

Artikel

Vorhersage der Mikrobiom-Wachstumsdynamik unter Umweltstörungen

Georg Sonne ¹ und Yi-Hui Zhou ^{2,*} ¹ Bioinformatik-Forschungszentrum, North Carolina State University, Raleigh, NC 27695, USA; 3litus@gmail.com² Abteilungen für Biowissenschaften und Statistik, North Carolina State University, Raleigh, NC 27695, USA

* Korrespondenz: yihui_zhou@ncsu.edu

Zusammenfassung: MicroGrowthPredictor ist ein Modell, das Long Short-Term Memory (LSTM)-Netzwerke nutzt, um dynamische Veränderungen im Mikrobiomwachstum als Reaktion auf unterschiedliche Umwelteinflüsse vorherzusagen. In diesem Artikel stellen wir die innovativen Funktionen von MicroGrowthPredictor vor, darunter die Integration von LSTM-Modellierung mit einer neuartigen Technik zur Schätzung von Konfidenzintervallen. Das LSTM-Netzwerk erfasst die komplexe zeitliche Dynamik von Mikrobiomsystemen, während die neuen Konfidenzintervalle ein robustes Maß für die Vorhersageunsicherheit bieten. Wir haben zwei Beispiele aufgenommen – eines, das die Zusammensetzung und Vielfalt der menschlichen Darmmikrobiota aufgrund wiederholter Antibiotikabehandlungen veranschaulicht, und das andere, das die Anwendung von MicroGrowthPredictor auf einem künstlichen Darmdatensatz demonstriert. Die Ergebnisse zeigen die verbesserte Genauigkeit und Zuverlässigkeit der LSTM-basierten Vorhersagen, die durch MicroGrowthPredictor ermöglicht werden. Die Einbeziehung spezifischer Metriken wie des mittleren quadratischen Fehlers bestätigt die Vorhersageleistung des Modells. Unser Modell birgt ein enormes Potenzial für Anwendungen in den Umweltwissenschaften, im Gesundheitswesen und in der Biotechnologie und fördert Fortschritte in der Mikrobiomforschung und -analyse. Darüber hinaus ist es bemerkenswert, dass MicroGrowthPredictor auf reale Daten mit kleinen Stichprobengrößen und zeitlichen Beobachtungen unter Umweltstörungen anwendbar ist, was seinen praktischen Nutzen in verschiedenen Bereichen gewährleistet.

Schlüsselwörter: Mikrobiomdynamik; Vorhersageunsicherheit; Umwelthanwendungen



Zitat: Sun, G.; Zhou, Y.-H.

Vorhersage der Wachstumsdynamik des Mikrobioms unter Umwelteinflüssen. *Appl. Microbiol.* **2024**, *4*, 948–958. <https://doi.org/10.3390/applmicrobiol4020064>

Wissenschaftlicher Redakteur: Bong-Soo Kim

Empfangen: 7. Mai 2024

Überarbeitet: 4. Juni 2024

Akzeptiert: 7. Juni 2024

Veröffentlicht: 10. Juni 2024



Copyright: © 2024 bei den Autoren.

Lizenznehmer MDPI, Basel, Schweiz.

Dieser Artikel ist ein Open Access-Artikel vertrieben unter den Bedingungen und

Bedingungen der Creative Commons

Namensnennunglizenz (CC BY)

(<https://creativecommons.org/licenses/by/4.0/>).

1. Einleitung

Das menschliche Mikrobiom, ein komplexes Ökosystem aus Billionen von Mikroorganismen, die im und auf dem menschlichen Körper leben, spielt eine entscheidende Rolle bei der Aufrechterhaltung der physiologischen Homöostase, der Stoffwechselfunktionen und der Immunreaktionen [1]. Störungen im Mikrobiom werden mit einer Vielzahl von Erkrankungen in Verbindung gebracht, die von Magen-Darm-Störungen bis hin zu systemischeren Erkrankungen wie Diabetes, Fettleibigkeit und sogar neurologischen Erkrankungen reichen [2]. Diese symbiotische Wirt-Mikroben-Interaktion unterstreicht die Notwendigkeit, die dynamische Natur des menschlichen Mikrobioms zu verstehen [3], insbesondere, wie es sich im Laufe der Zeit und als Reaktion auf verschiedene Umweltreize verändert [4,5].

Unter normalen Bedingungen besteht das Darmmikrobiom aus einer vielfältigen Bakteriengemeinschaft, wobei Firmicutes und Bacteroidetes die vorherrschenden Stämme sind. Umweltstörungen wie Ernährungsumstellungen, Antibiotikagebrauch und Schadstoffbelastung können die Zusammensetzung und Funktion des Mikrobioms erheblich verändern und potenzielle gesundheitliche Folgen haben. So kann beispielsweise eine Antibiotikabehandlung die mikrobielle Vielfalt drastisch reduzieren, was häufig zu einem Überwuchern resistenter Bakterien und einem Rückgang nützlicher Mikroben führt, was Stoffwechselprozesse und Immunfunktionen stören kann [6]. Das Verständnis dieser Populationsdynamik ist entscheidend für die Entwicklung von Strategien zur Abschwächung der negativen Auswirkungen solcher Störungen auf die menschliche Gesundheit.

Hochdurchsatz-Sequenzierungstechnologien, insbesondere die 16S rRNA-Sequenzierung, haben eine neue Ära in der Mikrobiomforschung eingeläutet und ermöglichen detaillierte Bewertungen der mikrobiellen Vielfalt und relativen Häufigkeit in verschiedenen menschlichen Populationen und unter verschiedenen Bedingungen [7]. Die durch diese Technologien generierten enormen Datenmengen bieten jedoch sowohl Chancen als auch Herausforderungen.

Eine der größten Herausforderungen besteht darin, die zeitlichen Muster zu entschlüsseln und zukünftige Zustände des Mikrobioms vorherzusagen, was für präventive und therapeutische Anwendungen im Gesundheitswesen von entscheidender Bedeutung ist.

Bei der prädiktiven Modellierung in der Biometrie wurden in der Vergangenheit verschiedene statistische Methoden eingesetzt, aber diese traditionellen Ansätze reichen oft nicht aus, um die hohe Dimensionalität und Nichtlinearität von Mikrobiomdaten zu bewältigen. Das Aufkommen des maschinellen Lernens, insbesondere des Deep Learning, bietet vielversprechende neue Ansätze für derart komplexe Daten [8]. Rekurrenente neuronale Netze (RNNs) [9] und ihre fortgeschrittene Variante, Long Short-Term Memory (LSTM)-Netze [10], eignen sich hervorragend für die Analyse und Vorhersage zeitlicher Abläufe und bieten einen hervorragenden Rahmen für die Modellierung der Mikrobiodynamik.

In dieser Studie stellen wir das Modell MicroGrowthPredictor vor, das darauf abzielt, die Leistungsfähigkeit von LSTM-Netzwerken zu nutzen, um Veränderungen im menschlichen Mikrobiom als Reaktion auf Umweltstörungen vorherzusagen, ein entscheidender Schritt in Richtung personalisierte Medizin und gezielter therapeutischer Eingriffe.

