Maschinelles Lernen in der Radiologie:

Mit wenig Trainingsdaten doch eine exakte Diagnose erhalten

In medizinischen Datensätzen überwiegen in der Regel die Röntgenbilder gesunder Menschen. Nur wenige Bilder weisen auf eine Krankheit hin. Wie können Machine-Learning-Modelle für die medizinische Bildanalyse ohne ausbalancierten Datensatz doch erfolgreich trainiert werden?

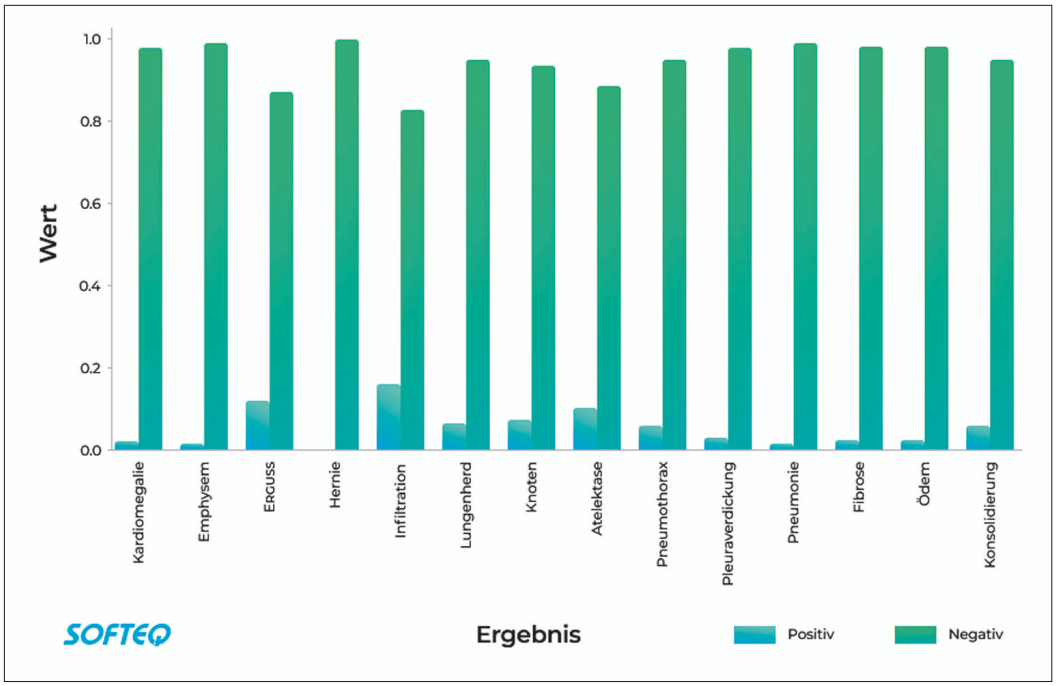
Röntgenaufnahmen des Brustkorbs helfen Ärzten bei der Diagnose vieler Erkrankungen, einschließlich Lungenentzündung und Lungenkrebs. Um zu verstehen, wie maschinelles Lernen den Prozess erleichtern könnte, schauen wir uns diese Röntgenbilder einmal genauer an (Bild 1).

Links sieht man ein Röntgenbild des Brustkorbs eines gesunden Patienten, rechts das eines Patienten mit bestätigter Lungenentzündung. Ärzte, ohne radiologischen Schwerpunkt, werden geschwollene und entzündete Bereiche in der Lunge eines Patienten auf diesen Bildern kaum erkennen. Selbst zertifizierte Radiologen können den Verdacht auf eine Lungenentzündung mit anderen Erkrankungen verwechseln – und es dauert etwa 20 Minuten, bis Ärzte die Krankheit bestätigen oder ausschließen können.

