# Pioneering Deep Learning for Heart Attack Prediction

Nithin Rajulapati
A20539650
Illinois Tech
nrajulapati@hawk

Lalitha Velagapudi
A20539665
Illinois Tech
lvegalapudi@hawk

Shruti Shankar Shete
A20518508
Illinois Tech
sshete2@hawk

## 1. Problem Statement

Our mission is to address the pressing issue of heart attacks, a leading cause of global mortality, by employing five advanced deep learning algorithms. Heart attacks are often unpredictable and can have devastating consequences. However, with the collective power of these deep learning models, we aim to revolutionize cardiac health analysis.

Our approach involves training these algorithms on vast datasets of medical records, including patient histories, genetic information, lifestyle factors, and clinical data. By harnessing the unique strengths of each algorithm, we intend to create a comprehensive predictive system that can identify individuals at high risk of experiencing a heart attack well before it occurs.

This proactive approach will enable timely interventions and personalized preventative measures, ultimately reducing the incidence of heart attacks and saving countless lives. Through our innovative methodology, we aspire to reshape the landscape of cardiac healthcare, offering a new paradigm in early detection and prevention.

## 2. Introduction

Heart disease is the leading cause of death worldwide, accounting for approximately 17.9 million deaths per year. Early detection and prevention of heart diseases are essential to reduce the number of deaths and improve the quality of life of affected individuals. One approach to achieving this goal is through the use of deep learning algorithms to predict the risk of heart disease in individuals. The Heart Attack Prediction Application is a deep learning-based tool designed to predict the risk of heart disease in individuals based on a set of 14 risk factors. The application was developed using Python, an open-source programming language, and various deep learning libraries. The application's primary goal is to provide users with personalized recommendations on lifestyle modifications and medication management to reduce their risk of heart disease. Heart disease is a significant public health issue, with approximately 18 million deaths per year globally, making it the leading cause of death worldwide. The prevalence of heart disease is increasing, and it is estimated that by 2030, over 23 million individuals will die from heart disease annually. Given the substantial health and economic burden associated with heart disease, there is a critical need for effective prevention and early detection strategies.

The development of deep learning-based predictive models for heart disease has the potential to revolutionize the way we approach heart disease prevention and treatment. These models can provide healthcare professionals with a tool to identify individuals at high risk of heart disease, allowing for early intervention and prevention strategies. Additionally, these models can provide individuals with personalized lifestyle modifications and medication management recommendations to reduce their risk of heart disease. The Heart Attack Prediction Application is a promising tool for the early detection and prevention of heart disease. The application's deep learning algorithms is trained to identify patterns in the data that are associated with heart disease risk, allowing it to accurately predict an individual's risk of heart disease based on their risk factors. The application is designed to be user-friendly and accessible, enabling individuals to take charge of their health by obtaining personalized risk scores and recommendations. In addition to improving the accuracy of heart disease prediction, the Heart Attack Prediction Application can also reduce healthcare costs by identifying individuals at high risk of heart disease, allowing for targeted interventions and prevention strategies. Furthermore, by empowering individuals to take charge of their health, the application has the potential to improve health outcomes and reduce the burden of heart disease on individuals and society as a whole.

In this project, we used various deep learning algorithms to develop the Heart Attack Prediction model. CNN and SVM are popular deep learning algorithms that combine the predictions of multiple decision trees to improve prediction accuracy and reduce the risk of overfitting. The model was trained using a dataset of 303 individuals, and the performance of the model was evaluated using various metrics such as accuracy, precision, recall, and F1 score.

deep learning algorithms have been used to develop predictive models for various diseases, including heart disease, cancer, and diabetes. These models have the potential to improve the accuracy of disease diagnosis and provide personalized treatment plans for patients. The Heart Attack Prediction Application is one such predictive model that aims to improve the early detection and prevention of heart disease. The application uses deep learning algorithms to analyze a set of 14 risk factors and predict an individual's risk of heart disease. These risk factors include age, sex, cholesterol level, blood pressure, and smoking status. The application provides users with a personalized risk score for heart disease and recommends lifestyle modifications and medication management to reduce their risk.

