

Abstract

For a fast and point of care (POC) diagnosis, a machine learning (ML) approach could be used to measure the concentration of digoxin and its hapten, digoxigenin, in human serum from an image readout system of gold nanoparticle lateral flow (LF) assay [1]. In this work an image processing method is developed that can be applied to separate the region of interest (ROI) from the raw data, pre-process the data, and perform ML to predict the concentration of digoxigenin solution from the image intensity.

To accomplish the whole process a three-step algorithm was developed. At first, a two-step cropping method was developed to perform the cropping and separation of the ROI from raw images. In the next step, pre-processing was done, which includes building a data frame from cropped images and apply some feature engineering to improve the performance of the ML model. Lastly, ML models were developed to predict the concentration of the digoxigenin solution. ML model then evaluated by different evaluation methods and after that validated by K-fold cross-validation. The *adjusted – R^2* value, accuracy of the model, for the final ML model is 81% and K-Fold cross-validation is 70.6%.

Keywords Python, Machine Learning, scikit-learn, scikit-image, Template matching, Lateral Flow Assay.

Zusammenfassung

Für eine schnelle und Point-of-Care-Diagnose (POC) könnte ein maschinelles Lernverfahren (ML) verwendet werden, um die Konzentration von Digoxin und seinem Hapten, Digoxigenin, in Humanserum aus einem Bildauslesesystem des Gold-Nanopartikel-Lateral-Flow (LF)-Tests [1] zu messen. In dieser Arbeit wird eine Bildverarbeitungsmethode entwickelt, die angewendet werden kann, um den interessierenden Bereich (ROI) von den Rohdaten zu trennen, die Daten vorzubereiten und eine ML durchzuführen, um die Konzentration der Digoxigeninlösung aus der Bildintensität vorherzusagen.

Um den gesamten Prozess durchzuführen, wurde ein dreistufiger Algorithmus entwickelt. Zunächst wurde ein zweistufiges Zuschneideverfahren entwickelt, um das Zuschneiden und Trennen des interessierenden Bereichs von den Rohbildern durchzuführen. Im nächsten Schritt wurde eine Vorverarbeitung durchgeführt, die das Erstellen eines Datenrahmens aus zugeschnittenen Bildern und das Anwenden von Feature-Engineering umfasst, um die Leistung des ML-Modells (Machine Learning) zu verbessern. Zuletzt wurden ML-Modelle entwickelt, um die Konzentration der Digoxigeninlösung vorherzusagen. Das ML-Modell wurde dann durch verschiedene Bewertungsmethoden bewertet und anschließend durch K-fache Kreuzvalidierung validiert. Der *adjusted* – R^2 Wert, Genauigkeit des Modells, für das endgültige ML-Modell beträgt 81% und die K-Fold-Kreuzvalidierung beträgt 70,6%.

Schlüsselwörter Python, Machine Learning, scikit-learn, scikit-image, Template matching, Lateral Flow Assay.