**Heart Disease Prediction System**

**Using Machine Learning and Deep Learning**

Project Report Submitted in Partial Fulfilment of the Requirements for the Degree of

**Bachelor of Technology**

***in***

**CSE**

Submitted by

Ankit Nainwal 1000014497

Kamlesh Singh 1000014214

Abhimat Kala 1000014969

***Under the Supervision of***

Miss Shreya Deoli

Assistant Professor

Computer Science and Engineering

May, 2024

**DECLARATION**

We declare that this written submission represents our ideas in our own words and where others’ ideas or words have been included, we have adequately cited and referenced the original sources. We declare that we have adhered to all principles of academics’ honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. I understand that any violation of the above will be cause for disciplinary action by the University and can also evoke penal action from the source which have thus not been properly cited or from whom proper permission has not been taken when needed. The plagiarism check report is attached at the end of this document.

Ankit Nainwal Signature and Date

Kamlesh Singh Signature and Date

Abhimat Kala Signature and Date



**CERTIFICATE**

This is to certify that the report entitled **“Heart Disease Prediction System Using Machine Learning and Deep Learning”** in partial fulfilment of the requirement for the award of the **Degree of Bachelor of Technology in Computer Science and Engineering** submitted to the **DIT University, Dehradun, Uttarakhand, India**, is and authentic record of bonafide work carried out by **Ankit Nainwal, Kamlesh Singh, Abhimat Kala**, under the supervision of **Miss Shreya Deoli**, Assistant Professor-CSE, School of Computing.

Miss Shreya Deoli

(Assistant Professor-CSE)

Date:

Place:

**ACKNOWLEDGEMENT**

At the outset, we would like to express our gratitude to the Lord Almighty for helping us to complete the project work successfully. We would like to express our sincere thankfulness to our supervisor Miss Shreya Deoli for providing her invaluable advice, remark, and proposals throughout this project phase. We would especially like to extend our gratitude to our mentor for constantly encouraging us to work harder and guiding us in approaching this project. Her encouragement guardianship, inspirational and timely suggestions at various stages of this phase of the project made it possible for us to complete this report. We also wish to extend our thanks to all the faculties for their insightful comments and constructive suggestions to improve the quality of this research work. We express deep and sincere gratitude to our honourable dean Dr. Debopam Acharya for providing us with the essential tools and services at DIT University.

Finally, our biggest thanks to our family for all the support they have shown us throughout our studies, research, and the culmination of years of distance learning. This accomplishment would not have been possible without them.

Ankit Nainwal 1000014497

Kamlesh Singh 1000014214

Abhimat Kala 1000014969

# Table Of Content

**Introduction**

1) Introduction to the System

2) Problem Definition

3) Aim

4) Objective

5) Goal

6) Need of System

**Hardware and Software requirement**

1) Introduction

2) System environment

3) Software requirement

4) Hardware requirements

**System Analysis**

1) Purpose

2) Project Scope

3) Existing System

4) Proposed System

**Implementation issues**

1. Min Max Scaling
2. XGB Classifier
3. Sequential Model
4. Voting Classivier
5. Saving Model

**Result Received**

1. All models achieved accuracy and prediction scores.

2. The confusion matrix depicting predictions from both the XGB Classifier and Sequential Model.

**Conclusion**

1. Features of “Heart Disease Prediction”
2. Benefits Accrued from “Heart DiseasePrediction”
3. Limitations of “Heart Disease Prediction”

**Bibliography**

**Introduction**

The Heart Disease Prediction System utilizes a range of Machine Learning and Deep Learning algorithms, developed in Python.  
Our system features a web page where users can input their health details. Leveraging the models we've employed, it predicts whether the individual has heart disease or not.

