

# Probabilities of Heart Attack Based on Risk Factors

## CS4442 Artificial Intelligence II

Thilakshanan Aran

*Department of Electrical and Computer  
Engineering*

*Western University*

[taran@uwo.ca](mailto:taran@uwo.ca)

Nicholas Moniz

*Department of Electrical and Computer  
Engineering*

*Western University*

[nmoniz5@uwo.ca](mailto:nmoniz5@uwo.ca)

Resche Balaganesh

*Department of Electrical and Computer  
Engineering*

*Western University*

[rbalagan@uwo.ca](mailto:rbalagan@uwo.ca)

Ayush Sharma

*Department of Electrical and Computer  
Engineering*

*Western University*

[ashar463@uwo.ca](mailto:ashar463@uwo.ca)

**Abstract** - Heart attacks remain to be one of the leading causes of death worldwide. In this study, we decided to build a reliable early detection system based on machine learning techniques. We first began by cleaning the dataset using label conversion and SMOTE. This was to ensure it improves the performance of the model and no bias toward a particular outcome. We then compared four models consisting of Logistic Regression, Random Forest, XGBoost, and Support Vector Machine (SVM) and were able to reach about 98.5% accuracy. Our goal is to enhance our dataset by getting close to real time data from wearable devices. We aim to improve our results with this new dataset.

**Index Terms**—heart attack prediction, machine learning, ensemble methods, data balancing, clinical diagnostics

## I. INTRODUCTION

There are many different factors that can cause a heart attack, making it extremely difficult to predict. This project aims to make a more reliable system to predict a person's chances to have a heart attack. Our data set includes many different features including age, blood sugar, heart rate, and many more [1]. This project will use logistic regression, random forest, XGBoost, and SVM to predict if a person is at high risk of having a heart attack.

## II. BACKGROUND & RELATED WORK

Many studies have tried to use machine learning algorithms to try and predict health conditions. This section will mainly highlight all the previous work/similar models found, literature analysis, and research gaps.

The use of AI in the field of healthcare was first approved by the FDA in 2017, this marks the beginning of many AI uses in healthcare [2]. Machine learning was first originally used to read data but has evolved into diagnosing diseases. Since then many modern hospitals have used and proven the capabilities of machine learning.

Previous predictive models for heart disease go over machine learning applications such as Logistic Regression, Naïve Bayes, and Random Forest [3]. These studies relied mostly on mining techniques to get the predictive parts from data. These studies provided a basis for researchers to then expand and build on the advanced methods.

The research mainly aims to improve the accuracy and quality of heart disease prediction models [3]. These models vary from traditional statistical models, Naive Bayes, CNNs, and many more. The earlier studies mainly depended on data mining techniques by getting predictive features based on any clinical data that was available. Logistic Regression is one of the earlier data mining techniques. It was used to model the probability of a binary outcome while taking into account various related factors. These factors included age, blood pressure, cholesterol levels, and output probability. Other models like Naive Bayes, decision trees, and neural networks were discussed.

Researchers in this field have also tried hybrid models such as using a convolutional neural network and a long short-term memory network together to build a hybrid model and many more[2]. These were all beaten by extreme gradient boost and random forest with an accuracy of 95.08% and 96.72% respectively. Though these models are very accurate they are intended to be used to help doctors make decisions not to replace their service.

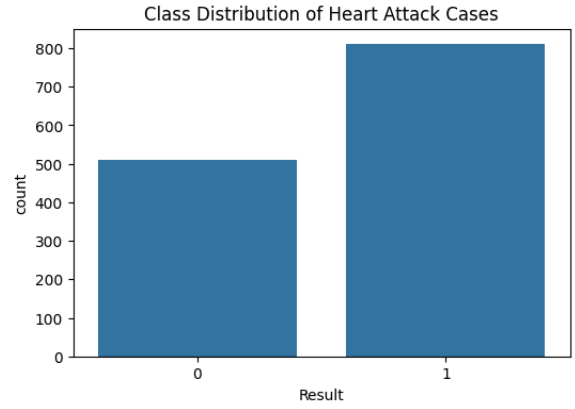
A massive issue with these models is population bias. Despite the models having high accuracy the data used to train the model itself may have bias. “Historically women and minority groups are underrepresented in cardiology”, thus the model’s predictions may not be accurate for these groups [4]. It has been found that similar commercial prediction models have shown large racial bias, the models were designed to “optimize hospital resources” which resulted in identifying postal codes as the best indicator of longer hospital stays. This resulted in the model giving higher scores than black patients as postal codes with lower incomes were predominantly black neighbourhoods. It is essential that the data set used to train the model does not contain population bias, to ensure the accuracy of the model.