2. Materialien und Methoden

2.1. Long Short-Term Memory (LSTM)-Modell

Das Long Short-Term Memory (LSTM)-Netzwerk, eine spezialisierte Form der rekurrenenden neuronalen Netzwerkarchitektur (RNN), wurde speziell dafür entwickelt, die Herausforderungen des Lernens aus sequenziellen Daten, insbesondere langfristigen Abhängigkeiten, zu bewältigen. Herkömmliche RNNs sind zwar theoretisch in der Lage, solche Abhängigkeiten zu verarbeiten, scheitern jedoch in der Praxis oft am Problem des verschwindenden Gradienten, bei dem während des Trainings mit jedem Zeitschritt Informationen verloren gehen. LSTM-Netzwerke sind darauf ausgelegt, diese Einschränkung zu überwinden, wodurch sie sich besonders für Anwendungen in verschiedenen Bereichen eignen, wie etwa Zeitreihenanalyse, Verarbeitung natürlicher Sprache und, was für unsere Arbeit relevant ist, Mikrobiomdatenanalyse.

LSTM-Netzwerke führen eine komplexere Zellstruktur ein als herkömmliche RNNs [11]. Jede LSTM-Zelle enthält Mechanismen, sogenannte Gates, die den Informationsfluss in die Zelle und aus ihr heraus regeln. Innerhalb einer LSTM-Zelle gibt es drei Arten von Gates (Abbildung 1A):

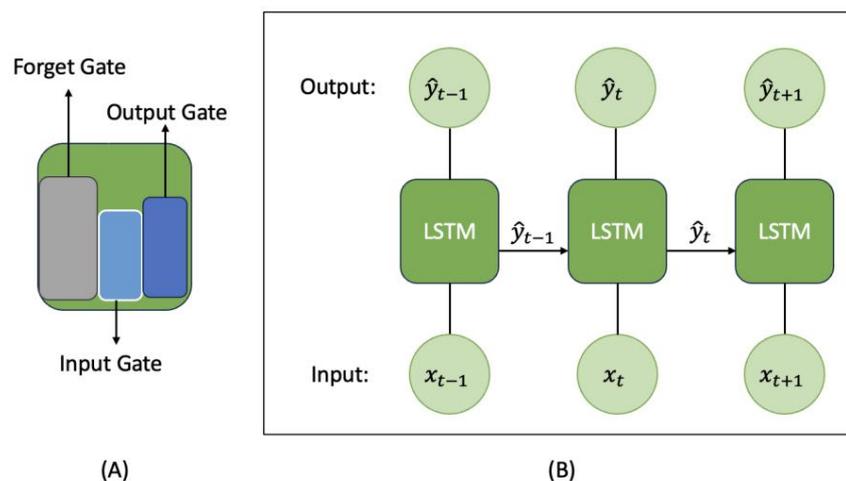


Abbildung 1. Architektur des Long Short-Term Memory (LSTM): **(A)** Eine Vergrößerung einer LSTM-Zelle, die ihre drei Gates zeigt: das Eingangsgate, das Forget-Gate und das Ausgangsgate. **(B)** Der Fluss der Eingangs- und Ausgangsdaten in einem LSTM-Netzwerk vom Zeitpunkt $t-1$ bis zum Zeitpunkt t .

- Eingabe-Gate: Moduliert die Menge neuer Informationen, die dem Zellzustand hinzugefügt werden.
- Vergessen-Gate: Bestimmt den Umfang der Informationen, die aus dem Zellzustand verworfen werden.

Das Vergessenstor hilft, irrelevante oder veraltete mikrobielle Informationen zu eliminieren und so nur die relevantesten Daten für eine genaue Modellierung beizubehalten.

Ausgabebtor: Steuert die Menge der von der Zelle auszugehenden Informationen. Bei Mikrobiomdaten hilft das Ausgabebtor dabei zu entscheiden, welche verarbeiteten mikrobiellen Informationen die Vorhersagen oder Analysen des Netzwerks bei jedem Zeitschritt beeinflussen sollen.

Diese Gates arbeiten zusammen, um den Zustand der Zelle zu aktualisieren und es dem LSTM zu ermöglichen, Informationen über lange Zeiträume hinweg sowohl zu speichern als auch zu vergessen (Abbildung 1B), was für das Erlernen langfristiger Abhängigkeiten entscheidend ist. Abbildung 1B veranschaulicht den Übergang von Daten durch das LSTM-Netzwerk von einem Zeitschritt zum nächsten. Sie zeigt die Eingabe- und Ausgabedaten, wie sie vom Zeitschritt $t - 1$ zum Zeitschritt t fließen. Bei jedem Zeitschritt werden die Eingabedaten zusammen mit dem Zellzustand des vorherigen Zeitschritts von der LSTM-Zelle verarbeitet. Diese Verarbeitung führt zu einem aktualisierten Zellzustand und einer Ausgabe, die dann an den nächsten Zeitschritt weitergegeben werden. Dieser sequentielle Mechanismus ermöglicht es dem LSTM-Netzwerk, zeitliche Abhängigkeiten effektiv zu handhaben und sicherzustellen, dass Informationen über verschiedene Zeitschritte hinweg weitergegeben und genutzt werden, um eine verbesserte Vorhersage und Analyse in Zeitreihenaufgaben zu ermöglichen.

Im Bereich der Mikrobiomanalyse ist das Verständnis zeitlicher Dynamiken und sequentieller Muster angesichts der Art der Entwicklung und Interaktion mikrobieller Gemeinschaften im Laufe der Zeit von entscheidender Bedeutung. Hier verwenden wir eine spezielle Notation, um die Mechanik des LSTM-Modells zu erläutern. Betrachten Sie einen Trainingsdatensatz $D = \{(x_t, y_t)\}$, wobei x_t den Vektor der relativen Häufigkeiten [12] aller mikrobiellen Taxa zum t -ten Zeitschritt und y_t bezeichnet die entsprechende gewünschte Ausgabe. Das LSTM nimmt diese Eingabesequenzen und verarbeitet sie durch seine komplexe Zellstruktur, wobei wertvolle zeitliche Abhängigkeiten in den Daten erfasst werden, die für genaue Vorhersagen und Analysen in Mikrobiomstudien entscheidend sind.

2.2. Modellstruktur zur Vorhersage des Mikrobiomwachstums

Das in dieser Studie verwendete LSTM-Modell ist sowohl auf Einfachheit als auch auf Leistung ausgelegt. Die Eingabeebene ist darauf ausgelegt, die relativen Häufigkeitsgrade von Taxa zu verarbeiten, und umfasst eine umfangreiche Reihe mikrobieller Taxa, die als x_t bezeichnet werden. Diese Ebene besteht aus n_{taxa} -Knoten, die jeweils die relative Häufigkeit eines bestimmten Taxons darstellen, und entspricht der Gesamtzahl der eindeutigen Taxa, die im Mikrobiom-Datensatz identifiziert wurden.

Was die Architektur betrifft, besteht unser Modell aus zwei verborgenen Schichten, die zwischen den Eingangs- und Ausgangsstufen angeordnet sind. Die primäre verborgene Schicht enthält ein LSTM mit n_h verborgenen Zuständen, das innerhalb einer einzigen Schicht funktioniert. Diese Konfiguration ist entscheidend, da sie es dem Modell ermöglicht, die der Eingangssequenz innewohnende zeitliche Dynamik zu erfassen und zu interpretieren, was den charakteristischen Speicherzellen des LSTM zu verdanken ist.

Um Überanpassung zu vermeiden und die Robustheit des Modells zu verbessern, wird im Anschluss an die LSTM-Schicht eine Dropout-Strategie implementiert. Diese Strategie, die von einer vorab festgelegten Dropout-Wahrscheinlichkeit p gesteuert wird, beinhaltet die willkürliche Deaktivierung von Knoten, wodurch die Generalisierungsfähigkeit des Modells gestärkt wird. Knoten, die nicht von Dropout betroffen sind, werden dann an die nächste Schicht übergeben – eine vollständig verbundene Schicht mit n_{FC} -Knoten.

