
DIAGNOSIS OF HEART DISEASE USING ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEM (ANFIS)

ERIN M. KUNZ, CS 229 SPRING 2019

ABSTRACT

Heart Disease is the leading cause of death for both men and women in the United States. Approximately 610,000 people die from heart disease each year, which is equivalent to 1 in every 4 deaths. Many health risk and lifestyle factors have been associated with the onset of heart disease including high blood pressure, high cholesterol, smoking, physical activity levels, sleep quality, stress and much more. Most prevention and treatment techniques are focused on these such lifestyle factors. For this reason, diagnosis and prognosis can be somewhat subjective, despite clinically available quantifications that may be more indicative of heart disease. In this study, a variety of machine learning techniques are used to predict the occurrence of heart disease given a variety of easily obtainable heart-health-related features. Furthermore, to incorporate the more subjective nonlinear decision-making that a medical professional might offer, and likelihood that a heart disease prediction may involve complex relationships between features, further analysis utilizing a fuzzy inference system is done.

INTRODUCTION AND RELATED WORK

More than 600,000 Americans die from heart disease every year and it is the leading cause of death globally. It is more common and has been increasing in many developed countries, particularly the United States. A variety of risk factors are linked to heart disease including diet, exercises, use of tobacco, alcohol and high blood pressure. It is estimated that up to 90% of heart disease may be preventable through improving these risk factors. Cardiovascular or heart disease, actually refers to a class of diseases that involve the heart and blood vessels, and includes several distinct types – coronary artery diseases, stroke, heart failure, hypertensive heart failure, rheumatic heart disease, and many more. For this reason, diagnosis can be even less linear based on lifestyle factors and available medical data. Being able to better predict and prevent heart disease could change or save the lives of millions of people. A variety of machine learning classification techniques may be able to better analyze the more complex relationships between medical data features and a possible heart disease diagnosis.

Several studies have been completed to define a classifier for heart disease utilizing machine learning techniques. As in [4] it has been proven that machine learning techniques such as logistic regression and neural networks can improve upon the established risk prediction methods for diagnosing heart disease. Ultimately, all prediction methods achieved an accuracy nearing 80%. While a great improvement from current medical standard, with heart disease affecting such a high portion of the population a misdiagnosis rate of higher than 20% is still a huge portion of the population and improving this accuracy rate even further could positively impact and possibly save many more lives.

Additionally, many studies have been done in predicting the diagnosis of other lifestyle-correlated illnesses utilizing fuzzy logic principles. Inference systems utilizing fuzzy logic have been used with success in predictions of both Diabetes and Hepatitis [2], [3]. Intuitively, a continuous scale between 0 and 1 makes sense as diseases like heart disease, diabetes, and hepatitis are known to be a gradual development as a result of long-term lifestyle factors and usually aren't instantaneous occurrences.

In this study the input will be 10-13 features of medical data that are easily obtainable and typically used for heart disease diagnosis and the output will be the prediction of heart disease (1) or no heart disease (0). We will run and tune a variety of classification algorithms to try and achieve the best accuracy possible. Furthermore, in this study we also attempt to incorporate nonlinear complexities through merging fuzzy logic principles with an artificial neural network in an Adaptive Network-Based Fuzzy Inference System, in seeking further improvement.

DATASET AND FEATURES

In this study, two datasets were considered. The first was a complete dataset of 303 subjects with 13 features including age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting electrocardiographic results, exercise induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels colored by flourosopy, presence of no defect, reversible defect, or fixed defect, and diagnosis of heart disease (0 or 1). The second was a dataset of 760 subjects with the same first 10 features as the first dataset and the diagnosis of heart disease. This dataset was collected from studies in Hungary, Switzerland, and Cleveland, OH. Both of these datasets were obtained from the UCI Machine Learning Repository [1]. Any missing feature entries in the datasets were imputed with mean values. Initially looking at the correlation matrix for the dataset with 10 features and the dataset with 13 features it appears there is no strong correlation between any one feature and a diagnosis of heart disease (target).

