

Technische Hochschule Mittelhessen

## **Akustische Covid-Diagnose durch Husten**

**im Modul**  
**Praktikum Künstliche Intelligenz (PKI)**

**Autorin:**

Paula Schwalm (Matrikelnummer: 347488)

# 1. Einleitung

„Die Art des Geräusches ändert sich, wenn Sie Covid haben – selbst wenn Sie noch asymptomatisch sind. [1]“

— Brian Subirana

The way you produce sound changes when you have Covid, even if you're asymptomatic. [2]

Die Infektionskrankheit COVID-19, verursacht durch das SARS-CoV-2-Virus, verbreitet sich seit ihrem Ausbruch im Dezember 2019 [3] weltweit und wurde am 11. März von der Weltgesundheitsorganisation (englisch: World Health Organization, WHO) als Pandemie eingestuft [4]. Erst im Jahr 2022 konnte die Pandemie durch die Entwicklung eines Impfstoffs eingedämmt [5] und seitdem die Anzahl der Infektionen deutlich reduziert werden. Das Virus weist eine hohe Symptomvielfalt auf und kann von milden Erkältungsercheinungen bis hin zu schweren Lungenentzündungen führen, die im schlimmsten Fall zum Tod führen können [6]. In einer Studie der WHO wurde festgestellt, dass 67,7 % der 55.924 untersuchten Fälle unter einem trockenen Husten litten. Angesichts der Häufigkeit dieses Symptoms haben verschiedene Forscherinnen und Forscher untersucht, ob ein neuronales Netzwerk entwickelt werden kann, das in der Lage ist, eine Audioklassifizierung (COVID-19 positiv oder negativ) anhand eines Hustengeräusches durchzuführen. Durch Arbeiten wie „Cough Audio Analysis for COVID-19 Diagnosis“ [7] oder „COVID-19 detection using cough sound analysis and deep learning algorithms“ [8] konnte gezeigt werden, dass ein solches Verfahren möglich ist und vielversprechende Ergebnisse liefert. Ein solches Verfahren bietet eine kostengünstige, nicht-invasive und schnelle Alternative zu herkömmlichen Testverfahren und zeigt, wie Deep Learning Technologien künftig in der Medizin zur Diagnose von Krankheiten eingesetzt werden könnten.

Ziel dieser Arbeit ist die Entwicklung von neuronalen Netzwerken für die vorgestellte Problemstellung, wobei der Fokus auf der Betrachtung von unterschiedlichen Lösungsansätzen liegt. Die Arbeit, die im Rahmen des Moduls „Praktikum Künstliche Intelligenz“ (PKI) an der Technischen Hochschule Mittelhessen (THM) im Wintersemester 2024/2025 entstanden ist, verfolgt dabei nicht das Ziel, den optimalen Ansatz zu ermitteln, sondern vielmehr unterschiedliche mögliche Lösungsansätze zu untersuchen. Um möglichst realitätsnahe Ergebnisse zu erzielen, wird darauf geachtet, die Daten nicht

vorzuverarbeiten, d. h. es werden keine Hintergrundgeräusche herausgefiltert oder ähnliche Verfahren angewendet.

## 2. Forschungsstand

Eine der ersten Forschungsergebnisse zur Erkennung von COVID-19 durch die Analyse von Hustengeräuschen wurde von Wissenschaftler:innen des Massachusetts Institute of Technology veröffentlicht [2]. Zur Erkennung entwickelten Sie ein Künstliche Intelligenz (KI) Sprachverarbeitungsframework, das akustische Biomarker-Funktionsextraktoren nutzt und auf einem Convolutional Neural Network (CNN) basiert. Dadurch konnten Sie eine Erkennungsquote von 98,5 % bei nachweislich infizierten Personen erreichen und 100 % bei symptomfreien Probanden [9], wobei das Modell an 4256 Probanden getestet worden ist. Aufgrund der vielversprechenden Ergebnisse wurden insbesondere im Zeitraum der Covid-19 Pandemie weitere Forschungsergebnisse veröffentlicht, die sich mit verschiedenen Ansätzen zur Audioklassifizierung von COVID-19 Hustengeräuschen beschäftigen. Beispielsweise wurde in dem Paper „Cough Audio Analysis for COVID-19 Diagnosis“ [10] eine Wettbewerbsanalyse von vier verschiedenen Modellen: Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks with Long Short-Term Memory und VGG-19 with Support Vector Machines durchgeführt, wobei das beste Ergebnis mit dem MLP-Modell erzielt werden konnte.

