Designing An ANFIS Classifier To Classify Heart Disease Prediction Classes

Name / ID

Date

# Introduction

This task models the ANFIS classifier, assesses its accuracy and other metrics (confusion matrix and histograms), then compare the results with the performance of the ANN.

# Methodology

## Data Description

The data used was obtained from the archives of UCI (2011), which was merged from 3 source, by region as shown in the table below:

Table 1: THe different datasets used on ANFIS modelling

|  |  |
| --- | --- |
| Region | Observations in Data |
| Cleveland | 303 |
| Hungary | 294 |
| Long beach | 200 |
| Switzerland | 123 |

Looking at the variables, there were 13 patients’ information and the target variable, Summary statistics for the data is shown below.

Table 2: Summary statistics of the data used

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | **Count** | **Mean** | **Std. Deviation** | **Min** | **25%** | **50%** | **75%** | **Max** |
|  | 920.0 | 53.510870 | 9.424685 | 28.0 | 47.0 | 54.0 | 60.0 | 77.0 |
|  | 920.0 | 0.789130 | 0.408148 | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
|  | 920.0 | 3.250000 | 0.930969 | 1.0 | 3.0 | 4.0 | 4.0 | 4.0 |
|  | 861.0 | 132.132404 | 19.066070 | 0.0 | 120.0 | 130.0 | 140.0 | 200.0 |
|  | 890.0 | 199.130337 | 110.780810 | 0.0 | 175.0 | 223.0 | 268.0 | 603.0 |
|  | 830.0 | 0.166265 | 0.372543 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
|  | 918.0 | 0.604575 | 0.805827 | 0.0 | 0.0 | 0.0 | 1.0 | 2.0 |
|  | 865.0 | 137.545665 | 25.926276 | 60.0 | 120.0 | 140.0 | 157.0 | 202.0 |
|  | 865.0 | 0.389595 | 0.487941 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |
|  | 858.0 | 0.878788 | 1.091226 | -2.6 | 0.0 | 0.5 | 1.5 | 6.2 |
|  | 611.0 | 1.770867 | 0.619256 | 1.0 | 1.0 | 2.0 | 2.0 | 3.0 |
|  | 309.0 | 0.676375 | 0.935653 | 0.0 | 0.0 | 0.0 | 1.0 | 3.0 |
|  | 434.0 | 5.087558 | 1.919075 | 3.0 | 3.0 | 6.0 | 7.0 | 7.0 |
|  | 920.0 | 0.995652 | 1.142693 | 0.0 | 0.0 | 1.0 | 2.0 | 4.0 |

From the table above, it is realised that there are many missing observations in most of the attributes. Since dropping them was not the optimal solution, filling in the observations with the mean of age groups was found to be the best imputation method.