The dataset employed in this project encompasses various parameters including age, gender, type of chest pain, resting blood pressure, cholesterol level, fasting blood sugar, results of resting electrocardiography, maximum heart rate achieved, presence of exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels, thalassemia status, and the indication of heart disease (target variable)

**Problem Definition:**

Detecting heart disease early can significantly improve treatment outcomes, yet it's often challenging to diagnose in its early stages. Doctors may lack resources for accurate and timely diagnosis. Thus, the goal of the Heart Disease Prediction System using Machine Learning and Deep Learning is to create a tool that accurately forecasts a patient's likelihood of developing heart disease. It will utilize a variety of clinical and demographic parameters. This system aims to be user-friendly and accessible to patients, doctors, and healthcare providers, aiding in the early detection and prevention of heart disease. Additionally, it will offer personalized care recommendations based on the patient's prediction results.

**Aim:**

The goal of the Heart Disease Prediction System employing Machine Learning and Deep Learning is to furnish a precise and effective means of assessing the risk of heart disease in patients. Leveraging machine learning algorithms, the system scrutinizes patient data, deriving predictions from diverse parameters. Its primary objective is to support doctors and healthcare practitioners in rendering prompt and precise diagnoses, thereby fostering enhanced patient outcomes and healthcare provision. Furthermore, the system endeavors to offer users an intuitive and accessible interface accessible via the internet, ensuring ease of use from any location.

**Objective:**

The objectives of the Heart Disease Prediction System using Gradient boosting classifier project are:

1. To develop a machine learning model that can accurately predict the presence of heart disease in patients based on their medical parameters.
2. To create a user-friendly web application that allows users to input their medical parameters and receive a heart disease prediction in real-time.

**Goal:**

The goal of the Heart Disease Prediction System using Machine Learning and deep Learning algorithm is to provide an accurate prediction of the likelihood of a patient having heart disease based on their medical history and certain parameters such as age, sex, blood pressure, cholesterol level, and other relevant factors. This system aims to help users in making a timely and accurate diagnosis of heart disease, leading to early treatment and better patient outcomes. Additionally, the system aims to provide a user-friendly interface for patients to assess their risk of heart disease and access information about recommended doctors in their area.

**Need of the System:**

The Heart Disease Prediction System is essential due to the rising incidence of heart diseases and the critical requirement for precise and timely diagnosis to enhance patient outcomes. Early detection and prediction of heart disease empower healthcare providers to formulate effective treatment strategies and mitigate adverse events. By furnishing an accurate forecast of heart disease likelihood, this system contributes to diminishing the mortality and morbidity rates associated with heart diseases. It enables doctors to make well-informed decisions regarding patient care. Furthermore, the system empowers patients to undertake preventive measures and embrace lifestyle changes to mitigate their risk of developing heart disease.

**PREDICTABLE ATTRIBUTE**

**Data Used :**

1. **Age**
2. **Sex**
3. **Chest Pain Type**

* **What Are the 4 Types of Angina?**

Understanding types of angina will help you determine how to respond

|  |  |  |  |
| --- | --- | --- | --- |
| **Types of Angina** | **Causes or Triggers** | **Symptoms** | **Treatment Options** |
| **1. Microvascular Angina** | * MVD * Spasms within the walls of tiny arteries that reduce blood flow to the heart * More common in women | * Pain may be more severe than other types * Pain that come with, fatigue, sleep problems * Pain may be noticed when experiencing stress | * Medications * Medical strategies that increase blood flow and reduce workload in the heart |
| **2. Stable Angina or Angina Pectoris** | * Physical activity or emotional stress * Exposure to very cold or hot temperature * Heavy meals * Smoking | * Pain often lasts 5 minutes or less * Pain may feel like indigestion or gas * Pain may feel like it spreads to the arms, back or other areas | * Rest * Medication (e.g., nitroglycerin) |
| **3. Unstable Angina or Acute Coronary Syndrome** | * Coronary arteries narrowed by fatty buildups (atherosclerosis) * Fatty buildups may rupture blood vessel and can cause blockage of blood | * Unexpected pain that occurs during physical activity or when you’re resting or sleeping * Pain that may last longer than stable angina * Pain that may get worse over time * Pain that doesn’t go away with rest or medication * Chest pain or discomfort | * Cardiac catheterization * Percutaneous coronary intervention (PCI) * Coronary artery bypass graft surgery |
| **4. Variant (Prinzmetal) Angina or Angina Inversa** | * stress, smoking, cocaine use or exposure to cold weather * Medicines that tighten your blood vessels. * It manily occurs in younger patients | * Severe chest pain or discomfort that occurs while resting, at night or early in the morning | * Medications (e.g., calcium antagonists and nitrates) |

1. Blood Pressure

The average blood pressure varies depending on factors such as age, gender, and overall health. However, a generally accepted average blood pressure for adults is around 120/80 mmHg.