### III. METHODOLOGIES

This section details the methods the team adopted to build, train, and test various machine learning models for predicting heart attacks. The first step was applying Data Preprocessing methods which include handling categorical data, checking for missing values, fixing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE) and splitting data into training and test sets [5]. Upon completion of data preprocessing the team will perform feature analysis and exploratory data analysis (EDA) to help understand relations and determine which features should be used to build models. The next step will be identifying models that can be used for predicting heart attacks as well as using hyperparameter tuning to help improve performance of models that may be under performing. Lastly, once all models made predictions on the test set the team will evaluate the performance of each model using different metrics.

#### A. Data Preprocessing and Exploratory Data Analysis

Before we could start building and training some models, it was important that we spend some time cleaning and exploring the dataset to ensure our dataset was machine-learning friendly. The first involved checking for any missing or null values and we then determined that the data was complete across all columns. One thing we did notice was that the result column used the words “positive” and “negative” to label heart attack cases. That kind of format does not play well with most machine learning algorithms, so we went ahead and converted those labels to binary values, with 1 representing positive and 0 representing negative. Once all the data has been cleaned we plotted a class distribution graph as shown in **Figure 1**. From this diagram we determined that the dataset exhibited a class imbalance, with approximately 61% of samples labeled as positive (heart attack) and 39% as negative.



**Figure 1.** Heart Attack Class Distribution

In order to combat this class imbalance and limit potential bias towards positive cases during model training, we oversampled the minority class (negative) of the training data using **SMOTE** (Synthetic Minority Over-sampling Technique). This remedied the imbalance for the minority class and achieved a balanced 50/50 class distribution, as seen in Figure 2 below.



**Figure 2.** Class Distribution after SMOTE

The dataset was then split into an 80% training set and a 20%. To help explore potential linear relationships between features and our target variable, we generated a correlation heatmap as seen in **Figure 3**.

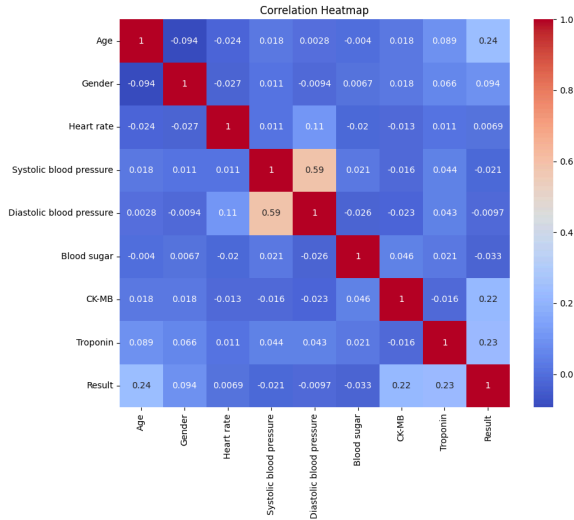


Figure 3. Heat Map

The heatmap shows that none of the features have a strong linear correlation with a heart attack happening. This further shows that linear models will not be able to identify helpful patterns, but there is a good chance that non-linear models will be able to perform well. This information was important in deciding which models to build.

### B. Model Selection

The first model the team decided on was Logistic Regression, a linear classifier very commonly used for binary classification problems [6]. The main goal of it is to model the probability of the target class using a logistic function and it performs the best when the features it's given have a strong linear relationship with the target variable you are trying to predict. Based on our correlation heatmap seen in **Figure 3**, we observed that none of the features had a strong linear correlation with heart attacks. As a result, we expect this model to underperform compared to more flexible, non-linear models. Although we are not confident in this model, we decided to include it in our study as a common baseline to compare against more complex classifiers.

The second model the team agreed on was a Random Forest Classifier. Random forests are an ensemble based model that constructs multiple decision trees and then averages their predictions [7]. Random Forests are very good at handling non-linear relationships and have the bonus of being highly resistant to noisy features as well as providing feature importance values. The team ended up selecting this model due to the reasons mentioned above and since we determined earlier that our features have weak linear correlation we expect this model to be one of the best.

Our third model we chose was the XGBoost (Extreme Gradient Boosting) classifier, a gradient boosted tree model known for being fast, regularized, and effective

on non-linear data. Unlike Random Forest that builds trees independently and in parallel, XGBoost builds trees sequentially by fixing the error of the ones that precede it [8]. We chose XGBoost as we expect it to be one of our high-performing models due to its ability to handle non-linear relationships. Moreover, it provides useful feature importance information and controls bias and variance by allowing accurate control of model complexity.

