

4. Zusammenfassung und Ausblick

Grundlage dieser Arbeit ist das Ziel, Mikroplastik von anderen Mikropartikeln unterscheiden zu können. Dazu entwickelte das Start-Up ████████ die notwendige Messhardware, die mit einer vereinfachten Variante der elektrischen Impedanzspektroskopie die komplexe Impedanz eines Partikels erfasst, das einen Durchflusssensor passiert. Mit diesen Daten wurden in früheren Arbeiten mit überwachtem Lernen bereits gute Klassifikationsergebnisse von über 90 % Richtigkeit erzielt. Da das überwachte Lernen jedoch immer durch menschliches Handeln beeinflusst und limitiert wird (manuelle Kennzeichnung der Ausgangswerte nötig), wurde in dieser Arbeit die Möglichkeit der Klassifikation mittels unüberwachtem Lernen untersucht. Dabei wurden drei Klassifizierungsaufgaben unterschieden: Zum ersten die Zuordnung zur richtigen Materialgruppe (anorganisch, biologisch und metallisch), zum zweiten die binäre Klassifikation der anorganischen Partikel in Kunststoff und Nicht-Kunststoff und zum dritten die detaillierte Klassifikation der anorganischen Partikel in Glas und die spezifischen Kunststoffarten.

Dazu wurden aus den Zeitreihendaten der Impedanzmessungen zunächst die Partikelpeaks identifiziert und anschließend die Informationen für die sechs Merkmale Betrags- und Phasenänderung, Betrags- und Phasensteigung und Betrags- und Phasenbreite bei 66 % der Höhe ermittelt. Bezogen auf die Materialgruppe waren diese Daten bei der Betrachtung der Impedanzänderung bis auf wenige Holzpartikel bereits optisch gut identifizierbar. Die Dichteregionen der einzelnen Materialgruppen variierten jedoch so stark, dass in dieser Arbeit die Weiterentwicklung des DBSCAN-Algorithmus, der HDBSCAN-Algorithmus, zur Klassifikation verwendet werden musste. Dieser findet Cluster mit unterschiedlichen Dichten, ohne vorher ein Maß für den notwendigen Abstand der Punkte zueinander definieren zu müssen.

Die Klassifikation wurde durchgeführt, indem der HDBSCAN-Algorithmus zunächst Gruppen im Datensatz identifizierte, diese anhand ihrer Schwerpunkte den richtigen Klassen zuordnete und schließlich jeden klassifizierten Punkt mit seiner richtigen Klasse verglich. Somit konnte ein quantitatives Maß für die Genauigkeit der Klassifizierung angegeben werden, ohne den Clusteralgorithmus mit zusätzlichen Daten zu verfälschen. Bezogen auf die Materialgruppen anorganisch, biologisch und metallisch konnte durch diese Vorgehensweise mit den beiden Merkmalen Betrags- und Phasenänderung eine Genauigkeit von 92,5 % erreicht werden. Lediglich nasse Holzpartikel, die sich nach kurzer Zeit wie anorganische Partikel verhalten, konnten nicht korrekt zugeordnet werden. Bei Hinzunahme der anderen vier Merkmale konnte kein Informationsgewinn festgestellt werden, da durch die Steigung die Korrektheit von 92,5 % zwar gleich blieb, aber durch die größere Anzahl der Rauschpunkte eine größere Unsicherheit entstand und sich die Genauigkeit durch die zeitliche Dauer der Peaks bei 66 % der Peakhöhe sogar auf 73,2 % verschlechterte. Auch die beiden unüberwachten Klassifikationsalgorithmen

ANN (72,5 %) und GMM (89 %) konnten die Genauigkeit des HDBSCAN-Algorithmus mit den zwei Merkmalen Betrags- und Phasenänderung nicht erreichen.

Ein anderes Ergebnis ergab sich bei der detaillierten Klassifikation der anorganischen Partikel, die rein optisch nicht wie bisher zugeordnet werden konnten. Hier stellte sich zunächst die Frage nach der Möglichkeit einer binären Einteilung in Kunststoff und Nicht-Kunststoff. Um die Ergebnisse gegen den Zufall abgrenzen zu können, wurden die Klassengrößen jeweils auf die Größe der kleineren Klasse (207) begrenzt, was zu einer binären Zufallsgenauigkeit von 50 % führte. Hier wurde die Genauigkeit bei Verwendung der zusätzlichen Merkmale deutlich verbessert. Lag die Genauigkeit des HDBSCAN-Algorithmus mit zwei Merkmalen (55,3 %) und vier Merkmalen (56,1 %) noch leicht über dem Zufall, so stieg sie mit allen sechs Merkmalen auf 74,9 %. Die Annahme, dass die ersten vier Merkmale (Impedanzänderung und Impedanzsteigerung) unwichtig sind, konnte jedoch widerlegt werden, da sich die Genauigkeit ohne diese Merkmale auf 69,5 % verschlechterte.

