

# Abstract

Orphan diseases are characterized by low incidence and complex pathology. The conventional examination methods are arduous but cannot guarantee the correctness of a diagnosis. The limited acquaintance with and knowledge about rare disorders, coupled with insufficient epidemiological evidence and clinical experiences, makes it challenging for general practitioners to identify rare diseases reliably. Medication mistakes based on an inaccurate diagnosis may cause adverse effects on the patient's health. With advancements in artificial intelligence, decision support systems are gradually gaining prominence in modern medical diagnosis. An innovative medical service model centered on patient symptoms provides technological advantages in improving the future direction of orphan disease diagnostic and treatment processes.

This research aims to amass a rare disease repository by collecting diagnosed patients' information and complementary therapeutic schemes, thereby providing elaborate references for disease diagnostics and reducing gaps in the interpretation of rare disease symptoms. A prospective study was created in cooperation with Medizinische Hochschule Hannover. The interdisciplinary research focuses on developing a computer-aided diagnostic support system based on statistical analysis and machine learning algorithms to optimize rare disease diagnosis processes. An ensemble of classification methods was conceptualized to discover the potential rules inherent in the patient questionnaire data, thus incorporating disease profile aggregation, pattern recognition, and information comparison. Moreover, an interactive data visualization platform was successively established, promoting intuitive understanding through diagnosis graphics that depict orphan diseases' specific characteristics. Data visualization strengthens the transmission efficiency of important information, motivates disease researchers to explore data, and provides in-depth insights into problematic connected symptoms. The practically realized concept reveals evidence to sensitize doctors to correct decision-making based on an internal evaluation for system-aided diagnosis. A self-adapting questioning mechanism results in personalized questionnaires that automatically analyze suspicious disease trends and accumulate experience.

An iterative validation process reveals that the classifier fusion method presented notable performance improvements compared to any single supervised classification model. The patient-oriented inquiry algorithm can efficiently reduce unnecessary questions while having a reliable diagnosis. By effectively combining theory with practice, the statistical learning method combined with the visualization module can systematically present disease-related symptoms patterns, thereby guiding medical practitioners in increasing their understanding of orphan disease characteristics, and providing new means for rare disorder diagnostics based on the system's learning results from previously diagnosed patients. The supplementary diagnosis prediction mechanism introduced in this doctoral dissertation is not confined to questionnaires on orphan diseases. Indeed, the methodological procedure can be applied to any survey with closed-ended questions for researchers interested in analyzing the predispositions and significant influence factors of different groups.

# Zusammenfassung

Seltene Krankheiten haben eine geringe Inzidenz und eine komplexe Pathologie. Die herkömmlichen Untersuchungsmethoden sind mühsam, können jedoch die Richtigkeit einer Diagnose nicht immer garantieren. Begrenztes Wissen und Kenntnisse über seltene Erkrankungen in Verbindung mit unzureichenden epidemiologischen Nachweisen und klinischen Erfahrungen machen es für Allgemeinmediziner schwierig, seltene Krankheiten zuverlässig zu identifizieren. Falsche oder suboptimale Medikationen, die auf einer ungenauen Diagnose zurückzuführen sind, können sich nachteilig auf die Gesundheit des Patienten auswirken. Im Zeitalter der computergestützten künstlichen Intelligenz werden Entscheidungssysteme zunehmend als Assistenz in der modernen medizinischen Diagnose verwendet. Das in dieser Arbeit entwickelte medizinische Servicemodell, das sich auf Patientensymptome konzentriert, bietet die technologischen innovativen Vorteile, um die zukünftige Ausrichtung für Diagnoseprozesse für seltene Erkrankungen und die Behandlung von Patienten zu verbessern.

Diese Forschung zielt darauf ab, eine zentrale Wissensbasis für seltene Krankheiten aufzubauen, indem Informationen diagnostizierter Patienten und ergänzende therapeutische Schemata zentral gesammelt werden, wodurch wertvolle Referenzen für die Diagnostik von Krankheiten bereitgestellt und Lücken bei der Interpretation seltener Krankheitssymptome geschlossen werden. In Kooperation mit der Medizinischen Hochschule Hannover wurde eine prospektive Studie durchgeführt. Die interdisziplinäre Forschung konzentriert sich auf die Entwicklung eines computergestützten diagnostischen Unterstützungssystems, das auf statistischen Analysen und Algorithmen für maschinelles Lernen basiert, um Prozesse zur Diagnose seltener Krankheiten zu optimieren. Ein Ensemble von Klassifizierungsmethoden wurde konzipiert, um die potenziellen Symptommuster zu ermitteln, die aus Daten eines Patientenfragebogens der MHH hervorgehen, und so die systematische Aggregation von Krankheitsprofilen, die Mustererkennung und den Informationsvergleich berücksichtigen. In diesem Zusammenhang wurde nacheinander eine interaktive Datenvisualisierungsplattform entwickelt, die ein intuitives Verständnis durch Diagnosegrafiken fördert, die die spezifischen Merkmale von seltenen Erkrankungen darstellen. Die Datenvisualisierung stärkt die Übertragungseffizienz wichtiger Informationen, motiviert Krankheits-

forscher zur Erforschung von Daten und bietet detaillierte Einblicke in Problemverbundene Symptome. Das praktisch realisierte Konzept ermöglicht belegbare Diagnosen, um Ärzte für eine korrekte Entscheidungsfindung auf der Grundlage einer internen Bewertung für eine systemgestützte Diagnose zu sensibilisieren. Ein dazu entwickelter Fragenalgorithmus führt zu personalisierten Fragebögen basierend auf verdächtigen Krankheitstrends, welche automatisch analysiert werden, um nicht relevante Symptomfragen für den Patienten auszugrenzen.

Ein iterativer Validierungsprozess zeigt, dass die Classifier-Fusion-Methode im Vergleich zu einem einzelnen überwachten Klassifizierungsmodell bemerkenswerte Leistungsverbesserungen aufweist. Der patientenorientierte Fragenstellungsalgorithmus kann unnötige Fragen effektiv reduzieren und gleichzeitig eine zuverlässige Diagnose liefern. Durch die effektive Kombination von Theorie und Praxis kann die statistische Lernmethode in Kombination mit dem Visualisierungsmodul krankheitsbezogene Symptommuster systematisch darstellen, wodurch Ärzte bei der Verständnisverbesserung der Merkmale seltener Krankheiten angeleitet werden und neue Mittel für die Diagnose seltener Erkrankungen auf der Grundlage des Systems bereitgestellt werden. Lernergebnisse von zuvor diagnostizierten Patienten. Der in dieser Dissertation vorgestellte ergänzende Diagnosevorhersagemechanismus ist nicht auf Fragebögen zu seltenen Krankheiten beschränkt. Tatsächlich kann das methodische Vorgehen auf jede Umfrage mit geschlossenen Fragen für Forscher angewendet werden, die daran interessiert sind, die Prädispositionen und signifikanten Einflussfaktoren verschiedener Gruppen zu analysieren.