Our fourth and final model that we plan to implement is a Support Vector Machine (SVM) using a radial basis function (RBF) kernel. SVMs are robust classifiers that try to find the optimal decision boundary between classes by maximizing the margin between support vectors [9]. We are aware that SVMs are sensitive to feature scaling and must be tuned for hyperparameters such as C and gamma when using non-linear kernels. We ended up choosing to include an SVM model because we are curious to see how it performs compared to tree-based models like Random Forests and XGBoost.

### C. Model Evaluation

In order to evaluate the performance of each model, we will calculate a range of classification metrics which include accuracy (the overall proportion of correct predictions), precision (the proportion of positive predictions that are actually correct), recall (the proportion of actual positives that are correctly identified), F1-score (the harmonic mean of precision and recall), and ROC-AUC (a measure of the model's ability to distinguish between classes across all thresholds) [10]. These metrics will provide us with a general idea of each model's performance and ability to determine which patients exhibited heart attacks. On top of these metrics, we will also plot each model's confusion matrix to analyze the distribution of true positives, false positives, true negatives, and false negatives. For models that underperform with default parameters, we will apply hyperparameter tuning, with a focus on maximizing the F1-score. This process will allow us to improve the performance of models that yield weak results using default parameters. We expect this method to be particularly useful for models like the SVM, which tend to benefit from parameter optimization and feature scaling.

## IV. RESULTS

### A. Logistic Regression

The performance of the Logistic Regression model for heart attack classification is summarized in **TABLE 1**, which presents five key evaluation metrics: Accuracy, Precision, Recall, F1 Score, and ROC AUC.

Evaluation Metrics Table

Metric	Score
Accuracy	0.7916666666666666
Precision	0.8741258741258742
Recall	0.7716049382716049
F1 Score	0.819672131147541
ROC AUC	0.8769063180827886

TABLE 1. Logistic Regression Metrics

This model achieved an accuracy of approximately 79.2%, suggesting that it is able to roughly predict four out of five cases correctly, regardless of whether the case is positive or negative. A precision of 87.4% further reinforces this model's ability to correctly predict which patients are at risk of a heart attack. The higher precision over accuracy suggests that this model is able to minimize false positives at the expense of low accuracy suggesting we are classifying too many positive heart attack cases as negative. We would prefer if our model is over cautious with false positives because heart attacks are dangerous and we do not want to miss patients who are at risk. The recall of 77.2% shows that the model was able to successfully identify most heart attack cases, though it still missed some (as reflected in the 37 false negatives shown in **Figure 5**). Lastly, the F1 score shows the balance between precision and recall and with a value of 81.97% it suggests solid overall performance in the model's ability to distinguish between patients who are likely and unlikely to suffer a heart attack.

**Figure 4** shows the ROC Curve of the model. The curve quickly ascends toward the top left corner which signifies high sensitivity and specificity.

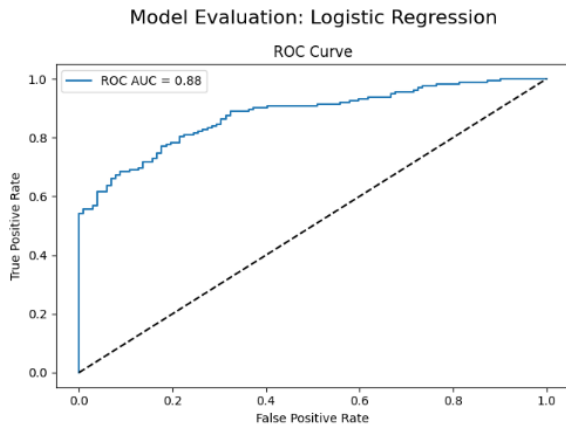


Figure 4. Logistic Regression ROC Curve

The confusion matrix shown below in **Figure 5** provides the model's final classification results. Looking at the figure, we can see that the model was able to correctly classify 84 non-heart attack cases (true negatives) and 125 heart attack cases (true positives), but misclassified 18 non-heart attack cases as heart attacks (false positives) and

37 heart attack cases as non-heart attacks (false negatives). While these results would be acceptable in some other scenarios, in a medical sense 37 false negatives suggests the model is missing too many heart attack cases which goes against this project's purpose of trying to detect heart attacks early and accurately in all patients.

