Cardiovascular Disease Diagnosis using Deep Neural Networks

Jai Sharma, Milind Maiti, Christopher Sun

July 2021

Contents

1 Abstract

Motivation Cardiovascular disease causes 25% of deaths in America (Heart Disease Facts). Specifically, misdiagnosis of cardiovascular disease results in 11,000 American deaths annually, emphasizing the increasing need for Artificial Intelligence to improve diagnosis (Sharkey 2019).

Objective The goal of our research was to determine the probability that a given patient has Cardiovascular Disease using 11 easily-accessible objective, examination, and subjective features from a data set of 70,000 people (Ulianova 2019). To do this, we compared various Machine Learning and Deep Learning models.

Methods Exploratory Data Analysis (EDA) identified that blood pressure, cholesterol, and age were most correlated with an elevated risk of contracting heart disease. Principal Component Analysis (PCA) was employed to visualize the 11-D data onto a 2-D plane, and distinct aggregations in the data motivated the inference of specific cardiovascular conditions beyond the binary labels in the data set.

To diagnose patients, several Machine Learning and Deep Learning models were trained using the data and compared using the metrics Binary Accuracy and F1 Score. The initial Deep Learning model was a Shallow Neural Network with 1 hidden layer consisting of 8 hidden units. Further improvements, such as adding 5 hidden layers with 8 hidden units each and employing Mini-Batch Gradient Descent, Adam Optimization, and He's Initialization, were successful in decreasing train times. These models were coded without the utilization of Deep Learning Frameworks such as TensorFlow. The final model, which achieved a Binary Accuracy of 74.2% and an F1 Score of 0.73, consisted of 6 hidden layers, each with 128 hidden units, and was built using the highly optimized Keras library (Keras Team).

Discussion and Conclusion While current industrial models require hundreds of comprehensive features, this final model requires only basic inputs, allowing versatile applications in rural locations and third-world countries. Furthermore, the model can forecast demand for medical equipment, improve diagnosis procedures, and provide detailed personalized health statistics.

2 Introduction

2.1 Motivations

Cardiovascular disease causes 1 out of every 4 deaths in America (Heart Disease Facts). The World Health Organization (WHO) 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 (Cardiovascular Diseases). 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 modern medical facilities to conduct the 8 to 10 tests required to diagnose heart disease, misdiagnosis of heart disease still causes more than 11,000 deaths each year in America (Sharkey 2019). Worse, in rural areas, because of the lack of rural medicine practitioners, patients at risk for heart disease cannot receive diagnoses at all, let alone diagnoses that are accurate and early in the prognosis of the disease. In addition, diagnoses from doctors are subject to the risks of subconscious biases, which can hamper the accuracy of the diagnosis and delay appropriate treatment.

Current Machine Learning and Deep Learning models still come short in one key area: Versatility. Industrial models and models with high accuracy conceived by researchers require a multitude of comprehensive features such as the result of a thallium stress test and the slopes of specific segments of a patient's electrocardiogram (Janosi, et al.). Furthermore, features such as blood glucose and serum cholesterol are provided on a continuous rather than discrete scale. Advanced features such as these are simply unattainable in developing countries due to the lack of expensive technology.

2.1.1 Current Methods of Heart Disease Diagnosis

The cardiovascular disease diagnosis process is quite a hassle for people who suspect that they have contracted heart disease. An accurate diagnosis requires around 10 clinical procedures, forcing potential patients to make several visits to hospitals (Whitworth 2018). More often than not, the diagnosis process can delay vital treatment for affected patients,

generating greater complications that could have been prevented earlier.

The following table lists common diagnosis procedures that physicians currently conduct (Whitworth 2018):

Procedure Name	Description
Electrocardiogram	Electrical signals from the heart are projected on a scale for physicians to identify irregular rhythms.
Cardiac MRI	Magnetic Resonance Imaging of the heart generates detailed views that can be examined for diagnosis.
Homocysteine	Quantity of the amino acid homocysteine is measured.
Lipid profile	Low and high density lipoprotein, cholesterol, and triglyceride levels in the bloodstream are measured.

Figure 1: The above table lists several common procedures for heart disease diagnosis.

2.2 Ongoing Research for Automated Diagnosis

Currently, most research regarding diagnosis of cardiovascular disease utilizing Artificial Intelligence models involves the use of a single cardiovascular disease data set from the UCI Machine Learning Repository called the Cleveland Database (Janosi, et al.). This data set contains 303 patients with 75 features[1]. Ongoing research studies involve training various Machine Learning models on this data set, achieving accuracies ranging around 85 to 90% (Krittanawong 2020). 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 an interesting new development which 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.

2.3 Project Goal

2.3.1 Data Set

Selecting a data set was a challenge because of the lack of open-source cardiovascular disease data sets. We decided to use a data set from Kaggle that contained 70,000 examples of patients. The data set contained 11 features (Ulianova 2019):

1. Age | integer (days)
2. Height | integer (cm)
3. Weight | float (kg)
4. Gender | 1 = Female, 2 = Male
5. Systolic blood pressure | integer (mmHg)
6. Diastolic blood pressure | integer (mmHg)
7. Cholesterol | 1 = normal, 2 = above normal, 3 = well above normal
8. Glucose | 1 = normal, 2 = above normal, 3 = well above normal
9. Smoking | 0 = non-smoker, 1 = smoker
10. Alcohol intake | 0 = no consumption of alcohol, 1 = consumption of alcohol
11. Physical activity | 0 = no regular physical activity, 1 = regular physical activity
12. Target Variable: Presence or absence of cardiovascular disease | 0 = healthy, 1 = unhealthy

[1] Not all features are provided for each patient

Determining Project Goals After having an expert manually classify a random sample of 100 patients in the data set, they achieved a 77% accuracy and 0.77 F1 Score; we regarded this classification performance as a proxy for Bayes Accuracy, or the best theoretical possible accuracy. We then decided to set a goal of 70% to 75% Binary Accuracy and 0.70 to 0.75 F1 Score for our predictive models.

