 doi:10.20944/preprints202207.0097.v1

# Leveraging Machine Learning and Model-Agnostic Explanations to Understand Automated Diagnosis of Cardiovascular Disease

**Christopher Sun, Jai Sharma, Milind Maiti**


## ABSTRACT

The pervasiveness of cardiovascular disease and physician misdiagnosis creates the urgent need for artificial intelligence models to improve diagnosis accuracy. The first objective of this study was to train machine learning models on publicly available data sets containing simple medical information of patients to diagnose cardiovascular disease. The Multilayer Perceptron (MLP) assembled for this task performed optimally with an $F_1$ score of 0.8968. This prompted the creation of an open-source, automated cardiovascular disease diagnosis tool, powered by the MLP. The second objective of this study was to employ a meta-learning methodology called Local Interpretable Model-Agnostic Explanations (LIME) to understand the impact of different features on the model's diagnosis in the form of marginal probabilities. K-Means Clustering was employed to segment the data into ten clusters, after which each data example was passed through LIME. The resulting histograms depict the complex relationship between feature, cluster, and impact on diagnosis. A series of P-values with contrasting orders of magnitude shows the nuances in the MLP's understanding of patients from different clusters. The results of meta-learning analysis reveal that the most important features for cardiovascular disease diagnosis are fasting blood sugar, type of chest pain, and slope of the ST segment on an electrocardiogram. Future experiments should replicate the novel methodology introduced in this study on data sets containing more specialized medical features in order to gain practical medical insights about different types of cardiovascular disease represented by each cluster. Finally, feature engineering pathways should be explored with consideration of these results to create versatile diagnosis models not only for cardiovascular disease, but adaptable to other diseases as well.


***K*eywords**: Cardiovascular Disease;· Model-Agnostic Meta-Learning

## 1 Introduction

### 1.1 Research Context

Cardiovascular disease causes one out of every four deaths in America [6]. The World Health Organization reports that 18 million people in the world die from cardiovascular disease complications annually, granting the disease the disreputable title of the leading cause of death [2]. Even highly developed nations with the best medical professionals using the best diagnosis and treatment technologies are susceptible to this issue that is evolving into a worldwide health crisis.

For example, although urban cities have the modern medical facilities to conduct procedures required to diagnose cardiovascular disease, misdiagnosis still causes more than 11,000 deaths each year [13]. Worse, in rural areas, because of the lack of rural medicine practitioners, patients at risk for the disease cannot receive diagnoses that are accurate and early in the prognosis of the disease [3]. The cardiovascular disease diagnosis process involves multitude of procedures for those suspecting they have contracted the disease. An accurate diagnosis requires around ten clinical procedures, forcing potential patients to make several visits to hospitals [16]. In worst cases, the diagnosis process can delay vital treatment for affected patients, generating greater complications that could have been prevented earlier. In addition, diagnoses from doctors are subject to the risks of subconscious biases [1].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

Most research regarding artificial-intelligence-assisted cardiovascular disease diagnosis involves the use of a single data set from the UCI Machine Learning Repository called the Cleveland Database, containing 303 patients and 75 features, though only 14 features have been used in the majority of research. [7]. There are a multitude of ongoing studies training various machine learning models on this data set, achieving accuracies ranging from 85% to 90% [9] [10]. Other research has focused on using software to synthesize patient data based on trends in real patient data. Due to the limited amounts of data that is available for cardiovascular disease diagnosis, data synthesis is a development that may result in more accurate models in the near future; however, it is too early to determine how effective this technique is at generalizing to real-world problems.

### 1.2 Objective

The objectives of this research are twofold. The first objective is to train machine learning and deep learning models to accurately diagnose cardiovascular disease, using the highest-performing model to create a final open-source product available for use by all. The second objective is to understand the process of learning of the highest-performing model to discover each feature's degree of importance in the diagnosis. Results from this analysis may motivate feature engineering pathways to create more versatile diagnosis models not limited to only cardiovascular disease, but including other diseases as well.

## 2 Materials and Methods

### 2.1 Data Used For Experimentation

The data set used for this research is a cumulative data set synthesized from patient records from four hospitals around the world available on the UCI Machine Learning Repository [7]. The data set contains 918 patient records and 11 features: age, gender, resting systolic blood pressure, cholesterol, presence of elevated fasting blood sugar, maximum heart rate, presence of exercise-induced angina, type of chest pain, resting electrocardiographic results, ST depression induced by exercise (oldpeak), and ST segment slope. The target variable is a binary classification of the presence or absence of cardiovascular disease.