* **Resting Blood Pressure:** Resting blood pressure refers to blood pressure measurements taken when a person is in a relaxed state, typically while sitting or lying down and not engaged in any physical activity. It is considered the baseline blood pressure level of an individual and is often used as an indicator of overall cardiovascular health. Resting blood pressure measurements provide important information about cardiovascular health and are used to assess the risk of developing conditions such as hypertension (high blood pressure).

**Normal Blood Pressure:** A normal blood pressure reading is typically below 120/80 mmHg.

**Elevated Blood Pressure:** A reading consistently higher than 120/80 mmHg but lower than 130/80 mmHg is considered elevated.

**Hypertension (High Blood Pressure):** Hypertension is diagnosed when blood pressure consistently measures at 130/80 mmHg or higher.

**Hypotension (Low Blood Pressure): Hypotension** occurs when blood pressure is consistently below 90/60 mmHg.

Blood pressure can fluctuate throughout the day due to various factors such as physical activity, stress, emotions, diet, and medications. It typically rises during periods of physical activity or stress and decreases during rest or relaxation.

1. Serum Cholesterol

Serum cholesterol refers to the level of cholesterol found in the blood serum, which is the clear, yellowish fluid that remains after blood has clotted and the clotting factors have been removed. Cholesterol is a type of lipid (fat) that is essential for the body's normal functioning. It plays a crucial role in building cell membranes, producing certain hormones, and aiding in the digestion of fats.

* Optimal Ranges

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Age and sex** | **Total serum cholesterol** | **HDL level** | **LDL level** | **Triglycerides** |
| all aged 19 years and younger | 170 mg/dl at most | at least 45 mg/dl | less than 100 mg/dl | less than 150 mg/dl |
| females aged 20 years and older | 125–200 mg/dl | at least 50 mg/dl | less than 100 mg/dl | less than 150 mg/dl |
| males aged 20 years and older | 125–200 mg/dl | at least 40 mg/dl | less than 100 mg/dl | less than 150 mg/dl |

1. **Fasting Blood Sugar :** Compares the fasting blood sugar value of an individual with 120mg/dl.  
   If fasting blood sugar > 120mg/dl then : 1 (true)  
   else : 0 (false)
2. **Exercise indused Agina :**   
   Exercise-induced angina, also known as angina pectoris or simply angina, is chest discomfort or pain that occurs during physical exertion or exercise. It's a symptom of underlying coronary artery disease
3. Resting Electrocardiographic Result: Displays resting electrocardiographic results

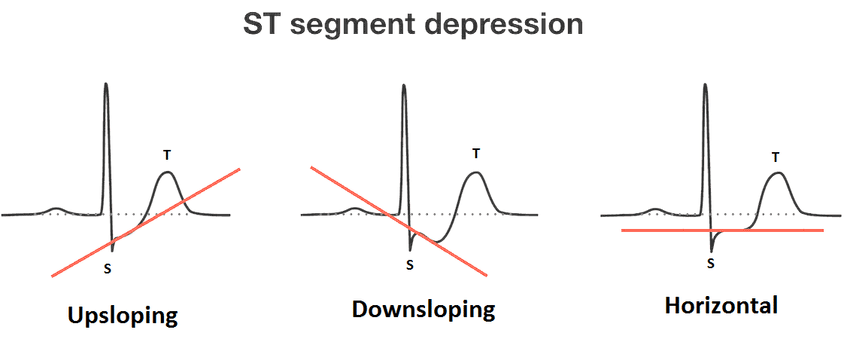
**Normal:** The ECG shows a regular heartbeat with no abnormal ST segment or T wave patterns. No arrhythmias are present.

