

Diplomprojekt – Ausschreibung

Entwicklung eines KI-Modells zur Detektion und Klassifikation pigmentierten Hautläsionen und dessen Integration in eine Mobile App

9. Juli 2024

1 Hintergrund zum Thema:

Die Klassifikation von pigmentierten Hautläsionen spielt eine entscheidende Rolle in der medizinischen Diagnostik, insbesondere bei der frühzeitigen Erkennung von Hautkrebs. Mit der zunehmenden Verfügbarkeit großer Datensätze und fortschrittlicher Algorithmen im Bereich des maschinellen Lernens eröffnen sich neue Möglichkeiten zur automatisierten Analyse von Hautläsionen. Der HAM10000-Datensatz ("Human Against Machine with 10000 training images") stellt eine umfassende Sammlung von dermatoskopischen Bildern dar, die zur Entwicklung und Evaluation von KI-Modellen genutzt werden können.

2 Fragestellungen und Aufgaben

Das Ziel dieses Projekts ist die Entwicklung eines leistungsfähigen Machine Learning Modells zur Klassifikation von Hautläsionen basierend auf dem HAM10000-Datensatz. Das Projekt umfasst die vollständige Umsetzung des Machine Learning Life Cycles, einschließlich Datenvorbereitung, Modelltraining, Modellbewertung, sowie Deployment und Integration in eine mobile Anwendung. Zusätzlich sollen, entlang des gesamten Machine Learning Life Cycles, didaktische Elemente (z.B. Jupyter Notebooks) erstellt werden, die eine Einführung in die einzelnen Stufen der Entwicklung eines KI-Modells bieten. Dieses Projekt bietet eine umfassende Einführung in den praktischen Einsatz von Machine Learning zur Lösung realer medizinischer Probleme und fördert gleichzeitig die Entwicklung von Fähigkeiten in den Bereichen Datenverarbeitung, Modellierung, Softwareentwicklung und mobile Anwendungsentwicklung. Die spezifischen Ziele sind:

1. Datenvorbereitung:

- Bereitstellung der Daten über eine Datenbank (strukturiert oder unstrukturiert).
- Datenbereinigung und -augmentation, um die Qualität und Diversität des Trainingsdatensatzes zu verbessern.

- Alternativ: Ergänzung des HAM10000 Datensatzes über weitere Datensätze (ISIC Challenge)
- Aufteilung der Daten in Trainings-, Validierungs- und Testdatensätze.

2. Modellentwicklung:

- Auswahl und Implementierung geeigneter Machine Learning Algorithmen, insbesondere Convolutional Neural Networks (CNNs). Es sollen unterschiedliche Ansätze untersucht werden wie z.B. eine direkte Klassifikation, oder ein zweistufiger Prozess wo zuerst die Läsion detektiert (Bounding Box, Semantic Segmentation, etc.) wird und anschließend klassifiziert.
- Training der Modelle auf dem HAM10000-Datensatz und Optimierung der Modellparameter (Hyperparameter-Tuning).

3. Modellbewertung:

- Evaluierung der Modellleistung anhand relevanter Metriken wie Accuracy, Precision, Recall, AUROC und F1-Score.

4. Deployment:

- Implementierung des trainierten Modells in einer serverseitigen Umgebung.
- Entwicklung einer RESTful API zur Bereitstellung des Modells als Webservice.
- Optional: Bereitstellung des Modells direkt am Mobile Device (z.B. über TensorFlow Lite oder PyTorch Edge)

5. Integration in eine mobile App:

- Entwicklung einer einfachen mobilen Anwendung, die es Nutzern ermöglicht, Bilder von Hautläsionen aufzunehmen und klassifizieren zu lassen.
- Integration der RESTful API in die mobile App zur Echtzeit-Bereitstellung der Klassifikationsergebnisse.
- Tech-Stack

3 Tech-Stack:

- Datenmanagement: PostgreSQL/MongoDB, Python, Pandas, Numpy, OpenCV
- Modellentwicklung und Evaluierung: TensorFlow Keras oder PyTorch, scikit-learn
- API-Entwicklung in Python: Flask oder FastAPI
- Mobile Development für Android: Flutter oder Kotlin