Die Literaturrecherche ergab, dass es verschiedene Ansätze zur Audioklassifizierung von COVID-19 Hustengeräuschen gibt, die sich vielen Faktoren wie der Vorverarbeitung der Daten, der Wahl und Komplexität des Modells und der Wahl des Datensatzes unterscheiden. Diese Faktoren beeinflussen die Genauigkeit der Modelle und zeigen, dass es keine allgemeingültige Lösung gibt. Aufgrund des Endes der Pandemie und der damit verbundenen sinkenden Relevanz von COVID-19 ist die Forschung in diesem speziellen Anwendungsfall rückläufig oder wird nicht mehr weiterverfolgt. Forschende und Unternehmen konzentrieren sich derzeit auf andere medizinische Anwendungsfälle, wie die Erkennung von Tuberkulose oder anderen Krankheiten durch KI-Modelle. So arbeitet Google beispielsweise an einem bioakustischen Modell, das Krankheiten, wie Tuberkulose, anhand von Geräuschen erkennen kann, um so eine Frühdiagnostik zu ermöglichen. FIXME.

### 3. Datensätze

Es konnten verschiedene Open-Source-Datensätze identifiziert werden, die Hustengeräusche von COVID-19-Patienten enthalten. Dazu gehören COUGHVID, Virufy, Coswara und der Datensatz der Cambridge University. Bei der Recherche konnte festgestellt werden, dass die Datensätze sich teilweise stark voneinander unterscheiden. Beispielsweise ist der Datensatz der Cambridge University, der den größten Datensatz (FIXME Daten) darstellt, nur für Forschungszwecke verfügbar. Der Datensatz Virufy enthält im Gegensatz dazu lediglich 16 Aufnahmen.

Die beigefügten Metadaten der Datenquellen unterscheiden ebenfalls voneinander. Beispielsweise enthält der Datensatz COUGHVID Informationen über das Alter, Geschlecht und den Gesundheitszustand der Probanden, während der Datensatz Virufy nur Informationen über eine vorliegende Erkrankung erhält. Der Datensatz FIXME differenziert neben dem Gesundheitszustand zusätzlich auch nach dem Schweregrad der Erkrankung. Alleine die Betrachtung der Datensätze zeigt, dass es kein einheitliches Datenschema im Kontext der Problemstellung gibt. Des Weiteren stellt sich die Frage, mit welchen Metadaten ein Modell trainiert werden sollte, um möglichst genaue Ergebnisse zu erzielen, aber auch ob die Metadaten überhaupt notwendig sind, um die Problemstellung mit ähnlichen Ergebnissen zu lösen.

Die Qualität der Hustengeräusche unterscheidet sich ebenfalls voneinander. Da es sich häufig um online gesammelte Daten handelt, die von Freiwilligen gesammelt wurden und von verschiedenen Geräten in unterschiedlichen Umgebungen aufgenommen wurden, was sich ebenfalls auf die Qualität des Modells auswirken kann. Hustengeräuschen haben jedoch den Vorteil, dass es sich dabei ein eher unwillkürliches Geräusch handelt, wodurch Einflussfaktoren wie Akzent oder Dialekte beim Husten weniger ins Gewicht fallen. Darüber hinaus können Hustengeräusche einfach und kostengünstig aufgenommen werden, was die Verwendung von Hustengeräuschen als Datengrundlage für die Entwicklung eines neuronalen Netzwerks attraktiv macht.

Alle hier genannten Datensätze bis auf den Datensatz der Cambridge University wurden im Rahmen dieser Arbeit auf ihre Eignung untersucht. Am Ende fiel die Entscheidung auf den Datensatz COUGHVID, da diese die größte Menge an Daten frei zur Verfügung stellt. Um die Datenvielfalt zu erhöhen, wurde ebenfalls der Datensatz Virufy verwendet. Auf die Verwendung der bereitgestellten Metadaten wurde verzichtet, um die Ergebnisse so realitätsnah wie möglich zu halten und

die Modelle möglichst wenig zu beeinflussen. Zum einen kann dadurch festgestellt werden, ob ein Modell allein anhand der Audiodateien in der Lage ist, eine Diagnose zu stellen, und zum anderen kann die Datenmenge so in Zukunft einfacher erweitert werden.

Somit liegen für die Entwicklung insgesamt FIXME Audio-daten vor, wobei FIXME Covid-19-positiv und FIXME Covid-19 negativ sind. Im Durchschnitt sind die Audiodaten FIXME Sekunden lang, wodurch sich eine Trainingsdauer von FIXME pro FIXME ergibt. Die Daten werden insgesamt in 80% Trainingsdaten und 20% Validierungsdaten aufgeteilt, da sich dabei um eine Empfehlung von FIXME handelt.

durchschnitts Sampling-Rate

Ein Audio-Signal ist eine eindimensionale Sequenz von Amplitudenwerten. CNNs sind jedoch für die Verarbeitung von 2D- oder 3D-Daten (z. B. Bilder oder Videos) optimiert. Audiodateien können unterschiedlich lang sein. MFCCs standardisieren die Eingabe und ermöglichen eine feste Eingabegröße für das Modell. Für einfache Aufgaben reicht oft 128x128.

### 4. Literaturverzeichnis

- [1] [Online]. Verfügbar unter: <https://www.heise.de/news/KI-soll-COVID-19-Infizierte-am-Husten-erkennen-4945746.html>
- [2] [Online]. Verfügbar unter: <https://www.bbc.com/news/technology-54780460>
- [3] [Online]. Verfügbar unter: <https://www.dzif.de/de/glossar/sars-cov-2>
- [4] [Online]. Verfügbar unter: <https://www.tagesschau.de/ausland/europa/coronavirus-317.html>
- [5] [Online]. Verfügbar unter: <https://www.mdr.de/wissen/ende-corona-pandemie-weltweit-zweitausenddreihundzwanzig-deutschland-endemischer-zustand-100.html>
- [6] [Online]. Verfügbar unter: <https://www.infektionsschutz.de/coronavirus/basisinformationen/symptome-und-krankheitsverlauf/>
- [7] Zugriffen: 3. Januar 2025. [Online]. Verfügbar unter: <https://www.dzif.de/de/glossar/sars-cov-2>
- [8] Zugriffen: 3. Januar 2025. [Online]. Verfügbar unter: <https://www.dzif.de/de/glossar/sars-cov-2>
- [9] [Online]. Verfügbar unter: <https://www.embs.org/ojemb/articles/covid-19-artificial-intelligence-diagnosis-using-only-cough-recordings/>

- [10] [Online]. Verfügbar unter: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9791965/>
- [11] [Online]. Verfügbar unter: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
- [12] [Online]. Verfügbar unter: <https://c4science.ch/diffusion/10770/>
- [13] [Online]. Verfügbar unter: <https://github.com/virufy/virufy-data>
- [14] [Online]. Verfügbar unter: [https://www.covid-19-sounds.org/en/blog/neurips\\_dataset.html](https://www.covid-19-sounds.org/en/blog/neurips_dataset.html)
- [15] [Online]. Verfügbar unter: <https://www.heise.de/news/Frueherkennung-Neue-Google-KI-soll-Krankheiten-hoeren-koennen-9853725.html>