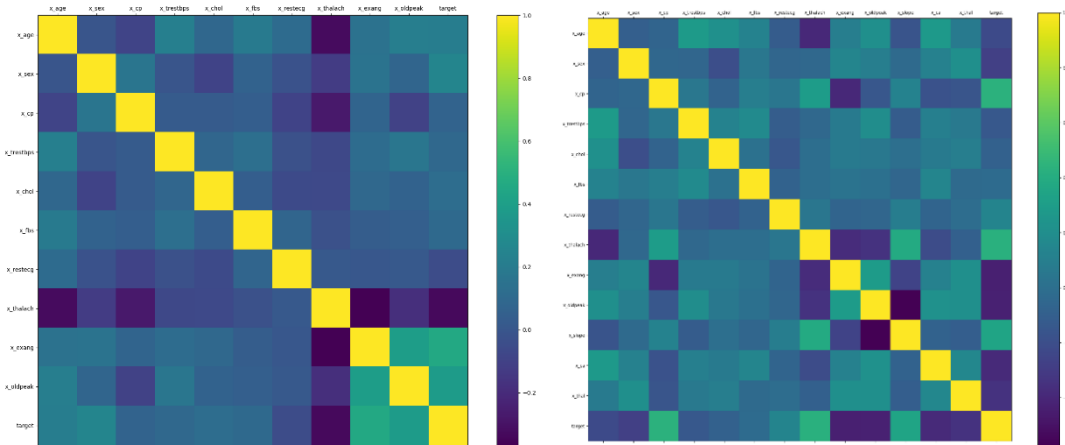


Figure 1: Correlation matrix for 10 features for the combined datasets (left), correlation matrix for all 13 features for the first dataset (right).

METHODS, EXPERIMENTS AND RESULTS

Four types of binary classifiers were initially tested on both datasets: logistic regression, K-Nearest Neighbor classifier, Support Vector Machine classifier and an artificial neural network. For all experiments data was randomly shuffled, normalized and split 80% for training and 20% for testing. All accuracies reported are for the test data.

LOGISTIC REGRESSION

Initially, logistic regression was applied to both datasets. Logistic regression is a nonlinear transformation of the linear regression model between the input variables and the binary class assignment. It maps this linear regression using a function like the sigmoid (S-shape) so that all probabilities are mapped between 0 and 1. Then the class of the input example is determined based on this probability (i.e. a value greater than 0.5 would denote class 1 and vice versa).

As a first phase classifier for both datasets a logistic regression strategy did fairly well with over 80% accuracy for both datasets. The larger dataset with 10 features achieved an accuracy of 0.809 and the smaller dataset with 13 features achieved an accuracy of 0.836. There was a nearly 3% improvement between the larger dataset with less features than the smaller dataset with more features, possibly indicating that the three additional features may provide additional relevant information for a diagnosis of heart disease. However, due to the smaller size of the dataset, more experiments will need to be run to confidently make this conclusion as this could also be a result of overfitting to this particular dataset and region of collection.

K-NEIGHBORS CLASSIFIER

Next, a k-nearest neighbors classifier was run on both datasets. In this type of classifier, the K nearest neighbors of a given data point are analyzed and the majority class of these K neighbors is assigned to the data point. The number of K nearest neighbors can be varied and in this study, it was varied from 1 to 100 for the larger dataset, finding a peak accuracy of 0.87 at K=47 nearest neighbors. For the smaller dataset it was varied from 1 to 20 finding a peak accuracy of 0.89 for K=10 nearest neighbors. For both datasets the accuracy of the classifier is improved from the logistic regression classifier. Again, we see a slightly superior performance from the smaller dataset with more features than the larger dataset with three fewer features.

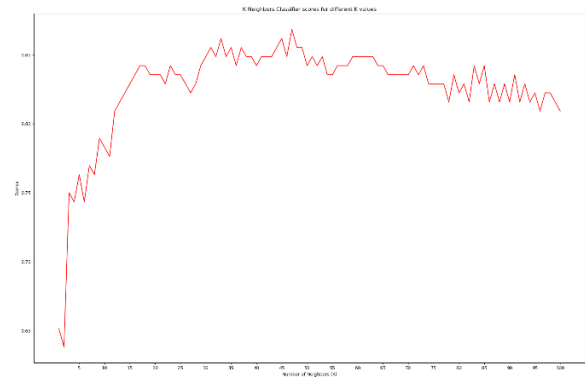
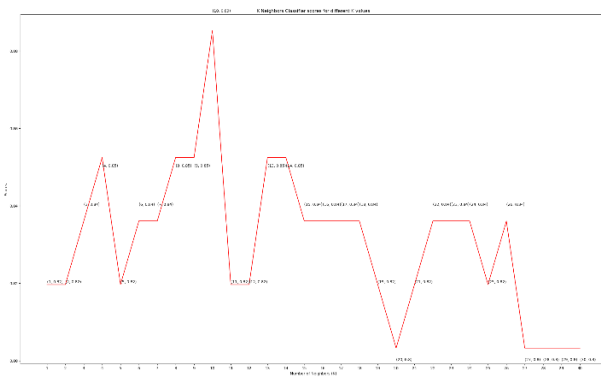