**ST-T wave abnormality:** The ECG indicates abnormalities suggestive of myocardial ischemia, injury, or repolarization issues.

**Left Ventricular Hypertrophy (LVH):** Signs consistent with an enlarged left ventricle are observed on the ECG, often seen in conditions like hypertension or cardiomyopathy.

1. Maximum Heart Rate Achieved : displays the max heart rate achieved by an individual.
2. Old Peak :  ST depression caused by activity in comparison to rest
3. Slope :

The slope of the peak exercise ST segment, often referred to as the ST segment slope, is an important electrocardiographic (ECG) parameter used in the assessment of cardiac function during exercise stress testing. It refers to the direction and steepness of the ST segment deviation from the baseline during the peak of exercise.



Type of slopes are :

**Upsloping ST Segment**: In some cases, particularly in individuals without CAD, the ST segment may show a slight upsloping pattern during exercise. This is generally considered a benign finding and not indicative of significant ischemia.

**Horizontal ST Segment:** A horizontal or flat ST segment during exercise suggests significant myocardial ischemia. This pattern is often associated with severe coronary artery stenosis or occlusion and indicates a high risk of coronary events.

**Downsloping** ST Segment: A downsloping ST segment is considered the most concerning pattern during exercise stress testing. It is highly suggestive of myocardial ischemia, especially if it occurs at a low workload or persists into the recovery phase. Downsloping ST depression is associated with a higher risk of adverse cardiac events and is often indicative of significant CAD

1. No. of Major Vessel: displays the value as integer or float.
2. Thal :   
   Thalassemia is a group of inherited blood disorders characterized by abnormal production of hemoglobin, the protein in red blood cells that carries oxygen. It is caused by mutations in the genes responsible for producing hemoglobin

**Normal:** No significant abnormalities related to thalassemia are observed.

**Fixed Defect:** Indicates a permanent or irreversible issue related to thalassemia, such as severe symptoms or complications.

**Reversible Defect:** Suggests a condition related to thalassemia that can be improved or corrected with appropriate treatment or intervention.

**Hardware and Software Requirements**

* Software requirement
* Hardware requirements

**Software Requirements:**

* Technology: Python Django
* IDE : Pycharm/Atom
* Client Side Technologies: HTML, CSS, JavaScript , Bootstrap
* Server Side Technologies: Python
* Data Base Server: Sqlite
* Operating System: Microsoft Windows/Linux

**Hardware Requirements:**

* Processor: Pentium-III (or) Higher
* Ram: 64MB (or) Higher
* Hard disk: 80GB (or) Highe

**System Analysis**

##### Purpose

* Project Scope
* Existing System
* System Overview

### Purpose:

The purpose of the Heart Disease Prediction System using Machine Learning and deep Learning algorithm is to provide an accurate prediction of the likelihood of a patient having heart disease based on their medical history and certain parameters such as age, sex, blood pressure, cholesterol level, and other relevant factors. This system aims to user in making a timely and accurate diagnosis of heart disease, leading to early treatment and better patient outcomes.

**Project Scope:**

The Heart Disease Prediction System holds vast potential given the global prevalence of heart disease, offering benefits to hospitals, clinics, and medical practitioners who can utilize it to forecast heart disease in their patients and implement preventive measures. Moreover, individuals concerned about their heart health can proactively utilize this system to mitigate the risk of heart disease.

In an industrial context, this system finds application in the healthcare and medical sectors. It facilitates swift and precise predictions about heart disease, thereby enhancing treatment outcomes and patient care. Additionally, pharmaceutical companies can leverage the prediction outcomes to develop novel drugs or treatments for heart disease.

**Existing System:** The current approach to Heart Disease Prediction typically involves manual diagnosis by medical professionals, relying on symptoms, medical history, and diagnostic tests. The process entails the following steps:

1. Patient Consultation: Healthcare professionals gather information about the patient's medical history, lifestyle, and heart disease-related symptoms.
2. Physical Examination: Vital signs like blood pressure and heart rate are assessed, and the patient's heart sounds are examined.
3. Diagnostic Tests: Based on the initial assessment, various diagnostic tests such as blood tests, electrocardiogram (ECG), stress tests, echocardiograms, or cardiac catheterization may be conducted.
4. Analysis and Diagnosis: Test results are analyzed alongside the patient's medical history and symptoms to make a diagnosis, relying on the expertise of healthcare professionals.
5. Treatment and Management: Treatment options, medications, lifestyle modifications, or further specialist consultations are recommended based on the diagnosis.