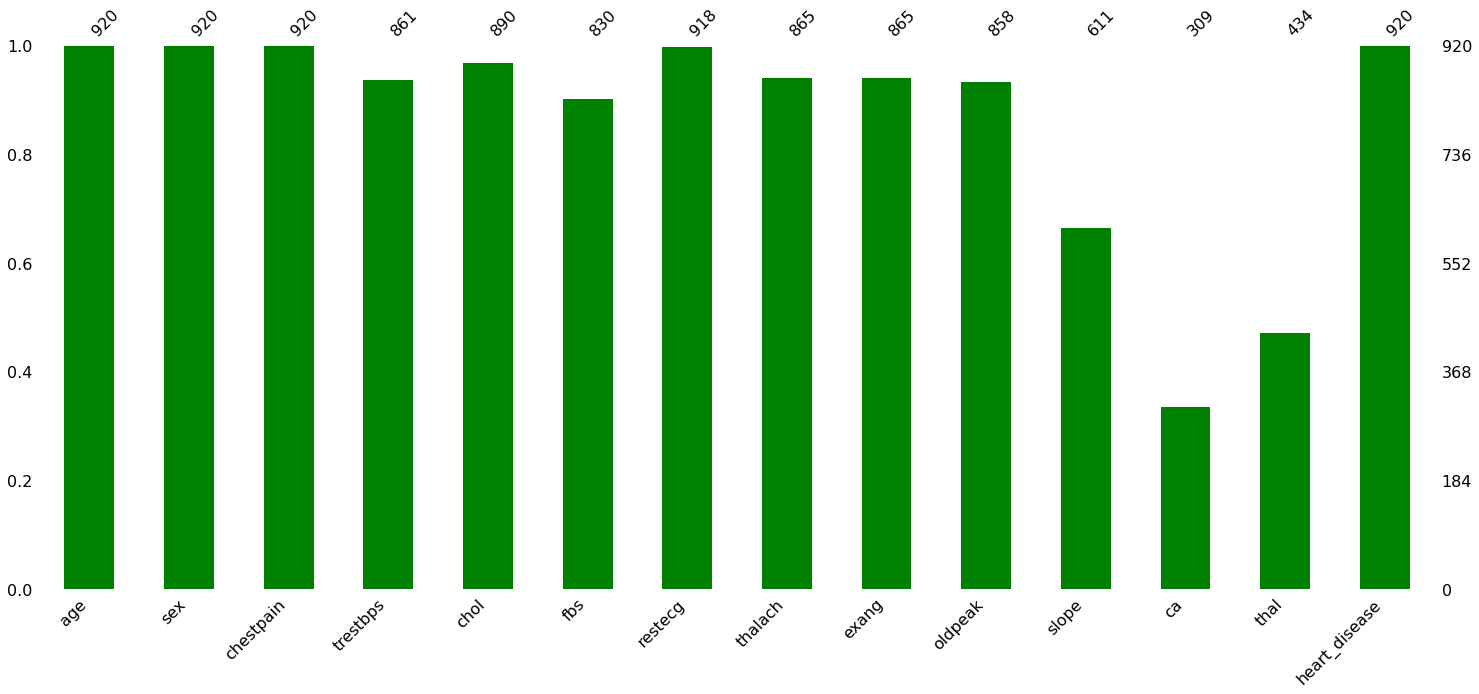


Figure 1: Missing observations per variable. Cleaning was required before modelling ANFIS

Looking at the target variable, predicted values range from 0 to 4, 0 being no presence of Heart Disease and 1,2,3,4 are the stages of Heart Disease, as seen in the figure below, showing the distribution of each class. The classes are seen to be unbalanced, and so, scaling was done to standard this distribution to avoid overfitting some classes.

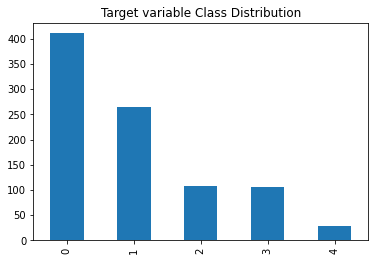


Figure 2: Class distribution in the heart disease target variable

To model ANFIS, it was seen necessary to have two classes in the target variable. Class 0 remained as it is, while classes 1 to 4 were merged into a class 1, for present of heart attack conditions. The distribution is show in the figure below:

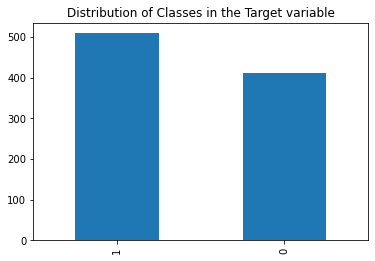


Figure 3:Class distribution in the heart disease target variable now with classes 1 to 4 merged

## The ANFIS and ANN Classifiers

The ANFIS is a hybrid intelligent system that uses both artificial neural networks and fuzzy logic. Given that the Takagi-Sugeno type inference model's parameters are changed using the ANN approach, ANFIS has the same capacity for learning from training data as ANN. As a consequence, a Fuzzy Inference System's replies may be described using language (FIS). By establishing the structure of an ANFIS classifier, five different levels are employed to demonstrate the concept of ANFIS structure. The ANFIS structure starts with the fuzzification layer, then moves on to the rule base layer in the second layer, membership functions (MFs) being normalised in the third layer, defuzzification in the fourth layer, and summing in the fifth layer. In a paper by Walia et al (2015), the construction of ANFIS is extensively discussed.

The basic architecture of the ANFIS is shown in Figure 3 with two inputs.