Figure 2: Score plots for variations in K neighbors for small dataset (top) and larger dataset (bottom)

SUPPORT VECTOR MACHINES

Next, a support vector machine classifier was tested to see if the accuracy could further be improved. Support vector machines is a machine learning technique that looks at separations in data when mapped to higher-dimensional space. It can utilize a variety of “kernel tricks” to assess not just linear relationships but also use polynomial, radial basis function or sigmoid feature mappings to find a split in higher dimensional space between binary classifications. In this study, linear, polynomial, radial basis function, and sigmoid feature mappings were tested on both datasets.

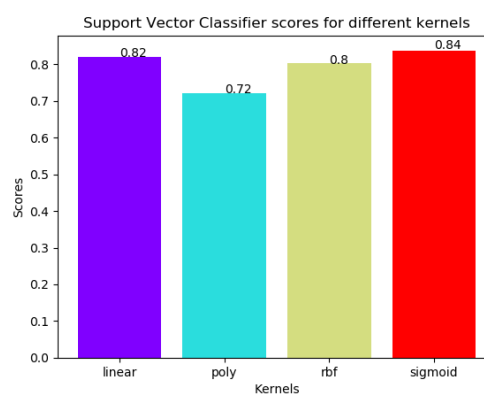
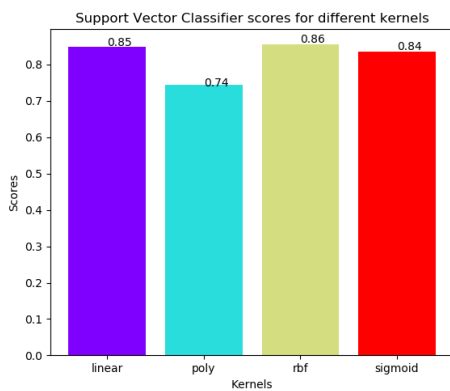


Figure 3: Score plots for different feature mapping for SVM classifier for large dataset (left) and smaller dataset (right)

The SVM classifier does not improve upon the 0.87 or 0.89 accuracy from the highest performing K-neighbors algorithm, however it is interesting to note that different feature mappings had different performance orders for the two datasets, and for the first time the classifier for the larger dataset with fewer features outperformed the classifier for the smaller dataset with more features. The radial basis function feature map for the large dataset classifier performed the best overall with an accuracy of 0.86, while the sigmoid function performed best for the smaller dataset with an accuracy of 0.84. This indicates that the three additional features of the smaller dataset may add complexities to the relationships between the two classes in higher dimensional space that may not be able to be entirely captured by these four feature mapping split strategies.

NEURAL NETWORK

The next experiment was to implement a neural network classifier on the two datasets to see if we could exceed the previous max accuracy of 0.87. Several iterations of a neural network were constructed varying numbers of hidden layers, numbers of neurons in each layer and batch size during training. All hidden and input layers used a ReLu activation function, and the output layer employed a sigmoid activation function. For both datasets the best performing neural network had three hidden layers with 6, 16, and 5 neurons and was trained for 100 epochs with a batch size of 2. This design achieved an accuracy of 0.787 for the smaller dataset and an accuracy of 0.803 for the larger dataset.

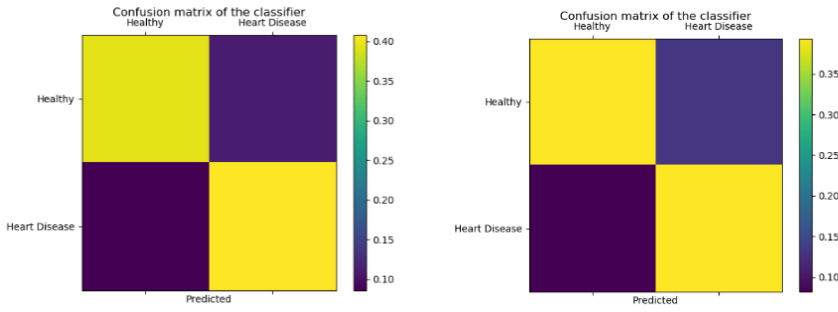


Figure 4: Normalized confusion matrixes of test data from neural network classifier for large dataset (left) and smaller dataset (right)

ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEM (ANFIS)

ANFIS combines the structure of a neural network with “fuzzy logic” principles that attempt to replicate the nonlinearities of a medical professional’s experiential knowledge when making a decision about a patients diagnosis based on medical data. This principle of fuzzy logic looks at a classification more as a sliding scale between 0 and 1, where there may be more of a continuity to a subject’s diagnosis of heart disease instead of a cut and dry binary classification.

ANFIS consists of two parameter types, premise and consequence which are tuned during the training of the algorithm. There are five layers to the ANFIS network structure. The first, denoted the fuzzification layer, uses membership functions to obtain fuzzy groups of the inputs. Premise parameters are tuned to determine the form of membership functions. The second layer, or rule layer, determines the firing strengths of the rules generated by the first layer. The normalization layer comes third and it calculates the normalized firing strengths belonging to each rule. In the fourth layer, or defuzzification layer, weighted values of each rule are calculated using a first order polynomial and the tuned consequence parameters. The fifth and output layer then obtains the actual output by summing the outputs from the defuzzification layer.

ANFIS Layer Calculations	
$\mu_{A_i}(x) = gbellmf(x; a, b, c) = \frac{1}{1 + \left \frac{x-c}{a} \right ^{2b}}$	$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$
$O_i^1 = \mu_{A_i}(x)$	
$O_i^2 = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad i = 1, 2.$	$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$
$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + w_4} \quad i \in \{1, 2, 3, 4\}$	

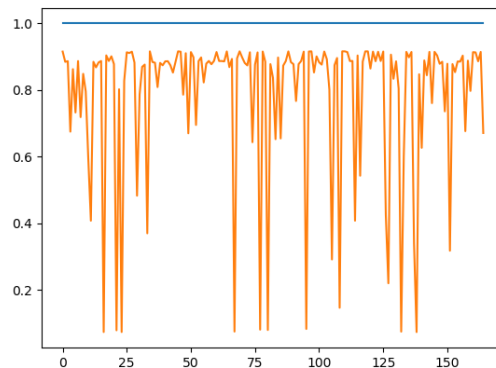
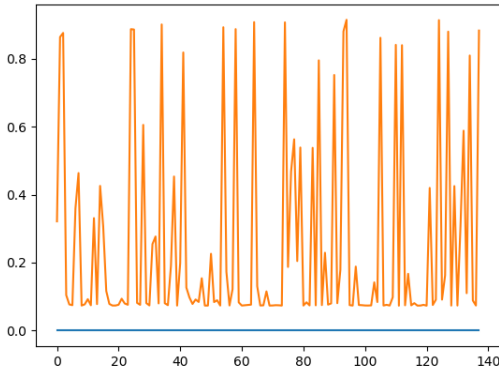


Figure 5: True labels vs. output of ANFIS for small dataset with 13 features. Healthy subjects (left) and subject with a heart disease diagnosis (right).

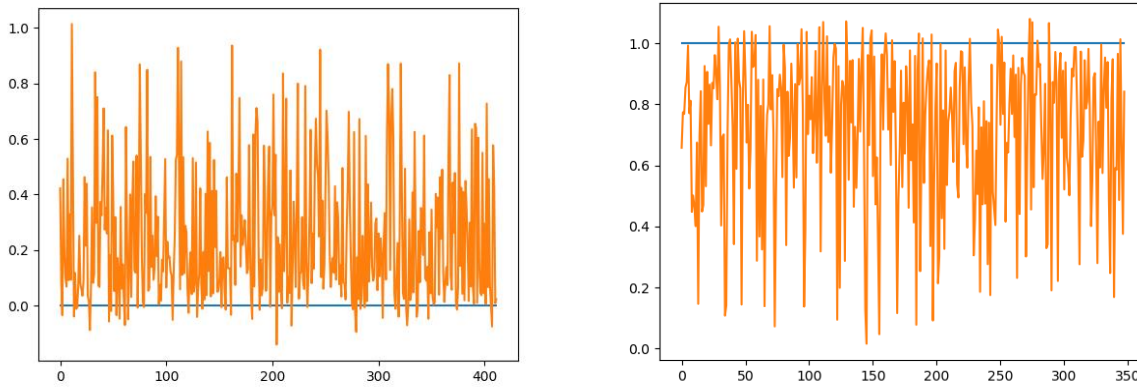


Figure 6: True labels vs. output of ANFIS for small dataset with 10 features. Healthy subjects (left) and subject with a heart disease diagnosis (right).

Accuracy for the ANFIS was determined utilizing the same threshold as binary classifiers in which a value greater than -0.5 was denoted as a heart disease diagnosis and vice versa. Figure 6 shows the outputs of ANFIS as compared to the test data. The ANFIS performs quiet well on the small dataset achieving an accuracy of 0.891, and performs on par with existing prediction methods on the larger dataset achieving an accuracy of 0.820. The small dataset ANFIS system was tuned to have 24 membership function rules and utilized a training learning rate of 0.07, while the ANFIS trained using the large dataset performed best with 32 membership function rules and a training learning rate of 0.05. Looking at the plots of the test data outputs for the small dataset, it appears that most misclassifications by ANFIS were quite confident. These particular data points should be further studied to better understand why ANFIS so confidently misclassifies them. The outputs for the larger dataset are a bit noisier in their classifications, which may be as a result of the smaller feature space.

CONCLUSION AND FUTURE WORK

DATASET	LR	KNN	SVM	NN	ANFIS
SMALL W/ 13 FEATURES	0.809	0.89	0.84	0.787	0.891
LARGE W/ 10 FEATURES	0.836	0.87	0.86	0.803	0.820

Generally performance was good for all types of classifiers tried. The best results achieving approximately 89% accuracy were the ANFIS algorithm and K-nearest neighbors with $K = 10$ for the smaller dataset with more features. While a few percent improvement from existing systems scoring just under 80% may not seem drastic, this can actually be quite substantial for a disease that affects such a high portion of the global population. It appears that either the extra three features or some other noncategorical attribute of the smaller dataset lead to a better performance in heart disease prediction overall. In future work, this could be explored by collecting more data inclusive of the three extra features and running the same experiments. Other data of interest could also be collected and analyzed in the future, including lifestyle factors such as diet, smoking history, body mass index, frequency of exercise and geographic region. While these may introduce more levels of uncertainty and imprecision, ANFIS may be able to handle these sorts of features better than traditional prediction approaches. Further experiments of the ANFIS could also be run. For example, it would be interesting to incorporate a medical professional's experiential knowledge by initially suggesting certain membership rules in the fuzzification layer. Lastly, further exploration could be done in combining fuzzy logic principles with one of the higher performing traditional classifiers, such as K-Nearest Neighbors.

Overall, the study was successful in achieving a slight improvement over existing prediction methods for heart disease diagnosis using the same feature set. ANFIS performed well but not drastically better or worse than any of the traditional classification algorithms tested. Future improvement will require further data collection with a bigger feature space and perhaps further understanding of the causes of heart disease to improve both feature selection and membership function generation in the ANFIS algorithm. Additionally, since the neural network classifier performed the worst, it may be beneficial to combining fuzzy logic principles with one of the higher-performing traditional classifiers in a new type of algorithm.

APPENDIX

All code can be found at <https://github.com/ekunz/CS-229-Project>.

ACKNOWLEDGEMENTS/REFERENCES

- [1] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [2] Ibrahim, R., Olawale, O. and Funmilayo, K. (2018). Diagnosis of Hepatitis using Adaptive Neuro-Fuzzy Inference System (ANFIS). *International Journal of Computer Applications*, 180(38), pp.45-53.
- [3] Dogantekin, E., Dogantekin, A., Avci, D. and Avci, L. (2010). An intelligent diagnosis system for diabetes on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System: LDA-ANFIS. *Digital Signal Processing*, 20(4), pp.1248-1255.
- [4] Weng, S.F., Reys, J., Kai, J., Garibaldi, J.M., Qureshi, N. (2017) Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*, 12 (4), art. no. 0174944.
- [5] B. K. Sarkar, "Hybrid model for prediction of heart disease." *Soft Computing*, 2019.
- [6] K. Brahim and A. Zell, "ANFIS-SNNS: Adaptive Network Fuzzy Inference System in the Stuttgart Neural Network Simulator," *Fuzzy-Systems in Computer Science*, pp. 117–127, 1994.