Project Goal The goal of our project was to predict heart disease with a 70% to 75% Binary Accuracy and 0.70 to 0.75 F1 Score on the validation set using Deep Neural Networks with 11 experimental, objective, and subjective features, thereby improving accessibility to diagnosis.

3 Methods

3.1 Exploratory Data Analysis

3.1.1 Removing Outliers

After importing the data set, we removed outlier data points that may have been caused by human error when creating the data set. As a systematic way to remove these outliers, we used the standard IQR Rule. We first calculate the 25th percentile for some feature. Let this be $q1$ for "quartile 1." Then, we calculate the 75th percentile for this feature. Let this be $q3$ for "quartile 3."

Using these, we define the iqr, or the "interquartile range," to be

$$iqr = q3 - q1$$

Then, we define an *upper* threshold and a *lower* threshold which we use to determine which values should be removed. We define the *upper* threshold to be

$$upper = q3 + 2 \times iqr$$

Similarly, we define the *lower* threshold to be

$$lower = q1 - 2 \times iqr$$

Any value greater than *upper* or less than *lower* is considered an outlier, and these examples are removed from the data set.

While the coefficient 1.5 instead of 2 is commonly used in statistics, we found that the value of 2 was more realistic to our specific data set in removing examples that we considered outliers.

3.1.2 Effect of Each Feature on Risk of Cardiovascular Disease

Procedure of Analysis for Discrete Features

We followed a common procedure to analyze the effect of each discrete feature of the data set on the increased risk of contracting cardiovascular disease. For example, we analyzed the discrete feature of smoking using the following procedure:

1. Let x be defined as:
$$x = \frac{\text{\# of smokers who have cardiovascular disease}}{\text{total \# of smokers}}.$$
2. Let y be defined as:
$$y = \frac{\text{\# of non-smokers who have cardiovascular disease}}{\text{total \# of non-smokers}}.$$
3. Notice that x is the experimental probability that a smoker has heart disease, and that y is the experimental probability that a non-smoker has heart disease. The value $x - y$ can then be computed to find the effect of smoking on one's chances of contracting heart disease.

We similarly computed this increased probability for other discrete features such as Gender, Physical Activity, and Cholesterol Levels.

Procedure of Analysis of Continuous Features

When analyzing discrete features, patients could be objectively divided into separate classes. However, no such guidelines exist for continuous features such as Age, Height, and Blood Pressure.

To divide patients based on age, we considered the median of all of the ages in the data set. If a given patient was older than or the same age as the median, then they were considered "old." Conversely, if a given patient was younger than the median, then they were considered "young". This allowed us to treat the continuous features as discrete features.

We repeated this procedure for the other continuous features, including Height, Weight, Systolic Blood Pressure, and Diastolic Blood Pressure.

3.2 Principal Component Analysis (PCA) and K-Means Clustering

3.2.1 PCA

To better understand the data set visually, we employed Principal Component Analysis (PCA) to visualize the 11-dimensional data on a 2-dimensional plane. PCA works by finding the optimal 2-dimensional plane onto which the data will be projected such that the variance retained is maximized.

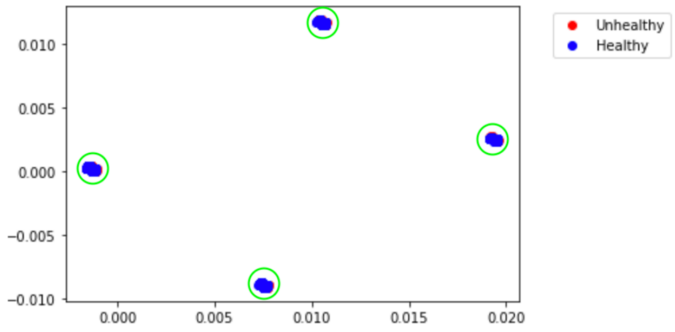


Figure 2: The above graph is a visualization of our Data Set with 71.51% variance retained.

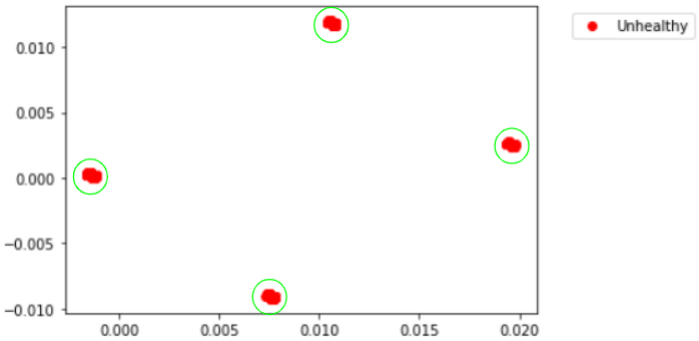


Figure 3: The above graph is a visualization of only the unhealthy patients using PCA.