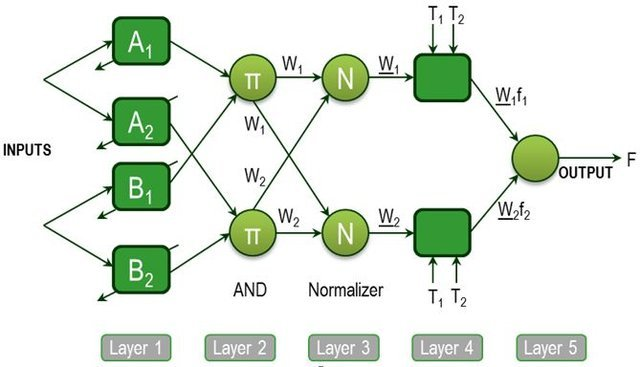


Figure 4: ANFIS Architecture

There are two parts to the ANFIS classifier layout: creation and training. The building process includes steps like separating the input space, selecting the kind and quantity of MFs for inputs, creating fuzzy rules, selecting the premise and conclusion parts of fuzzy rules, and selecting the starting values for MF parameter. An ANFIS classifier's initial build should make advantage of training data patterns. The ANFIS classifier's anticipated output and inputs are represented by these data patterns. The size of the input-output data pattern is significant when data production is costly. It is necessary to divide the input-output data into rule patches in order to build the ANFIS classifier. Numerous methods, including as fuzzy c-means (FCM), subtractive clustering, and grid partitioning, may be used to do this (Guillaume, 2001).

Grid division is only appropriate for problems with few input variables (Walia, et al., 2015), e.g., fewer than 6. A classifier with six inputs needs 729 rules compared to a classifier with three inputs and three fuzzy sets for each input (36). Evidently, the practical limit of the typical ANFIS classifiers is low-dimensional modelling.

As a result, ANFIS was only built using six of the thirteen variables that were included in the whole dataset. A preview of the training dataset is shown below:

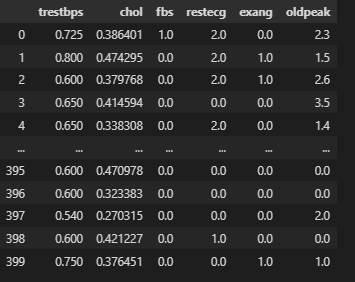


Figure 5: Head and Tail of the X train dataset

## Evaluation Metrics

In this section, we assess the Confusion matrix for the ANFIS classifier and the membership histograms. For both ANFIS and ANN, we use 15 and a batch size of 5. The MAE was used to assess accuracy between the two models.

Examining the distribution of the data, membership functions are required for each input in order to decide the function to use in ANFIS. For the histograms in figure 5, we see that all inputs are fuzzy enough to be included in the model. Second, we check the accuracy of the ANFIS model. It was seen to achieve a test accuracy of 86.5% at its best, and an MAE of 0.2318 at the 15 and 3 inputs as is discussed below.

The confusion matrix showed the results in Table 3. This shows a greater number of True Positives (TPs) showing that the model is predicting most of the classes correctly. Looking at figure 6 and 7, we can see that the Loss for ANFIS is lower in the validation set, meaning that less error are made in the classification.

However, when we reduce the input to 3 for both ANFIS and ANN, the ANFIS model performs better than ANN, with a percentage accuracy of 86.5%

