

Literatur-Seminar-Arbeit

Vorhersagen von Verkehrsunfällen mithilfe künstlicher neuronaler Netze

Erik Rohr

Fachbereich Informatik (02) - Hochschule Bonn-Rhein-Sieg

9. Juli 2024

Betreuerin: Doerthe Vieten

Abstrakt

« Kurz beschreiben »

Inhaltsverzeichnis

1	Einleitung	4
1.1	Theorie	4
1.1.1	Künstliche neuronale Netze (KNN)	4
1.2	Historische Forschungslage	5
1.2.1	Smith et al.	5
1.2.2	Edmond Chin-Ping Chang	5
1.3	Aktuelle Forschungslage	6
1.3.1	Banerjee et al.	6
1.3.2	Maurya et al.	6
1.3.3	Zohra et al.	6
2	Methodik	6
2.1	Die PRISMA-Leitlinie	7
2.2	Forschungsfragen	7
2.3	In- und Exklusionskriterien	7
2.4	Prozess der Recherche	8
3	Ergebnisse	8
3.1	Qian et al.	8
3.2	Das et al.	9
3.3	Bao et al.	9
4	Diskussion und Fazit	11
4.1	Diskussion	11
4.2	Limitationen	12
4.3	Fazit	13
4.4	Ausblick auf zukünftige Forschung	13

1 Einleitung

Jedes Jahr sterben weltweit ca. 1,19 Millionen Menschen in Verkehrsunfällen (RTA, engl.: „road traffic accident“) [1]. Eine umfassende Analyse vorliegender Verkehrsdaten im Hinblick auf potentielle Bedrohungen kann dazu beitragen, Baumaßnahmen unfallminimierend zu gestalten.

Vorangehende Analysen von Verkehrsdaten, die eine Vielzahl an „Machine Learning“-Modelle nutzten, haben ergeben, dass eine Vielzahl an Faktoren existieren, die das Risiko auf RTA erhöhen, wie Wetterbedingungen, Straßenkonditionen, Zustand des Fahrers, Lichtverhältnisse, Tageszeit und die Verkehrsdichte [2, 3]. „Machine Learning“-Modelle (ML) sind darauf ausgelegt, nicht anhand von spezifischen Anweisungen, sondern allein durch das Erkennen von Mustern und Abhängigkeiten mithilfe komplexer Funktionen [3] in Daten Erkenntnisse zu ziehen und Entscheidungen sowie Vorhersagen treffen zu können [4].

Unter den ML-Modellen werden künstliche neuronale Netze (KNN) für diesen Anwendungsfall bevorzugt, da diese keine zugrundeliegenden Beziehungen zwischen den Eingangsvariablen benötigen und darauf ausgelegt sind, mit historischen Daten Vorhersagen zu tätigen [3].

In dieser Arbeit wird eine systematische Literaturrecherche (SLR) zur Vorhersage von RTAs mithilfe KNNs dargestellt. Die Ergebnisse werden im Hinblick auf deren Mehrwert in der Minimierung von Straßenverkehrsunfällen durch unsichere Planung der Architektur untersucht und eingeordnet.

1.1 Theorie

Zunächst werden die theoretischen Grundlagen sowie die Forschungslage von KNN ausgeführt, um die Ergebnisse der Literaturrecherche in den Kontext einordnen zu können.

1.1.1 Künstliche neuronale Netze (KNN)

Künstliche neuronale Netze (KNN) sind ein Teilbereich der ML-Modelle. Sie können Entscheidungen auf ähnlicher Art und Weise treffen, wie das menschliche Gehirn [5]. Das Modell besteht aus einer Vielzahl von Schichten mit unterschiedlich vielen Knoten. Diese Knoten sind mit anderen Knoten aus den benachbarten Schichten verbunden und besitzen sogenannte „weights“ und „bias“ [5]. Eingabedaten werden von der „input layer“ (Deutsch: „Eingangsschicht“) entlang aller hidden layer (Deutsch: „verborgene Schicht“) durchgereicht, verarbeitet und schließlich in der „output layer“ (Deutsch:

„Ausgangsschicht“) aggregiert ausgegeben [5].

Zur Konfigurierung von KNN können unter anderem die Anzahl der Schichten, die Optimierungsverfahren (engl.: „optimizers“) (OV), die Aktivierungsfunktion (engl.: „activation function“) (AF) und die Verlustsfunktion (engl.: „loss function“) frei gewählt werden [6].