3.2.2 Clustering

Notice how the data seems to be grouped into distinct clumps after PCA. This fact motivated the use of K-Means Clustering to group the data into 8 clusters. This specific number of clusters was chosen using the Elbow Method, as shown below.

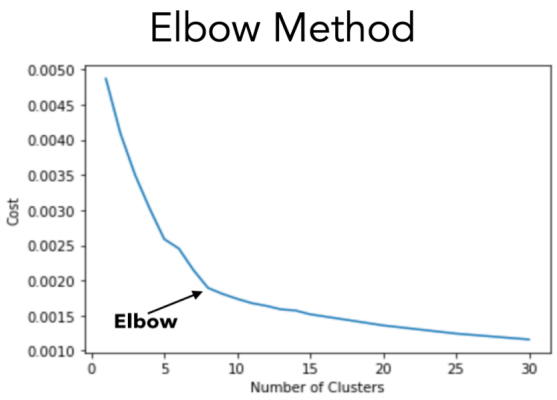


Figure 4: We used the Elbow Method to find the optimal number of clusters, as shown above.

Even though all patients suffered from cardiovascular disease, we hypothesized that patients in each cluster suffered from distinct cardiovascular conditions. By analyzing the data from each centroid, which represents the average medical characteristics of all patients assigned to a given cluster, we inferred the specific diseases that the patients suffered from.

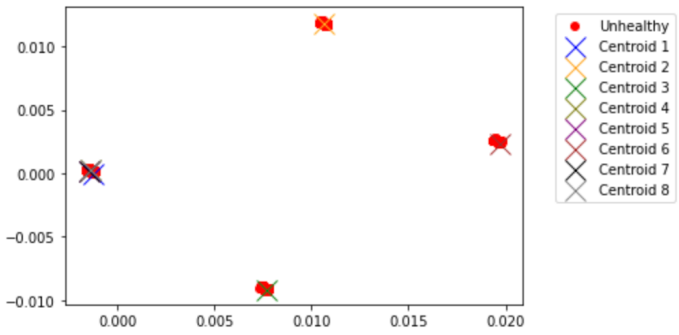


Figure 5: The above image is a clustering of the Unhealthy patients into 8 clusters.

3.3 Machine Learning Models

The final goal of our project was to create Deep Neural Network models to diagnose cardiovascular disease. We first trained several Machine Learning models to act as Standards of Comparison for our future Deep Learning models. The following is a list of the Machine Learning models used:

- Logistic Regression
- Random Forest
- Extreme Gradient Boosting
- Support Vector Machine

We used the metrics Binary Accuracy and F1 Score to evaluate each of these models, as mentioned previously.

The Support Vector Machine yielded the greatest performance[2], achieving a Binary Accuracy of 73.3% and a F1 Score of 0.712. As such, we chose this model to serve as the Standard of Comparison for our later Deep Learning Models.

[2]See the results of the other Machine Learning models in the Results section.

3.4 Deep Learning Models

3.4.1 Shallow Neural Network

The initial model, a Shallow Neural Network trained using Batch Gradient Descent, consisted of 1 hidden layer with 8 hidden units. This model achieved a Binary Accuracy of **64.1%** and an F1 Score of **0.621**.

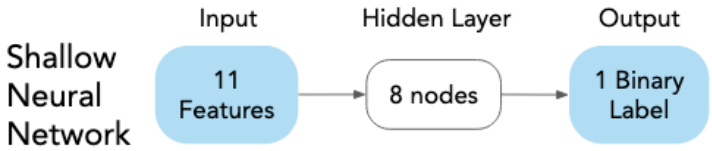


Figure 6: The above image is a flowchart of the Shallow Neural Network architecture.

3.4.2 Deep Neural Network

We added 5 additional hidden layers to the initial model, each layer consisting of 8 hidden units. This model achieved a Binary Accuracy of **69.7%** and a F1 Score of **0.719**.

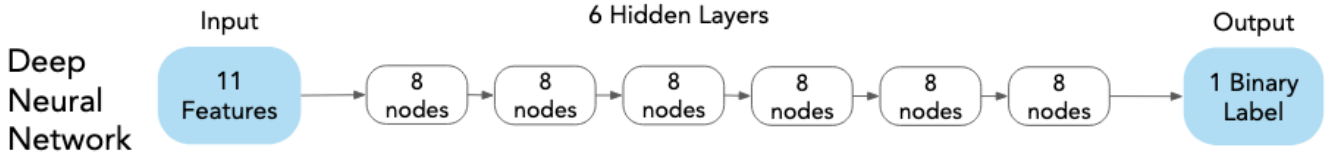


Figure 7: The above image is a flowchart of the Deep Neural Network architecture.

