Model Drift

Nhut Hoa Huynh[1]

[1] Department Informatik, HAW Hamburg, Berliner Tor 7, 20099 Hamburg
nhuthoa.huynh@haw-hamburg.de

Abstrakt. Das Papier beschreibt das Konzept Modelldrift und ihre potenziellen Auswirkungen. Es werden verschiedene Arten der Modelldrift untersucht, darunter die Datendrift und die Konzeptdrift. Außerdem werden die Ursachen der Modelldrift und Methoden zur Erkennung und Behandlung dieses Phänomens untersucht. Darüber hinaus enthält die Arbeit ein praktisches Beispiel für die Erkennung von Modelldrift unter Verwendung von „evidently“.

Keywords: Modeldrift, Datendrift, Konzeptdrift, Kovariate Drift, Feature Drift, Label Drift, plötzliche Drift, allmähliche Drift, wiederkehrende Drift, Überwachung von ML- Modellen, Retraining, Model-Tuning, Transfer-Lernen, evidently.

1 Einleitung

1.1 Ziele der Arbeit

Ziel dieser Arbeit ist es, Einblicke in Modelldriften zu gewinnen. Es werden verschiedene Arten von Modelldriften sowie deren mögliche Ursachen untersucht. Eine wichtige Fragestellung ist außerdem, wie man Modelldriften erkennen kann. Darüber hinaus wird untersucht, welche Maßnahmen zur Behandlung von Modelldriften eingesetzt werden können. Um das Fachwissen zu vertiefen, wird auch ein Praxisbeispiel durchgeführt. Durch diese Arbeit sollen wichtige Erkenntnisse gewonnen werden, um in der Praxis erfolgreich mit Modelldriften umzugehen.

1.2 Gliederung der Arbeit

Diese Arbeit ist in mehrere Abschnitte unterteilt, um eine systematische Untersuchung von Modelldriften zu ermöglichen. Kapitel 2 befasst sich mit den Grundlagen der Modelldrift. Hier werden die wichtigsten theoretischen Konzepte erläutert. In Kapitel 3 werden Methoden zur Erkennung und Behandlung von Modelldriften vorgestellt. In Kapitel 4 werden Fallbeispiele vorgestellt. Schließlich werden in Kapitel 5 die Ergebnisse der Arbeit zusammengefasst und kritisch reflektiert. Darüber hinaus wird ein Ausblick auf zukünftige Forschung gegeben.

2 Grundlagen

2.1 Definition & Auswirkungen von Modelldriften

Der Begriff Modelldrift (auch „model decay“) beschreibt das Phänomen, dass die Leistung von Modellen für maschinelles Lernen mit der Zeit verschlechtert wird. Wenn ein Modell von Modelldrift betroffen ist, kann dies dazu führen, dass die Vorhersagen des Modells ungenau oder fehlerhaft werden.

Die ungenauen oder fehlerhaften Vorhersagen der Modelle können schwerwiegende Folgen haben. Unternehmen, die sich auf die Vorhersagen ihrer Modelle verlassen, können möglicherweise schlecht planen oder ineffiziente Kampagnen durchführen. In der Produktion können fehlerhafte Modelle dazu führen, dass Geschäfts- oder Produktionsprozesse scheitern oder die Kosten steigen. Daher ist es wichtig, die Modelldrift zu erkennen und zu beheben, um eine hohe Modellleistung und -genauigkeit zu gewährleisten.

2.2 Arten von Modelldriften

Modelldriften können in verschiedene Arten unterteilt werden, die sich in ihrer Ursache unterscheiden. Eine mögliche Unterscheidung ist die Datendrift und die Konzeptdrift.

Bei der Datendrift ändern sich die Eigenschaften bzw. die Verteilung der Variablen innerhalb der Datenpopulation, was sich auf die Vorhersagen des Modells auswirken kann. Beispiele für Datendrift sind Kovariate-/Feature-Drift, bei der sich die Verteilung der Eingaben des Modells ändert, oder Label-Drift, bei der sich die Verteilung der Ausgaben des Modells ändert.

Die Konzeptdrift hingegen beschreibt eine Veränderung der Beziehung oder Abhängigkeit zwischen den Eingabevariablen und den Zielvariablen. Konzeptdrift kann in verschiedenen Formen auftreten, z. B. als plötzliche Drift, allmähliche Drift oder wiederkehrende Drift, die jeweils unterschiedliche Auswirkungen auf die Modellleistung haben können.

Bei einer plötzlichen Drift tritt ein neues Konzept innerhalb eines kurzen Zeitraums auf und kann die Leistung des Modells abrupt beeinflussen. Bei der allmählichen Drift hingegen wird das alte Konzept schrittweise durch ein neues Konzept ersetzt, was zu einer langsamen Veränderung der Modellleistung führen kann. Die wiederkehrende Drift beschreibt eine Situation, in der ein altes Konzept nach einer gewissen Zeit der Abwesenheit wieder auftaucht und zu einer erneuten Veränderung der Modellleistung führen kann.

2.3 Ursachen von Modelldriften

Es gibt viele mögliche Ursachen, die zu Modelldrift führen können.

Eine der Hauptursachen für Modelldrift sind, wie bereits erwähnt, Veränderungen in den Daten (Datendrift). Solche Veränderungen können saisonale Schwankungen

sein, wie z. B. die Verkaufszahlen in der Weihnachtszeit, oder Veränderungen im menschlichen Verhalten, wie z. B. im Konsum- oder Kaufverhalten. Auch technologische Veränderungen, wie die Einführung neuer Datenquellen oder -formate, können zu Veränderungen in den Daten führen.

Eine weitere mögliche Ursache für Modelldrift sind Veränderungen im Kontext (Kontextdrift). Technologische Änderungen wie die Einführung neuer Geräte oder Betriebssysteme können ebenfalls eine solche Kontextdrift verursachen. Auch Änderungen der Geschäftsstrategie oder der Benutzeranforderungen können zu einer Kontextänderung führen. Zum Beispiel kann ein Unternehmen beschließen, sich auf eine neue Benutzerzielgruppe zu konzentrieren. Auch wenn sich zum Beispiel die Anforderungen an ein Produkt oder eine Dienstleistung ändern, führt dies zu einer Kontextdrift.

3 Methoden zur Erkennung und Behandlung von Modelldriften

3.1 Methoden zur Erkennung von Modelldriften

Die regelmäßige Überwachung von ML- Modellen ist entscheidend, um Modelldrift frühzeitig zu erkennen. Es können beispielsweise die Dateneigenschaften, die Beziehungen zwischen den Variablen oder auch die Leistungsmetriken überwacht werden.

Die Überprüfung der Datendrift bezieht sich auf die Überwachung der Dateneigenschaften, um festzustellen, ob sich die Verteilungen oder deskriptiven Statistiken, z. B. Min, Max, Median, Mittelwert usw., signifikant verändert haben. Es gibt verschiedene Hypothesentests zur Feststellung von Verteilungsänderungen, z. B. Kolmogorov-Smirnov-Test, Populationsstabilitätsindex usw.

Bei der Konzeptdrift-Erkennung werden dagegen Veränderungen in der Beziehung zwischen Input- und Output-Variablen überwacht. Dies kann durch die Überwachung der Korrelation zwischen den Eingabe- und Ausgabevariablen geschehen.

Auch die Überwachung von Vorhersagefehlern oder Leistungsmetriken wie Konfusionsmatrix, Genauigkeit, Rückruf und F1-Score können zur Erkennung von Modelldrift verwendet werden.

3.2 Methoden zur Behandlung von Modelldriften

Wenn ein Machine-Learning-Modell von Modelldrift betroffen ist, gibt es mehrere Möglichkeiten, das Problem zu lösen.

Eine Möglichkeit ist, das Modell neu zu trainieren („Retraining“). Dabei wird das Modell mit einem neuen Datensatz oder aktualisierten Daten neu trainiert, um sicherzustellen, dass es die neuesten Daten und Trends berücksichtigt.

Eine andere Möglichkeit ist „Model-Tuning”. Dabei werden beispielsweise Merkmale, Hyperparameter oder die Architektur des Modells angepasst, um eine bessere Anpassung an die aktuellen Daten zu erreichen und die Modelldrift zu verringern.

Eine weitere Option ist die Entwicklung eines völlig neuen Modells. Dies kann notwendig sein, wenn sich die Daten so stark ändern, dass das aktuelle Modell nicht mehr geeignet ist.

Alternativ kann ein Teil des Modells für Teilergebnisse verwendet werden, die keine Modelldrift enthalten. Ein Beispiel hierfür ist das Transfer-Lernen. Dabei wird ein bereits trainiertes Modell als Ausgangspunkt genommen, um ein neues Modell zu trainieren. Das bereits trainierte Modell wird als Vorlage verwendet, um schneller ein neues Modell zu erstellen, das besser zu den neuen Daten passt.

4 Fallbeispiele für die Erkennung von Modelldriften

4.1 Übersicht

Es gibt verschiedene Python-Bibliotheken, die für die Erkennung von Modelldrift verwendet werden können. Beispiele sind „scikit-multiflow“ und „alibi-detect“.

„scikit-multiflow“ bietet Funktionen zur Erkennung von Drift in Datenströmen, einschließlich ADWIN, EDDM und HDDM_A. Während „alibi-detect“ eine Bibliothek ist, die sich auf die Erkennung von Ausreißern konzentriert.

Eine weitere Python-Bibliothek zur Erkennung von Modelldrift ist „evidently“. „evidently“ bietet umfassende Funktionen zur Überwachung der Datendrift, einschließlich der Überprüfung von Veränderungen in der Verteilung numerischer, kategorialer und textueller Features im Zeitverlauf. Darüber hinaus bietet „evidently“ auch Funktionen zur Analyse der Leistung von Klassifikations- und Regressionsmodellen.

In diesem Papier wird „evidently“ als Fallbeispiel vorgestellt. Das Fallbeispiel umfasst 4 kleinere Experimente:

- Erkennung von Datendrift in einer Klassifikationsaufgabe mit dem Brustkrebs-Datensatz
- Erkennung von Datendrift in einer Regressionsaufgabe mit dem Housing-Datensatz
- Analyse der Leistung eines Klassifikationsmodells mit dem Brustkrebs-Datensatz
- Analyse der Leistung eines Regressionsmodells mit dem Bike-Sharing-Datensatz

Jedes Experiment wird wie folgt durchgeführt.

4.2 Erkennung von Datendrift in einer Klassifikationsaufgabe mit dem Brustkrebs-Datensatz

Zunächst werden die Daten geladen. Der Datensatz wird dann zufällig neu sortiert, um eine gleichmäßige Verteilung der Klassen und anderer Merkmale zu gewährleisten. Dann werden die Daten in die „alten“ und die „neuen“ Datenpunkte aufgeteilt.

Anschließend wird mit dem Dashboard von evidently eine Datendrift-Analyse durchgeführt. Das Dashboard wird mit zwei Tabs erstellt: dem DataDriftTab, der die Unterschiede in der Verteilung aller Merkmale zwischen zwei Teilen des Datensatzes zeigt, und dem CatTargetDriftTab, der die Unterschiede in der Verteilung der Zielklassen zwischen den beiden Teilen des Datensatzes zeigt.

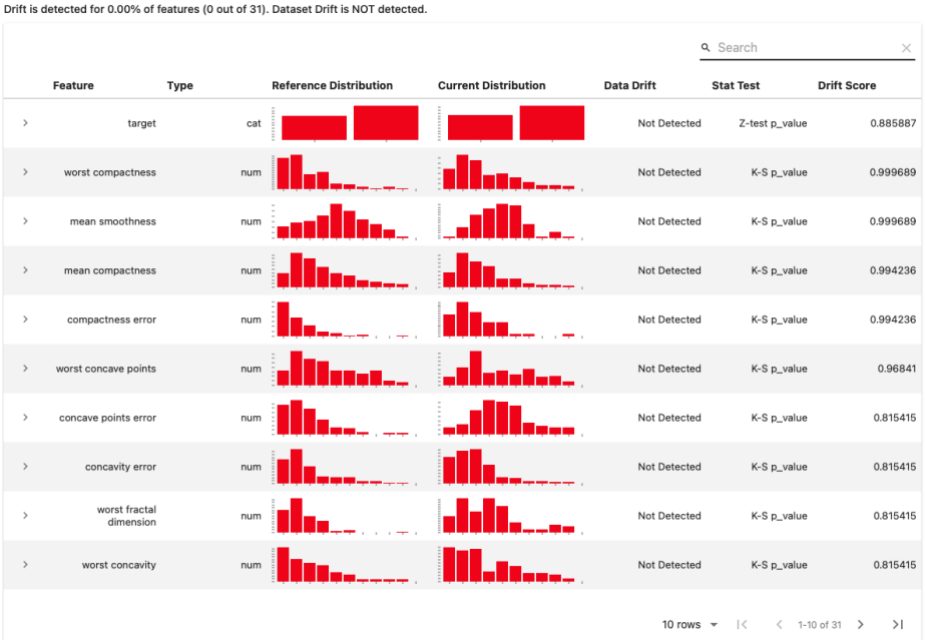


Abbildung 1: Data Drift Tab

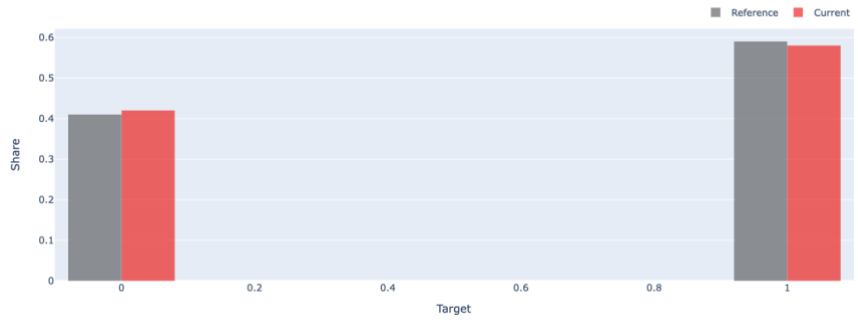


Abbildung 2: Cat Target Drift Tab

Hier wird keine Modelldrift erkannt.

4.3 Erkennung von Datendrift in einer Regressionsaufgabe mit dem Housing-Datensatz

Zunächst wird der Datensatz mit fetch_california_housing aus der Scikit-learn-Bibliothek geladen und wieder in die „alten“ und die „neuen“ Daten aufgeteilt. Die Datenpunkte von old_samples werden aus den ersten 15.000 Datensätzen ausgewählt, die Datenpunkte von new_samples aus den restlichen Datensätzen. Die Auswahl erfolgt nach dem Zufallsprinzip, aber mit einem festen Seed (random_state=0), um die Reproduzierbarkeit der Ergebnisse zu gewährleisten.

Hier müssen zusätzlich das Zielattribut und die numerischen Merkmale definiert und in einem ColumnMapping-Objekt gespeichert werden. Das ColumnMapping-Objekt wird später an das Dashboard übergeben, um die Merkmale des Datensatzes zu definieren.

Schließlich wird das Dashboard erstellt und die Datendriftanalyse durchgeführt. Das Dashboard enthält ebenfalls zwei Tabs: das DataDriftTab und das NumTargetDriftTab.

Abbildung 3: Data Drift Tab

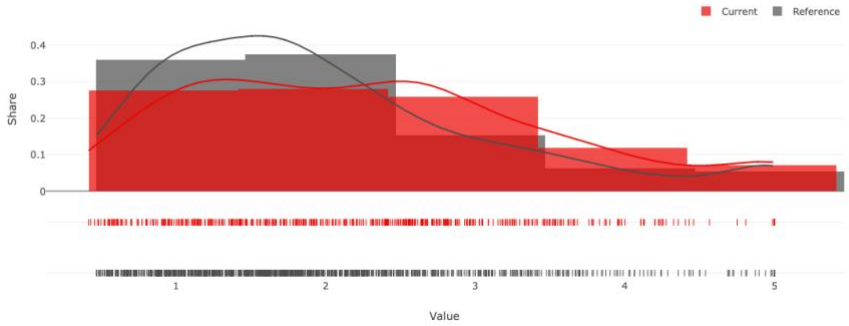


Abbildung 4: Num Target Drift Tab

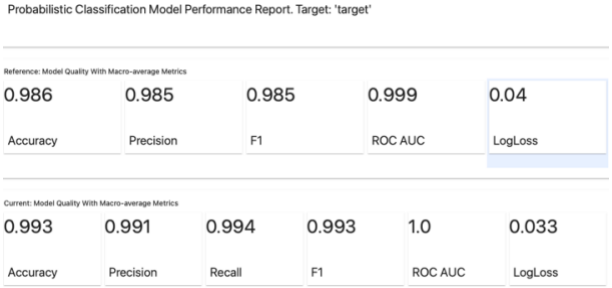
Hier werden Modelldriften erkannt.

4.4 Analyse der Leistung eines Klassifikationsmodells mit dem Brustkrebs-Datensatz

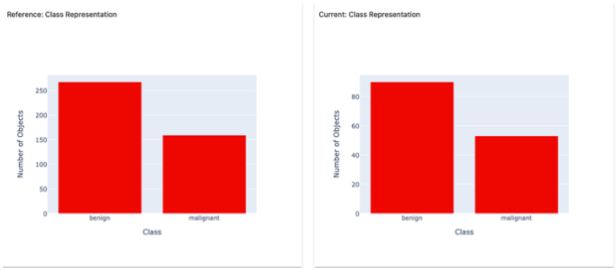
Der Datensatz wird neu geladen und in Trainings- und Testdaten unterteilt. Darüber hinaus werden die Daten auch in sogenannte Referenz- und Produktionsdaten unterteilt, wie im Fall von alten und neuen Daten bei der Datendrift-Erkennung.

Ein logistisches Regressionsmodell wird danach erstellt und mit beiden Datensätze trainiert.

Das Dashboard wird zum Schluss erstellt, um die Leistung des Modells für beide die Referenz- und Produktionsdaten zu vergleichen. Das Dashboard enthält nur einen ProbClassificationPerformanceTab, der zunächst Qualitätsmetriken für Referenz- und Produktionsdatensätze wie zum Beispiel die Genauigkeit, die Präzision, der F1-Score, die ROC-Kurve enthält.



Der Tab enthält außerdem Class Representation Diagramme, die die Anzahl der Objekte jeder Klasse in den Referenz- und Produktionsdatensätzen zeigen.



Der Tab zeigt zusätzlich jede der Modellqualitätsmetriken für die einzelnen Klassen an.

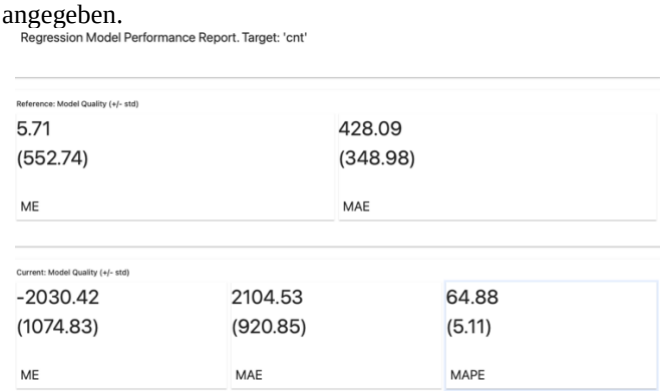
Abbildung 5: Prob Classification Performance Tab

Hier kann festgestellt werden, dass die Leistung des Modells stabil bleibt.

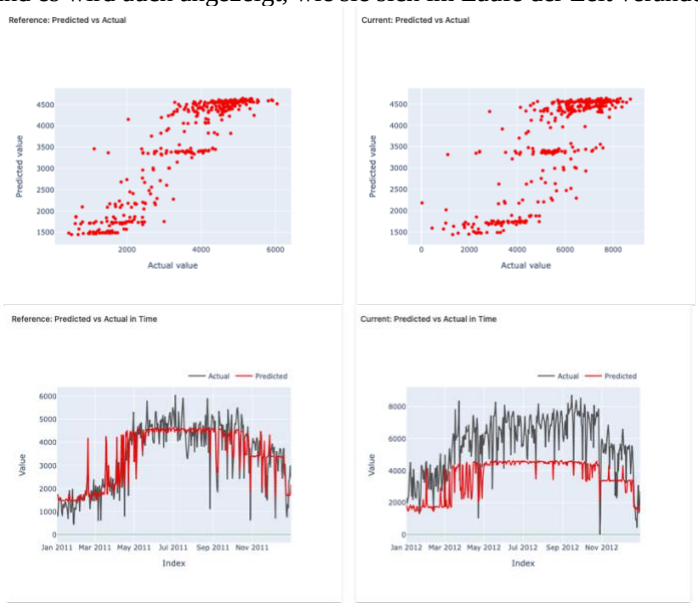
4.5 Analyse der Leistung eines Regressionsmodells mit dem Bike-Sharing-Datensatz

Zunächst wird der Datensatz heruntergeladen und ebenfalls in Trainings- und Testdaten sowie Referenz- und Produktionsdaten aufgeteilt. Anschließend wird ein Random-Forest-Regressionsmodell erstellt und trainiert.

Schließlich wird ein Dashboard mit nur einem Tab „RegressionPerformanceTab“ zur Leistungsvisualisierung von Regressionsmodellen erstellt. Der Tab zeigt zunächst sowohl für die Referenz- als auch für die Produktionsdatensätze einige Standardmodellqualitätsmetriken, die für Regressionsmodelle relevant sind z.B. ME, MAE und MAPE. Für jede Qualitätskennzahl wird auch eine Standardabweichung ihres Wertes angegeben.



Dann werden zusätzlich die Vorhersagen und die tatsächlichen Werte grafisch dargestellt und es wird auch angezeigt, wie sie sich im Laufe der Zeit verändern.



Der Tab stellt außerdem den Fehler des Modells und seine Veränderung im Laufe der Zeit visuell dar.

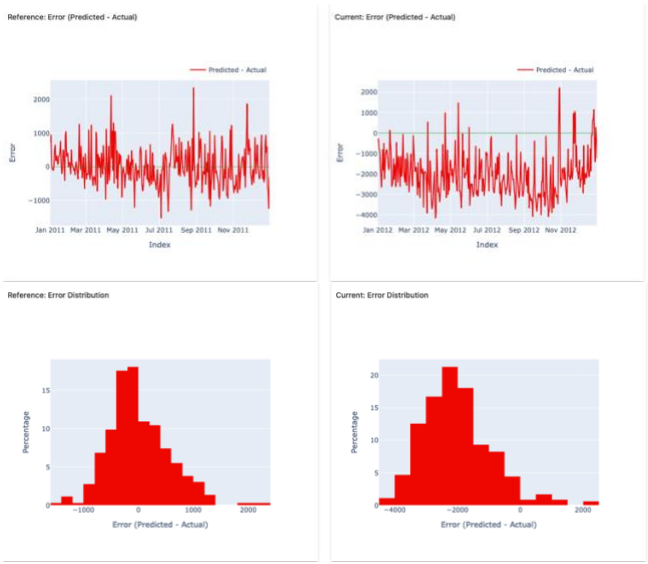


Abbildung 6: Regression Performance Tab

Hier ist zu erkennen, dass die Leistung des Modells zurückgegangen ist.

5 Zusammenfassung und Ausblick

5.1 Zusammenfassung

Die Arbeit klärt die Definition von Modelldrift und ihre möglichen Auswirkungen: Modelldrift ist das Phänomen, bei dem die Leistung von Modellen mit der Zeit abnimmt und zu ungenauen oder fehlerhaften Vorhersagen führt.

In dem Papier werden auch verschiedene Arten von Modelldrift untersucht: Datendrift und Konzeptdrift sind die Hauptarten von Modelldrift.

Die Arbeit untersucht auch die Ursachen sowie die Erkennungs- und Behandlungsmethoden. Ursachen können Änderungen der Daten und des Kontexts sein. Zur Behandlung von Modelldrift können verschiedene Maßnahmen ergriffen werden, wie z. B. Retraining, Tuning oder die Erstellung eines komplett neuen Modells.

Das Papier enthält auch ein praktisches Beispiel für die Erkennung von Modelldrift mit „evidently“.

5.2 Ausblick auf zukünftige Forschung

In der Zukunft gibt es viele Möglichkeiten für die Forschung im Bereich der Modelldrift. Eine Möglichkeit könnte die Anwendung von KI-Technologien zur Vorhersage von Modelldrift sein. Es könnte sein, Modelldrift mithilfe von KI-Systemen vorherzusagen, bevor sie auftritt.

Darüber hinaus könnten auch adaptive Modelle weiterentwickelt werden, die sich an veränderte Daten anpassen und so Modelldrift vermeiden können.

Schließlich könnten auch Untersuchungen zur Kombination von Methoden zur Erkennung und Behandlung von Modelldrift mit anderen Techniken des maschinellen Lernens von Interesse sein. Insbesondere könnten hier Ansätze untersucht werden, die auf der Verwendung von Ensembles oder transferiertem Lernen basieren, um die Modellleistung und -stabilität zu erhöhen.