This system heavily relies on medical professionals' expertise for accurate diagnosis and lacks the utilization of machine learning algorithms or automated prediction models to assist in the diagnosis process.

However, the Heart Disease Prediction System employing Gradient Boosting Classifier aims to augment the existing system by integrating machine learning techniques. It harnesses historical patient data, including age, sex, cholesterol levels, blood pressure, and other pertinent features, to train a model capable of predicting heart disease likelihood. This system aids healthcare professionals in making informed decisions, potentially enhancing diagnosis accuracy.

**System overview:**

We've created a system where users can input their details on a webpage to ascertain if they have heart disease. To design the user interface, we opted for the Streamlit library in Python, renowned for its seamless deployment of machine learning models. In order to ensure precise predictions, we compared various machine learning and deep learning models. Following thorough evaluation, the XGB Classifier emerged as the most effective, yielding superior results. Hence, we utilized the XGB Classifier to generate outputs for our project.

# **Implementation issues**

# Min-max scaling

Min-max scaling, also known as normalization, is a technique commonly used in data preprocessing. It is used to transform numerical features into a specific range, typically between 0 and 1. Min-max scaling can be useful in various situations, such as:

Machine Learning Algorithms: Many machine learning algorithms perform better when the input features are normalized. By scaling the features to a specific range, you can prevent any particular feature from dominating the learning process. This is especially important when working with algorithms that are sensitive to the scale of the data, such as k-nearest neighbors (KNN) and support vector machines (SVM).

Neural Networks: Deep learning models, such as neural networks, often benefit from input data that is scaled between 0 and 1. Scaling the features can speed up the convergence of the training process and improve the stability of the model.

Distance-Based Algorithms: Distance-based algorithms, like KNN, calculate the distance between data points. If the features have different scales, features with larger values can dominate the distance calculations, leading to biased results. Min-max scaling can help to alleviate this issue by putting all features on a similar scale.

Visualization: When visualizing data, it is often easier to interpret features that are on the same scale. By using min-max scaling, you can ensure that the features are within a consistent range, making it easier to compare and understand the visualizations.

However, it is important to note that min-max scaling may not always be the best choice. If your data contains outliers, they can distort the scaling and affect the results. In such cases, you might consider using other scaling techniques, such as standardization (Z-score scaling) or robust scaling, which are more robust to outliers.

The formula for min-max scaling is:

𝑋scaled=𝑋−𝑋min𝑋max−𝑋min*X*scaled​=*X*max​−*X*min​*X*−*X*min​​

where:

* 𝑋*X* is the original feature value.
* 𝑋min*X*min​ is the minimum value of the feature.
* 𝑋max*X*max​ is the maximum value of the feature.
* 𝑋scaled*X*scaled​ is the scaled feature value.

# Voting Classifier

ML | Voting Classifier using Sklearn

A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output.

It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on the highest majority of voting. The idea is instead of creating separate dedicated models and finding the accuracy for each them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class.

Voting Classifier supports two types of votings.

**Hard Voting:** In hard voting, the predicted output class is a class with the highest majority of votes i.e the class which had the highest probability of being predicted by each of the classifiers. Suppose three classifiers predicted the output class(A, A, B), so here the majority predicted A as output. Hence A will be the final prediction.

**Soft Voting:** In soft voting, the output class is the prediction based on the average of probability given to that class. Suppose given some input to three models, the prediction probability for class A = (0.30, 0.47, 0.53) and B = (0.20, 0.32, 0.40). So the average for class A is 0.4333 and B is 0.3067, the winner is clearly class A because it had the highest probability averaged by each classifier.