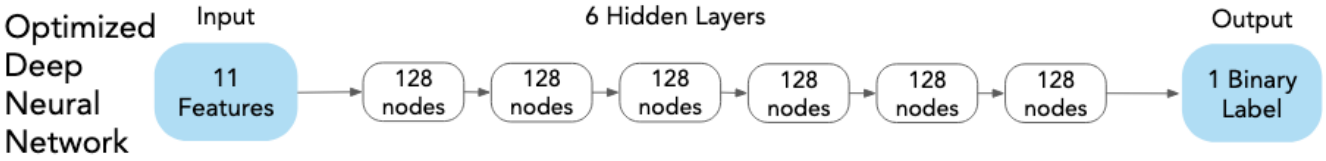
3.4.3 Deep Neural Network with Advanced Techniques

After noticing the abysmal train times and non-convergent behavior of the Shallow Neural Network and the Deep Neural Network models[3], we implemented Adam optimization, which significantly improved train times (Kingma, et al. 2015). Additionally, He's initialization was utilized to mitigate the problems of Vanishing/Exploding Gradients (Ng). As a result, the Deep Neural Network with Advanced Techniques[4] was able to converge to a relatively low cost in a relatively few number of iterations of training. This model achieved a Binary Accuracy of **72.8%** and an F1 Score of **0.731**.

All of the models listed above were coded without the utilization of Deep Learning Frameworks such as TensorFlow or Keras.

3.4.4 Keras Implementation (Final Model)

The final model utilized the TensorFlow Framework to implement a Neural Network that consisted of 6 hidden layers each with 128 hidden units. This model applied techniques including Adam Optimization, Xavier's Initialization, Batch Normalization, Early Stopping, and Dropout Regularization (Ng). This model achieved a Binary Accuracy of **74.2%** and an F1 Score of **0.726**.



[3]Both models were trained for 100,000 iterations without significant decrease in cost. See detailed graphs in the Results section.
[4]The Deep Neural Network with Advanced Techniques used the same architecture as the Deep Neural Network in the previous section.

Figure 8: The above image is a flowchart of the Keras Implementation (final model).

4 Results

4.1 Exploratory Data Analysis

Through Exploratory Data Analysis, we determined the effect of each of the features on heart disease. The following bar graph summarizes the results:

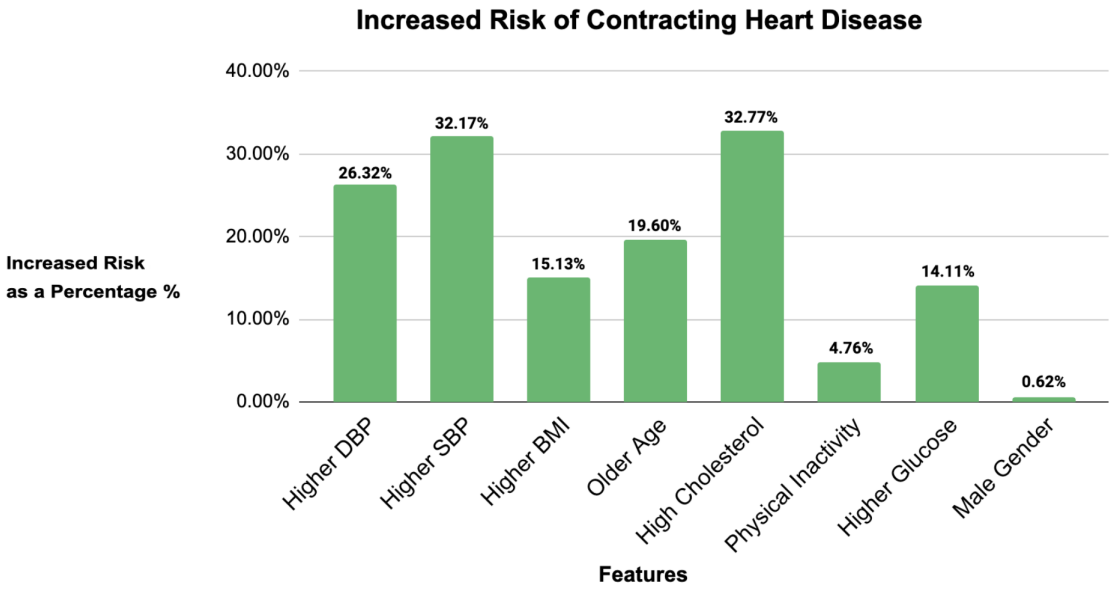


Figure 9: The above bar graph lists the increased chance of cardiovascular disease for each feature.

In conclusion, Cholesterol, Blood Pressure, and Age most significantly affect risk for heart disease.

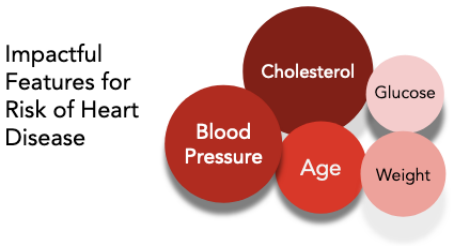


Figure 10: The above visual displays the most significant features for heart disease diagnosis.

4.2 Clustering

Centroids The following table exhibits the centroids for each of the data's 8 clusters:

Cluster	# of Patients	Age	Gender	Height	Weight	SBP	DBP	Chol.	Glucose	Smoking	Alcohol	Physical Activity	Inference
1	898	54.5	2.00	169.8	78.0	133.5	84.5	1.13	1.04	0.00	0.00	1.00	Early stage of heart disease
2	142	54.7	1.49	165.7	81.1	136.5	86.5	1.74	1.37	0.00	1.00	0.84	Cirrhosis
3	134	53.9	1.85	169.3	79.5	135.5	85.2	1.56	1.27	1.00	0.00	0.80	Atherosclerosis
4	16215	56.2	1.26	163.2	78.3	134.7	84.7	2.44	2.70	0.00	0.00	0.84	Diabetic Dyslipidemia
5	157	54.3	1.00	161.6	73.2	131.8	83.5	1.00	1.05	0.00	0.00	1.00	Early stage of heart disease
6	12703	53.7	1.92	169.9	80.3	138.0	86.8	1.58	1.24	1.00	1.00	0.81	Atherosclerosis
7	32	54.8	1.32	164.4	75.4	131.6	83.5	1.19	1.09	0.00	0.00	0.00	Edema
8	1891	55.8	1.15	161.6	76.9	135.5	85.1	2.59	1.00	0.00	0.00	0.91	Hypercholesterolemia

Figure 11: The above table lists the centroid data for each of the 8 clusters.

