**Mini Project Report(KCS-554)**

**on**

**Multiple Diseases Prediction System**

**Submitted in partial fulfillment for award of**

**BACHELOR OF TECHNOLOGY**

**Degree**

**In**

**COMPUTER SCIENCE & ENGINEERING**



**2022-23**

**Under the Guidance of: Submitted By:**

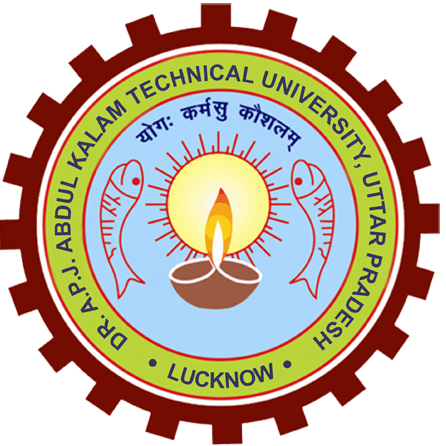
**Mr.Vineet Srivastava** **Kaushki Gupta (119)**

**Assistant Professor**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**RAJ KUMAR GOEL INSTITUTE OF TECHNOLOGY**

**DELHI-MEERUT ROAD, GHAZIABAD**



**Affiliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow**

**November 2022**

TABLE OF CONTENTS

**CHAPTER NO. TITLE PAGE NO.**

**SYNOPSIS iii**

**LIST OF TABLES ix**

**LIST OF FIGURES x**

**1. INTRODUCTION** 1

1.1 DESCRIPTION 1

1.2 PROBLEM STATEMENT 2

1.3 PROPOSED SYSTEM 2

**2. LITERATURE REVIEW** 3

**3. SYSTEM ANALYSIS** 4

**3.1 FUNCTIONAL REQUIREMENT 4**

**3.2 NON – FUNCTIONAL REQUIREMENT 4**

**4. DESIGN FLOW** 5

**4.1 BLOCK DIAGRAM 5**

**4.2 DESIGN FLOW 6**

**5. CONCLUSION** 7

**6. FUTURE SCOPE** 7

**7. REFERENCES** 8

**LIST OF TABLES**

**CHAPTER NO. TABLE NO. TITLE PAGE NO.**

2 Table 2.1 Audio Stream with Encoded 32

4 Table 4.1 Guitar.wav 68

4 Table 4.2 Jhankar.wav 69

4 Table 4.3 Mohabatten.wav 70

4 Table 4.4 Pahalanasha.wav 70

4 Table 4.5 Mean Opinion Score MOS 72

LIST OF FIGURES

**CHAPTER NO. TITLE PAGE NO.**

1 Figure 1.1: Categories of Steganography 6

2 Figure 2.1: Microsoft RIFF File Format 23

2 Figure 2.2: Encoding Message HEY in 16 bit CD Quality

Audio steganography at LSB 31

1. Figure 2.3: Sampling of the size wave followed by

Quantization Process 32

2 Figure 2.4: Hiding 1 bit of image in 16 bit quantized audio

sample 33

1. Figure 2.5: Hiding 4 bit of Image in 16 bit Quantized audio signal 33
2. Figure 2.6 : Illustrated the original signal and encoded signal

of phase coding Technique. 39

## CHAPTER 1

**INTRODUCTION**

In this digital world, data is an asset, and enormous data was generated in all the fields. Data in the healthcare industry consists of all the information related to patients. Here a general architecture has been proposed for predicting the disease in the healthcare industry. Many of the existing models are concentrating on one disease per analysis. Like one analysis for diabetes analysis, one for cancer analysis, one for skin diseases like that. There is no common system present that can analyze more than one disease at a time. Thus, we are concentrating on providing immediate and accurate disease predictions to the users about the symptoms they enter along with the disease predicted. So, we are proposing a system which used to predict multiple diseases by using Django. In this system, we are going to analyze Diabetes, Heart, and malaria disease analysis. Later many more diseases can be included. To implement multiple disease prediction systems we are going to use machine learning algorithms, and Django. Python pickling is used to save the behavior of the model. The importance of this system analysis is that while analyzing the diseases all the parameters which cause the disease is included so it is possible to detect the disease efficiently and more accurately. The final model's behavior will be saved as a python pickle file.

* 1. **DESCRIPTION**

A lot of analysis over existing systems in the health care industry considered only one disease at a time. For example, one system is used to analysediabetes, another is used to analyse diabetes retinopathy, and another system is used to predict heart disease. Maximum systems focus on a particular disease. When an organization wants to analyse their patient’s health reports then they have to deploy many models. The approach in the existing system is useful to analyse only particular diseases. In multiple diseases prediction system a user can analyse more than one disease on a single website. The user doesn’t need to traverse different places in order to predict whether he/she has a particular disease or not. In multiple diseases prediction system, the user needs to select the name of the particular disease, enter its parameters and just click on submit. The corresponding machine learning model will be invoked and it would predict the output and display it on the screen.