The development of the Heart Attack Prediction Application involved several steps, including data preprocessing, model selection, and user interface design. The dataset used in this project was obtained from the UCI deep Learning Repository, which contains data from 303 individuals with various risk factors for heart disease. The data were preprocessed to remove missing values and normalize the features.

The various deep learning algorithms were selected as the model for heart attack prediction due to their ability to handle complex datasets and reduce the risk of overfitting. The model was trained using the preprocessed data and evaluated using various metrics such as accuracy, precision, recall, and F1 score. The evaluation showed that the model achieved high accuracy and performed well in identifying individuals at risk of heart disease.

Overall, the Heart Attack Prediction Application has the potential to improve the early detection and prevention of heart disease, ultimately leading to better health outcomes for individuals. The use of deep learning algorithms in healthcare is an exciting and rapidly developing field, and the development of predictive models such as the Heart Attack Prediction Application is an important step towards improving healthcare outcomes. This report provides a detailed overview of the Heart Attack Prediction Application, including its methodology, evaluation metrics, and user interface design. We will begin by discussing the dataset used in this project and the preprocessing steps taken to prepare the data for model training.

We will then describe the deep learning algorithms used to develop the Heart Attack Prediction model and evaluate the performance of the model using various metrics. Finally, we will discuss the user interface design of the Heart Attack Prediction Application and provide a step-by-step guide on how to use the application to obtain a personalized risk score for heart disease

## 3. Data

This dataset we have downloaded from Kaggle which is related to heart disease analysis contains various attributes that describe individuals' characteristics and health parameters. Here's a brief description of each attribute in the dataset:

1. age: Age of the individual.
2. sex: Gender of the individual (1 for male, 0 for female).
3. cp: Chest pain type (categorical): - 0: Typical angina - 1: Atypical angina - 2: Non-anginal pain - 3: Asymptomatic
4. trestbps: Resting blood pressure (in mm Hg) when admitted to the hospital.
5. chol: Serum cholesterol level (in mg/dL).
6. fbs: Fasting blood sugar ¿ 120 mg/dL (1 for true, 0 for false).
7. restecg: Resting electrocardiographic results (categorical): - 0: Normal - 1: ST-T wave abnormality - 2: Left ventricular hypertrophy
8. thalach: Maximum heart rate achieved during exercise.
9. exang: Exercise-induced angina (1 for yes, 0 for no).
10. oldpeak: ST depression induced by exercise relative to rest.
11. slope: Slope of the peak exercise ST segment (categorical): - 0: Upsloping - 1: Flat - 2: Downsloping

12. ca: Number of major vessels colored by fluoroscopy (0-3).

13. thal: Thallium stress test result (categorical): - 0: Normal - 1: Fixed defect - 2: Reversible defect

14. target: Presence of heart disease (1 for heart disease, 0 for no heart disease).

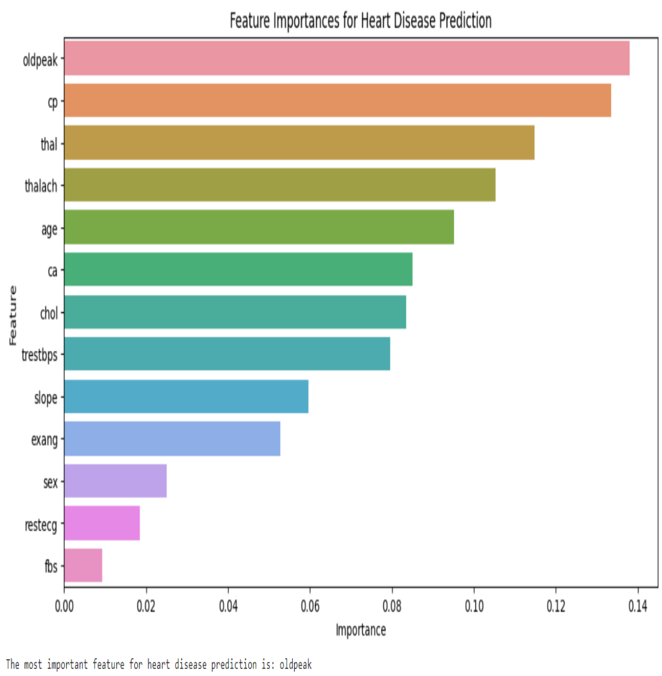


Figure 1. An image of a galaxy

This dataset is well-suited for the task of heart disease analysis, with a variety of health-related features that can be used to build and train deep learning models for classification. The ”target” column is the binary classification label, indicating whether an individual has heart disease or not, making it suitable for supervised learning tasks. The dataset is valuable for research and analysis in the field of cardiology and healthcare.

