Master Thesis

Lateral Flow Image Analysis with Python

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%.

Introduction

Fast and easy to use point of care (POC) devices, which could be operated by patients him or herself for the determination of drag or measurement of the concentration of metabolic, is an essential process in the area of Precision Medicine. This easy to use system much be cost-effective, easy to operate, and should not require any additional equipment to measure or process the information [1].

To solve this issue a machine learning approach could be used to predict the concentration of digoxin and it's hapten, digoxigenin, from image intensity. Digoxin and it's hapten, digoxigenin are used for the treatment of tachycardia with small concentration, $0.5 - 2.0$ ng. mL^{-1}, in human serum solution. When the concentration of digoxin in serum solution is above 2.5 ng. mL^{-1}, it becomes very toxic, which should be avoided [1].

Material & Methods

To calculate the concentration of digoxin from image intensity a three steps algorithm in Python programming was developed. In the first step, the region of interest (ROI), control line and test line, from raw images was extracted. This step requires an additional two steps. At first, a built-in function, match template, from scikit-learn library is used to extract LF from raw images. Next, several conditions and masking ideas used to extract ROI.

After that, these images then converted into a data frame (DF). To improve machine learning (ML) model performance, a couple of feature engineering (FE) was performed. Transformation of data using logarithm & square root and merging similar data based-

Verfasser: Md. Abdullah-al-Kamran Ripon; 255630
Betreuer: Prof. Dr. Matthias Kohl
        Prof. Dr. Hans-Peter Deigner

on the feature. After performing these FE, ML model performance improve significantly.

In the last step, ML model was developed to predict the concentration of digoxin from image intensity. The Performance of ML model was then evaluated and validated.

Results

In the data frame (DF) we created ten features from cropped images. Before selecting a final ML model, we used all the features form DF and created two ML models to test and analyse the performance of DF. Very first ML model which is used unmodified DF motivates to perform FE on the DF.

Several evaluation methods used to evaluate the performance of the ML model. Here is the result from evaluation matrix

Table 1: Evaluation Matrix for three ML model

	First	Second	Final
MSE	393.16	39.62	68.72
RMSE	19.83	6.30	8.29
Mean AE	11.22	3.94	5.91
Median AE	6.52	1.32	3.67
R^2	0.66	0.93	0.88
Adj R^2	0.55	0.62	0.81

From the above table, we can say final ML model perform much better than first two ML mode. Both R^2 and adjusted R^2 are pretty high and in the same range. That's tell the goodness of fit of ML model.

Evaluation Matrix is not only evaluation we perform on ML model. There is some other evaluation also perform. Some of them are, residual plot, quantile-quantile plot, histogram of residual, scatter plot of predicted vs actual value, distribution plot of predicted and actual value. Here is some result from evaluation.

Figure 1-a shows scatter plot between-

actual vs predicted value with a regression line. Fig. 1-b shows distribution plot of actual value, redline, and predicted value, blueline.
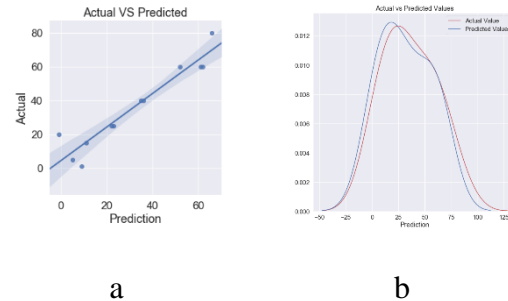

a   b

Figure 1: a. Scatter plot between predicted and actual, b. distribution plot: predicted and actual

Discussion

In the area of precision medicine, it is essential to have a measurement system of drug and other metabolite concentration at a high speed and easy to use point of care (POC) technique. It is very important to monitor the concentration of digoxin in human serum solution [1].

For monitoring the digoxin concentration in a fast and easy to use POC fashion, a machine learning approach could be used to measure the concentration of digoxin and its hapten, digoxigenin, in human serum [1].

Conclusion

This work demonstrates a method of preparing data to analyse and develop a machine learning model to predict the concentration of digoxin from the intensity of the image. This idea could be used to further develop a smartphone application which enables a first and easy to use POC measurement technique in the field of precision medicine.

Reference

1. Ruppert, C., Phogat, N., Laufer, S., Kohl, M. and Deigner, H.P., 2019. A smartphone readout system for gold nanoparticle-based lateral flow assays: application to monitoring of digoxigenin. Microchimica Acta, 186(2), p.119. https://doi.org/10.1007/s00604-018-3195-6

Verfasser: Md. Abdullah-al-Kamran Ripon; 255630
Betreuer: Prof. Dr. Matthias Kohl
                  Prof. Dr. Hans-Peter Deigner