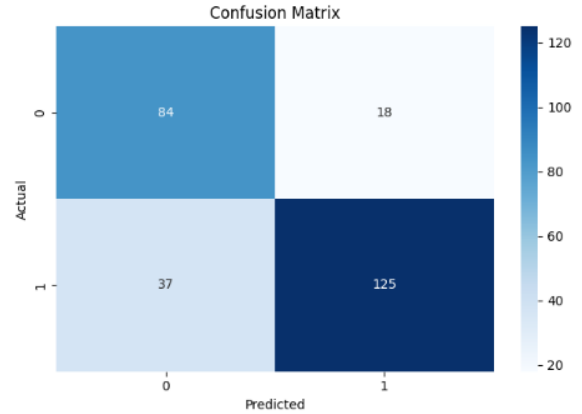
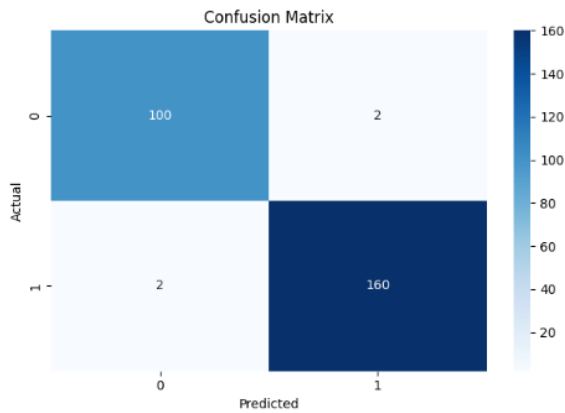


Figure 5. Logistic Regression Confusion Matrix

As mentioned above, these results may be acceptable in different scenarios but in a medical sense the main goal is to reduce the number of false negatives. We want to be over cautious with false positives then miss potential heart cases. Also from what we saw in our heat map in **Figure 3** it makes sense why this model may of under performed. Logistic regression is a linear model and as we saw earlier in the heat map none of the features showcased strong linear correlation for a heart attack. This shows that we need more advanced models capable of capturing nonlinear relationships to accurately predict the outcome of a heart attack.

## B. Random Forest

Compared to the Logistic Regression, the Random Forest demonstrated exceptional results in classifying heart attack cases. The classification results can be seen in the confusion matrix below in **Figure 6**. From the confusion matrix we can see the model was able to correctly predict 100 true negatives and 160 true positives, while only misclassifying 2 false positives and 2 false negatives which suggests very strong performance.



**Figure 6.** Random Forest Confusion Matrix

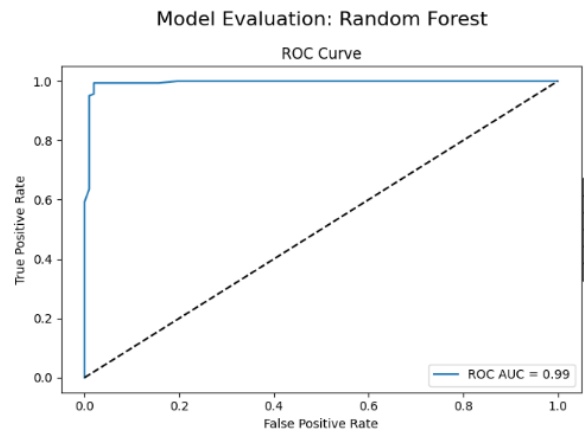
Given the model's performance, it is further reinforced in **TABLE 2**. We can also see that the accuracy for the model was about 98.5%, with precision, recall, and a F1 score of 98.8%. These results show that its predictions are for the most part perfect. Given these high scores, the model is accurate when predicting the positive cases with a rare chance for the model to miss patients with a high risk of a heart attack.

**Evaluation Metrics Table**

Metric	Score
Accuracy	0.9848484848484849
Precision	0.9876543209876543
Recall	0.9876543209876543
F1 Score	0.9876543209876543
ROC AUC	0.994795449043815

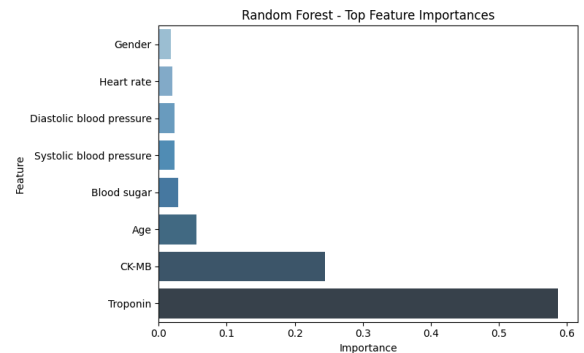
**TABLE 2.** Random Forest Metrics

As shown in **Figure 7**, the ROC Curve supports how well the model is performing. With an ROC AUC score of around 0.995, the curve almost touches the top-left corner of the graph which is something you would expect to see in a model that performs well. The shape of the curve indicates that the model can correctly tell the difference between patients who did and didn't have heart attacks without much compromise between catching real cases and avoiding false alarms.