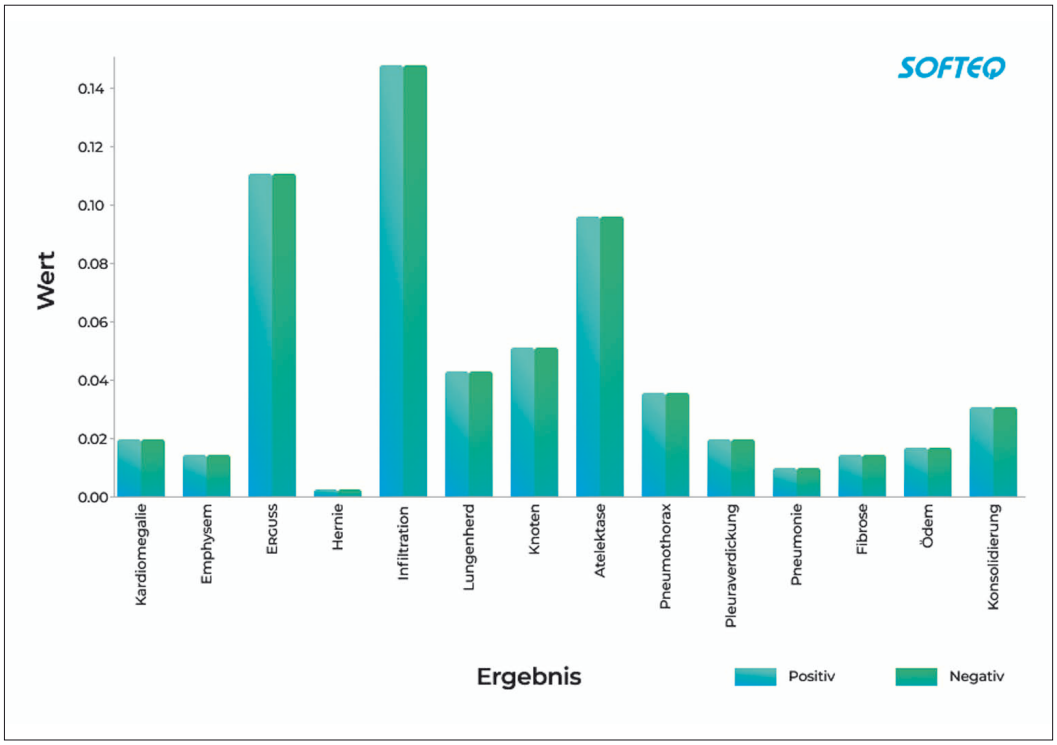
ML-Modelle können diese Aufgabe in nur 10 Sekunden erledigen. Die Zeit ist entscheidend, wenn es um eine dringend benötigte Behandlung geht.

ML-Algorithmen in der medizinischen Bildgebung: Herausforderungen

- Zu kleine Datensätze: Um die neuronalen Netzwerke in smarten Lösungen zur medizinischen Bildgebung zu trainieren, benötigt man Tausende von kommentierten Röntgenbildern. Aufgrund der DSGVO-Bestimmungen ist es jedoch oft schwer, diese Bilder zu erhalten. Um dieses Problem zu lösen, können Spezialisten für maschinelles Lernen vorhandene Algorithmen zur Bildklassifizierung wiederverwenden (Transfer-Lernen) und die Größe eines Datensatzes künstlich erweitern, indem sie Röntgenbilder modifizieren (Datenerweiterung).
- Überlappende Symptome von Lungenkrankheiten: Ein eigenstän-



Softeq
www.softeq.com



diges ML-Modell zu entwickeln, das mehrere Krankheiten mit überlappenden Symptomen unterscheidet, ist keine triviale Aufgabe. Entwickler lösen das Problem mithilfe einer Multi-Task-Verlustfunktion, die die Genauigkeit von Algorithmen mit Multi-Task-Learning (MTL)-Fähigkeiten schätzt. Durch Modifizierung der Verlustfunktion ist es möglich, die Leistung von ML-Algorithmen in Bezug auf eine bestimmte Bildklasse und auch auf den gesamten Datensatz gleichzeitig zu bewerten.

- Unbalancierte Datensätze: In medizinischen Datensätzen überwiegen in der Regel die Bilder gesunder Menschen gegenüber den Röntgenbildern, die auf eine Krankheit hinweisen. Dies liegt in der Natur der Sache. Entwickler benötigen jedoch eine gleiche Anzahl beider Bildtypen, um die Genauigkeit von ML-Algorithmen zu erhöhen. Wie können die Herausforderungen beim Training der Algorithmen für medizinische Bildgebung bewältigt werden?

Mit weniger mehr erreichen

Intelligente Bildgebungslösungen mit unausgewogenen Datensätzen: Um ML-Algorithmen zu trainieren, eine Lungenentzündung genau zu diagnostizieren und den Ort und das Ausmaß der Infektion zu bestimmen, benötigen Entwickler eine gleiche Menge an Röntgenbildern mit positiven sowie negativen Befunden. Diese stellen zwei verschiedene Bildklassen dar. Wie schwierig es ist, diese ausgewogenen Datensätze zu erhalten, zeigt das Balkendiagramm in Bild 2, das auf einem Datensatz des National Institute of Health (NIH) basiert. Hier ist das Verhältnis zwischen positiven und negativen Röntgenbefunden abgebildet.

