beitrag.

Ebenso bedanke ich mich bei den vielen unbekanntenen Patientinnen und Patienten, die die Daten ihrer Krankheitsgeschichte für die Forschung zur Verfügung gestellt haben.

Von meiner Frau Katrin Holzäpfel und meiner Tochter Anne Lena bekam ich großartige Unterstützung, viel Rückmeldung und die Motivation diese Arbeit erfolgreich abzuschließen.