### 2.2 Data Preprocessing

Prior to experimentation, categorical features (i.e. type of chest pain and resting electrocardiographic results) were one-hot encoded as binary subfeatures. A patient's type of chest pain was separated into four binary features: typical angina, atypical angina, nonanginal pain, and asymptomatic. Resting electrocardiographic results were separated into three features binary: normal results, ST-T wave abnormality, and left ventricular hypertrophy. After one-hot encoding, the data set contained eighteen features.

The data set also contained invalid cholesterol measurements recorded as 0 mg/dL. These instances were interpolated using K-Nearest Neighbors (KNN) interpolation. The K-value used to train the interpolation algorithm was determined by training a KNN Classifier on the remaining data examples that contained valid cholesterol measurements. The K-value that resulted in the highest prediction accuracy was 30, so this was used accordingly for interpolation.

Furthermore, one resting systolic blood pressure measurement in the data set was recorded as 0 mmHg. This invalid measurement was imputed with the median resting systolic blood pressure, 130 mmHg.

After necessary imputations and interpolations, the data set was divided into train and validation sets with a 75-25% split.

### 2.3 Exploratory Data Analysis

Tests for statistical inference were conducted to gauge the validity of the data set and provide insights regarding whether specific manifestations of features are more or less conducive to the presence of cardiovascular disease. First, the data set was split into two subsets by patient diagnosis. These subsets were then treated as separate, independent populations for the purpose of the statistical tests.

For numerical features, two-sample t-tests were conducted for the difference in means. For binary features, two-proportion z-tests were conducted for the difference in proportions. The null hypothesis was the absence of a statistical difference in the mean or proportion of the features between patients with and without cardiovascular disease. For categorical features, chi-squared tests for homogeneity were conducted to find the association between categories and the presence of cardiovascular disease. The null hypothesis was that the populations represented by the two data sets constituted the same distribution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

### 2.4 Predictive Modeling

#### 2.4.1 Metrics

Three metrics were used to assess the performance of the predictive models: binary accuracy, the $F_\beta$ score with a $\beta$ value of 1, and the AOC-ROC Curve.

Let $m$ be the number of patients in the data set. Let $x$ be the number of patients who are correctly classified by the Machine Learning model. Then, the binary accuracy is then defined as $\left(100 \cdot \frac{x}{m}\right)\%$.

Let $tp$ denote the true positives, the number of times that a patient with cardiovascular disease is correctly classified. Let $fp$ denote the false positives, the number of times the model incorrectly classifies a healthy patient as having cardiovascular disease. Let $tn$ denote the true negatives, the number of times that a patient without cardiovascular disease is correctly classified. Let $fn$ denote the false negatives, the number of times the model incorrectly classifies a cardiovascular disease patient as healthy. Then, the $F_1$ score is defined as the harmonic mean of precision and recall, where precision equals $\frac{tp}{pp}$ and recall equals $\frac{tp}{rp}$ :

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}.$$

The receiver operating characteristic (ROC) curve is a metric often used in clinical application that gauges the relationship between the rate of true positives and the rate of false positive at different discrimination thresholds. The area under the ROC curve (AUC) is a number between 0 and 1, with a higher AUC implying more optimal performance and a more convex ROC curve.

#### 2.4.2 Machine Learning Models

Four machine learning models were trained to classify the presence of cardiovascular disease, namely the Logistic Regression, Random Forest, Extreme Gradient Boosting, and Support Vector Machine classifiers.[1] The Random Forest and Extreme Gradient Boosting classifiers contained 100 trees. The Support Vector Machine used the radial basis function kernel.

#### 2.4.3 Deep Learning Model

The deep learning model constructed for this task was a Multilayer Perceptron (MLP) consisting of 4 dense layers with 64 hidden units each and an output layer with one node. Batch Normalization and Dropout regularization with a dropout rate of 0.4 were used to regularize the model. Additionally, Xavier's initialization was used to mitigate the problems of vanishing and exploding gradients [11]. The model was trained using Adam optimization and early stopping until the validation set accuracy plateaued, with a mini-batch size of 64 examples and a learning rate of 0.005. [8].

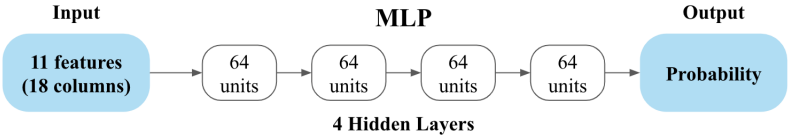



Figure 1: The Multilayer Perceptron received an input shape of 18 and outputted the patient diagnosis. The model contained 14,857 parameters.

The research methodologies described henceforth will concern the deep learning model, our final predictive model.