Aus der Grafik wird deutlich, dass es an Röntgenbildern von gesunden Lungen nicht mangelt, es aber zu wenige Röntgenbilder von erkrankten Lungen gibt. Lungenhernien werden z. B. auf nur 0,2 % der Röntgenbilder beobachtet. Auch bei Lungeninfiltraten, die auf eine Lungenentzündung, Nokardiose oder Tuberkulose hinweisen können, stehen positive Röntgenbefunde in einem Verhältnis 1:5 zu negativen.

Um die Datensätze auszugleichen, können wir jedes Beispiel aus jeder Klasse mit dem entsprechenden Gewichtungskoeffizienten (Wpos, Wneg) multiplizieren. Mit folgender Formel können wir die Gewichtungskoeffizienten berechnen. Freqp und Freqn stehen für die Häufigkeit der positiven und negativen Röntgenergebnisse im Datensatz:

Wpos × Freqp = Wneg × Freqn

So können wir die Parameter Freqp und Freqn berechnen:

positive_frequencies =
np.sum(labels, axis=0) / N

negative_frequencies =
1 - positive_frequencies

Diese Methode hilft, das Ungleichgewicht in den Trainingsdatensätzen zu beseitigen (Bild 3).

Nun müssen wir die Genauigkeit des ML-Modells bewerten. Und hier kommt die Verlustfunktion ins Spiel. Wenn das Modell Ergebnisse vorhersagt, die weit vom Zielwert entfernt sind, „bestrafen“ wir den Algorithmus, bis er richtige Prognosen stellt.

Binäre Kreuzentropie-Funktion

Es gibt verschiedene Arten von Verlustfunktionen, die beim maschinellen Lernen verwendet werden. Für unsere Zwecke werden wir die binäre Kreuzentropie-Funktion verwenden. Sie misst die Leistung von Klassifizierungsmodellen, deren Ausgabe zwischen null und eins liegt.

Jetzt können wir die Kreuzentropie-Verlustfunktion modifizieren:

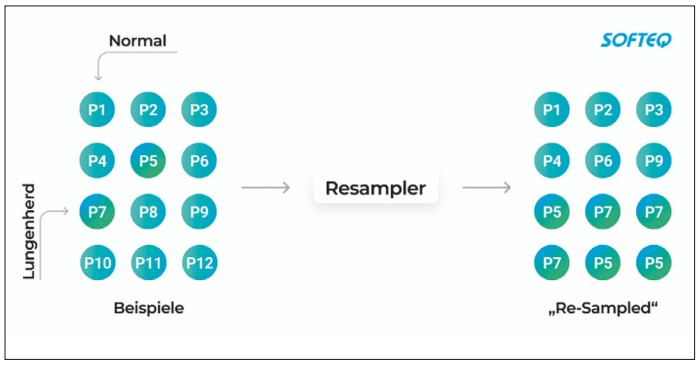
Lcross-entropy(xi) =
-(yi*log(f(xi)) + (1-yi)*log(1-f(xi)))

Hier sind xi und yi die Eingabe und das Label, und f(xi) ist die Ausgabe des ML-Modells, in unserem Fall die Wahrscheinlichkeit, dass ein Patient tatsächlich eine Lungenentzündung hat.

Nach der Berechnung der Gewichtungskoeffizienten ermitteln wir den endgültigen gewichteten Verlust für jeden Trainingsfall:

Lcross-entropy(x) = -(wp*y*log(f(x)) + wn*(1-y)*log(1 -f(x)))

„resampeln“

Eine weitere Möglichkeit, die Genauigkeit eines ML-Modells zu erhöhen, besteht darin, Datensätze so lange zu „resampeln“, bis wir eine gleiche Menge an positiven und negativen Röntgenbildern erhalten (Bild 4):

Um ein Ergebnis zu erhalten, kombinieren wir die Bilder und wählen die gleiche Menge an negativen und positiven Röntgenbildern aus. Nun können wir die Verlustfunktion verwenden, ohne den Gewichtungskoeffizienten für die Bildklassen zu berechnen. Dieser Ansatz hat jedoch einen erheblichen Nachteil. Da wir eine begrenzte Anzahl von Röntgenbildern aus demselben Datensatz wiederverwenden, können die Algorithmen unterdurchschnittlich abschneiden, wenn sie mit Bildern aus einem anderen Datensatz konfrontiert werden. ◀