**Figure 7.** Random Forest ROC Curve

On top of making accurate predictions, the Random Forest model provided the weights of features that mattered the most. The detailed chart of the weights can be seen below in **Figure 8**. From the graph, we can see that Troponin levels stood out as the top contributor, followed by CK-MB, which were the two top indicators for heart attacks. Other features like age, blood sugar, and blood pressure had smaller roles but still played a minor role in the model's decision making process. To our surprise, we weren't expecting gender and heart rate to have little effect in classifying heart attacks, so we are interested to see if these features will matter more in different models.



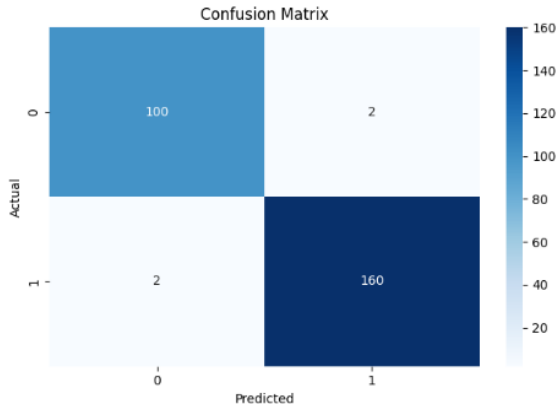
**Figure 8.** Random Forest Feature Importance

In summary, we saw that the Random Forest model performed exceptionally well in classifying heart attacks which is backed by the models near perfect accuracy and precision. Additionally, we were able to extract the key feature weights that the model used in the decision making process so this gave us a better idea of which factors indicate the presence of a heart attack. The Random Forest model was able to perform exceptionally well compared to other models due to its ability to identify intricate patterns in our dataset as well as distinguish between variables that may or may not indicate a heart attack.



### C. XGBOOST

The XGBoost model returned results that closely resembled the Random Forest model. Like the Random Forest model the XGBoost achieved the same success and was able to classify heart attacks effectively. As seen in the confusion matrix in **Figure 8**, the model correctly classified 100 non-heart attack cases (true negatives) and 160 heart attack cases (true positives), with only 2 false positives and 2 false negatives.



**Figure 9.** XGBoost Confusion Matrix

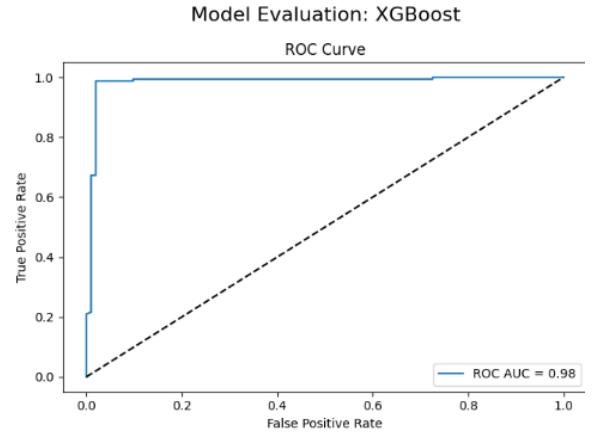
The evaluation metrics in **TABLE 3** explain why the XGBoost model performed exceptionally well. As seen in the table, the model achieved an accuracy, precision, recall and F1-score close to 99%. which solidifies that XGBoost is reliable at both identifying heart attack cases as well as minimizing false positives and negatives.

