

Retrospektive Erkennung von ARDS mit Hilfe einer Support Vector Machine

(Bachelorarbeit)



MESUT IBRAHIM TASTEKIN

Motivation

Akutes Lungenversagen (engl. Acute Respiratory Distress Syndrome [ARDS]) ist eine schwere Beeinträchtigung der Lungenfunktion und hat eine hohe Sterblichkeitsrate, welche durch die häufig zu späte oder ausbleibende Diagnose begründet ist. Die fehlenden Diagnosen wirken sich auch auf die verfügbaren Sekundärdatenbanken aus, welche für die weitere Erforschung des Krankheitsbildes genutzt werden. Um die Forschung zu unterstützen, ist eine hohe Qualität des Labelings der Patienten fundamental. Aus diesem Grund ist eine retrospektive Klassifizierung der Patientengruppen wichtig. Hierfür kann eine Support Vector Machine (SVM) eingesetzt werden, welche in vielen Bereichen der Medizin für binäre Klassifizierungsaufgaben eingesetzt werden.

Stand der Technik

SVMs werden in mehreren wissenschaftlichen Arbeiten für die Klassifizierung von Krankheiten genutzt. In einer Arbeit wird beispielsweise eine SVM für die Früherkennung und Klassifikation von Covid-19 verwendet, während in einer anderen Forschungsarbeit Patienten mit Schilddrüsenunterfunktion mit Hilfe einer Mehrklassen-SVM, also eine Kombination aus mehreren binären SVMs, klassifizieren und mit anderen Machine Learning Verfahren vergleichen.

Am Lehrstuhl Informatik 11 wurde bereits ein Bayes'sches Netz für die Erkennung von ARDS implementiert. Darüber hinaus wird die Nutzung von Random-Forest Algorithmen für dieses Problem evaluiert. In der Literatur wurde bereits die Anwendung einer SVM für die Erkennung von ARDS getestet. Da der verwendete Datensatz und die genutzten Features für das Training der SVM nicht veröffentlicht sind, ist die weitere Nutzung dieses Ansatzes nicht gewährleistet, dient aber dennoch als erster Ansatz für die Erstellung einer SVM mit den am Lehrstuhl zur Verfügung stehenden Datensätzen.

Zielsetzung

Ziel dieser Bachelorarbeit ist die Implementierung einer SVM zur retrospektiven Klassifizierung von ARDS bei intensivmedizinischen Patientendaten. Hierzu werden die Daten aus den verschiedenen Sekundärdatenbanken so aufbereitet, dass mit diesen die Erstellung einer SVM möglich ist. Hierbei dient ein Teil der Daten dem Training und der restliche Teil der Evaluation. Die entwickelte SVM soll dazu dienen, die ARDS-Diagnose von Patienten in den Datenbanken zu verifizieren.

Geplante Vorgehensweise

Nach einer anfänglichen Literaturrecherche erfolgt die Methodenidentifikation, bei der u.a. bereits vorhandene Bibliotheken für die SVM untersucht werden. Für die Implementierung der SVM werden mit Hilfe verschiedener Feature Selection Verfahren relevante Parameter identifiziert. Schließlich folgt eine Evaluation mit Hilfe der bereitstehenden Datenbanken und k-fold cross validation. Abschließend werden die Ergebnisse mit den verwandten Arbeiten verglichen.