Die zweite verborgene Schicht verwendet die Aktivierungsfunktion Rectified Linear Unit (ReLU) für Datenpunkte, die aus der vollständig verbundenen Schicht stammen. Dies verleiht dem Modell wesentliche Nichtlinearität und bereitet es darauf vor, komplexe Muster innerhalb des Datensatzes zu erkennen. Vorhersagen werden auf Grundlage der Ausgabe dieser Schicht formuliert.

Zusammenfassend lässt sich sagen, dass unser MicroGrowthPredictor-Modell zur Vorhersage der Mikrobiodynamik speziell entwickelte Schichten integriert, die jeweils darauf ausgelegt sind, die nuancierten zeitlichen Dynamiken in Mikrobiomdaten zu interpretieren. Die Architektur beginnt mit einer Eingabeschicht, die n_{taxa} taxa-repräsentative Knoten enthält, und geht in ein einschichtiges LSTM mit n_h verborgenen Zuständen über.

Obwohl nicht explizit beschrieben, gehen wir davon aus, dass die LSTM-Schicht die herkömmliche Zusammensetzung von LSTM-Zellen beibehält, einschließlich Eingabe-, Vergessens- und Ausgabe-Gates für eine effektive Informationsübertragung. Diese Struktur ist entscheidend dafür, dass das Modell lernen und die langfristigen Abhängigkeiten sequenzieller Daten bewahren kann.

Nach der LSTM-Schicht wird eine Dropout-Technik mit einer bestimmten Wahrscheinlichkeit p angewendet, die als Regularisierungsmechanismus dient und das Risiko einer Überanpassung mindert. Anschließend wird eine vollständig verbundene Schicht mit n_{FC} -Knoten eingeführt, die in einer dichten Schicht gipfelt, die nichtlineare Abhängigkeiten in den Daten erfassen kann. Die letzte Phase des Modells enthält eine ReLU-Aktivierungsfunktion, die Nichtlinearität einführt und die Komplexität des Modells für eine detaillierte Dateninterpretation erhöht. Diese Phase ist entscheidend für die Gestaltung des endgültigen Ergebnisses und gewährleistet präzise und flüssige Vorhersagen in der sich dynamisch verändernden Landschaft der M

2.3. Training des MicroGrowthPredictor-Modells

Wenn Sie mit Zeitreihendaten arbeiten und LSTM verwenden, um die Auswirkungen von Umweltstörungen vorherzusagen, muss die Kreuzvalidierung die den Daten innewohnenden zeitlichen Abhängigkeiten berücksichtigen. Um robuste und genaue Vorhersagen zu gewährleisten, haben wir eine Zeitreihen- Kreuzvalidierungsmethode mit einem Rollfenster-Ansatz verwendet. Der Datensatz wurde ohne Mischen in K aufeinanderfolgende Faltungen unterteilt. Für jede Faltung k wurde das Modell an den ersten k Faltungen trainiert und an der $k + 1$ Faltung getestet, wobei dies wiederholt wurde, bis jede Faltung als Testsatz diente. Diese Methode stellt sicher, dass zeitliche Abhängigkeiten eingehalten werden, und vermeidet Datenlecks.

Für jede Faltung wurden Bewertungsmaßstäbe, wie etwa der mittlere quadratische Fehler (MSE), erfasst und die durchschnittliche Leistung aller Faltungen berechnet, um die Robustheit des Modells zu beurteilen.

2.4. Vorhersageintervall

Während traditionelle Ansätze zur Festlegung von Konfidenz- oder Vorhersageintervallen in Deep-Learning-Modellen aufgrund der Nichtlinearität und der komplexen Architektur dieser Modelle vor erheblichen Herausforderungen stehen, haben jüngste Fortschritte den Weg für robustere Lösungen geebnet. Ein solcher Fortschritt ist die Arbeit von [13], in der das Monte-Carlo-(MC) -Dropout-Framework genutzt wurde, um eine Methode einzuführen, die zwar effektiv ist, aber Raum für weitere Verfeinerung und Anwendung in neuen Bereichen wie der Mikrobiomdatenanalyse bietet.

Unsere Forschung baut auf dieser grundlegenden Arbeit auf und wendet das Prinzip stochastischer Dropouts nach jeder verborgenen Schicht in der neuronalen Netzwerkarchitektur an. Wir erweitern dieses Konzept jedoch, indem wir den Dropout-Prozess und die anschließende Interpretation der Modellergebnisse speziell auf die Eigenschaften und Komplexitäten von Mikrobiomdaten zuschneiden. Diese Anpassung ermöglicht nicht nur die theoretische Interpretation der Modellausgabe als Zufallsstichprobe aus der posterioren prädiktiven Verteilung, sondern berücksichtigt auch das einzigartige Datenverhalten in Mikrobiomstudien.

Der Prozess der Konstruktion einer empirischen Verteilung der vorhergesagten Werte durch Behandlung jeder Vorhersage während des Dropouts als Stichprobe aus der zugrunde liegenden Datenverteilung stellt in unserer Studie einen differenzierten Ansatz dar. Er weicht von klassischen Techniken ab, indem er einen Einblick in die Vorhersagefähigkeiten und Unsicherheiten des Modells bietet, die speziell auf den Mikrobiomkontext abgestimmt sind, und so die Robustheit der auf diesen Vorhersagen basierenden Entscheidungsfindung stärkt.

In unserem Ansatz kennzeichnen wir die Testdaten, die vorhergesagt werden sollen, mit dem hochgestellten \tilde{y} . Grundlage des Vorhersageintervalls ist die bedingte Wahrscheinlichkeit $p(y | \tilde{x}, \tilde{y}, D)$. Diese Wahrscheinlichkeit kann als Integral des Produkts von $p(y | \tilde{x}, \tilde{y}, \tilde{y})$ und $p(\tilde{y}|D)$ über den Parametervektor \tilde{y} ausgedrückt werden, der wie folgt bezeichnet wird:

$$p(y | \tilde{x}, \tilde{y}, D) = \int_{\tilde{y}} p(y | \tilde{x}, \tilde{y}, \tilde{y}) p(\tilde{y}|D) d\tilde{y}.$$

\tilde{y} stellt den Parametervektor des Deep-Learning-Modells dar, und $p(\tilde{y}|D)$ entspricht der Posterior-Verteilung. Allerdings ist es im Allgemeinen nicht möglich, eine analytische Form für $p(y | \tilde{x}, \tilde{y}, \tilde{y})$ abzuleiten. Um diese Herausforderung zu überwinden, wird in Ref. [14] eine Näherungstechnik vorgeschlagen, die eine Variationsverteilung verwendet, die als $q(\tilde{y})$ bezeichnet wird. Folglich ergibt sich die folgende Näherung:

$$p(y | \tilde{x}, \tilde{y}, D) \approx \int_{\tilde{y}} p(y | \tilde{x}, \tilde{y}, \tilde{y}) q(\tilde{y}) d\tilde{y} \approx \frac{1}{K} \sum_{k=1}^K p(y | \tilde{x}, \tilde{y}, \tilde{y}^k), \quad (1)$$

wobei $\tilde{y}^k \sim q(\tilde{y})$. Diese endgültige Näherung, die durch die Stichprobennahme von $\{\tilde{y}^k\}_{k=1, \dots, K}$ aus der Variationsverteilung $q(\tilde{y})$ erreicht wird, verwendet die Technik der Monte-Carlo-Integration. Darüber hinaus ist diese Näherung gleichbedeutend mit der Implementierung des in [13] eingeführten Monte-Carlo-Dropout -Algorithmus. Im Wesentlichen dient für einen gegebenen Testdatenpunkt (\tilde{x}, \tilde{y}) , die prädiktive Verteilung als \tilde{y} wird bei \tilde{x} mehrfach ausgewertet \tilde{y} Ausgabe y der resultierenden empirischen Schätzung von $p(y | \tilde{x}, \tilde{y}, D)$. Vorhersageintervalle

Erfassen Sie die Variabilität, die aus zwei Hauptquellen stammt: Modellunsicherheit (\hat{y}_1) und inhärentes Rauschen (\hat{y}_2).