4.3 Machine Learning Models

The following bar graph summarizes the performance of the different Machine Learning predictive models using the metrics Binary Accuracy and F1 Score[5]:

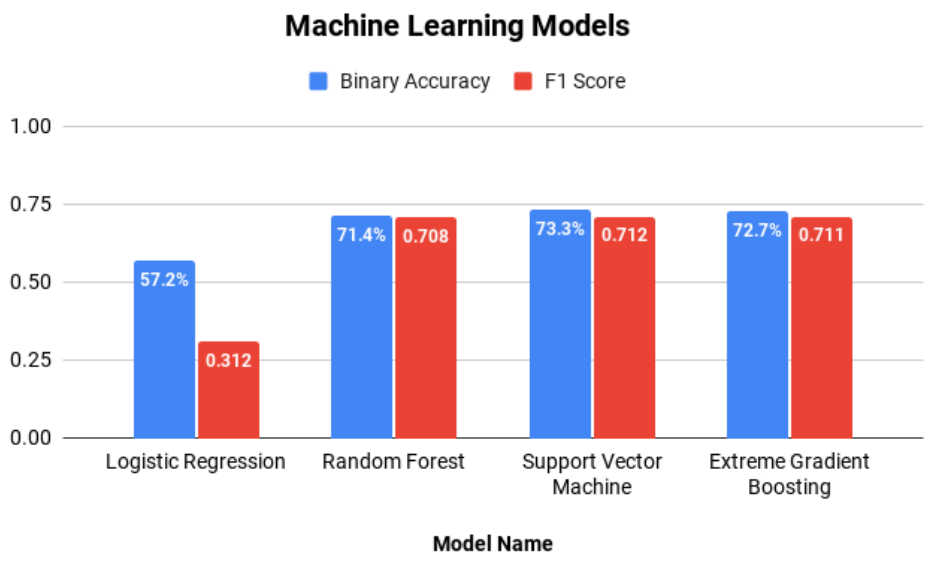


Figure 12: The above table summarizes the performance of each of the Machine Learning models using the metrics Binary Accuracy and F1 Score.

The best Machine Learning model was the Support Vector Machine. Therefore, we used this model as a Standard of Comparison for our later Deep Learning models.

[5]As mentioned before, the Machine Learning models used were Logistic Regression, Random Forest, Extreme Gradient Boosting, and Support Vector Machine. We used these Machine Learning models as Standards of Comparison for our later Deep Learning models.

4.4 Deep Learning Models

We trained several generations of predictive Deep Learning models. The following graphs plot the Binary Cross-Entropy Cost (y axis) against the Number of Iterations Trained (x axis) for each of the Deep Learning models:

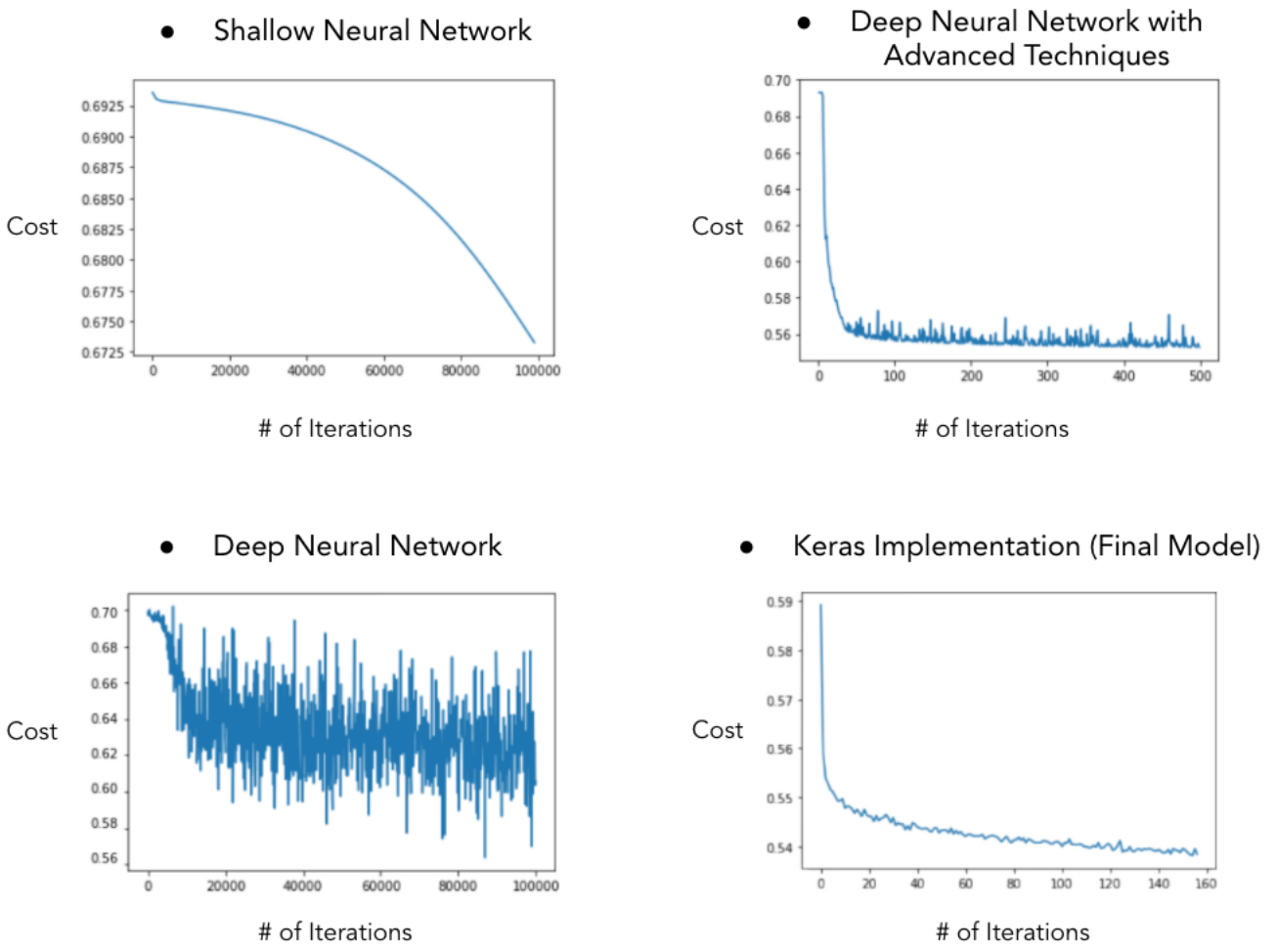


Figure 13: The above images exhibit the decrease in cost over iterations trained for all of the Deep Learning models.

Notice how the first two Deep Learning models do not converge to a relatively low cost even after $100,000$ iterations of training. Given this, we implemented advanced techniques such as Adam Optimization, He's Initialization, and more in the final two models. Therefore, the final two models converged to a relatively low cost in only a few hundred iterations.

The following bar graph summarizes the performance of the different Deep Learning predictive models:

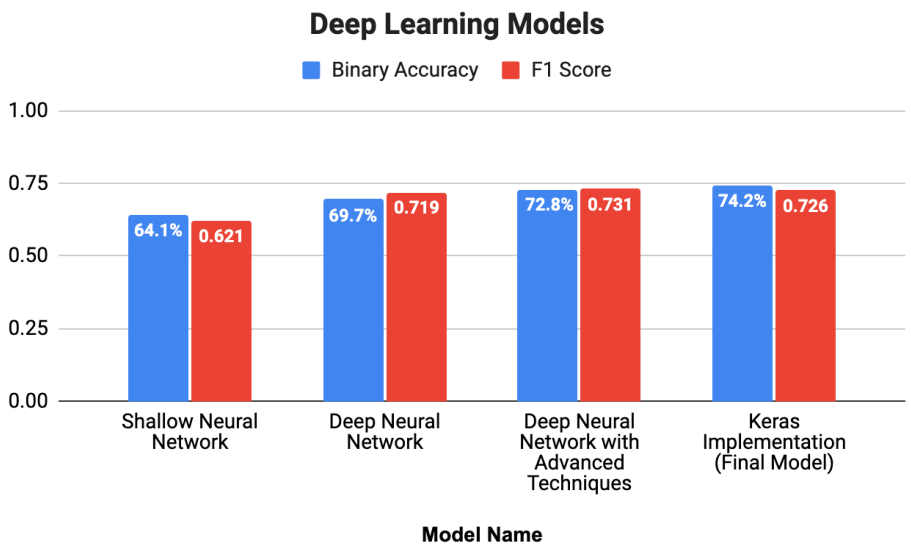


Figure 14: The above table summarizes the performance of each of the Deep Learning models using the metrics Binary Accuracy and F1 Score.

Both the Deep Neural Network with Advanced Techniques and the Keras Implementation performed better than the Standard of Comparison, which was the Support Vector Machine.

Additionally, both of these models also fulfilled our final goal for this project, which was to achieve a Binary Accuracy of 70% to 75% and a F1 Score of 0.70 to 0.75.

5 Discussion

5.1 Clustering

Cluster 1 Cluster 1 comprises male patients who have an average age of 54.5 years. The calculated BMI of these patients is 27.1, indicating classification in the overweight category. The patients' average blood pressure (BP) of 134.5/84.5 mmHg is elevated, however not yet at hypertensive levels. Apart from these mild health troubles, all patients in Cluster 1 engage in physical activity, abstain from smoking and alcohol, and have normal cholesterol and glucose levels. Therefore, we infer that Cluster 1 represents patients who suffer from an **early stage of heart disease**.

Cluster 2 Cluster 2 comprises male and female patients who have an average age of 54.7 years. Patients have a BMI of 29.5 and BP of 136.5/86.5 mmHg, meaning that they are close to obesity and hypertension. The defining factor of Cluster 2 is that all patients consume alcohol. Given this fact, along with their moderately elevated cholesterol levels, we infer that Cluster 2 represents patients who suffer from **preexisting liver conditions such as cirrhosis, which are heavily correlated with heart failure**.