# XGB Classifier

The XGB Classifier, short for Extreme Gradient Boosting Classifier, is a machine learning algorithm known for its high performance and efficiency in handling large datasets. It belongs to the family of gradient boosting algorithms, which work by iteratively training weak learners (decision trees, in the case of XGB) to correct the errors made by the preceding models. This iterative process continues until the algorithm converges to a strong learner that provides accurate predictions.

XGB Classifier utilizes a boosting technique that combines the predictions from multiple weak learners to make a final prediction. It focuses on reducing errors by adjusting the weights of misclassified data points in subsequent iterations, resulting in a more accurate model.

One of the key advantages of XGB Classifier is its ability to handle complex datasets and capture intricate patterns in the data. It also offers various hyperparameters that can be tuned to optimize model performance, such as learning rate, tree depth, and regularization parameters.

Overall, the XGB Classifier is a powerful and versatile algorithm suitable for a wide range of classification tasks, including but not limited to heart disease prediction. Its effectiveness lies in its ability to produce highly accurate predictions while efficiently handling large-scale datasets.

# Sequence models

**Sequence models** are a specialized category of machine learning models crafted for handling tasks involving sequential data, where the arrangement of elements in the input holds significance. Such data spans textual data, time series data, audio signals, video streams, or any other ordered data, with varying lengths and interdependent sequence elements. Unlike conventional machine learning algorithms, sequence models are engineered to tackle data that demonstrates dependencies rather than being independently and identically distributed (i.i.d.).

Key Sequence Models:

Recurrent Neural Networks (RNNs): RNNs serve as a foundational form of sequence model. They process sequences incrementally while maintaining an internal hidden state that encodes information about preceding elements in the sequence. This mechanism enables RNNs to capture dependencies across different time steps. However, traditional RNNs grapple with the "vanishing gradient" issue, which hampers their ability to effectively capture long-range dependencies.

Long Short-Term Memory (LSTM) Networks: LSTMs, a variant of RNNs, are engineered to address the vanishing gradient problem. They incorporate specialized memory cells and gating mechanisms, enabling them to retain and leverage information across extended sequences proficiently.

Gated Recurrent Units (GRUs): GRUs represent another variant of RNNs, akin to LSTMs but featuring a simpler architecture. They employ gating mechanisms to regulate information flow within the network. GRUs exhibit computational efficiency compared to LSTMs while still adeptly capturing dependencies in sequential data.

Transformer Models: Transformers stand as a more contemporary and highly efficient architecture for sequence modeling. They leverage a self-attention mechanism to process sequences concurrently, facilitating the capture of long-term dependencies within data. Transformers have enjoyed significant success, particularly in natural language processing (NLP) tasks, leading to the development of prominent models like BERT and GPT.

# Saving trained model with pickle

The pickle module can be used to serialize and deserialize the Python objects. Pickling is the process of converting a Python object hierarchy into a byte stream, while Unpickling is the process of converting a byte stream (from a binary file or other object that appears to be made of bytes) back to an object hierarchy.

For saving the ML models used as a pickle file, you need to use the Pickle module that already comes with the default Python installation.

To save your iris classifier model you simply need to decide on a filename and dump your model to a pickle file like this:

import pickle

# save the iris classification model as a pickle file

model\_pkl\_file = "iris\_classifier\_model.pkl"

with open(model\_pkl\_file, 'wb') as file:

pickle.dump(model, file)

As you can see the file is opened in wb (write binary) mode for saving the model as bytes. Also, the dump() method stores the model in the given pickle file.

You can also load this model using the load() method of the pickle module. Now you need to open the file in rb (read binary) mode to load the saved model.

# load model from pickle file

with open(model\_pkl\_file, 'rb') as file:

model = pickle.load(file)

# evaluate model

y\_predict = model.predict(X\_test)

# check results

print(classification\_report(y\_test, y\_predict))

Once loaded you can use this model to make predictions.

Pros of the Python pickle approach

* 1. Pickling comes as the standard module in Python which makes it easy to use for saving and restoring ML models.
  2. Pickle files can handle most Python objects including custom objects, making it a versatile way to save models.
  3. For small models, pickle approach is quite fast and efficient.
  4. When an ML model is unpickled, it is restored to its previous state, including any variables or configurations. This makes Python pickle files one of the best alternatives for saving ML models.