Ein Optimierungsverfahren ist eine Funktion oder ein Algorithmus, welches die Gewichte der Knoten und die Lernrate (engl.: „learning rate“) anpasst, um den Verlust zu minimieren, die Genauigkeit zu maximieren und die benötigte Trainingszeit exponentiell zu reduzieren [6]. Beispiele für solche Optimierungsverfahren sind der „gradient descent“, „RMS prop (Root Mean Square)“ und „adam’s optimizer“ [6].

Die AF entscheidet, wie der kalkulierte Ausgangswert eines Knotens zu interpretieren ist. Dies fügt dem Netz eine Nicht-Linearität hinzu [6]. Somit können die verborgenen Schichten jeweils für andere Bereiche der Interpretation der Daten stehen. Linearität würde auf der anderen Seite bedeuten, dass das Netz in eine Funktion zusammengefasst werden könnte [6]. Beispiele sind die binäre „step function“, die Sigmoid-Funktion und die „RELU“-Funktion [6].

Eine Verlustsfunktion (VF) misst, wie genau das KNN den Datensatz modelliert, indem es den vorliegenden Verlust (engl.: „loss“) berechnet [6]. Gängige Funktionen sind die „Mean Squared Error“-Funktion und die „Binary Cross Entropy“-Funktion [6].

1.2 Historische Forschungslage

1.2.1 Smith et al.

Die Anwendung eines neuronalen Netzes ist an einer Simulation einer „nine intersection“ in Manhattan, New York untersucht worden. Es ergab sich eine 10-prozentige Verringerung in der Wartezeit von Fahrzeugen im Vergleich zur Nutzung der bereits existierenden, „in-place“ Strategie [7].

1.2.2 Edmond Chin-Ping Chang

Die Anwendung von einem neuronalen Netz auf „traffic engineering“ wird mithilfe verschiedener Sensoren und historischer Daten erforscht. Die Effektivität wird mit einem konventionellen Algorithmus verglichen und anschließend werden zwei neuronale Netze trainiert, um die Herangehensweisen beider Mittel zu kombinieren ([8] S. 642). Man hat herausgefunden, dass trotz dem Verlust geringfügiger Genauigkeit das Zwei-Neuronale-Netz-Modell sehr geeignet für das Verarbeiten von Verkehrsunfallsdaten ist ([8] S. 645).

1.3 Aktuelle Forschungslage

1.3.1 Banerjee et al.

Eine Vielzahl an Klassifizierern (engl.: „classifiers“) sind mit KNN-Modellen bezüglich Vorhersage der Mortalität in RTAs verglichen worden. Darunter gehören Random Forest, Support Vector Machine, K-Nearest Neighbor Classifier, AdaBoost Classifier, XGBoost Classifier. Hierbei hat das KNN mit 7 „hidden layer“, 1 „input layer“, Adams Optimizer und einer „dropout class“ in der Eingangsschicht die beste Genauigkeit von 84,36% erzielt [9].

1.3.2 Maurya et al.

Vorhersehen von Kraftfahrzeuggeschwindigkeiten dienen als Basis für ein fortgeschrittenes Verkehrsmanagementsystem [10]. Hierbei wurden Lineare Regression, „random forest“-Regression, „Decision Tree“ und ein KNN miteinander Vergleichen und eine „Performance Analyse“ durchgeführt. KNN und die „random forest“-Regression haben mit einem „R squared“-Wert ¹ von jeweils 0,9301 und 0,9642 das beste Fitting erzielt [10].

1.3.3 Zohra et al.

Anhand den Datensätzen aus „US Accidents (2016-2023)“ wurden ein KNN, ein Random Forest Klassifizierer und eine logistische Regression trainiert und getestet. Hier wurden jeweils Genauigkeitswerte von 81,1%, 90,7% und 87,03% erreicht [12].

2 Methodik

Im Folgenden wird die Vorgehensweise bei der Literaturrecherche beschrieben. Die Vorgehensweise orientiert sich an der PRISMA-Leitlinie (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). PRISMA hat sich ursprünglich als effektiver Vorreiter bzgl. SLR im medizinischen Sektor etabliert, mit dem Fachkräfte auf dem aktuellen Stand der Wissenschaft bleiben und bestehende Vorschriften aktualisiert werden [13]. Entnehmend der zahlreichen SLR-Veröffentlichungen in Bereichen der Informatik (siehe Datenbanken ACM/IEEE), hat sich PRISMA als hilfreiches Mittel zur SLR auch in anderen Wissenschaftsbereichen bewiesen und wird daher als Instrument der SLR in dieser Arbeit verwendet.