Die folgenden Schritte beschreiben den Prozess: Berechnen Sie für jeden einzelnen Datenpunkt \hat{y} im x des Testsatzes die entsprechende Ausgabe \hat{y} , indem Sie jeden Knoten mit einer bestimmten Ausfallwahrscheinlichkeit p zufällig ausschließen. Wiederholen Sie diesen Prozess B -mal, um eine große Anzahl vorhergesagter Werte \hat{y} zu erhalten, von denen jeder aufgrund des zufälligen Ausfalls von Knoten variiert. Berechnen Sie als Nächstes die Modellunsicherheit \hat{y}_1 , indem Sie die durchschnittliche quadrierte Differenz zwischen jedem vorhergesagten Wert \hat{y} und dem Mittelwert aller vorhergesagten Werte $\bar{\hat{y}}$ berechnen. Dies ist die Formel $\hat{y}_1 = \frac{1}{B} \sum_{b=1}^B (\hat{y}_b - \bar{\hat{y}})^2$. Um das inhärente Rauschen mit Hilfe von \hat{y}_1 zu quantifizieren, berechnen Sie die durchschnittliche quadrierte Differenz zwischen jedem vorhergesagten Wert \hat{y} und seinem entsprechenden wahren y -Wert y_j unter Verwendung des Testdatensatzes der Länge V . Dies ergibt das inhärente Rauschen \hat{y}_2 , berechnet als $\hat{y}_2 = \frac{1}{V} \sum_{j=1}^V (\hat{y}_j - y_j)^2$. Berechnen Sie durch Kombinieren der Modellunsicherheit und des inhärenten Rauschens \hat{y} als Quadratwurzel der Summe von \hat{y}_1 und \hat{y}_2 , d. h. $\hat{y} = \sqrt{\hat{y}_1 + \hat{y}_2}$. Bestimmen Sie abschließend die oberen und unteren Grenzen des Vorhersageintervalls, indem Sie $z\hat{y}/2$ mal \hat{y} zum mittleren vorhergesagten Wert $\bar{\hat{y}}$ addieren und davon subtrahieren. Hier stellt $z\hat{y}/2$ den z-Score dar, der mit dem gewünschten Konfidenzniveau $(1 - \hat{y})100\%$ verbunden ist. Ein formaler Algorithmus ist in Algorithmus 1 aufgeführt.

Algorithmus 1: LSTM-Neuralnetzwerk und Vorhersageintervall. , p , t , nh , nf , c

Erforderlich: x , y , \hat{y} , ja

Sicherstellen: \hat{y} , U , L

- 1,
 - 2 **Wiederholung** $z_1 \hat{y} x$ von der LSTM-Schicht mit t und nh ;
 - 3 $z_2 \hat{y} z_1$ durch zufälligen Dropout mit p ; $z_3 \hat{y} z_2$ von der
 - 4 vollständig verbundenen Schicht mit nf c -Knoten;
 - 5 Wenden Sie ReLU auf z_3 an;
 - 6 $y^* \hat{y} z_3$ von der Ausgabeebene;
 - 7 Bewerten Sie y^* mit y ;
 - 8 Update \hat{y} für Modell $m\hat{y}$; **9 bis** zur letzten Epoche; **10 für** $i = 1$ bis B **do** $\hat{y} m\hat{y}(x \hat{y})$ mit
 - 11 y^* zufälligem Dropout;
 - i **12 Ende**
 - 13 Berechnen Sie $\bar{\hat{y}}$ und \hat{y} ;
 - 14 U , L $\hat{y} \bar{\hat{y}} \pm z\hat{y}/2 \times \hat{y}$;
-

2.5. Parameterabstimmung

Um die Leistung unseres MicroGrowthPredictor-Modells zu optimieren, verwenden wir einen zweistufigen Abstimmungsprozess.

Im ersten Schritt wählen wir auf der Grundlage vorläufiger Experimente die Anzahl der verborgenen Einheiten in der LSTM-Schicht (nh) und der vollständig verbundenen Schicht (nf c) vor. Anschließend untersuchen wir verschiedene Kombinationen der Dropout-Wahrscheinlichkeit (p) und der Sequenzlänge (T), die die Anzahl der vorherigen Datenpunkte darstellt, die als Merkmale für die Vorhersage verwendet wurden. Die Leistung des Modells wird durch Berechnung des mittleren quadratischen Fehlers (MSE) in einem separaten Testdatensatz bewertet, und wir wählen die Kombination aus p und T aus, die diesen Fehler minimiert.

Sobald die optimale Dropout-Wahrscheinlichkeit und Sequenzlänge bestimmt sind, fahren wir mit dem zweiten Schritt fort, in dem wir die Anzahl der Knoten sowohl in der LSTM- als auch in der vollständig verbundenen Schicht feinabstimmen. Für jede Architekturkombination trainieren wir das Modell mehrmals mit unterschiedlichen Initialisierungen, um Variationen zu berücksichtigen, die durch zufällige Dropouts und anfängliche Gewichtungseinstellungen entstehen. Wir berechnen den MSE für jeden Trainingslauf und wählen die Architektur aus, die den geringsten Fehler im Testdatensatz ergibt.

Dieser rigorose Feinabstimmungsprozess stellt sicher, dass unser MicroGrowthPredictor-Modell optimal für den jeweiligen Datensatz konfiguriert ist und so seine Vorhersageleistung verbessert.

3. Ergebnisse

In dieser Studie wenden wir das MicroGrowthPredictor-Modell und das zugehörige Tuning-Verfahren auf zwei unterschiedliche Datensätze an: den Datensatz zum Antibiotikum Ciprofloxacin (Cp) aus [15] und den in [16] beschriebenen Datensatz zum künstlichen Darm. Beide Datensätze bieten Einblicke in die zeitliche Dynamik des Mikrobioms unter verschiedenen Umwelteinflüssen.

3.1. Ciprofloxacin Dataset Reference

[15] unterstreicht die signifikanten Veränderungen der Zusammensetzung und Vielfalt der menschlichen Darmmikrobiota durch wiederkehrende Antibiotikabehandlungen. Diese Forschung umfasste eine eingehende Überwachung der Bakteriengemeinschaften im distalen Darm bei drei Probanden (D, E und F). Stuhlproben wurden über zehn Monate hinweg regelmäßig gesammelt, was insgesamt 52–56 Proben pro Person ergab. Innerhalb dieses Zeitraums wurden jedem Probanden zwei separate 5-Tage-Behandlungen mit dem Antibiotikum Ciprofloxacin (Cp) im Abstand von 6 Monaten verabreicht. Intensive Probenentnahme – täglich über zwei 19-tägige Zeiträume, die mit jeder Cp-Behandlung zusammenfielen – lieferte eine detaillierte Perspektive des Mikrobioms während der Antibiotikaexposition. Außerhalb dieser Zeiträume wurden Proben entweder wöchentlich oder monatlich entnommen, um die mikrobielle Zusammensetzung in Abwesenheit einer Behandlung zu erfassen.