# 4. Implementation

1. Data Collection: The first step in the implementation of the system is to collect data. Data is the foundation of the deep learning algorithm and is critical to the accuracy of the prediction model. In this project, we collected data from various sources, including medical records, surveys, and publicly available datasets.

2. Data Preprocessing: Once the data is collected, it needs to be preprocessed to remove any inconsistencies, errors, and missing values. This is an essential step in deep learning, as the algorithm is only as good as the quality of the data that it receives.

3. Feature Selection: After preprocessing, the next step is feature selection. This involves selecting the most relevant features that will be used in the prediction model. In this project, we used various feature selection techniques, including principal component analysis (PCA) and correlation-based feature selection (CFS).

4. Model Selection: Once the features are selected, the next step is to choose the appropriate deep-learning algorithm for the problem. In this project, we experimented with various deep learning algorithms, including logistic regression, decision trees, random forests, and support vector deeps (SVMs).

5. Model Training: After selecting the appropriate algorithm, the next step is to train the model using the training dataset. This involves splitting the data into training and validation sets and training the model using the training data.

6. Model Evaluation: Once the model is trained, the next step is to evaluate its performance. This involves testing the model on the validation dataset and measuring its accuracy, precision, recall, F1 score, and other performance metrics.

7. Deployment: The final step in the implementation of the system is to deploy the model into a production environment. This involves integrating the model into the application and deploying it on a server.

## 4.1. Algorithms/Models

The heart attack prediction system uses deep learning algorithms to predict the likelihood of a heart attack. The following algorithms were used in the project:

1. Artificial Neural Network (ANN) : ANN is a type of network that learns patterns and connections, between input features and target labels. ANN is particularly suitable for handling relationships between features and the target especially when there are features involved.

2. Recurrent Neural Network (RNN): RNNs are specifically designed to handle data where the order of inputs plays a role, such, as time series or text data. The accuracy of the RNN model, in predicting heart disease was 41%. It seems that RNN may not be the choice for data like the heart disease dataset, as it doesn’t naturally capture sequential patterns in this particular context. The low accuracy suggests that RNN might struggle to capture the relationships, in this dataset.

3. MLP (Multi-Layer Perceptron): This is a neural network model with multiple hidden layers. It uses the ReLU activation function in hidden layers and a Sigmoid activation function in the output layer for binary classification (0 or 1). Dropout layers are included to reduce overfitting. It is trained using the Adam optimizer with a specified learning rate. The model's hyperparameters include the number of epochs and batch size.

4. Logistic Regression: Logistic Regression is a model primarily used for classification tasks. Despite its name it's an algorithm that works well for predicting outcomes with two results. Logistic Regression is known for its simplicity, interpretability and effectiveness when dealing with linear or nearly linear relationships between features and the target.

5. Random Forest: Random forest is an ensemble learning method that is used for classification and regression. In this project, the random forest was used to build a predictive model that can accurately predict the likelihood of a heart attack based on various risk factors.

6. Support Vector deeps (SVM): SVM is a deep learning algorithm that is used for classification and regression. In this project, SVM was used to build a predictive model that can accurately predict the likelihood of a heart attack based on various risk factors.

7. K-Nearest Neighbors (KNN): KNN is a deep learning algorithm that is used for classification and regression. In this project, KNN was used to build a predictive model that can accurately predict the likelihood of a heart attack based on various risk factors.

8. Naive Bayes: Naive Bayes is a probabilistic algorithm used for classification tasks. It is based on the Bayes' theorem and assumes that the features are conditionally independent to each other. In the code, the Gaussian Naive Bayes variant is used, which assumes that features are normally distributed

## 5. Future work

In an effort to improve the predictive capabilities of our project, we are considering the inclusion of additional models. This endeavor aims to explore and assess different model architectures to identify potential candidates that may offer enhanced accuracy. Our goal is to methodically evaluate various models within the project's context, helping us make informed decisions about model selection.