**Evaluation Metrics Table**

Metric	Score
Accuracy	0.9848484848484849
Precision	0.9876543209876543
Recall	0.9876543209876543
F1 Score	0.9876543209876543
ROC AUC	0.9842350520455095

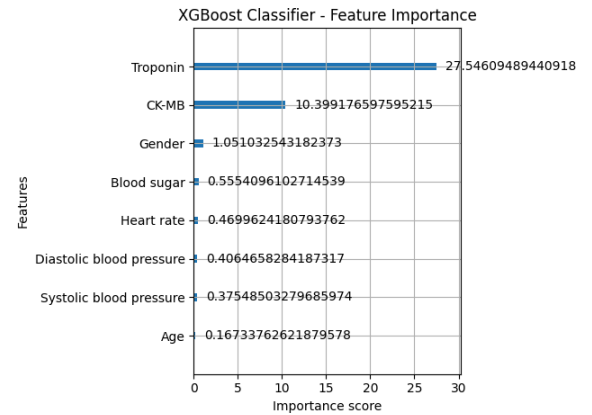
**TABLE 3.** XGBoost Metrics

**Figure 10** showcases the ROC curve for the XGBoost model. The AUC score was about 0.98 which indicates that this model is good at distinguishing between positive and negative cases across various thresholds. The curve on the graph also hugs the top left corner which suggests a strong balance between sensitivity and specificity.



**Figure 10.** XGBoost ROC Curve

The main difference between XGBoost and Random Forest lies in feature importance, as seen in **Figure 11**. Both models ended up identifying Troponin and CK-MB as the two most important features and used them as main factors in the decision making process. The difference lies in how both models handled the other less significant features. The feature importance of XGBoost was different then the Random Forest feature importance when it came to Age and Gender. Both models ranked the other features the exact same but had significant different positions for Age and Gender. The difference in feature importance shows that although models may yield similar results the decisions they make to get those results may be due to different reasons as we see with Age and Gender.

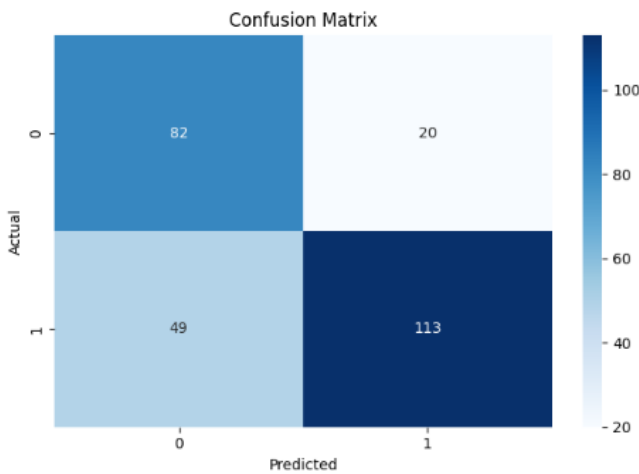


**Figure 11.** XGBoost Feature Importance

Overall, our XGBoost model matched the Random Forest in predictive performance and offers another robust option for heart attack classification with the only difference being the feature weights it uses to make decisions.

#### D. SVC

Initially, our SVC model didn't perform as well as the other models. This can be seen in the confusion matrix in ( **Figure 12** ), evaluation metrics (**Table 4** ), and ROC curve (**Figure 13** ). Before we did any tuning, the model was able to correctly identify 82 negative cases and 113 positive cases. However, it also misclassified 20 cases as false positives and more importantly it missed 49 actual heart attack cases as false negatives. As shown in Figure 16, it reached an accuracy of 73.9%, with a precision of 84.96%, recall of 69.75%, and an F1 score of 76.61%. The ROC AUC score of 0.846 (**Figure 13**) further indicated that the model struggled to distinguish between the classes in its default state.

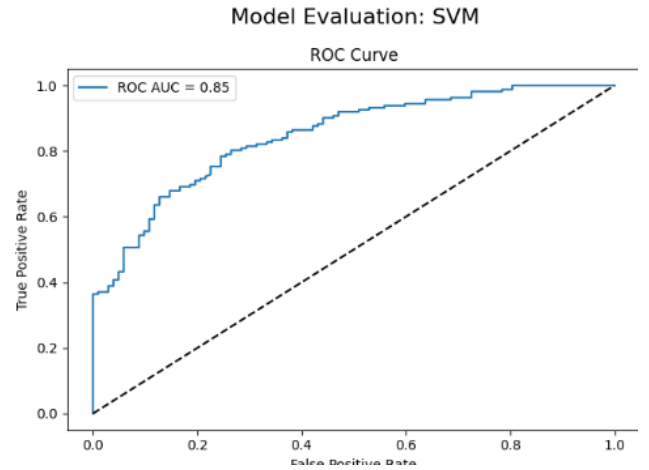


**Figure 12.** SVC Confusion Matrix

**Evaluation Metrics Table**

Metric	Score
Accuracy	0.7386363636363636
Precision	0.849624060150376
Recall	0.6975308641975309
F1 Score	0.7661016949152543
ROC AUC	0.8463749697409829