Cons of the Python Pickle Approach

1. If you unpickle untrusted data, pickling could pose a security threat. Unpickling an object can execute malicious code, so it’s crucial to only unpickle information from reliable sources.
2. Pickled objects’ use may be constrained in some circumstances since they cannot be transferred between different Python versions or operating systems.
3. For models with a big memory footprint, pickling can result in the creation of huge files, which can be problematic.
4. Pickling can make it difficult to track changes to a model over time, especially if the model is updated frequently and it is not feasible to create multiple pickle files for different versions of models that you try.

Pickle is most suited for small-size models and also has some security issues, these reasons are enough to look for another alternative for saving the ML models. Next, let’s discuss Joblib to save and load ML models.

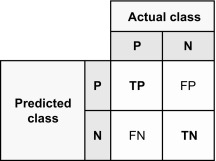
# Result Received

We have used multiple machine learning model like

A screenshot of a computer

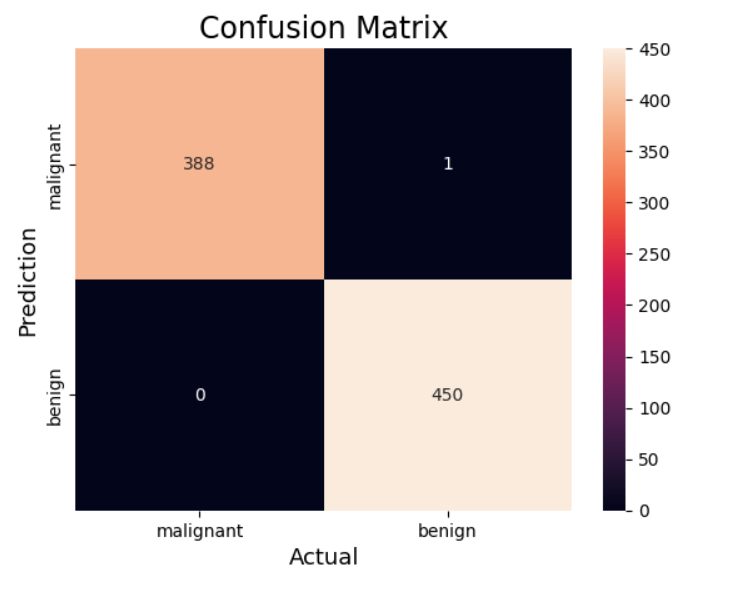
Description automatically generated

Confusion matrix :  
A confusion matrix represents the prediction summary in matrix form. It shows how many prediction are correct and incorrect per class. It helps in understanding the classes that are being confused by model as other class.



TP here stands for True Positive predictions, for a binary classification problem like classifying the fraudulent transactions as 1, TP will give the count of the number of 1 s that were correctly classified as 1, i.e., number of fraudulent transactions that were classified as fraudulent. TN stands for true negative predictions, i.e., number of 0 s, non-fraudulent transactions, classified as 0. FP (False Positive) is the count of number of non-fraudulent transactions that were classified as fraudulent and FN (False Negative) is the count of number of fraudulent transactions that were classified as non-fraudulent.

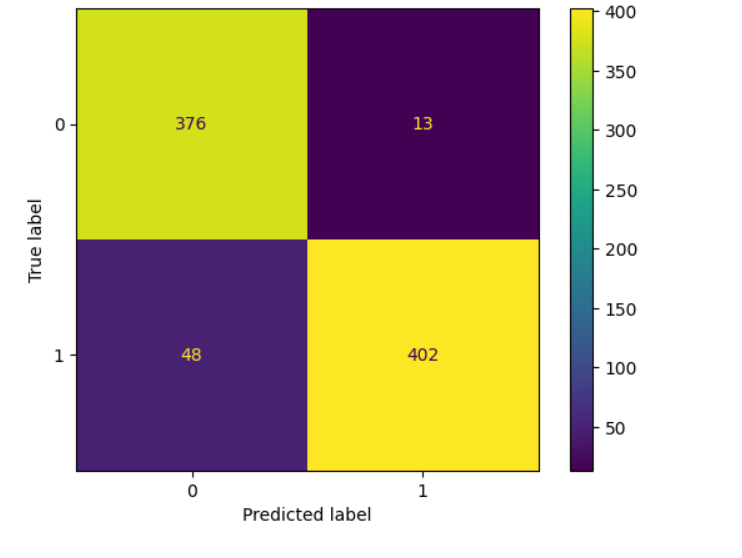
Confusion Matrix of xgb classifier:



**Malignant** is an adjective that’s defined as “disposed to cause harm, suffering, or distress deliberately; feeling or showing ill will or hatred.” It describes something that is very dangerous or harmful in influence or effect.

**Benign** is an adjective that means having a kindly or gracious disposition. Benign can also refer to “showing or expressive of gentleness or kindness” or “favorable; propitious.”

Confusion Matrix in Deep learning:



# User Screens

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**CONCLUSION**

**Advantages & Limitations**

**Advantages of the "Heart Disease Prediction System":**

Early Detection: This system identifies heart diseases early, enabling prompt treatment and minimizing complications.

Precision: Powered by the Gradient Boosting Classifier algorithm, it ensures accurate predictions.

Time Efficiency: Providing quick predictions, it saves time for medical practitioners and patients.

Cost Savings: Early detection reduces treatment and hospitalization expenses, while also easing the workload of healthcare professionals, leading to financial savings.

Enhanced Healthcare: By offering precise predictions, it supports healthcare providers in delivering superior patient care.

**Limitations of the "Heart Disease Prediction System":**

Data Quality Dependency: Accuracy relies heavily on data quality; incomplete or biased datasets can lead to inaccurate predictions.

Restricted Feature Set: The predictive efficacy depends on including relevant features during model training.

Overfitting or Underfitting: Performance may suffer from overfitting or underfitting, resulting in suboptimal predictions.

Interpretability Challenges: Gradient Boosting Classifiers may lack interpretability compared to simpler models.

Medical Expertise Dependency: The system complements medical expertise but does not replace it; final decisions should be made by qualified practitioners based on thorough patient evaluations.

In summary, our investigation has revealed that the XGB Classifier stands out as the most effective among the range of machine learning models and deep learning techniques we explored. Its proficiency in accurately predicting heart disease risk in patients marks a significant advancement. Through the analysis of patient data using machine learning and deep learning algorithms, our system endeavors to support both users and healthcare professionals in achieving timely and precise diagnoses. This, in turn, holds promise for early disease detection, improved patient outcomes, and enhanced healthcare delivery standards. Furthermore, the system's intuitive user interface, accessible from anywhere via the internet, amplifies its potential impact on healthcare accessibility and efficiency.

Key advantages of the system include its capacity to efficiently analyze large datasets, its heightened prediction accuracy, and its user-friendly interface, which enables users to register, log in, make predictions, review prediction history, and access pertinent information about healthcare providers. Administrative functions, including user account management, prediction result tracking, and other administrative tasks, are also streamlined.

While the system does have its limitations, such as reliance on precise and comprehensive datasets, ongoing research and advancements in machine learning techniques hold promise for addressing these challenges and refining its performance. Maintaining the privacy and security of patient data, alongside adherence to relevant regulations and ethical guidelines, remains paramount.

The Heart Disease Prediction System bears significant potential for the medical industry, healthcare professionals, and individuals seeking to assess their heart disease risk. It serves as an efficient and accessible platform for early detection and personalized interventions, thereby contributing to improved patient outcomes and a reduction in heart disease-related complications.

Looking ahead, further enhancements to the system could involve integrating advanced algorithms, collaborating with other healthcare systems and technologies, and broadening its capabilities to offer more tailored risk assessments and preventive measures. By continually enhancing and updating the system, we can advance towards more effective management and prevention of heart disease, ultimately fostering better overall health and well-being for individuals.

**Bibliography**