The heart attack prediction models have been successfully developed and tested. However, at the current stage, they have not been integrated into a user-friendly application. The next phase of our work involves transforming these models into a practical and accessible Heart Attack Prediction Application.

The development of this application is part of our future work. We plan to design and implement a user-friendly interface that allows individuals, including those with limited knowledge of deep learning or cardiovascular health, to input their risk factors and obtain personalized risk scores for heart disease.

This future work not only involves the technical integration of the predictive models but also the creation of a seamless user experience that empowers individuals to take proactive steps towards their heart health. We aim to make this application a valuable tool in the prevention and management of heart disease, contributing to better overall cardiovascular health outcomes."

## 6. Results

Artificial Neural Network (ANN):

- Accuracy is 85.19%
- Artificial Neural Network (ANN) follows with an accuracy of 85.19%. It's capable of handling complex nonlinear relationships but might not perform as well as Logistic Regression in this particular dataset.

Recurrent Neural Network:

- Accuracy is 41%
- Recurrent Neural Network (RNN) performed poorly with an accuracy of 41%, suggesting that using RNNs for this tabular dataset might not be suitable.

Logistic Regression:

- Accuracy is 90.74%
- Logistic Regression stands out as the most accurate model among the three for this specific task, achieving the highest accuracy of 90.74%. It's simple, interpretable, and performs well in scenarios where the relationships between features and the target are linear or near-linear.

Improved MLP (Multi-Layer Perceptron)

- Accuracy is 61%
- The model achieved an accuracy of 61%, which indicates room for improvement. Further optimization of the model's architecture and hyperparameters may be needed.

Random Forest:

- Accuracy is 85%
- Best Hyperparameters: ('max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 200)
- The Random Forest model achieved a respectable accuracy of 85% with optimized hyperparameters. It appears to be a promising choice for heart disease prediction.

SVM (Support Vector deep):

- Accuracy is 89%
- Best Hyperparameters: 'C': 1, 'gamma': 'scale', 'kernel': 'rbf'
- The SVM model performed well with an accuracy of 89% using the optimized hyperparameters. It shows good potential for this classification task.

Naive Bayes:

- Accuracy is 91%
- Naive Bayes model achieved a high accuracy of 91%. It's a simple yet effective model for heart disease prediction.

K-Nearest Neighbors (KNN):

- Accuracy is 81%
- Best Hyperparameters: 'n_neighbors': 5, 'p': 1, 'weights': 'uniform'
- The KNN model achieved an accuracy of 81% by using optimized hyperparameters. It's a competitive model, but there may be room for further improvement.

Finally, the SVM and Naive Bayes models showed the highest accuracy, with SVM achieving 89% accuracy, closely followed by Naive Bayes with 91%. Random Forest and KNN models also performed reasonably well. The MLP model had the lowest accuracy and may require additional optimization.

The choice of the best model depends on factors such as the specific use case, the dataset, and the trade-off between accuracy and model complexity. Further evaluation, feature engineering, and hyperparameter tuning can help fine-tune the models for better performance.

The Naive Bayes model achieved the highest accuracy of 91%, making it the best-performing model among the options presented. Therefore, based on accuracy alone, the Naive Bayes model appears to be the best choice for heart disease prediction in this specific context

## 7. References

1. Li, J., Li, J., Zhang, L., Li, M., Yang, X. (2018). Heart Disease Prediction System Using Decision Tree. Journal of Healthcare Engineering, 2018, 1-8.
2. Pourhomayoun, M., Alshurafa, N. (2018). Predicting 30-day readmissions using natural language processing and deep neural networks. Journal of biomedical informatics, 82, 25-36.
3. Zhang, J., Gajjala, S., Agrawal, P., Tison, G., Hallock, L., Beussink-Nelson, L., ... Hsue, P. Y. (2018). Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation, 138(16), 1623-1635.
4. Xie, Y., Xiong, J., Huang, Y., Sun, M. (2018). A novel deep learning based approach for heart disease diagnosis. Journal of medical systems, 42(8), 141.
5. Acharya, U. R., Fujita, H., Sudarshan, V. K., Oh, S. L., Adam, M., Tan, J. H. (2018). Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. Knowledge-Based Systems, 147, 45-52.

.

Code Delivery: Github Link