Cluster 3 Cluster 3 comprises mostly male patients with an average age of 53.9 years. These patients have an elevated BP of 135.5/85.2 mmHg and an elevated BMI of 27.7. Cholesterol and glucose levels are only slighted elevated, and are therefore not extremely influential. Cluster 3 is distinctly characterized by the fact that all represented patients smoke, but do not consume alcohol. We hypothesize that this habitual smoking exacerbates the progression of **atherosclerosis as well as pulmonary disease**.

Cluster 4 Cluster 4 comprises mostly female patients who are 56.2 years old on average. The most striking quality of this cluster is the overwhelmingly high cholesterol and glucose levels that all patients possess. This fact is a convincing indicator of atherosclerosis. Considering the patients' BMI of 29.4 and their slightly elevated BP of 134.7/84.7 mmHg, it becomes increasingly clear that C**luster 4 represents patients who suffer from diabetic dyslipidemia that leads to cardiovascular complications such as coronary artery disease**. This inference is highly plausible because a sizeable 16,215 patients out of 35,000 are grouped into Cluster 4, and diabetes is a frequent precursor to cardiovascular disease.

Cluster 5 Cluster 5 comprises female patients who have an average age of 54.3 years. Cluster 5 displays interesting parallels with Cluster 1, with opposite gender and only mild variations in blood pressure and BMI setting these two classes of patients apart. Patients have an overweight 28.0 BMI and an elevated BP of 131.8/83.5 mmHg. Because patients abstain from smoking and alcohol, and engage in physical activity, we hypothesize that **Cluster 5 represents patients who have a very early stage of cardiovascular disease**.

Cluster 6 Cluster 6 comprises mostly male patients who have an average age of 53.7 years. Their BMI of 27.8 and BP of 138/86.8 mmHg classifies them as overweight and nearly hypertensive. Having moderate cholesterol levels and slightly above normal glucose levels, these patients are not extremely unhealthy, however their habitual smoking and alcoholism accounts for their heart disease. Given that all patients in Cluster 6 smoke and consume alcohol, we hypothesize that **Cluster 6 represents patients who suffer from prominent atherosclerosis which aggravates their cardiovascular disease**.

Cluster 7 Cluster 7 comprises mostly female patients who have characteristics similar to patients in Cluster 1. These patients are 55.8 years old on average, have a 27.9 BMI, have an elevated BP of 131.6/83.5 mmHg, and have normal glucose and cholesterol levels. However, no patients in Cluster 7 engage in physical activity, which indicates the likely possibility of **fatigue or edema in the lower extremities**. These symptoms are correlated with **early stages of heart failure**, which is our inference for the disease represented by Cluster 7.

Cluster 8 Cluster 8 comprises mostly female patients who are aged 55.8 years on average. With a BMI of 29.4 and elevated BP of 135.5/85.1 mmHg, these patients are similar to patients in other clusters, but one facet of their physical characteristics clearly differs: abnormally high cholesterol levels. The juxtaposition between patients' astoundingly elevated cholesterol levels and completely healthy glucose levels reveals the **presence of hypercholesterolemia**. The inferred excess of low-density lipoproteins is a destructive contributor to Cluster 8 patients' cardiovascular disease.

5.2 Machine Learning and Deep Learning

Our final Deep Learning model (Keras Implementation) performed better than the Machine Learning models by 1% in binary accuracy and 0.01 in F1 score, which met our project goal.

Final Model

	Our Goal	Final Model
Binary Accuracy	70% to 75%	74.20%
F1 Score	0.70 to 0.75	0.73

Figure 15: The above image compares our project goal to our actual model performance.

However, this is not as much of an improvement of the Machine Learning models as we had hoped due to the relatively small data set, lack of computation power, and simple model that we could train and access.

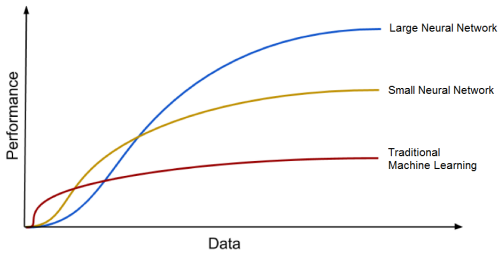


Figure 16: The above image demonstrates the performance of Machine Learning models compared to Deep Learning models based on the amount of data. (Oppermann 2019).

Since our data set comprised only 70,000 patients, we believe our Neural Network did not have enough data to achieve a more significant improvement over a relatively simpler Machine Learning model such as Random Forest.

5.2.1 Analysis of Performance

Comparison to Industrial Models Our best Deep Learning model achieved a Binary Accuracy of 74.2%, compared to the 91.1% achieved by industrial researchers using ML and DL models on the Cleveland Heart Disease data set[6](Liaqat, et al.). We analyze the possible reasons for our model's inferiority in this section.

Size and Features While our model's accuracy is inferior to aforementioned industrial model accuracy, it is important to remember that our data set comprises 70,000 patients, compared to 300 patients provided in the Cleveland heart data set (Janosi, et al.). In addition, our data set provides only 11 simplistic features compared to the 75 more comprehensive features provided in the Cleveland Heart Disease data set, which makes it much more challenging for our models to achieve a similar accuracy (Janosi, et al.).