Table 3:Confusion matrix for the ANFIS classifier

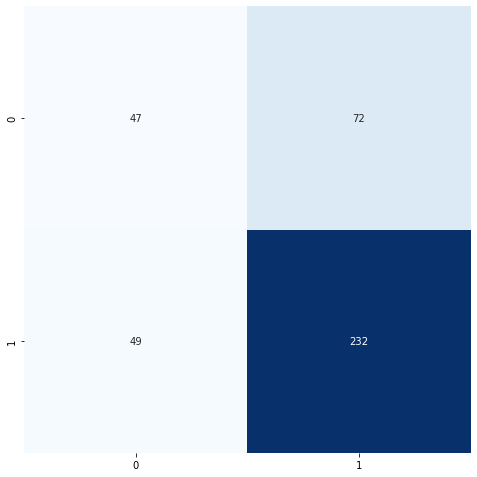


Table 4: Confusion matrix for the ANN classifier

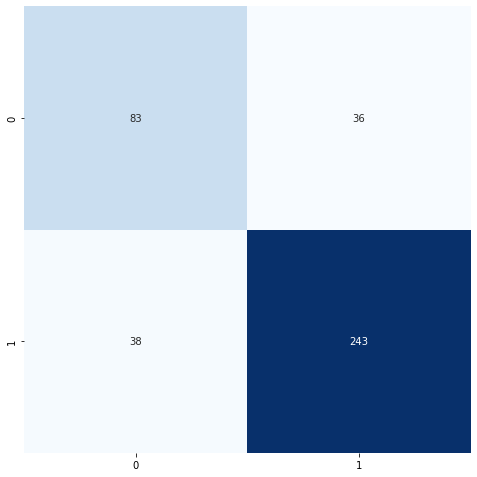


Figure 6: Membership functions showing the fuzziness of each of the 6 inputs

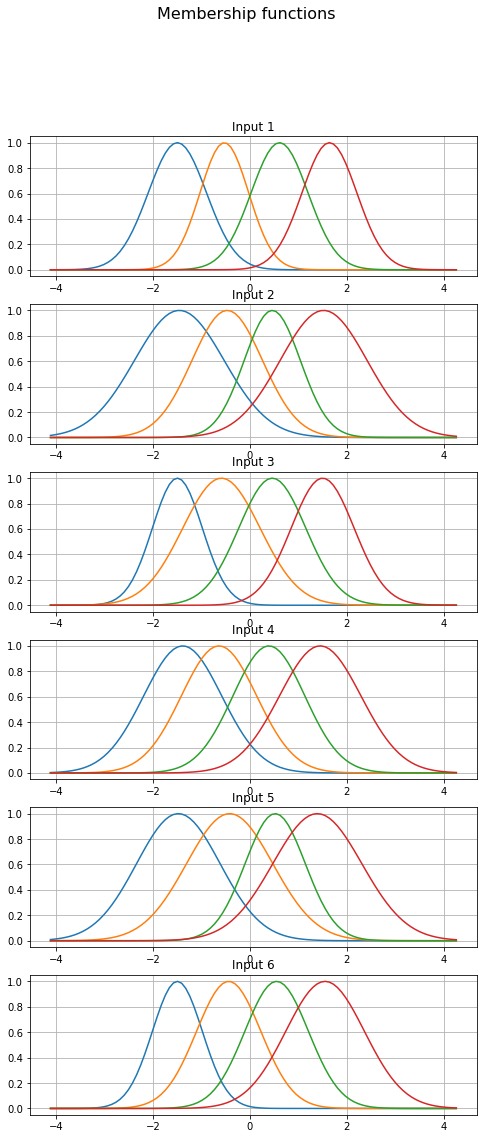


Figure 7: ANFIS Evaluation: Loss, MAE, MSE, Validation Metrics. The comparison show that the ANFIS model is performing well, however, overfitting with inputs =6, since the validation loss is lower than the training one.

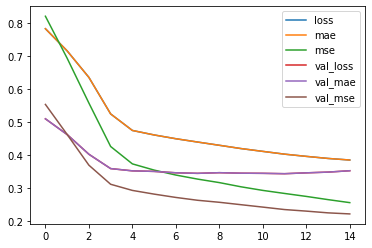
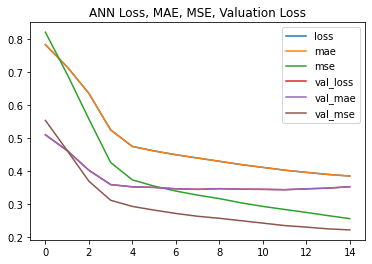


Figure 8: ANN Evaluation: Loss, MAE, MSE, Validation Metrics. The comparison show that the ANN model is performing quite well, just like the ANFIS model, but ANN is overfitting more than ANFIS does. However, more evaluation is needed, and thus, the classification report and accuracies were obtained.



# Model Comparison (With ANN Algorithm)

### Classification Report and Accuracy.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Inputs | Model | Target Class | Recall | Precision | F1-score | Accuracy |
| 6 | ANFIS | Class 0  Class 1 | 0.70  0.86 | 0.69  0.87 | 0.69  0.87 | 69.75% |
| ANN | Class 0  Class 1 | 0.39  0.83 | 0.49  0.76 | 0.44  0.79 | 81.50% |
| 3 | ANFIS | Class 0  Class 1 | 0.73  0.86 | 0.72  0.89 | 0.74  0.87 | 86.5% |
| ANN | Class 0  Class 1 | 0.39  0.86 | 0.54  0.79 | 0.45  0.81 | 77.75% |

### Discussion

ANN performs better than ANFIS on many inputs, since ANFIS is well designed for fewer inputs than the ANN is; meaning ANFIS is a better classifier than ANNN with . This is also seen in the accuracy level of ANN (81.5%) than that of the ANFIS which is lower at 69.75% when the inputs are many (). Also, looking at the Confusion matrix, ANFIS classifier more and than the ANN classifier.

Overall, the ANFIS is preferred since any classifier of binary outputs that has less inputs and is not overfitting, is the most preferred model.

# References

Guillaume, S., 2001. Designing fuzzy inference systems from data: An Interpret ability-oriented review. *IEEE Transactions on Fuzzy Systems,* 9(3), pp. 426-443.

UCI, 2011. *Machine Learning Datasets, Heart Disease.* [Online]   
Available at: https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/  
[Accessed 27 November 2022].

Walia, N., Singh, H. & Sharma, A., 2015. ANFIS: Adaptive Neuro-Fuzzy Inference System- A Survey. *International Journal of Computer Applications,* 123(13), pp. 32-38.