### 2.5 Local Interpretable Model-Agnostic Explanations

After assembling our final predictive model to accurately diagnose cardiovascular disease, it was important for us to understand the individual predictions of the artificial neural network in an effort to demystify how the model learned. To do so, we leveraged model-agnostic explanations using the LIME framework developed by Ribeiro et al. to gauge the relative important of each feature for the model to arrive at a diagnosis [12]. As the task at hand is a classification problem, LIME deciphers these marginal probabilities through sparse linear explanations and the submodular pick

[1]The machine learning model achieving the highest $F_1$ score acted as a standard of comparison for the MLP.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

algorithm [12]. After feeding each data example through the LIME explainer, the result is a catalog of marginal probabilities, negative and positive, that each feature contributes to the model's diagnosis for that example.

#### 2.5.1 K-Means Clustering

Patients would not be comparable with respect to feature impact on diagnosis unless they were grouped into larger, more representative categories. Hence, we used model-agnostic explanations in conjunction with K-Means Clustering, as we hypothesized that the cardiovascular disease patients in the data set could be grouped into distinct clusters, where patients within in each cluster are more so homogeneous than patients belonging to different clusters.[2] Medically, these clusters could represent the presence of more nuanced cardiovascular conditions, or different types of cardiovascular disease (e.g. rheumatic heart disease, congestive heart failure, etc.). In turn, this allowed us to compare the results of the LIME explainer across patients of different clusters.

The optimal number of clusters was ten, as determined by the elbow method [4]. For each patient in each cluster, we computed the marginal probability provided by each feature to the diagnosis of the patient. We did not differentiate between different manifestations of features (e.g. presence vs. absence of exercise-induced angina), as we applied the absolute value to the resulting series of probabilities to understand the magnitude of the impact of different features on the diagnosis. This allowed us to convert bimodal distributions to near-normal distributions that could be better analyzed.

For each feature, we created 2-D and 3-D histograms that compare the distribution of the feature impact on diagnoses of patients between all ten clusters. This created a total of 36 histograms containing ten distributions each.

Furthermore, for each cluster, we created histograms that compare the distribution of feature impact on diagnoses of patients in that cluster across all features. This created a total of ten histograms containing 18 distributions each.

## 3 Results and Interpretation

### 3.1 Exploratory Data Analysis

Comparative analysis between the two subgroups of healthy and unhealthy patients showed that the manifestations of each feature between groups was statistically significant.

Patients with cardiovascular disease possessed a higher mean age, higher mean resting blood pressure, higher mean cholesterol, lower mean maximum heart rate, and higher mean oldpeak depression than their healthy counterparts. These differences in means were all statistically significant at the $\alpha = 0.01$ significance level.

The proportion of patients with cardiovascular disease that were male, had elevated fasting blood sugar levels, and suffered from exercise-induced angina was higher than that of patients without cardiovascular disease. These proportions were all statistically significant at the $\alpha = 0.01$ level.

Finally, the chi-squared tests showed that the deviation between observed and expected values of patients with different types of chest pain, electrocardiographic results, and ST segment slope, against a null hypothesis of homogeneity between healthy and unhealthy patients, was statistically significant at the $\alpha = 0.01$ level.

[2] Only patients with cardiovascular disease were clustered.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

Table 1: Statistical Difference of Feature Manifestation Between Healthy and Unhealthy Patients

| **Feature** | **P-value** |
|---|---|
| Age | 2.9259e-18 |
| RestingBP | 1.4038e-4 |
| Cholesterol | 2.7921e-3 |
| MaxHR | 6.0683e-37 |
| Oldpeak | 8.0281e-41 |
| ChestPain | 8.0837e-58 |
| ECG | 4.2292e-3 |
| ST_Slope | 5.1676e-78 |
| Sex | 1.0759e-20 |
| FastingBS | 2.7815e-16 |
| ExerciseAngina | 5.2677e-51 |

### 3.2 Predictive Models

Table 2: Model Metrics

| | **Binary Accuracy** | **$F_1$ Score** | **AUC** |
|---|---|---|---|
| Logistic Regression | 0.8565 | 0.8696 | 0.9199 |
| Random Forest | 0.8522 | 0.8640 | 0.9176 |
| Extreme Gradient Boosting | 0.8696 | 0.8790 | 0.9092 |
| Support Vector Machine | 0.8652 | 0.8755 | 0.9132 |
| Multilayer Perceptron | 0.8870 | 0.8968 | 0.9161 |

The machine learning model achieving the highest $F_1$ score was the Extreme Gradient Boosting Classifier. The Extreme Gradient Boosting Classifier outperformed the Support Vector Machine by 0.44 percentage points and 0.45 $F_1$ score. However, the Extreme Gradient Boosting Classifier possessed the lowest AUC, while the Logistic Regression Classifier surprisingly outperformed other machine learning models to score the highest AUC.