**TABLE 4.** SVC Metrics



**Figure 13.** SVC ROC Curve

The current implementation of the SVC model is prone to sensitivity with the default hyperparameters. As a result we conducted a grid search using the parameter space outlined below in **Figure 14** with the goal to find the best combination that yields the highest F1 score. This included varying C, gamma, kernel, and degree values to find an optimal configuration. Once the grid search was complete we determined the best performing set of hyperparameters that yielded the highest F1 score to be: svm\_C = 50, svm\_gamma = 1, svm\_kernel = 'poly', and svm\_degree = 3. We will use these updated hyperparameters to train a new SVC model with the hope of it producing better results.

```
param_grid = {
    'svm_C': [0.1, 1, 10, 50],
    'svm_gamma': ['scale', 0.01, 0.1, 1],
    'svm_kernel': ['rbf', 'poly'],
    'svm_degree': [2, 3]
}
```

**Figure 14.** SVC HyperParameter Tuning Grid

Looking at the results after tuning the SVC we can see that the new model classified a total of 93 true negatives, 144 true positives, 9 false positives, and 18 false negatives. This is all illustrated in the confusion matrix (**Figure 15**), the metrics (**Table 5**) and the ROC curve (**Figure 16**). After seeing the model's results, the accuracy was 89.8%, precision of 94.1%, recall of 88.9% and a F1 score of 91.4% (Figure 21). Furthermore looking at the ROC AUC it jumped to 0.934 as seen in **Figure 21**. This shows a big leap in the classifier's ability to differentiate the patients with risk vs the patients with no risk of a heart attack.

## V. CONCLUSIONS AND FUTURE WORK

### A. CONCLUSION

In this study, we used different machine learning algorithms to help predict heart attack risks given clinical factors. The algorithms we used for this were Logistic Regression, Random Forest, XGBoost, and Support Vector Machines (SVM).

After analyzing the Logistic Regression and the initial SVM before we tuned it, we saw that they struggled to capture the complex relationships between features and the outcome of a heart attack. For the SVM model, initial results were not promising, but after fine tuning hyperparameters and properly scaling the features, its performance improved. However, this performance was overshadowed by the non-linear Random Forest and XGBoost algorithm models. Both of the tree based models performed with an accuracy of about 98.5%, as well as a precision, recall, and f1-score roughly around 98.8%. These scores demonstrate the power of ensemble methods in capturing complex, non-linear relationships frequently present in real-world clinical data.

After looking over the results, the models clearly outlined that the most informative features were Troponin and CK-MB levels. These seem to be what mattered most for predicting heart attack risk. This finding also aligns with well established clinical knowledge, reinforcing the medical credibility of our model outputs. Having knowledge of these biomarkers can help demonstrate how we can use machine learning to improve clinical decision making of various different models.

Our research shows the potential that AI holds in the field of medicine. As these models advance and are trained on more accurate data we can one day begin to save lives by predicting heart attacks in advance.

### B. FUTURE WORK

To further enhance and extend the predictive performance demonstrated in this study, several promising avenues for future research have been identified. Future studies should include extending the dataset to encompass more diverse patient populations, namely those groups that have historically been underrepresented in clinical research. This is necessary to reduce existing biases and improve model generalizability and fairness across diverse patient populations. As highlighted in recent findings, “Women and minority groups are historically under-represented in cardiology, and the bulk of current evidence-based medicine might not necessarily apply to these populations” [4].

Additionally, the integration of other types of granular and continuous physiological monitoring data, such as that obtained from wearable health sensors, can

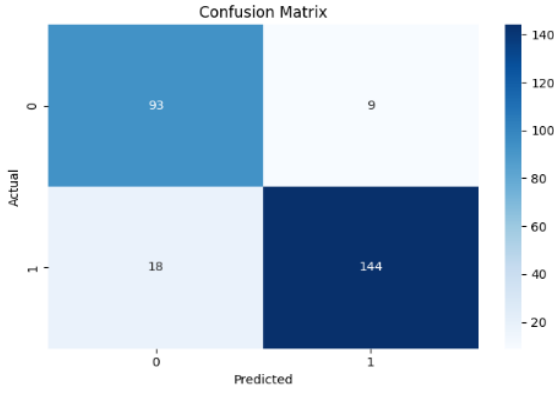


Figure 15. Tuned SVC Confusion Matrix

Metric	Score
Accuracy	0.8977272727272727
Precision	0.9411764705882353
Recall	0.8888888888888888
F1 Score	0.9142857142857143
ROC AUC	0.9343984507383201

