**Logo, company name

Description automatically generated**Graduation Project

**Prediction and Classification Techniques Applied to Medical Patient Data**

**Team Members:**

**Names ID Department**

Osama Ali Elgendy202018546IS

Eman Ashraf Bahgat 202017236CS

Mohamed Abdulaziz Hosni202017167CS

Mahmoud Aldalil Abdelkarim202001501CS

Mahmoud Ragab Saber 202006723CS

Maryam Mohamed Ali 202016808 CS

Merna Abdo Youssef 202018975 CS

**Supervised By:**

Dr. Shreen Taie

Dr. Albeer Shawky

**August-2023**

**Acknowledgment**

**Acknowledgment**

Although it was a challenging journey, we are here standing and discussing this project that helped us to be patient and hardworking. So, we would like to thank our supervisors.

Dr/ Shreen Taie

Dr/ Albeer Shawky

for their encouragement, guidance, patience, and support throughout the journey.

Also, we want to thank our university professors who truly cared to teach us anything, academically.

We also want to send a special thanks to:

Eng/ Hageer

Eng/ Ghada Abdullah

for their efforts and patience with us.

And in the general life. To our parents who raised us with love and helped us to become who we are. And to everyone who left the smallest, beautiful impression that kept us believing in the good in this world.

**Abstract**

**Abstract**

The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time. In this paper different machine learning algorithms and deep learning are applied to compare the results and analysis of the UCI Machine Learning Heart Disease dataset. The dataset consists of 14 main attributes used for performing the analysis. Various promising results are achieved and are validated using accuracy and confusion matrix. The dataset consists of some irrelevant features which are handled using Isolation Forest, and data are also normalized for getting better results. And how this study can be combined with some multimedia technology like mobile devices is also discussed. Using deep learning approach.

**Chapter one**

Introduction

**1.1 Overview**

The main objective behind data mining applications is to determine that the data, fact, number, text, etc., can be processed by a software system that yields useful knowledge. Data mining is an interactive and iterative process. It is the detection of automatic and semi-automatic correlation and pattern changes, anomalies, various structures and also events in the data. The main purpose of implementing data mining classification techniques on a mental health care dataset is to develop an automated tool to identify, identify and disseminate relevant mental health care information. This paper aims to assist healthcare professionals in decision making by performing a predictive analysis of a mental healthcare dataset using Wicca classifiers. We mainly implemented 3 classifiers - support vector machine, XGBoost, and Random Forest.

SVM is the main idea behind SVMs is to find a hyperplane that maximally separates the different classes in the training data. This is done by finding the hyperplane that has the largest margin, which is defined as the distance between the hyperplane and the closest data points from each class.

XGBoost for “Extreme Gradient Boosting” is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models. It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction.

Random forest is a machine learning algorithm that uses a collection of decision trees providing more flexibility, accuracy, and ease of access in the output. This algorithm dominates over decision trees algorithm as decision trees provide poor accuracy as compared to the random forest algorithm. In simple words, the random forest approach increases the performance of decision trees. It is one of the best algorithms as it can use both classification and regression techniques.

**1.2 Problem Definition**

The importance of early detection of cardiovascular diseases and cancer is what makes the difference between life and death, by recognizing the early signs of heart disease and cancer such as heart attack.

With the help of our project, we aim at early detection of these diseases to help both the doctor and the patient, so you will have a better chance of treatment, improvement in condition and avoidance of serious complications, leading to a reduction in the number of deaths from cardiovascular diseases and cancer.

**1.3 The proposed model**

Nowadays, many heart’s patients suffer from chronic diseases. Heart diseases that can kill a person fast, although there are many solutions to solve this problem. Among these solutions, which require a quick test on the patient's future condition as a result of the signals coming from the heart. We have to add a web application to do a comprehensive examination for 2 diseases including (..................). We look forward to continuing the path of researchers who have been busy with this technology and development on it to show more accurate results to help the doctor in the examination and give the appropriate medication before this process occurs.

That will help us to rescue our patient on early stages and reduce death cases which happen because of Congestive Heart Failure or Atrial Fibrillation.

**1.4 Project Scope**

With the help of our project, we aim to early detection of these diseases in order to help both the doctor and the patient, and thus you will have a better chance of treatment, improving the condition and avoiding serious complications, resulting in a reduction in the number of deaths resulting from these diseases and emergency rooms in hospitals that can be followed up quickly for operations if necessary. As well as Sports Field, then direct transfer to the hospital for treatment.

**Chapter Two**

Literature Review

**2.1 Overview**

This chapter is about the previous applications that are similar to our proposed model and nearly the same objectives.

**2.2 HEART DISEASE ANALYSIS & PREDICTION**

On this site, for example, they present the results of the analysis based on a giant database, just have to enter some data.

<https://byte7.github.io/projects/heart_disease_analysis/index.html>

**2.3 MYELIN**

Myelin is a hospital management information system designed to manage the work of medical, administrative, and financial cycle in hospitals. Myelin helps to transfer work style within these hospitals from the traditional paper-based methods to modern technological methods to save time and efforts to raise productivity and the profitability.

<https://www.codezone-eg.com/en/product/myelin?gclid=CjwKCAiA0JKfBhBIEiwAPhZXDxKPPbhPGgERKJ0KaAsg0rObob2A3rWysxRlvAYAIGHbrZYAAcYEBBoCvxYQAvD_BwE>

**2.4 Vezeeta**

Vezeeta.com is the leading digital healthcare booking platform and practice management software in MENA. They are pioneering the shift to automated physician, clinic and hospital bookings making healthcare easily accessible in the region.

<https://www.vezeeta.com/en>

**Chapter Three**

PLANNING & SELECTION

Assess scope and complexity of project then establish procedures to support activities Here we define. Clear, discrete activities and the work needed to complete each activity (Table 3,1)

Table 3.1: Project plan

|  |  |  |
| --- | --- | --- |
| Task | Start Time | End Time |
| Improve Team skills | 15-Sep | 15-Oct |
| Do surveys | 16-Oct | 25-Oct |
| Select project | 26-Oct | 31-Oct |
| Do Analysis | 1-Nov | 7-Nov |
| Download dataset | 8-Nov | 10-Nov |
| Work on frontend | 11-Nov | 5-Dec |
| Exams | | |
| Start to work on data set | 2-Feb | 12-Feb |
| Build the model |  |  |
| Connect the model with the website |  |  |
| Test |  |  |
| Modify & maintain |  |  |
| Final Release |  | |

As a completing step we used Gantt chart which is a diagram that represent our tasks and duration and dependency of each on other (Figure 3.1)

**Chapter** **Four**

**System Analysis**