¹Bestimmtheitsmaß. Stellt die Anpassungsgüte einer Regression dar [11]

2.1 Die PRISMA-Leitlinie

PRISMA beinhaltet eine aus 27 Stichpunkten bestehende „Checkliste“ und ein 4-phasiges Fluss-Diagramm [13]. Da diese allerdings ursprünglich für die Medizin entwickelt wurden, werden jene auf die Informatik angepasst als Basis der SLR verwendet. Mithilfe dieser Elemente kann sowohl Literatur systematisch für die Aufnahme in einer SLR auf Eignung geprüft als auch der Prozess erleichtert und standardisiert werden.

Zunächst werden Forschungsfragen aufgestellt, die im Laufe der Recherche beantwortet werden sollen. Im Anschluss wird dann die Literatur anhand der ausgewählten Such-Strategie ausgewählt. Die gesammelte Literatur wird dann durch Ein- und Ausschlusskriterien auf Eignung für die SLR durch die Checkliste geprüft. Schließlich werden aus der eingeschlossenen Literatur qualitative und quantitative Daten extrahiert, mit anderer Literatur verglichen und in der eigentlichen Literaturanalyse zusammengetragen.

2.2 Forschungsfragen

Um die Literaturrecherche zu systematisieren, werden laut PRISMA [13] Forschungsfragen aufgestellt, damit ein systematisches Full-Text-Screening für die Eignung möglich ist. In Rahmen dieser Arbeit wurden folgende Fragen aufgestellt:

1. Welche Netztopologie sind verwendet worden?
2. Welche Genauigkeit konnte erzielt werden?
3. Wie viele Knoten sind in den verborgenen Schichten verwendet worden?
4. Wie ist der Lernprozess realisiert worden? („optimizer“, „loss function“ usw.) ?
5. Welche Eingabevariablen sind gewählt worden?

2.3 In- und Exklusionskriterien

Um gezielt die Potenz künstlicher neuronaler Netze für die Verkehrsanalyse zu untersuchen, sind spezifische Ein- und Ausschlusskriterien gewählt worden. Somit wird die Anzahl der aus der Suche ergebnen Artikel reduziert und verschafft einen besseren Überblick.

Die für diese Arbeit relevanten Artikel sind auf solche beschränkt worden, die

1. eine Verkehrsdatenanalyse untersuchen und die Daten mit einer KNN auswerten,
2. ggf. einen quantitativen Vergleich verschiedener ML-Modelle durchführen oder

3. quantitative Merkmale der KNN aufzählen.

Zu letzterem Punkt gehören Merkmale wie gewählte Netztopologie, Anzahl der Schichten, Anzahl der Knoten pro Schicht, gewähltes Optimierungsverfahren, die gewählte AF Genauigkeit und Eingangsvariablen.

2.4 Prozess der Recherche

Für die Literaturrecherche sind die Datenbanken ACM Digital Library und die IEEE Explore verwendet worden, da diese überwiegend Artikel in der Informatik veröffentlicht haben. Hierbei wurden die folgenden Suchstrings verwendet:

ACM Digital Library

```
[All: "neural network?"] AND [ [All: "traffic flow"] OR [All: "traffic control"] OR  
[All: "accident"] ] AND [E-Publication Date: (01/01/2023 TO 12/31/2024)]
```

IEEE Explore

```
("All Metadata": "artificial neural network?") AND ("All Metadata:"traffic")  
AND ("All Metadata:"control" OR "All Metadata:"accident")
```

Bei der „IEEE-Explore“-Datenbank wurde als Zeitraum-Filter der 01.01.2023 bis zum 31.12.2024 gewählt, da dieser nicht in den Such-String mit eingebunden werden konnte. Zusätzlich ist die Suche auf den Inhaltstyp „Research Article“ und die Verfügbarkeit „Open Access“ begrenzt worden, soweit von den Einstellungen beider Datenbanken möglich.