Versatility and Applications Despite our inferior accuracy, we believe that our model can be used in more versatile applications in cardiovascular disease diagnosis due to the simplistic features it takes in as inputs. For example, the model can be used in rural areas or third-world countries, locations that may not have the assets for immaculate facilities and comprehensive equipment.

Validity Another advantage of our Deep Learning model is that it is trained on only real patients, as opposed to multiple scientific studies that have been published using synthesized examples generated based on real patient data (Muhammad, et al. 2020). Although this process of data generation is promising for other fields of research, our sole use of data from real patients allows the confirmation of our data set's validity.

6 Applications and Recommendations for Future Research

This project has numerous applications which can also serve as recommendations for future research:

Improving Patient Diagnosis The most obvious yet the most important application of this project is improving patient diagnosis. In rural areas and third-world countries, easily accessible features can be collected and used to diagnose high risk patients for heart disease. These patients can then be given the appropriate medical treatment, which could potentially save or prolong their life.

[6]This data set is publicly available from the UCI Machine Learning Repository.

Personalized Statistics We could also provide people with personalized statistics to make them more aware about how they treat their bodies.

"If you lose 8 lbs, you could lower your risk for Cardiovascular Disease by 15%."

Figure 17: The above image is an example of a personalized statistic.

Life and Health Insurance Applications of our work even extend to life and health insurance. Our predictive models can be used to evaluate the risk for heart disease for a given person, after which the person can be charged a personalized life and health insurance rate.

Demand Forecasting Our predictive models can also be used to forecast the demand for medical equipment in hospitals by predicting how many patients are likely to contract heart disease. In fact, by taking the medical features of hospital patients as inputs, our model allows for the effective allocation of hospital resources.

Long Term Health Monitoring Rather than just being used once for diagnosis, the predictive models can be used over a long period of time to monitor the cardiovascular health of a given patient. This tool can be used by doctors and even patients themselves to track health.

Tangible Product Finally, we can create a tangible product such as a mobile app to allow patients to diagnose themselves with cardiovascular disease from the comfort of their own homes.

7 References

Beckerman, James. Heart Failure: Causes, Symptoms, Types, and Stages. 24 Aug. 2020, Heart Disease: Types, Causes, and Treatments.,
www.medicalnewstoday.com/articles/237191.

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

"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.

"Heart Failure: Causes, Symptoms, Types, and Stages." WebMD, WebMD, 12 Nov. 2020,
www.webmd.com/heart-disease/heart-failure/heart-failure-overview.

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

Keras Team. "Keras Documentation: The Sequential Model." Keras,
keras.io/guides/sequential_model/.

Kingma*, Diederik P., and Jimmy Lei Ba. "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION." 2015, pp. 1–15., doi:
https://arxiv.org/pdf/1412.6980.pdf.

Krittanawong, Chayakrit., Virk, Hafeez Ul Hassan., et al., "Machine learning prediction in cardiovascular diseases: a meta-analysis," 16 Oct. 2020,
https://doi.org/10.1038/s41598-020-72685-1

"Machine Learning." Coursera,

www.coursera.org/learn/machine-learning.

Muhammad, Y., Tahir, M., Hayat, M. et al. "Early and accurate detection and diagnosis of heart disease using intelligent computational model." 12 November 2020. https://doi.org/10.1038/s41598-020-76635-9

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

Oppermann, Artem. "Deep Learning Models Scale Better with a Larger Amount of Data." Towards Data Science, Medium, 29 Oct. 2019, towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac.

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

Ulianova, Svetlana. Cardiovascular Disease Dataset. 20 Jan. 2019, www.kaggle.com/sulianova/cardiovascular-disease-dataset.

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

Acknowledgements We would like to acknowledge Dr. Shiqi Zhang from Binghamton University and give many thanks for his immense favors in the publication of this paper.

A Model Metrics and Cost Function

The Machine Learning and Deep Learning models were compared primarily using two methods: F1 Score and Binary Accuracy.

Binary Accuracy 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. The Binary Accuracy is then defined as

$$\boxed{\text{Binary Accuracy} = \left(100 \cdot \frac{x}{m}\right)\%}.$$

Binary Accuracy ranges from 0% to 100%, where 100% corresponds to perfect model performance.

F1 Score Let tp be the number of true positives, or the number of times that a positive example is correctly classified (positive examples are heart disease patients).

Let rp be the number of total cases in which the person has heart disease, regardless of the prediction (real positives).

Let pp be the total number of predicted positives, regardless of the actual label.

Then define

$$\text{precision} = \frac{tp}{pp}$$

and

$$\text{recall} = \frac{tp}{rp}.$$

Finally, the F1 Score of the model is defined as the harmonic mean of precision and recall, or:

$$\boxed{\text{F1 score} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}}.$$

By this definition, the F1 Score ranges from 0 to 1, where 1 corresponds to perfect model performance.

Cross-Entropy Cost Function When training the Deep Learning models, we used the Cross-Entropy Cost Function, defined as follows:

$$J(\theta) = \frac{1}{m} * \sum_{i=1}^{m} \left[y^{(i)} \log\left(h(x^{(i)})\right) + (1 - y^{(i)}) \log\left(1 - h(x^{(i)})\right) \right],$$

where θ contains the parameters for the Neural Network, m is the number of examples in the data set, $y^{(i)}$ is the ground truth label for example i, $x^{(i)}$ is the input for example i, $h(x^{(i)})$ is the prediction for example i, and log is the natural logarithm with base $e \approx 2.71828\ldots$