The Multilayer Perceptron outperformed all machine learning models in binary accuracy and $F_1$ score, but fell short in terms of the AUC metric. This means that though the MLP achieved the highest diagnosis accuracy, there was as steeper trade-off between the model's sensitivity and specificity. Despite this, the MLP was the considered the final model because of an optimal $F_1$ score, and as a result, was used to create the open-source diagnosis tool. The confusion matrix of the MLP is shown below.

Table 3: Confusion Matrix (Validation Set)

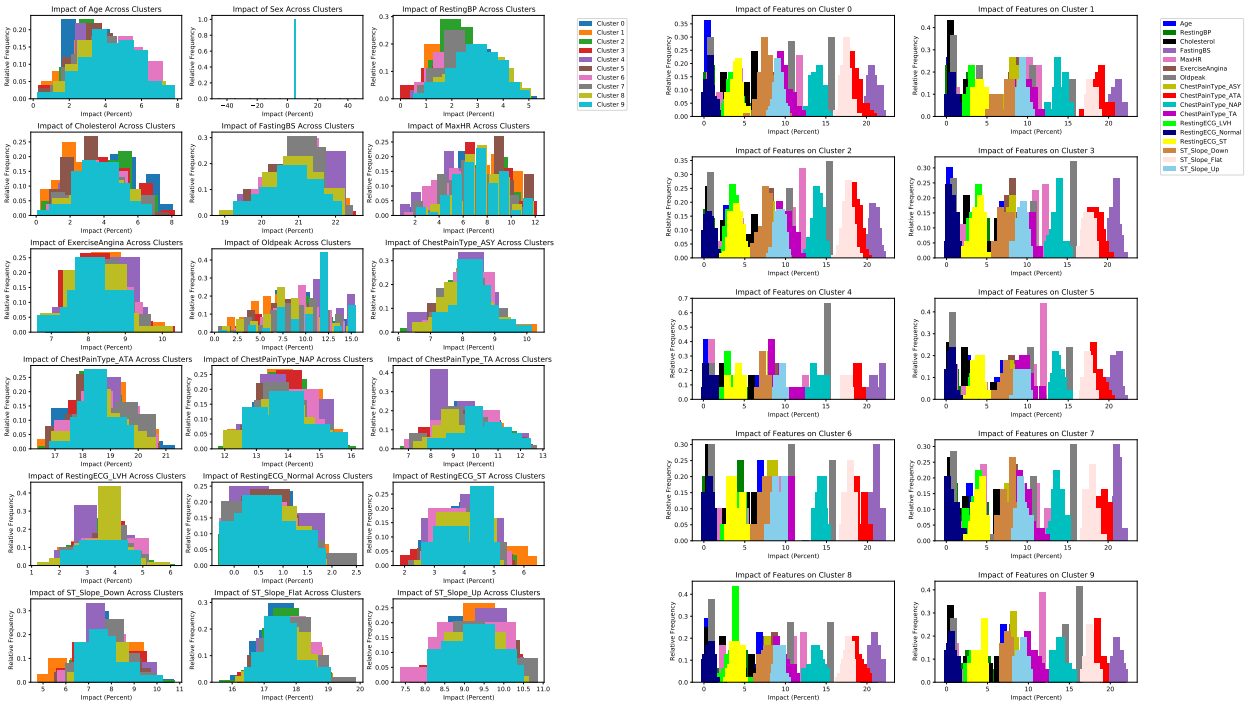
| | Actual: Disease | Actual: No Disease | |
|---|---|---|---|
| Predicted: Disease | TP = **113** | FP = **18** | 131 |
| Predicted: No Disease | FN = **8** | TN = **91** | 99 |
| | 121 | 109 | 230 |

#### 3.2.1 Cardiovascular Disease Diagnosis Tool

We have built an open-source cardiovascular disease diagnosis tool powered by the MLP, that has practical applications in households and in rural clinics where only simple feature information can be obtained [14]. We believe this resource can be a source of confirmation for physician diagnoses as well as an informative medical guide in the absence of physicians.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

### 3.3 Clustering and Local Interpretable Model-Agnostic Explanations




(a) Differential impact of each feature across all clusters. (b) Differential impact on each cluster across all features.

Figure 2: (a) shows how each cluster reacts differently to the same features. (b) ranks features in order of impact, for each cluster.

Figure 2a reveals that the distributions of feature impact take on different shapes and spreads across clusters, implying that the MLP did not look at patients from different clusters through the same lens even when patients were represented by similar feature information. For some plots (e.g. FastingBS, ChestPainType, and ST_Slope), the magnitude of feature impact is very clearly nonzero, implying these features are universally important for cardiovascular disease diagnosis. Furthermore, the second histogram in the first row of Figure 2a shows that the sex of a patient had no impact on the MLP's perception of that patient's cardiovascular condition. This was the same for patients in all clusters. Hence, the MLP was able to eliminate gender bias from its diagnoses.

Figure 2b[3] reveals that across all clusters, the most and least impactful features tended to remain the same. Namely, fasting blood sugar and chest pain type were the most important for the diagnosis, while resting electrocardiographic results were the least important, an interesting phenomenon due to the fact that ECG signals have been shown to possess high correlations with cardiovascular abnormalities [5].

#### 3.3.1 Continuous Features

For continuous features, initial histograms (before applying absolute value) showed that there seemed to be three distinct outcomes of feature impact: positive marginal probability with a large magnitude, negative marginal probability with a large magnitude, and near-zero marginal probability. The practical interpretation of this result is that depending on the manifestation of the continuous feature, the model either considers the feature as crucial biomarker for the diagnosis, or overlooks the feature. To combat this issue of non-normal distributions, we conducted random sampling to obtain the sampling distribution of continuous features' impact scores. The sample size $n$ was 4, and even though this sample size is statistically unsusceptible by the Central Limit Theorem, it was the best choice for our particular application, due to smaller cluster sizes. The process of sampling is illustrated below.

[3]Figure 2b does not plot the impact of sex on the diagnosis, for the reasons discussed in the former paragraph.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

**Algorithm 1** Continuous Features Procedure of Analysis

**procedure** SAMPLING DISTRIBUTION
$\bar{x} \leftarrow \{\}$
$n \leftarrow 4$
$m \leftarrow 100$

**for** $i \in \{0, 1, ..., m\}$ **do**
$K \leftarrow$ shuffled distribution
$\bar{x} \leftarrow \bar{x} || \frac{\sum_{j=0}^{n} K_j}{n}$
**end for**

$\mu \leftarrow \mu_{\bar{x}}$
$\sigma \leftarrow \sigma_{\bar{x}} \cdot \sqrt{n}$
**end procedure**

The resulting means and standard deviations were used to conduct the significance tests that resulted in the p-values shown in the Appendix. The figures above also depict the resulting sampling distributions of continuous features, not the original probabilities obtained through LIME.

## 4 Conclusions

Through a comprehensive machine learning study, we have demonstrated a feature extraction methodology to diagnosis heart disease in patients using simple medical data. The LIME and K-Means Clustering approaches can be applied to other larger data sets to detect nuanced relationships between different types of cardiovascular disease. We conducted two rounds of analysis to gauge feature impact: an initial data-driven exploratory analysis and a final model-agnostic interpretation to understand the wiring of automated diagnosis models.

The data set we selected for this study lacks the complexity of features for the resulting centroids to have any discrete medical implications, since physicians cannot differentiate between clusters using only simple feature information. Therefore, future research can focus on creating more detailed data sets involving measurements such as troponin and MRI results, so that the K-Means Clustering methodology can be validated and used to make an in-depth diagnosis of patients represented by the individual centroids.

**Research Applications** In rural areas and third-world countries, easily accessible features can be collected and fed through the diagnosis tool to diagnose high risk patients for heart disease in advance [14]. The diagnosis tool can also be used over a long period of time to monitor patients' cardiovascular health. These patients can then be given the appropriate medical treatment. Accurate predictive models may also be used to forecast demand for medical equipment in hospitals by predicting how many patients are likely to contract heart disease, allowing for the effective allocation of hospital resources.

We have made the code for the research in this paper open-source and available on GitHub [15].

## References

[1] Beery, T. A. (1995). Gender bias in the diagnosis and treatment of coronary artery disease. Heart & Lung, 24(6), 427-435.

[2] "Cardiovascular Diseases." World Health Organization, WHO, www.who.int/health-topics/cardiovascular-diseases/

[3] Celermajer, D. S., Chow, C. K., Marijon, E., Anstey, N. M., & Woo, K. S. (2012). Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection. Journal of the American College of Cardiology, 60(14), 1207-1216.

[4] Cui, M. (2020). Introduction to the k-means clustering algorithm based on the elbow method. Accounting, Auditing and Finance, 1(1), 5-8.

[5] Hammad, M., Maher, A., Wang, K., Jiang, F., & Amrani, M. (2018). Detection of abnormal heart conditions based on characteristics of ECG signals. Measurement, 125, 634-644.

[6] "Heart Disease Facts." Centers for Disease Control and Prevention, U.S. Department of Health; Human Services, 8 Sept. 2020, www.cdc.gov/heartdisease/facts.htm.

[7] Janosi, Andras, et al. "UCI Machine Learning Repository." Heart Disease Data Set, archive.ics.uci.edu/ml/datasets/Heart+Disease.

[8] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[9] Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., ... & Tang, W. H. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific Reports, 10(1), 1-11.

[10] Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. Scientific reports, 10(1), 1-17.

[11] Ng, Andrew, and deeplearning.ai. Deep Learning Specialization. www.coursera.org/specializations/deep-learning?utm_source=gg.

[12] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[13] Sharkey, L. "Misdiagnosis of the 'Big Three' Results in 'Serious Harm."' Medical News Today, MediLexicon International, 22 July 2019, www.medicalnewstoday.com/articles/325811.

[14] Sharma, J., Sun, C., & Maiti, M. Cardiovascular Disease Diagnosis Tool. 26 Jun. 2022. https://tinyurl.com/cardiovasculardiseasediagnosis

[15] Sun, C., Sharma, J., & Maiti, M. (2022). csun365/Cardiovascular-Disease: Code, data, and results for the paper "Leveraging Machine Learning and Model-Agnostic Explanations to Understand Automated Diagnosis of Cardiovascular Disease." GitHub. https://github.com/csun365/Cardiovascular-Disease

[16] Whitworth, Gerhard. "How Is Heart Disease Diagnosed?", Healthline Media, 29 Sept. 2018, www.healthline.com/health/heart-disease/tests-diagnosis.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 July 2022 doi:10.20944/preprints202207.0097.v1

# Appendix

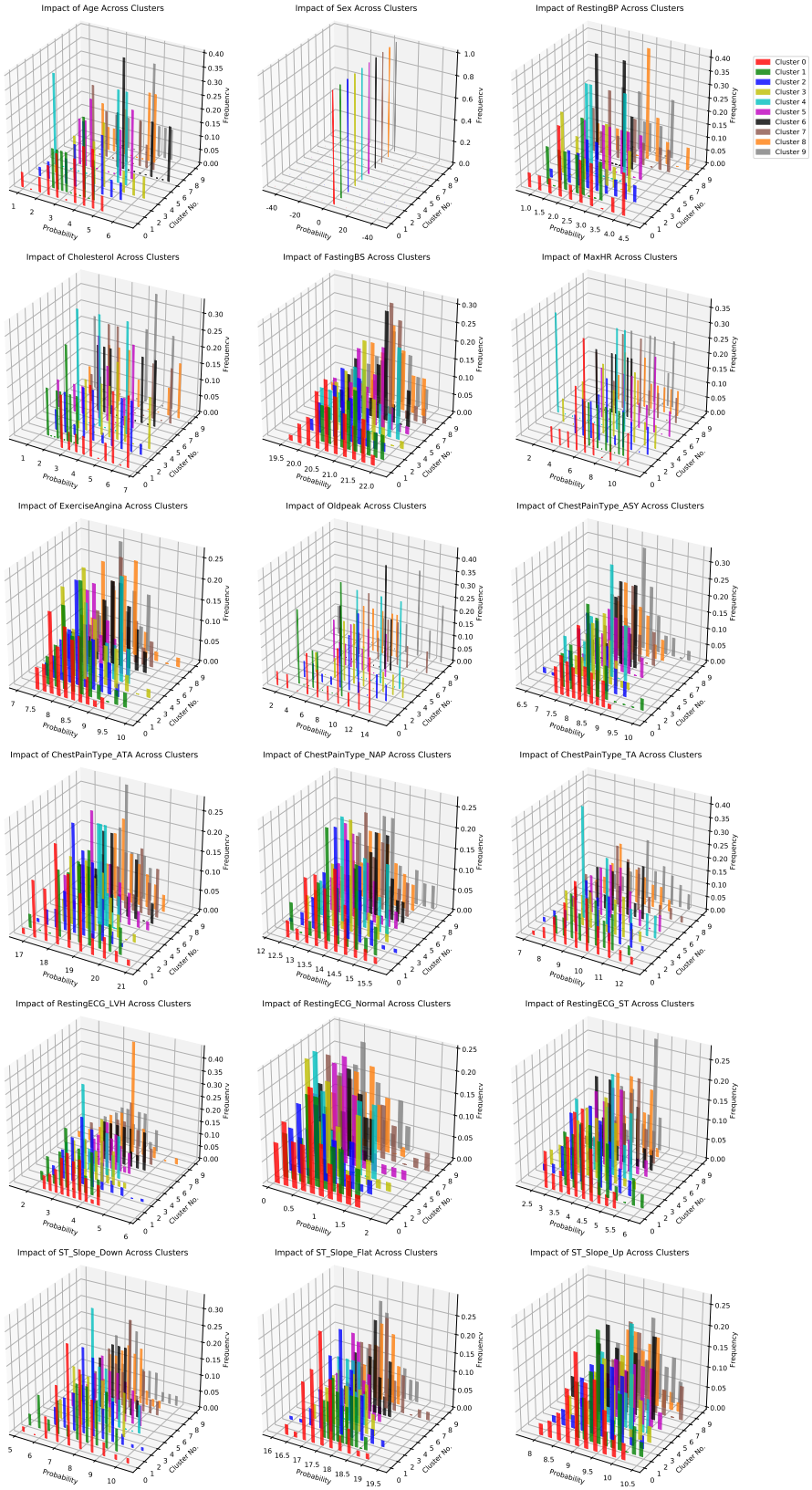



Figure 3: 3-D representation of Figure 2a.

Table 4: Cluster Centroids

| | Cluster | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Feature** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **Age** | 57.2468 | 54.0667 | 52.3759 | 58.4528 | 60.1667 | 58.2000 | 55.0000 | 56.8776 | 59.1667 | 58.6944 |
| **Sex** | 0.8571 | 0.9333 | 0.8722 | 0.9245 | 0.9167 | 0.9200 | 0.9500 | 0.9184 | 0.9583 | 0.8889 |
| **RestingBP** | 134.6753 | 128.8667 | 133.8647 | 138.3585 | 129.0000 | 133.1600 | 133.8500 | 133.6939 | 140.4167 | 131.9167 |
| **Cholesterol** | 255.9874 | 238.3922 | 251.3479 | 242.9226 | 249.0278 | 254.2980 | 246.4200 | 245.4646 | 253.0542 | 235.3213 |
| **FastingBS** | 0.2338 | 0.4667 | 3.045e-16 | -1.522609e-16 | 0.5000 | 1.000 | 0.4000 | 0.2653 | 1.0000 | 0.3611 |
| **MaxHR** | 134.8701 | 141.3333 | 122.2180 | 122.2642 | 127.0000 | 121.4400 | 144.5000 | 134.1837 | 121.8958 | 127.1389 |
| **ExerciseAngina** | 0.5974 | 0.4333 | 0.7143 | 0.8679 | 0.8333 | 0.5600 | 0.1500 | 0.4490 | 0.5208 | 0.7778 |
| **Oldpeak** | 1.2299 | 0.8433 | 1.2549 | 1.4094 | 2.1833 | 1.0300 | 1.1000 | 1.0878 | 1.0104 | 2.3389 |
| **ChestPainType_ASY** | 0.9091 | 0.9667 | 0.9474 | 0.7547 | 0.8333 | 0.9600 | 1.5454e-16 | -1.5454e-16 | 0.7708 | 0.8889 |
| **ChestPainType_ATA** | 0.0909 | 0.0333 | 0.0526 | 0.03774 | 5.7041e-18 | 0.0400 | 0.0000 | 3.4225e-17 | 0.0833 | 0.0278 |
| **ChestPainType_NAP** | -9.8864e-17 | -4.9432e-17 | -7.4148e-17 | 0.2076 | 0.1667 | -7.4148e-17 | -2.4716e-17 | 1.0000 | 0.1458 | 0.0833 |
| **ChestPainType_TA** | -3.5296e-17 | 5.8827e-18 | -2.3531e-17 | -5.8827e-18 | 0.0000 | -1.7648e-17 | 1.0000 | 5.8827e-18 | 5.8827e-18 | 0.0000 |
| **RestingECG_LVH** | 1.0000 | 9.5141e-17 | -1.1893e-16 | 1.4271e-16 | 0.0000 | 1.1893e-16 | 0.2500 | 0.2449 | 1.4271e-16 | 0.3333 |
| **RestingECG_Normal** | -2.4454e-16 | 1.0000 | 1.0000 | 3.2605e-16 | 8.1514e-17 | 1.0000 | 0.5500 | 0.7551 | 1.6301e-16 | 0.6667 |
| **RestingECG_ST** | 0.0000 | 4.6288e-17 | 9.2576e-17 | 1.0000 | 1.0000 | -1.8515e-16 | 0.2000 | -9.2576e-17 | 1.0000 | -9.2576e-17 |
| **ST_Slope_Down** | 1.6523e-16 | 0.0000 | 1.6523e-16 | 0.000000e+00 | 1.0000 | 0.0000 | 0.0500 | 0.0000 | 1.3769e-17 | 1.0000 |
| **ST_Slope_Flat** | 0.7143 | 0.0000 | 1.0000 | 0.8113 | 0.0000 | 1.0000 | 0.7500 | 0.8776 | 0.8750 | 0.0000 |
| **ST_Slope_Up** | 0.2857 | 1.0000 | 0.0000 | 0.1887 | 0.0000 | 1.3791e-16 | 0.2000 | 0.1224 | 0.1250 | 1.0343e-16 |
| **No. of Patients** | 77 | 30 | 133 | 53 | 12 | 50 | 20 | 49 | 48 | 36 |

Table 5: Statistical Significance of Feature Impact Across Clusters

| | Cluster | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Feature** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **Age** | 1.784339e-20 | 5.176378e-24 | 1.050402e-30 | 3.165768e-24 | 1.111984e-27 | 8.815899e-25 | 6.212547e-36 | 3.773383e-28 | 9.563693e-24 | 3.910492e-26 |
| **Sex** | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **RestingBP** | 1.221084e-26 | 3.266029e-26 | 3.089457e-29 | 9.913707e-21 | 7.224249e-30 | 1.520888e-29 | 8.238147e-27 | 4.328722e-31 | 1.256784e-35 | 2.953715e-31 |
| **Cholesterol** | 9.992876e-30 | 2.013162e-18 | 2.192333e-25 | 8.839342e-29 | 5.073344e-30 | 8.222597e-21 | 1.041601e-27 | 4.314337e-23 | 9.368246e-28 | 1.634394e-24 |
| **FastingBS** | 2.056140e-07 | 3.648067e-01 | 6.119374e-218 | 4.962892e-86 | 4.959277e-01 | 1.881636e-81 | 1.744204e-01 | 2.248206e-04 | 3.179422e-75 | 4.803726e-02 |
| **MaxHR** | 7.217434e-29 | 4.188556e-35 | 1.095027e-28 | 1.149748e-29 | 6.405235e-23 | 6.457849e-39 | 2.952127e-29 | 3.607672e-28 | 1.139184e-34 | 2.785763e-33 |
| **ExerciseAngina** | 3.770294e-02 | 2.521199e-01 | 6.781612e-08 | 4.367523e-10 | 5.670481e-03 | 1.974922e-01 | 2.110576e-04 | 2.430420e-01 | 4.037579e-01 | 1.196730e-04 |
| **Oldpeak** | 6.218910e-26 | 3.164580e-20 | 2.699646e-26 | 8.016411e-28 | 1.192472e-39 | 1.900990e-25 | 2.354649e-29 | 2.120115e-25 | 1.260829e-23 | 5.432951e-32 |
| **ChestPainType_ASY** | 2.169778e-20 | 1.015570e-14 | 2.430519e-47 | 5.302092e-05 | 5.563999e-03 | 3.747514e-22 | 2.412855e-26 | 2.307849e-60 | 3.884564e-05 | 7.253679e-09 |
| **ChestPainType_ATA** | 4.557646e-20 | 8.052819e-15 | 4.446854e-48 | 1.467667e-23 | 9.016133e-23 | 6.502432e-22 | 8.765439e-28 | 1.435190e-66 | 4.574066e-14 | 5.707229e-19 |
| **ChestPainType_NAP** | 5.967558e-104 | 2.693517e-39 | 9.845650e-182 | 1.231368e-06 | 4.717041e-03 | 5.497802e-69 | 3.676480e-30 | 3.701042e-66 | 3.439817e-09 | 3.986025e-11 |
| **ChestPainType_TA** | 9.963864e-76 | 8.989114e-32 | 2.821890e-126 | 9.649986e-51 | 1.939466e-11 | 3.832829e-50 | 1.551301e-20 | 5.342433e-46 | 1.692628e-48 | 2.956311e-37 |
| **RestingECG_LVH** | 3.409057e-63 | 3.547298e-23 | 1.947605e-101 | 8.559162e-46 | 1.390350e-09 | 1.534574e-47 | 1.319598e-02 | 6.386633e-05 | 1.097565e-34 | 1.911974e-02 |
| **RestingECG_Normal** | 5.393960e-19 | 1.008798e-09 | 2.007298e-33 | 4.838461e-13 | 1.481524e-04 | 1.213668e-14 | 5.841670e-02 | 3.101253e-03 | 4.449664e-17 | 2.537983e-02 |
| **RestingECG_ST** | 1.807510e-65 | 2.952998e-24 | 3.993066e-103 | 5.875667e-41 | 1.052601e-10 | 1.468051e-41 | 6.365470e-04 | 1.198070e-41 | 1.630726e-41 | 2.964184e-31 |
| **ST_Slope_Down** | 5.527833e-75 | 2.321342e-25 | 4.119160e-128 | 1.192835e-50 | 1.020696e-12 | 3.843365e-49 | 3.454020e-08 | 1.507370e-49 | 6.854263e-52 | 1.023695e-36 |
| **ST_Slope_Flat** | 4.193851e-05 | 1.238418e-50 | 2.579762e-201 | 1.578560e-07 | 6.352636e-22 | 1.610716e-82 | 9.463875e-03 | 9.266382e-11 | 2.074121e-10 | 6.976832e-56 |
| **ST_Slope_Up** | 4.815477e-05 | 1.573282e-40 | 1.010266e-168 | 2.949047e-07 | 1.672172e-16 | 1.418837e-63 | 1.536529e-03 | 5.715186e-11 | 1.910088e-10 | 1.922593e-46 |





doi:10.20944/preprints202207.0097.v1