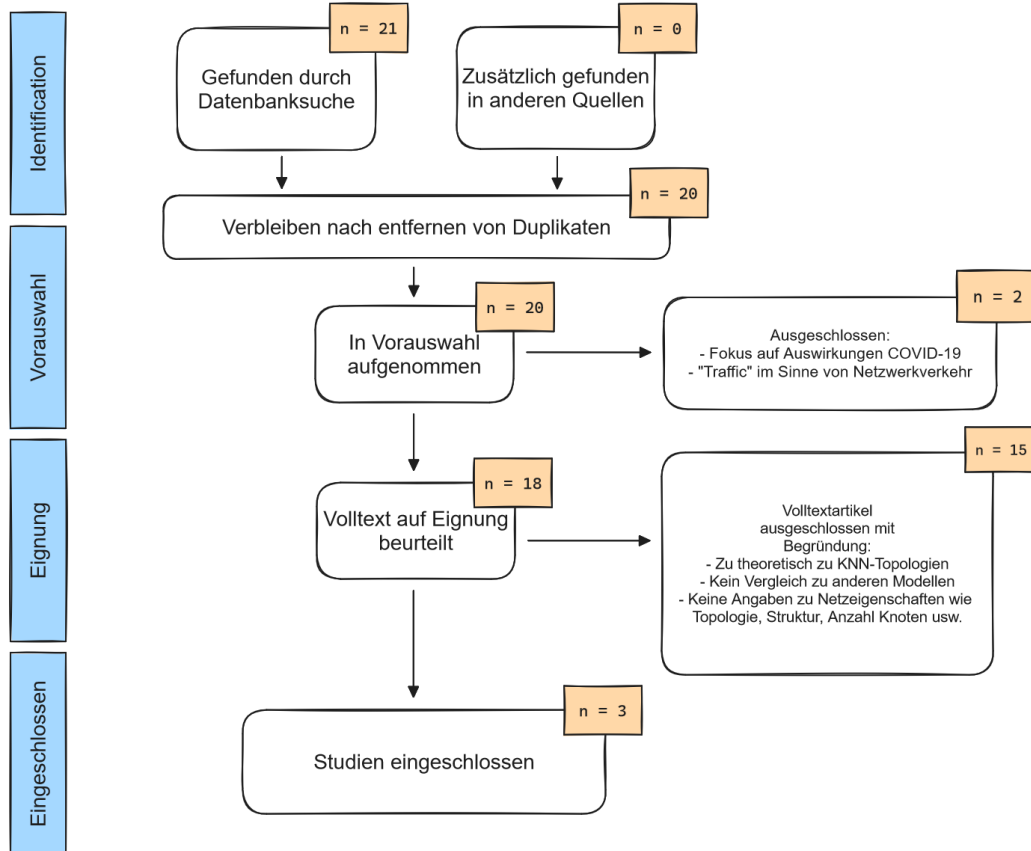
Die zusammengetragenen Artikel wurden im Anschluss mithilfe des PRISMA Fluss-Diagramms [13] auf Eignung überprüft (Abb. 1) und in einer Tabelle zusammengefasst (Tab. 1).

3 Ergebnisse

3.1 Qian et al.

Qian et al. haben eine Kombination aus einem „classification and regression trees“ (CART) und einem „back propagation neural network“ (BPNN) entwickelt und mit einer logistischen Regression, AdaBoost und dem naiven Bayes Klassifizierern verglichen. Die

Abbildung 1: SLR nach PRISMA



Untersuchung ergab, dass das CART-BPNN-Modell mit einer Genauigkeit von 78,56% am besten abschnitt [3].

3.2 Das et al.

In dieser Studie sind verschiedene KNN-Modelle verschieden-parametrisiert verglichen worden. Hierbei wurden Modelle mit unterschiedlichen OV's und AF's trainiert und ausgewertet. Das Modell mit einer Softmax-AF und Adams Optimierungsverfahren erzielte die höchste Genauigkeit von 94,62% [2].

3.3 Bao et al.

Ein KNN mit einer und zwei verborgenen Schicht(en), allerdings mit einer insgesamt gleichbleibenden Anzahl an verborgenen Knoten, sind miteinander verglichen worden. Der Vergleich ergab, dass zwar das Modell mit zwei Schichten geringfügig besser ab-

schnitt als das Modell mit einer Schicht, allerdings um mehr als das 20-fache an Trainingszeit benötigte. Daher ist man zum Schluss gelangt, dass eine verborgene Schicht für diesen Anwendungsfall besser geeignet ist.