TABLE 5. Tuned SVC Metrics

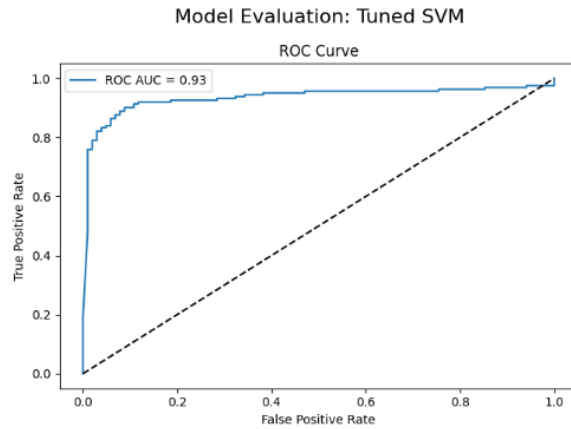


Figure 16. Tuned SVC ROC Curve

These results validate our earlier assumption that the poor initial performance was largely due to untuned parameters. Once we optimized the hyperparameter, our SVC model emerged as a competitive model that is capable of matching the performance of tree based models. Although the current hyperparameter set yields good results we do believe it is possible to improve this model even more by trying a broader range of hyperparameters, which include experimenting with different kernels, using higher-degree polynomials, or using different feature combinations. However, due to constraints in both time and computing resources, we were unable to perform a more exhaustive search.



substantially enhance real-time predictive capabilities and facilitate enhanced early detection of heart attacks. Integration of this information would provide an improved and earlier assessment of patient health condition, which may result in more rapid and effective clinical interventions.

Additional research work should also go in-depth into sophisticated interpretability techniques for sophisticated predictive models like XGBoost and Random Forest. Creating advanced interpretation methods will not only provide deeper insight into the decision-making processes inherent in these models but also increase clinical confidence and allow for more widespread acceptance of AI-based diagnostics among medical professionals.

Lastly, future endeavors must focus on developing welcoming, safe, and very efficient deployment platforms specifically tailored to the healthcare environment. Robust deployment architectures are paramount to being able to seamlessly incorporate these forecasting models into the clinical workflow so that they can become realistic, reliable, and valuable tools that can meaningfully help and complement patient care in real-life health care environments.

## REFERENCES

- [1] S. Brar, "Heart Attack Dataset," *kaggle*  
<https://www.kaggle.com/datasets/sukhmandeepsinghbrar/heart-attack-dataset>
- [2] A. Al Ahdal *et al.*, "Monitoring Cardiovascular Problems in Heart Patients Using Machine Learning," *Journal of Healthcare Engineering*, vol. 2023, p. e9738123, Feb. 2023, doi: <https://doi.org/10.1155/2023/9738123>
- [3] S. Kumar, A. Sharma, N. Kaur, L. Pawar, and R. Bajaj, *Optimized Predictive Models in Health Care Using Machine Learning*, 1st ed. Newark: John Wiley & Sons, Incorporated, 2024. doi: 10.1002/9781394175376
- [4] E. Tat, D. L. Bhatt, and M. G. Rabbat, "Addressing bias: artificial intelligence in cardiovascular medicine," *The Lancet Digital Health*, vol. 2, no. 12, pp. e635–e636, Dec. 2020, doi: [https://doi.org/10.1016/s2589-7500\(20\)30249-1](https://doi.org/10.1016/s2589-7500(20)30249-1).
- [5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 16, pp. 321–357, Jun. 2002, doi: <https://doi.org/10.1613/jair.953>.
- [6] V. Kanade, "What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices," *Spiceworks*, Apr. 08, 2022.  
<https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/>
- [7] "Ensemble Methods: Random Forest," *Scikit-learn*, [Online]. Available: <https://scikit-learn.org/stable/modules/ensemble.html#random-forests>
- [8] Jason Brownlee, "A Gentle Introduction to XGBoost for Applied Machine Learning," *Machine Learning Mastery*, Aug. 16, 2016.  
<https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>
- [9] scikit learn, "1.4. Support Vector Machines," *Scikit-learn.org*, 2018.  
<https://scikit-learn.org/stable/modules/svm.html>
- [10] "3.3. Metrics and scoring: quantifying the quality of predictions — scikit-learn 0.24.1 documentation," *scikit-learn.org*.  
[https://scikit-learn.org/stable/modules/model\\_evaluation.html#classification-metrics](https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics)