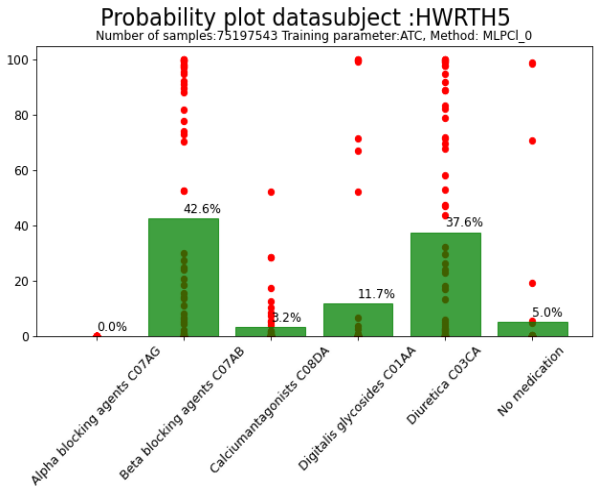
Output of the LorAIne software package. A visualization of the cardiac medication classification problem

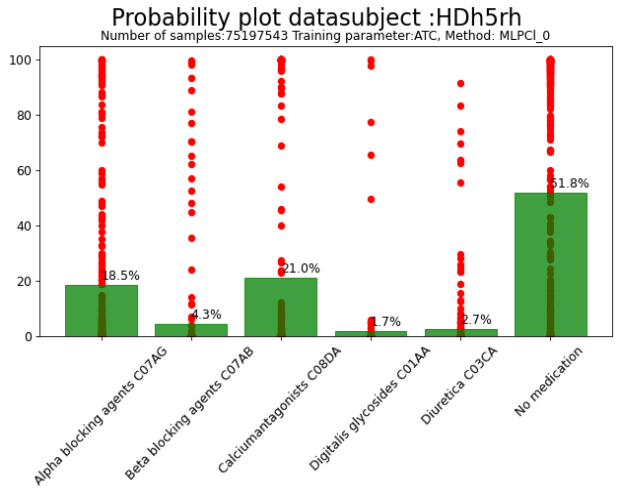
February 2023, rev A

Introduction
A study conducted to classify cardiac medication on ATC level based on fitness tracker data is conducted by Annovating. Based on over 75 million data points (HR and HRV) and a total number of 30k person days, the LorAIne software was utilized to classify datasubjects based on the use of cardiac medication on ATC main group level. Classification is used to categorize the subject bound fitness tracker time series data into selected ATC classes. The study started with predicting the class of a set of given data points converted into 539 features. The classes to be identified are referred to as ATC-labels. The main goal of the present study is to identify whether it is possible to classify a string of time series fitness tracker data into main cardiac ATC classes. It should be stated clearly that more subject medication records are required to validate any of the below demonstrated indicative outputs of the LorAIne software package.

Medication safety and times series data
As far as medication safety is concerned, first exercises conducted within the preset study started with identifying the use of cardiac medication at first and specifying into ATC-level as a secondary challenge. A future phase would also be the application of time related ML routines like Recurrent Neural Networks (RNN) i.e networks capable of handling long-term dependencies, e.g algorithms like LSTM. LSTM algorithms belong to a series of advanced RNN's, sequential networks that allow information to persist over time.

Multilayer Perceptron Algorithm
The feedforward artificial neural network applied within LorAIne is known as the multilayer perceptron (MLP). It is responsible for producing a set of outputs based on a collection of features as input. A multi-layer perceptron is distinguished by having many layers of input nodes coupled as a directed graph between the input and output layers. Within the MLP algorithm backpropogation is utilized in order to train the network. The multi-layer perceptron algorithm is composed of multiple layers of perceptrons and encompasses one hidden layer. A probability plot based on a MLPClassifier rendered on a dataset of 75 million data points reveals for each individual data subject the probability of the use of cardiac medication on ATC main group level.

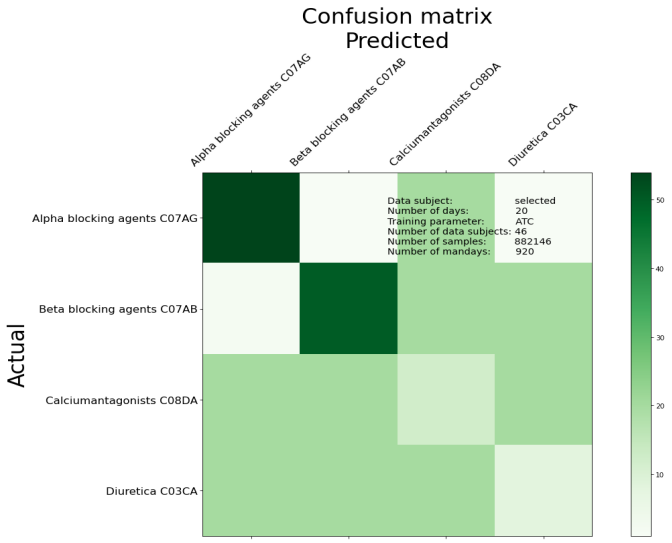


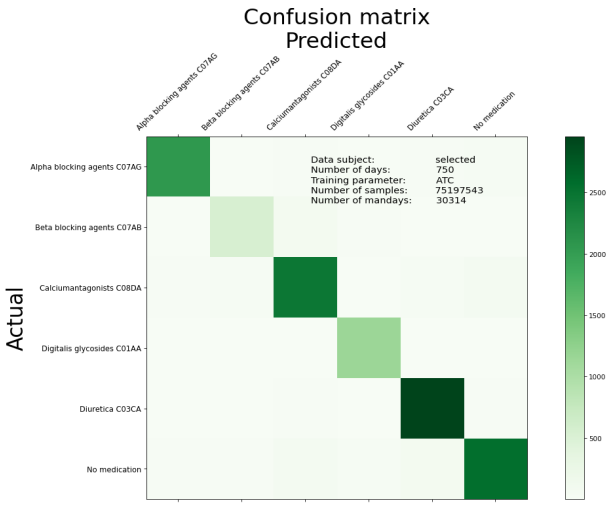


In the left figure above it is clearly revealed that the prevailing probability for subject with id 'HWRTH5' might be a combination of Diuretica CO3CA (37,6%) in combination with Beta blockers CO7AB (47,6%). Data subject with id 'HDh5rh' to the right shows a more ambivalent picture of potential medication consumption with a prevailing prediction that no medication is taken at all. In red, classifications are visualized of separate subject feature records covering one manday of fitness tracker data. More data will be required to validate any of the presented preliminary outputs of the LorAIne software package.

Boosting Algorithms
Boosting algorithms applied within the LorAIne software package create an ensemble model by combining several weak decision trees sequentially. Boosting algorithms assign weights to the output of individual trees and give incorrect classifications from the first decision tree a higher weight and deliver input to the next tree. After numerous cycles the boosting method combines these weak rules into a single powerful prediction rule. LorAIne deploys the XGBoosting

algorithm, well tested and a.o applied for game solving routines. Results of the XGBoosting algorithm can be visualized in a confusion matrix as presented below.

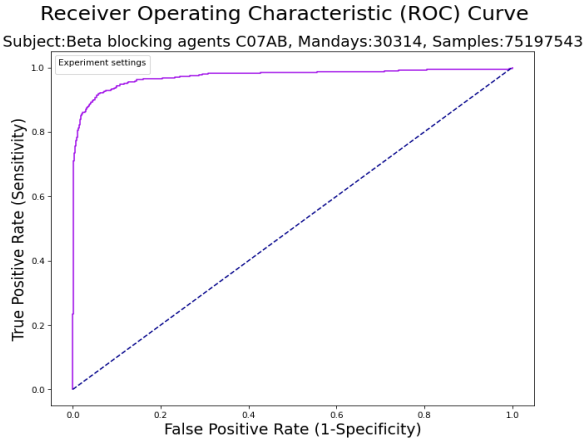


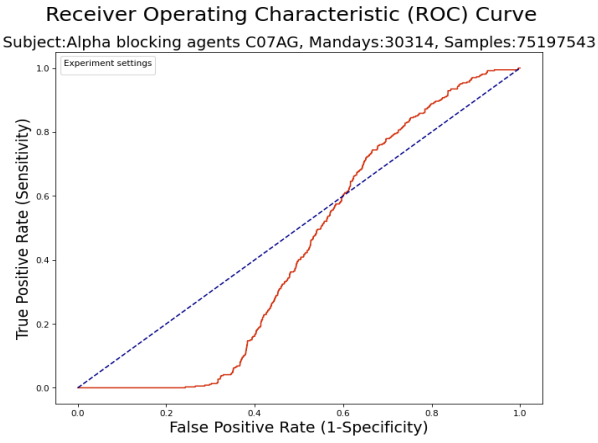


The figure to the left is based on 880k samples and 920 mandays of measurements whereas the figure to the right is based on 75 million data points and 30.314 mandays. It reveals the improved predictability of the datasubjects based on ATC codes when significantly increasing the sample volume and the number of mandays.

Preliminary AUCROC results

The AUCROC curve is a performance measurement suitable for the ATC classification problem. ROC is a probability curve and AUC represents the degree of separability. It tells how much the model is capable of distinguishing between classes. The higher the AUC, the better the model is capable at predicting the specific ATC class. AUC values range between 0 (useless) and 1 (perfect prediction) and demonstrate the model's predictability. Any AUC value exceeding 0,5 (the diagonal of the 1x1 plot) demonstrates any value.





A ROC curve as presented to the left does depict the sensitivity and selectivity of the ML-model to identify medication consumption from fitness tracker data. With an AUC score equalling ~0,5, these models do not have any practical use. A low performance (AUC<0,5) as presented in the picture to the right visualizing the ability to detect e.g Alpha blocking agents can be based on several factors. Factors like for example;lack of appropriate subject records, weak labeling, inadequate hyperparameter tuning or inadequate ML-model selection. During the course of this study the factors mentioned will be further optimized aiming at AUC levels exceeding 0.85

Annovating B.V. info@annovating.com, tel 0031(0)619674344, KvKh 69005087