Tabelle 1: Ergebnisse der Literaturrecherche

Autoren	Qian et al. [3]	Das et al. [2]	Bao et al. [14]
Topologie	BPNN	FFNN	BPNN
AF	Softmax	Softmax	Sigmoid
VF	Cross-Entropy	Cross-Entropy	Levenberg Marquardt
Verborg. Knoten	3	o.A.	388
Lernrate	0.01	o.A.	0,005
OV	Gradient Descent	Adam	o.A.
Genauigkeit	78,56%	94,62%	69,95%

Anmerkung: „o.A.“ = „ohne Angabe“

Tabelle 2: Eingabevariablen der KNN-Modelle

Autoren	Eingabevariablen
Qian et al.	Minute, Stunde, Jahr, Monat, Tag, Polizeieinsatz, Gebietsart, Fahrbahngefahren, Straßensonderbedingungen, Straßenoberflächenbedingungen, Wetterbedingungen, Lichtverhältnisse, Fußgängerüberweg vorhanden, Straßenummer und -Klasse, Kreuzungsverwaltung, Geschwindigkeitsbegrenzung, Straßentyp, Wochentag, Anzahl Verunglückte, Anzahl Fahrzeuge, Längen- und Breitengrad
Das et al.	Jahreszeit, Fahrzeugart des Unfallbeteiligten, Gebietsart, Aufprallart, Unfallsart, Alter des Beteiligten „a.m. peak“, „a.m. off-peak“, „p.m. peak“, „p.m. off-peak“
Bao et al.	Kreuzungsart, Unfallsortart, Straßentyp, Mittelstreifenart, Nicht-verkehrsbezogene Mittelstreifenart, Straßenmaterial, Straßenkonditionen, Bürgersteigkonditionen, Lichtverhältnisse, Wetterbedingungen, Sichtweite, Unfallsart

Die Ergebnisse der Studien fallen sehr unterschiedlich aus. Sowohl Qian et al. als auch Das et al. befürworten die Nutzung von KNN in der Durchführung und Auswertung von Verkehrsanalysen ([3] S. 2744, [2] S. 5), allerdings empfiehlt Bao et al., dass zukünftige Studien sich intensiver mit den einzelnen Faktoren der Verkehrsanalyse befassen sollten, da Verkehrsanalysen zu komplex sind, diese in einem Modell zusammenzufassen ([14] S.

784).

Diese Rückschlüsse werden auch von den erreichten Genauigkeiten der KNN-Modelle reflektiert: Absteigend sortiert erreichte Das et al. eine Genauigkeit von 94,62%, Qian et al. 78,56% und von Bao et al. einer Genauigkeit von 69,95% [2, 3, 14].

Entnehmend der Tabelle 1 gibt es dennoch einige Gemeinsamkeiten zwischen den Studien. Sowohl Qian et al. als auch Bao et al. verwendeten für ihre Netztopologie ein BPNN ([3] S. 2739, [14] S. 782), wohingegen Das et al. als besten Kandidat ein „feed forward neural network“ (FFNN) nominierte ([2] S. 2). Bezüglich den Aktivierungs- und Verlustfunktionen verwenden Qian et al. und Das et al. jeweils die Softmax-AF und die Cross-Entropy-VF ([3] S. 2742, [2] S. 2). Bao et al. haben hier hingegen eine Sigmoid-Funktion als AF und den „Levenberg Marquardt“-Algorithmus als VF verwendet ([14] S. 783). Sowohl in der gewählten Lernrate, dem gewählten Optimierungsverfahren und den gewählten Eingabevariablen unterscheiden sich die drei Studien maßgeblich. Qian et al. nutzten 3 verborgene Knoten mit einer Lernrate von 0,01, während Bao et al. 388 verborgene Knoten ([14] S. 783) und eine Lernrate von 0,005 für ihre Analysen verwendeten. Das et al. hat zu beiden Attributen keine Werte angegeben.

Der Tabelle 2 entnehmend sind die Eingabevariablen der Modelle divers. Qian et al., Das et al. und Bao et al. haben hierbei jeweils 31, 6 und 12 verschiedene Variablen für ihre Modelle gewählt.

4 Diskussion und Fazit

Im Rahmen der SLR werden im Folgenden die Ergebnisse der Studien kritisch untersucht, Limitationen sowohl der Studien als auch dieser SLR aufgezeigt und zusätzlich zum Fazit eine Empfehlung für zukünftige Forschungen im Bereich der Verwendung von KNN für Verkehrsanalysen ausgesprochen.

4.1 Diskussion

Alle drei inkludierten Studien sind zu maßgeblich unterschiedlichen Ergebnissen auf unterschiedlichen Wegen gekommen. Das et al. und Bao et al. stechen mit Genauigkeiten von jeweils 94,62% und 69,95% aus den Rechercheergebnissen heraus. Im Rahmen einer Verkehrsanalyse sollten die verwendeten Modelle so zuverlässig wie möglich sein, allerdings besteht die Gefahr, dass das trainierte Modell einem sog. „over-fitting“ (OF) oder „under-fitting“ (UF) unterliegt. OF bezeichnet eine zu starke Anpassung des Modells an die Trainingsdaten, weswegen es neue Daten nicht korrekt auswertet [15]. Dies kann

unter anderem passieren, falls die Trainingsdaten für den gewählten Klassifizierungskontext irrelevante Datensätze beinhalten (auch genannt „verrauschte Daten“), die Größe der Trainingsdaten zu klein, oder die Parameter des Modells zu komplex gewählt worden sind [15]. Das Pendant dazu ist ein Vorliegen eines UF. Hierbei ist das Modell zu simpel und kann Muster dementsprechend nicht auffassen und verinnerlichen [16].

Bao et al. erreichte auf dem Testdatensatz eine Genauigkeit von 69,95% mit einem „1-Hidden-Layer“-Modell. Im Gegensatz zu diesem Datensatz erreichte dieses Modell eine Genauigkeit von 76,1% auf den Trainingsdaten ([14] S. 784). Es ist anzunehmen, dass das Modell mit 388 verborgenen Knoten zu komplex gewählt und daher einem OF unterliegen könnte. Im Vergleich dazu erreichten Das et al. eine Genauigkeit von 94,62%. Die hohe Differenz zu den anderen beiden Studien liegt vermutlich in der Tatsache, dass das Modell auf einer kleineren Menge an Eingabevariablen trainiert wurde (siehe Tabelle 2). Hierbei haben Das et al. mit 10 Eingabevariablen die wohl simpelste Eingangsschicht gewählt. Vergleichsweise haben Bao et al. zwei mehr und Qian et al. das Dreifache an Eingangsvariablen verwendet, wodurch die Komplexität der Modelle ansteigt [15].

Aussagen über die gewählte Netztopologie im Bezug auf die Genauigkeitsdifferenz können aufgrund fehlender Angaben von Das et al. nicht getroffen werden. Es ist unklar, wie Das et al. die verborgene(n) Schichten konfiguriert haben. Das et al. haben eine relativ ähnliche Konfiguration gewählt wie Qian et al., allerdings mehr als 16% höhere Genauigkeit erzielen können. Durch diese Ähnlichkeiten ist unklar, ob die Eingangsvariablen oder die Gestaltung der verborgenen Schichten des KNN bedeutender für die Genauigkeit nach dem Training der Modelle sind.

Bao et al. erläutert, dass ihr Modell aufzeigt, dass die gewählten Eingangsvariablen zu komplex gewählt wurden und weitere Forschung diesbzgl. benötigt sei ([14] S. 784), lässt allerdings die gewählten Netz-Topologie-Parameter außer acht in dieser Ausführung. Interessant wäre zu wissen, wie die Netze mit unterschiedlichen Anzahlen von verborgenen Knoten abschneiden.

4.2 Limitationen

In dieser Arbeit werden lediglich drei Studien verglichen und eignet sich daher nicht, eine allgemeingültige Aussage über die Legitimität künstlicher neuronaler Netze im Bezug auf Verkehrsdatenanalyse zu treffen. Außerdem wurden nur Artikel im Zeitraum vom 01.01.2023 bis zum 01.05.2024 betrachtet, weswegen nur ein kleiner Einblick in die Forschung bereitet werden konnte.

Aufgrund der beschränkten Anzahl der Artikel konnten außerdem keine Aussage über neuere Netztopologien getroffen werden.

4.3 Fazit

Alle drei Studien befürworten die Verwendung von KNN in Verkehrsdatenanalysen, um die Verkehrsplanung sicherer gestalten zu können und dem überliegenden Ziel der Verkehrssicherheit einen Schritt näher zu kommen. Qian et al. verglichen zusätzlich KNN-Modelle mit anderen Regressionsmodellen und kamen zum Schluss, dass KNNs die zuverlässigste Art von ML-Modellen ist, um mit historischen Daten Vorhersagen zu tätigen.