Zur Veranschaulichung konzentrieren wir uns auf Thema D. Unser Optimierungsprozess umfasst die Erstellung eines Konturdiagramms des mittleren quadratischen Fehlers (MSE) gegenüber verschiedenen Werten der Dropout-Wahrscheinlichkeit p und der Anzahl der Zeitschritte. Abbildung 2 visualisiert diese Beziehung und leitet uns bei der Auswahl einer optimalen Kombination zur Verfeinerung des MicroGrowthPredictor-Modells. Das Konturdiagramm des mittleren quadratischen Fehlers wird mit der Dropout-Wahrscheinlichkeit p auf der x-Achse und der Anzahl der Zeitschritte auf der y-Achse dargestellt. Im Konturdiagramm gilt: Je dunkler die Schattierung, desto kleiner ist der Fehler. Wir verwenden eine Optimierungsfunktion, um die beste Kombination aus Dropout-Wahrscheinlichkeit und Sequenzlänge zu ermitteln.

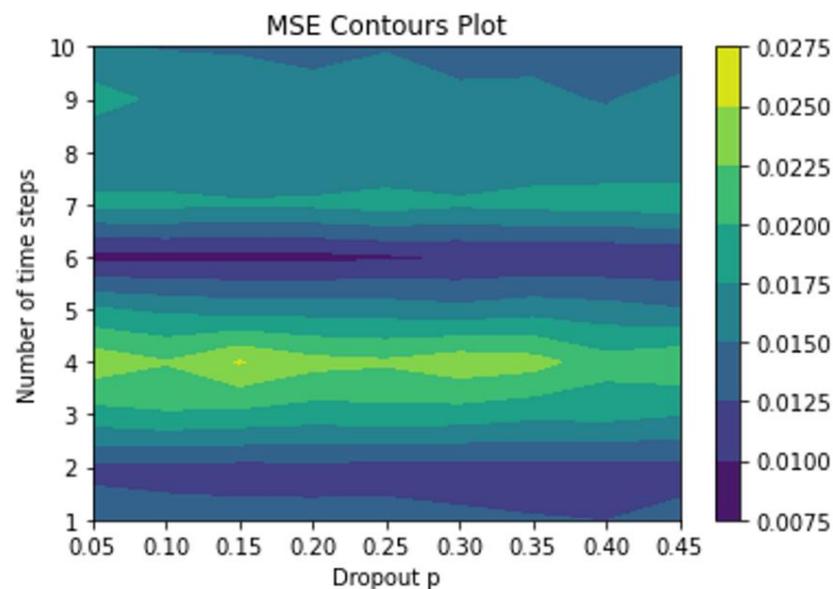


Abbildung 2. Konturdiagramm des mittleren quadratischen Fehlers über p und t für Subjekt D EU766613: Je dunkler das Konturdiagramm ist, desto kleiner ist der Fehler. Wir können die beste Kombination aus Dropout-Wahrscheinlichkeit und Sequenzlänge ermitteln.

Anschließend verlagert sich unser Fokus auf die Ermittlung der optimalen Knotenzahl sowohl für die LSTM- als auch für die vollständig verbundenen Schichten, wie in Abbildung 3 dargestellt. Die x-Achse stellt die Anzahl der verborgenen Zustände in der einzelnen LSTM-Schicht dar, und die y-Achse stellt die Anzahl der Knoten in der vollständig verbundenen Schicht dar. Unterschiedliche Kombinationen führen zu Änderungen des mittleren quadratischen Fehlerwerts. Das Konturdiagramm bietet eine direkte Darstellung des kleinsten MSE, der durch den dunkelsten Bereich in der Abbildung angezeigt wird.

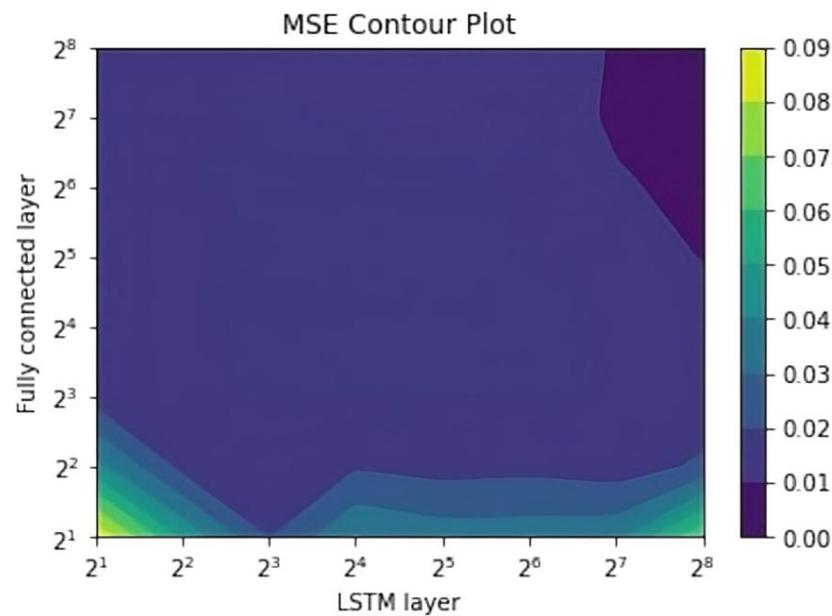


Abbildung 3. Konturdiagramm des mittleren quadratischen Fehlers über n_f und n_h für Subjekt D EU766613: Die x-Achse stellt die Anzahl der verborgenen Zustände in der einzelnen LSTM-Schicht dar, und die y-Achse stellt die Anzahl der Knoten in der vollständig verbundenen Schicht dar. Bei unterschiedlichen Kombinationen ändert sich der mittlere quadratische Wert. Das Konturdiagramm gibt uns im Wesentlichen einen direkten Eindruck vom kleinsten MSE, der durch den dunkelsten Bereich in der Abbildung dargestellt wird.

Durch diese systematische Untersuchung bleibt unser Ziel unverändert: Wir wollen eine Konfiguration ermitteln, die die Fehler im Testdatensatz minimiert und so die Wirksamkeit des MicroGrowthPredictor verbessert.

Es ist wichtig zu beachten, dass wir in unseren Trainingsdatensatz zwei Drittel der beobachteten Daten aufgenommen haben, um eine solide Grundlage für das Modell zu schaffen. Insbesondere gab es zwei Datenpunkte, die der Antibiotikaverabreichung für jeden Patienten entsprachen. Einer dieser Punkte wurde in den Trainingsdatensatz aufgenommen, während der andere für den Vorhersagedatensatz reserviert war. Basierend auf unseren Beobachtungen zeigte die Reaktion auf das erste Antibiotikum im Vergleich zum zweiten eine verzögerte Reaktion. Diese Beobachtung erklärt, warum unsere vorhergesagten Daten in Abbildung 4 ein verzögertes Muster aufweisen.

Die zeitlichen Einblicke, die durch die Visualisierung der Verläufe der relativen Mikrobiomhäufigkeit gewonnen wurden, waren entscheidend für das Verständnis der Dynamik von Mikrobiomveränderungen und ihrer möglichen Auswirkungen auf die Gesundheit des Wirts. Zur weiteren Erläuterung zeigt Abbildung 4 eine Analyse und Vorhersage der relativen Häufigkeit von Bacteroid EU766613 für Subjekt D unter Verwendung der oben genannten optimalen Parameter. Die Intervalle der Antibiotikaverabreichung werden durch eine blaue gepunktete vertikale Linie gekennzeichnet, während die rot gepunktete Abgrenzung die Trainings- und Testzeiträume trennt. In unserer Studie zu wiederholten Antibiotikabehandlungen legen wir Wert auf die Einbeziehung umfangreicher Daten zu Antibiotikainterventionen, um die Vorhersagekraft unseres Modells zu stärken. Dieser datengesteuerte Ansatz verbessert die Genauigkeit nachfolgender Behandlungsvorhersagen und bietet ein wichtiges Instrument zur Bekämpfung der Antibiotikaresistenz durch informierte, strategische Anwendung von Therapien.

Die Visualisierung unterstreicht die Fähigkeit des MicroGrowthPredictor -Modells, die Dynamik des Mikrobioms zu verstehen und Vorhersagen zu formulieren, die auf diesen identifizierten Mustern basieren. Dies wird erreicht, indem das Modell über 200 Epochen mit einer Lernrate von 0,001 trainiert wird. Darüber hinaus beträgt der mittlere quadratische Fehlerverlust für die Trainingsdaten 0,00081 und für die Testdaten 0,01021.

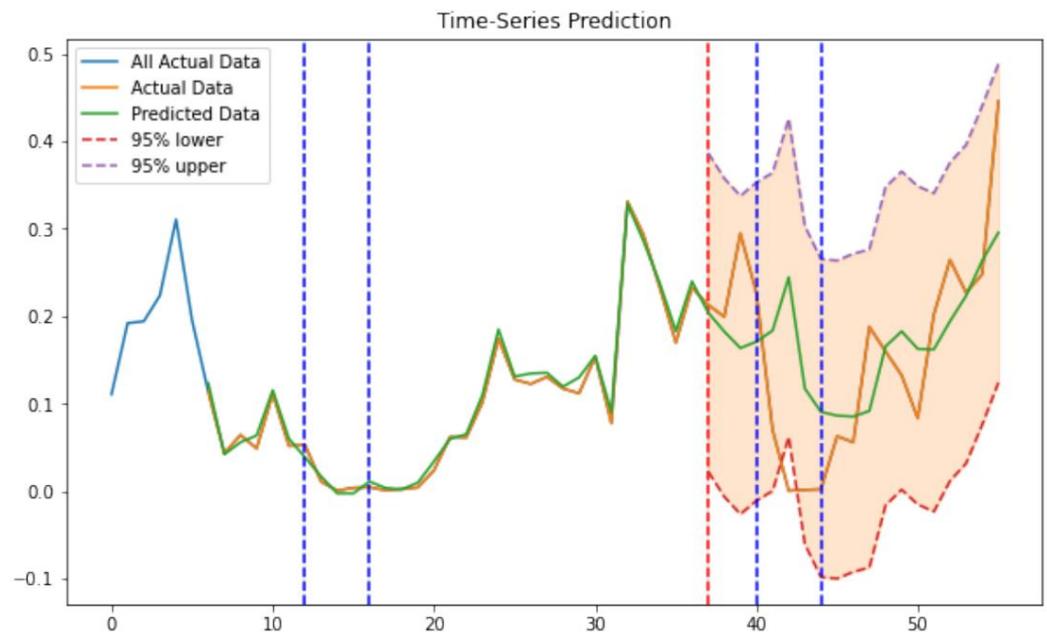


Abbildung 4. Verläufe der relativen Häufigkeit des Bacteroids EU766613 für Subjekt D. Die gewählten Parameter sind $p = 0,05$, $t = 6$, $n_h = 256$ und $n_f = 256$. Die blauen vertikalen Bänder stellen die beiden Antibiotikabehandlungszeiträume dar und die rote gepunktete Linie teilt die Daten in Training und Test auf.

3.2. Künstlicher Darm-Datensatz

Der von [16] bereitgestellte Datensatz umfasst zeitaufgelöste Messungen der Darmmikrobiota aus einem künstlichen menschlichen Darm. Diese Daten, die sowohl täglich als auch stündlich erfasst werden, stammen aus einem künstlichen Darm, der mithilfe von kontinuierlichen anaeroben Bioreaktorsystemen konstruiert wurde, wodurch eine genaue Darstellung der Dynamik der menschlichen Darmmikrobiota gewährleistet wird. Über einen Monat wurden vier Ex-vivo-Gefäße kultiviert, die jeweils mit einem identischen menschlichen Stuhlinokulum initialisiert wurden. Um die Genauigkeit des Experiments zu gewährleisten, wurden Schlüsselparameter wie pH-Wert, Temperatur, Medienzufuhr und Sauerstoffkonzentration strikt eingehalten. Am 23. Tag wurde die mikrobielle Dynamik gezielt durch die Einführung eines *Bacteroides ovatus*-Bolus stimuliert, eines aus dem Stuhlspender isolierten Stamms. Unvorhergesehene Unterbrechungen der Futterzufuhr in zwei Gefäßen zwischen Tag 11 und 13 führten jedoch zu ungeplanten mikrobiellen Variationen. Insbesondere beobachteten wir signifikante Veränderungen in der Population der Rikenellaceae, einer Bakterienfamilie, die für ihre Rolle im menschlichen Darmmikrobiom bekannt ist. Rikenellaceae sind am Abbau komplexer Kohlenhydrate beteiligt und spielen eine entscheidende Rolle bei der Aufrechterhaltung der Darmgesundheit und der Stoffwechselfunktionen. Die Veränderungen in dieser Population sind besonders interessant, da sie Erkenntnisse darüber liefern können, wie Ernährungsunterbrechungen und mikrobielle Einführungen die Stabilität und Funktion der Darmmikrobiota beeinflussen.

In diesem Beispiel dient das erste Schiff als Trainingsset, während das zweite Schiff als Testset fungiert. Unser MicroGrowthPredictor-Tool, das mit einer optimalen Dropout-Wahrscheinlichkeit (p) von 0,25 konfiguriert wurde, nutzte die vorangegangenen fünf Zeitpunkte, um vier Parameter zu identifizieren und optimale Vorhersagen zu erzielen. Die vollständig verbundene Schicht war mit 256 Knoten ausgestattet, und die LSTM-Schicht umfasste 128 Knoten. Das Modell wurde 800 Epochen lang trainiert. Der mittlere quadratische Fehler für die Trainingsdaten beträgt 0,00057 und für die Testdaten 0,01456. Ohne Verwendung unseres Vorhersagemodells hat ein verallgemeinertes additives Modell (GAM) einen MSE von 0,0048 für die Trainingsdaten, was ungefähr 8,42-mal höher ist. Die Leistung bei den Testdaten ist deutlich schlechter, daher werden sie nicht zum Vergleich einbezogen.

Die in Abbildung 5 visualisierten Verläufe der relativen Mikrobiomhäufigkeit liefern wichtige Einblicke in die Dynamik der Mikrobiomveränderungen im Laufe der Zeit. Die blaue Linie in Abbildung 5 stellt alle tatsächlichen Daten dar, während die orange Linie gleichzeitig mit der vorhergesagten Linie (grün) hervorgehoben wird. In unserem Deep-Learning-Algorithmus haben wir die vorherigen fünf Zeitpunkte verwendet, um den nächsten vorherzusagen. Bemerkenswerte Variationen, insbesondere für Rikenellaceae,

wurden aufgrund der Störung der ersten beiden Gefäße zwischen Tag 11 und 13 beobachtet. Diese Visualisierungen zeigen signifikante Verschiebungen in mikrobiellen Populationen und unterstreichen die Genauigkeit des Modells bei der Erfassung zeitlicher Veränderungen. Die beobachteten Muster stimmen mit unseren statistischen Analysen überein und bestätigen wesentliche Änderungen in der Mikrobiomzusammensetzung während dieser Zeit. Diese Ausrichtung stärkt unser Verständnis der Mikrobiodynamik und ihrer Reaktionen auf experimentelle Bedingungen.