Die zweite Frage an dieser Stelle war die Möglichkeit einer detaillierten Klassifikation der anorganischen Partikel in Glas und die spezifischen Kunststoffarten. Hier wurde die Klassengröße auf 64 Partikel begrenzt, um das Ergebnis gegen die Zufallsgenauigkeit (16,7 %) abgrenzen zu können. Das Ergebnis betrug dann für die ersten beiden Merkmale 26,8 %, unter Hinzunahme der Betrags- und Phasensteigerung 23,2 % und für alle Merkmale 34,2 %. Das bedeutet bei Verwendung aller Merkmale eine Verdopplung der Genauigkeit gegenüber dem Zufall. Hier wurde auch das Ergebnis betrachtet, wenn die Peaksteigungen weggelassen werden, da diese einen scheinbar negativen Einfluss auf die Genauigkeit haben. Die erhoffte Verbesserung der Genauigkeit konnte allerdings nur bis zu 0,3 % nachgewiesen werden. Nicht vergessen werden darf an dieser Stelle jedoch, dass ausschließlich anorganische Partikel betrachtet wurden. Sollten alle in Kapitel 3.2 als anorganisch klassifizierten Partikel untersucht werden, würde sich voraussichtlich ein größerer Fehler durch die Klassifizierungen ziehen.

Insgesamt zeigt diese Arbeit das Potential des unüberwachten Lernens für die Klassifikation von Mikropartikeln. Zwar konnten die Ergebnisse des überwachten Lernens nicht übertroffen werden, aber gerade bei der Klassifikation der Materialgruppen anorganisch, biologisch und metallisch sowie bei der Unterscheidung zwischen Kunststoff und Nicht-Kunststoff wurden positive Ergebnisse erzielt. Für die weitere Untersuchung dieses Themas ist es jedoch unerlässlich, einen größeren Datensatz einzubeziehen, um die Klassengrößen nicht auf sehr kleine Werte beschränken zu müssen und um stärkere Tendenzen in den Daten erkennen zu können. Ein weiterer offener, aber sehr bedeutender Punkt ist die Angabe der Unsicherheit von Ergebnissen, die durch maschinelles Lernen generiert wurden. Derzeit gibt es vereinzelte Bemühungen, Konzepte zu entwickeln, die die Unsicherheit von überwachten Lernsystemen

quantifizieren können. Beim unüberwachten Lernen ist dies jedoch wesentlich schwieriger, da das Modell nicht auf vorher bekannte Ausgangswerte trainiert wird, sondern ausschließlich nach Mustern und Abhängigkeiten in den Daten sucht. Ein möglicher Ansatz wäre die Angabe der durchschnittlichen Distanz der Datenpunkte eines Clusters zum Clusterzentrum, die im vorliegenden Anwendungsfall allerdings mit einer Dichteabhängigkeit angegeben werden müsste, da, wie in dieser Arbeit gezeigt wurde, die Datenpunkte verschiedener Materialien unterschiedliche Dichtewerte besitzen. Des Weiteren könnte bei der Zuordnung der gefundenen Cluster zu den Materialgruppen die Distanz der Schwerpunkte der Cluster zu den wahren Materialgruppen verwendet werden, um Aussagen über die Unsicherheit des Klassifikationsergebnisses treffen zu können. Dies würde jedoch eine fundierte Kenntnis der Schwerpunkte der Materialgruppen voraussetzen, was sich aufgrund unterschiedlicher Partikelgrößen und der Menge der benötigten Datenpunkte als schwierig erweisen könnte. Deshalb könnte zukünftig an einer alternativen Zuordnung der gefundenen Cluster zu den gewünschten Klassen geforscht werden, da diese sicherlich noch andere Identifikationsmöglichkeiten als den Schwerpunkt ihrer Merkmale besitzen.