Allerdings besteht bis heute die Herausforderung, optimale Parameter von KNN angepasst auf die Verkehrsanalyse ausfindig zu machen und effektiv einzusetzen.

4.4 Ausblick auf zukünftige Forschung

Es wird empfohlen, näher auf die Parameter der gewählten KNN-Topologien einzugehen und im Falle eines Vergleichs verschiedener Modelle diese auch ausführlich auszuführen. Zukünftige Forschung sollte zusätzlich neuere Topologien erforschen, wünschenswert auch eigenständige Modelle entwickeln.

Literatur

- [1] World Health Organization (WHO). *Road Traffic Injuries*. 2023. URL: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.
- [2] * Surojit Das et al. „Machine Learning Based Approach for Predicting the Impact of Time of Day on Traffic Accidents“. In: *2023 26th International Conference on Computer and Information Technology (ICCIT)*. 2023, S. 1–5. DOI: 10.1109/ICCIT60459.2023.10441325.
- [3] * Ruyi Qian und Xin Wang. „Prediction of accident severity based on BP neural networks“. In: *2023 35th Chinese Control and Decision Conference (CCDC)*. 2023, S. 2739–2744. DOI: 10.1109/CCDC58219.2023.10327407.
- [4] Systemanalyse Programmentwicklung (SAP). *Was ist Machine Learning?* 2024. URL: <https://www.sap.com/germany/products/artificial-intelligence/what-is-machine-learning.html>.

- [5] International Business Machines Corporation (IBM). *Was ist ein neuronales Netz?* 2024. URL: <https://www.ibm.com/de-de/topics/neural-networks>.
- [6] Deeksha Goplani. *Activation functions, loss functions & Optimizers*. 2023. URL: <https://medium.com/@deeksha.goplani/activation-functions-loss-functions-optimizers-6bd0316898ae>.
- [7] Richard H. Smith und Daniel C. Chin. „Evaluation of an adaptive traffic control technique with underlying system changes“. In: *Proceedings of the 27th Conference on Winter Simulation*. WSC '95. Arlington, Virginia, USA: IEEE Computer Society, 1995, S. 1124–1130. ISBN: 0780330188. DOI: 10.1145/224401.224783. URL: <https://doi.org/10.1145/224401.224783>.
- [8] Edmond Chin-Ping Chang. „A Neural Network Approach to Freeway Incident Detection“. In: *The 3rd International Conference on Vehicle Navigation and Information Systems*, 1992, S. 641–647. DOI: 10.1109/VNIS.1992.640268.
- [9] Amitayas Banerjee et al. „Comparative Analysis of Machine Learning and ANN models for Mortality prediction in RTAs“. In: *2023 OITS International Conference on Information Technology (OCIT)*. 2023, S. 698–702. DOI: 10.1109/OCIT59427.2023.10431379.
- [10] Ayush Maurya et al. „A Key Factor in Traffic Management - Vehicle Speed Prediction Using Machine Learning“. In: *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. 2023, S. 1–6. DOI: 10.1109/ICCAKM58659.2023.10449542.
- [11] Jim Frost. *How to interpret R-squared in Regression Analysis*. o. A. URL: <https://statisticsbyjim.com/regression/interpret-r-squared-regression/>.
- [12] Ennaji Fatima Zohra et al. „Accident Severity Prediction using Machine Learning: A case study on the US Accidents Dataset“. In: *2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2023, S. 242–246. DOI: 10.1109/SITIS61268.2023.00044.
- [13] David Moher et al. „Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement“. In: (2009). DOI: 10.1371/journal.pmed.1000097.
- [14] * Chun Bao et al. „Research on Classification and Prediction of General Traffic Accidents on National and Provincial Highways Based on BP Neural Networks“. In: *2024 4th International Conference on Neural Networks, Information and Com-*

munication Engineering (NNICE). 2024, S. 781–784. DOI: 10.1109/NNICE61279. 2024.10498278.

[15] Amazon Web Services (AWS). *Was ist Overfitting?* 2023. URL: <https://aws.amazon.com/de/what-is/overfitting/>.

[16] International Business Machines Corporation (IBM). *What is underfitting?* 2024. URL: <https://www.ibm.com/topics/underfitting>.

Abbildungsverzeichnis

1	SLR nach PRISMA	9
---	---------------------------	---

Tabellenverzeichnis

1	Ergebnisse der Literaturrecherche	10
2	Eingabevariablen der KNN-Modelle	10