Entgegen der Vorstellung, dass mehr Daten zu besseren Vorhersagen führen, ergab unser Experiment mit zusätzlichen Trainingsschiffen (darunter 1, 3 und 4) zur Vorhersage des zweiten Schiffs einen mittleren quadratischen Testfehler von 0,0265, also fast das Doppelte des ursprünglichen Testfehlers. Interessanterweise beträgt die Korrelation zwischen dem vorhergesagten Wert und dem tatsächlichen Wert beim Testen des Gefäßes 2 0,70 und ist damit 18 % höher als bei der Einbeziehung der Gefäße 1, 3 und 4.

Dies deutet darauf hin, dass eine sorgfältige Ausgewogenheit bei der Auswahl der Trainingsdaten von entscheidender Bedeutung für die Erzielung genauer Vorhersagen ist.

Im Bereich wissenschaftlicher Studien herrscht oft die Überzeugung vor, dass die Genauigkeit durch die Einbeziehung von mehr Datensätzen oder Informationen für das Training erhöht wird. Eine kritische Überlegung ergibt sich jedoch, wenn sich die Umgebung, in der das Modell trainiert wird, erheblich von der Umgebung unterscheidet, in der es zum Testen eingesetzt wird. Diese Diskrepanz in den Umgebungsbedingungen kann zu unvorhergesehenen Störungen und Herausforderungen führen.

In unserem Experiment wurde die anfängliche Annahme, dass mehr Trainingsdaten (einschließlich der Schiffe 1, 3 und 4) die Vorhersagen zwangsläufig verbessern würden, durch die beobachteten Ergebnisse in Frage gestellt. Die Unterbrechungen der Futterzufuhr in den ersten beiden Gefäßen zwischen Tag 11 und 13 führten zu Abweichungen in der mikrobiellen Dynamik, die durch die zusätzlichen Trainingsdaten nicht ausreichend erfasst wurden. Die unvorhergesehenen Unterbrechungen unterstreichen, wie wichtig es ist, die Trainingsdaten an die in der Testumgebung zu erwartenden Bedingungen und Störungen anzupassen.

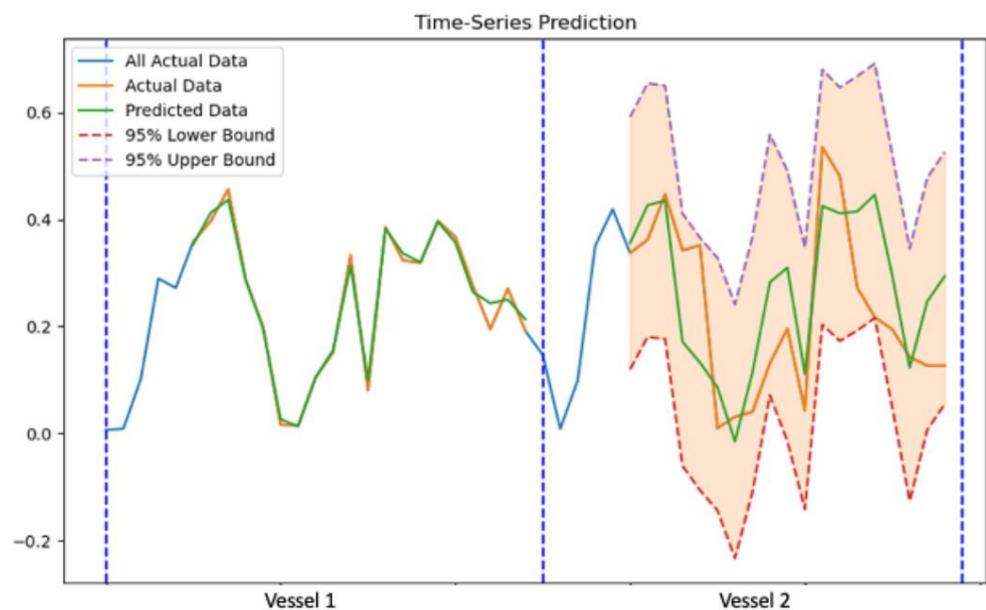


Abbildung 5. Verläufe der relativen Häufigkeit von Rikenellaceae in den Gefäßen 1 und 2. Der gesamte Verlauf von Gefäß 2 wird durch das Modell MicroGrowthPredictor vorhergesagt, das mit Daten aus Gefäß 1 trainiert wurde. Für die Testdaten von Vessel 2 werden Konfidenzintervalle bereitgestellt. In diesem Experiment wurde die optimale Dropout-Wahrscheinlichkeit p von 0,25 verwendet. Das Modell nutzte die vorangegangenen fünf Zeitpunkte, um vier Parameter zu identifizieren und optimale Vorhersagen zu erzielen. Die vollständig verbundene Schicht war mit 256 Knoten ausgestattet, und die LSTM-Schicht umfasste 128 Knoten. Das Modell wurde 800 Epochen lang trainiert.

Es ist zwar verlockend anzunehmen, dass eine größere Stichprobe automatisch zu besseren Vorhersagen führt, der Schlüssel liegt jedoch in der Relevanz der Trainingsdaten für die Testbedingungen. In Fällen, in denen die Testdaten unterschiedliche Umweltunterbrechungen oder Störungen beinhalten, kann dies zu erheblichen Abweichungen führen.

Das blinde Einbeziehen unterschiedlicher Datensätze kann zu suboptimalen Vorhersagen führen. Das empfindliche Gleichgewicht zwischen der Menge und Relevanz der Trainingsdaten ist entscheidend, um die Anpassungsfähigkeit des Modells an reale Szenarien sicherzustellen.

4. Diskussion

Das MicroGrowthPredictor -Modell nutzt Erkenntnisse aus Beobachtungen, die zeigen, dass wiederholte Antibiotikabehandlungen die mikrobielle Darmgemeinschaft stören und sich auf die Vielfalt und Häufigkeit bestimmter Bakteriengruppen auswirken. Durch die Analyse der Daten sagt das Modell genau voraus, wie sich das Mikrobiom im Laufe der Zeit als Reaktion auf eine Antibiotikastörung verändern wird. Dies ermöglicht ein tieferes Verständnis der Auswirkungen von Antibiotika auf die Darmmikrobiota und der möglichen Folgen für die menschliche Gesundheit.

Darüber hinaus wird die Vielseitigkeit des Modells durch seine Anwendung auf einen künstlichen Darmdatensatz demonstriert. Erkenntnisse aus dieser kontrollierten Umgebung zeigen die Anpassungsfähigkeit von MicroGrowthPredictor an verschiedene Mikrobiomsysteme. Der künstliche Darmdatensatz validiert die Vorhersagefähigkeiten des Modells unter bestimmten Bedingungen und unterstreicht seine Fähigkeit, komplexe zeitliche Dynamiken zu erfassen. Dies macht das Modell wertvoll für das Verständnis der Wirkung von Antibiotika und für breitere Anwendungen in den Umweltwissenschaften, im Gesundheitswesen und in der Biotechnologie.

Unsere Methode befasst sich mit realen Problemen, bei denen begrenzte Stichprobengrößen aufgrund logistischer, ethischer oder finanzieller Herausforderungen eine Einschränkung darstellen. Indem wir Methoden entwickeln und validieren, die mit begrenzten Daten gut funktionieren, bieten wir praktische Lösungen für solche Situationen.

Im Gegensatz zu vielen Black-Box-Modellen bietet unser Ansatz klare Einblicke in die Art und Weise, wie Umweltstörungen mikrobielle Populationen im Laufe der Zeit beeinflussen. Dies ist für das Verständnis biologischer Prozesse und die Entwicklung gezielter Interventionen von entscheidender Bedeutung. Insbesondere diskutieren wir das Potenzial dieses Ansatzes, durch die Vorhersage individueller Reaktionen auf Ernährungsumstellungen, Antibiotikabehandlungen und probiotische Interventionen zu personalisierten Behandlungsplänen beizutragen.

Zusammenfassend erweist sich MicroGrowthPredictor als leistungsfähiges Tool, das traditionelle Modellierungsansätze übertrifft. Das Modell basiert auf Erkenntnissen aus Daten und nicht auf der direkten Integration von Wissen. Es integriert LSTM-Netzwerke mit Konfidenzintervallschätzung, um zu einem ganzheitlichen Verständnis der Mikrobiodynamik beizutragen. Die erfolgreiche Anwendung des Modells sowohl auf reale menschliche Darmmikrobiota als auch auf künstliche Darmdatensätze unterstreicht seine Wirksamkeit und potenzielle Wirkung. Wir gehen davon aus, dass MicroGrowthPredictor eine entscheidende Rolle bei der Weiterentwicklung der Mikrobiomforschung spielen wird, wertvolle Erkenntnisse bietet und zu fundierten Entscheidungen in verschiedenen Bereichen beiträgt.

Beiträge der Autoren: Konzeptualisierung, Y.-HZ; Methodik, GS und Y.-HZ; Validierung, GS und Y.-HZ; Schreiben – Originalentwurf, GS und Y.-HZ; Schreiben – Überprüfung und Bearbeitung, GS und Y.-HZ; Visualisierung, GS und Y.-HZ; Überwachung, Y.-HZ; Projektverwaltung, Y.-HZ; Mittelbeschaffung, Y.-HZ. Alle Autoren haben die veröffentlichte Version des Manuskripts gelesen und stimmen ihr zu.

Finanzierung: Diese Forschung wurde von der US-Umweltschutzbehörde (Zuschussnummer 84045001), dem National Institute of Health P30ES025128 und dem Engineering Research Centers Program der National Science Foundation im Rahmen der NSF-Kooperationsvereinbarung Nr. EEC-2133504 finanziert.

Datenverfügbarkeitserklärung: Die Daten sind im Artikel enthalten.

Interessenkonflikte: Die Autoren erklären, dass die Forschung ohne kommerzielle oder finanzielle Beziehungen durchgeführt wurde, die als potenzieller Interessenkonflikt ausgelegt werden könnten.

Verweise

1. Altve, S.; Yildiz, HK; Vural, HC Interaktion der Mikrobiota mit dem menschlichen Körper bei Gesundheit und Krankheit. *Biosci. Microbiota Food Health* **2020**, *39*, 23–32. [[CrossRef](#)] [[PubMed](#)]
2. Smith, J.; Johnson, M. Mikrobiodynamik unter Umweltstörungen. *J. Microbiome Res.* **2022**, *10*, 123–145.
3. Brown, EM; Sadarangani, M.; Finlay, BB Die Rolle des Immunsystems bei der Steuerung der Interaktionen zwischen Wirt und Mikrobe im Darm. *Nat. Immunol.* **2013**, *14*, 660–667. [[CrossRef](#)] [[PubMed](#)]
4. Candela, M.; Biagi, E.; Maccaferri, S.; Turrone, S.; Brigidi, P. Die Darmmikrobiota ist ein plastischer Faktor, der auf Umweltveränderungen reagiert. *Trends Microbiol.* **2012**, *20*, 385–391. [[CrossRef](#)]

5. Uhr, GT; Dohnalová, L.; Thaiss, CA Die Dimension der Zeit in Wirt-Mikrobiom-Interaktionen. *mSystems* **2019**, *4*, e00216-18.
[\[Querverweis\]](#) [\[PubMed\]](#)
6. Willing, BP; Russell, SL; Finlay, BB Das Gleichgewicht verschieben: Auswirkungen von Antibiotika auf den Mutualismus zwischen Wirt und Mikrobiota. *Nat. Rev. Microbiol.* **2011**, *9*, 233–243. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Brown, E.; Williams, D. Prädiktive Modellierung des Mikrobiomwachstums mithilfe von LSTM-Netzwerken. *J. Comput. Biol.* **2021**, *45*, 321–335.
8. Ching, T.; Himmelstein, DS; Beaulieu-Jones, BK; Kalinin, AA; Do, BT; Way, GP; Ferrero, E.; Agapow, PM; Zietz, M.; Hoffman, MM; et al. Chancen und Hindernisse für Deep Learning in Biologie und Medizin. *JR Soc. Interface* **2018**, *15*, 20170387.
[\[Querverweis\]](#) [\[PubMed\]](#)
9. Medsker, LR; Jain, L. Rekurrierende neuronale Netzwerke. *Des. Appl.* **2001**, *5*, 2.
10. Graves, A.; Graves, A. Langzeit-Kurzzeitgedächtnis. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Deutschland, 2012; S. 37–45.
11. Yu, Y.; Si, X.; Hu, C.; Zhang, J. Eine Übersicht über rekurrierende neuronale Netzwerke: LSTM-Zellen und Netzwerkarchitekturen. *Neuronale Berechnung.* **2019**, *31*, 1235–1270. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Zhou, YH; Gallins, P. Eine Übersicht und ein Tutorial zu maschinellen Lernmethoden zur Vorhersage von Eigenschaften des Mikrobiomwirts. *Front. Genet.* **2019**, *10*, 579. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhu, L.; Laptev, N. Tiefe und zuverlässige Vorhersage für Zeitreihen bei Uber. In *Proceedings der 2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, Orleans, LA, USA, 18.–21. November 2017; IEEE: Piscataway, NJ, USA, 2017; S. 103–110.
14. Gal, Y.; Ghahramani, Z. Dropout als bayesianische Näherung: Darstellung von Modellunsicherheit beim Deep Learning. In *Proceedings of the International Conference on Machine Learning*, PMLR, New York, NY, USA, 20.–22. Juni 2016; S. 1050–1059.
15. Dethlefsen, L.; Relman, DA Unvollständige Genesung und individuelle Reaktionen der menschlichen distalen Darmmikrobiota auf wiederholte Antibiotikastörung. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4554–4561. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Silverman, JD; Durand, HK; Bloom, RJ; Mukherjee, S.; David, LA Dynamische lineare Modelle leiten Design und Analyse von Mikrobiota-Studien im künstlichen menschlichen Darm. *Microbiome* **2018**, *6*, 202.

Haftungsausschluss/Anmerkung des Herausgebers: Die in allen Veröffentlichungen enthaltenen Aussagen, Meinungen und Daten sind ausschließlich die der einzelnen Autoren und Mitwirkenden und nicht die von MDPI und/oder den Herausgebern. MDPI und/oder die Herausgeber lehnen jegliche Verantwortung für Personen- oder Sachschäden ab, die aus den im Inhalt erwähnten Ideen, Methoden, Anweisungen oder Produkten resultieren.