* 1. **PROBLEM STATEMENT**

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example first is for liver analysis, one for cancer analysis, one for lung diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients’ health. When an organization wants to analyse their patient’s health reports, they have to deploy many models which in turn increases the cost as well as time Some of the existing systems consider very few parameters which can yield false results.

* 1. **PROPOSED SYSTEM**

In multiple disease prediction, it is possible to predict more than one disease at a time. So the user doesn’t need to traverse different sites in order to predict the diseases. We are taking three diseases that are Liver, Diabetes, and Heart . As all the three diseases are correlated to each other. To implement multiple disease analyses we are going to use machine learning algorithms and Django. When the user is accessing this API, the user has to send the parameters of the disease along with the disease name. Django will invoke the corresponding model and returns the status of the patient.

## CHAPTER 2

**LITERATURE REVIEW**

According to the paper focuses about as diabetes is one of the dangerous diseases in the world , it can cause many varieties of disorders which includes blindness etc .In this paper they have used machine learning techniques to find out diabetes disease as it is easy and flexible to forecast whether the patient has illness or not . Their aim of this analysis was to invent a system that can help the patient to detect the diabetes disease of the patient with accurate results. Here they used mainly 4 main algorithms Decision Tree , Naïve Bayes , and SVM algorithms and compared their accuracy which is 85%,77%, 77.3% respectively . They also used ANN algorithm after the training process to see the reactions of the network which states whether the disease is classified properly or not . Here they compared the precision recall and F1 score support ad accuracy of all the models[1] .

The main aim of the paper is ,as heart plays an important role in living organisms. So the diagnosis and prediction of heart related disease should be perfect and correct because it is very crucial which can cause death cases related to heart .So Machine learning and Artificial Intelligence supports in predicting any kind of natural events .So in this paper they calculate accuracy of machine learning for predicting heart disease using k-nearest neighbor ,decision tree, linear regression and SVM by using UCI repositor dataset for training and testing . They also compared the algorithm and their accuracy SVM 83 %,Decision tree 79%,Linear regression 78%,k-nearest neighbour 87%[2]. 3. The system defines that liver diseases is causing high number of deaths in India and is also considered as a life threating disease in the world. As it is difficult to detect the liver disease at early stage .So using automated program using machine learning algorithms we can detect the liver disease accurately .They used and compared SVM ,Decision Tree and Random forest algorithm and measures precision, accuracy and recall metrics for quantitative measurement. The accuracy are 95%,87%,92% respectively[3] .

## CHAPTER 3

**SYSTEM ANALYSIS**

**3.1 FUNCTIONAL REQUIREMENT**

• The system allows the patient to predict the disease

• The user adds the input for the particular disease and based on the trained model of the user input the output will be displayed .

**3.2 NON FUNCTIONAL REQUIREMENT**

• The website will provide range of the values during the prediction of the disease.

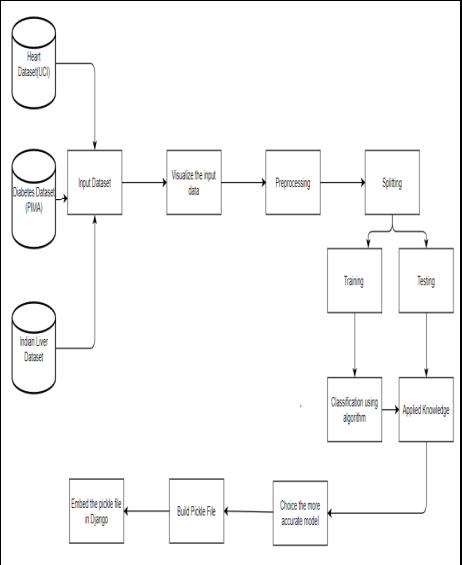
• The website should be reliable and consistent.

## CHAPTER 4

**DESIGN FLOW**

**4.1 BLOCK DIAGRAM**

We have experimented on three diseases that is heart,daibetes and liver as these are correlated to each other.The first step is to the dataset for heart disease,daibetes disease and liver disease we have imported the UCI dataset,PIMA dataset and Indian liver dataset respectively.Once we have imported the dataset then visualization of each inputed data takes place.



After visualization pre-processing of data takes place where we check for outliers , missing values and also scale the dataset then on the updated dataset we split the data into training and testing .Next is on the training dataset we had applied k-nearest neighbor, xgboost and random forest algorithm and applied knowledge on the classified algorithm using testing dataset. After applying knowledge we will choose the algorithm with the best accuracy for each of the disease Then we build a pickle file for all the disease and then integrated the pickle file with the Django framework for the output of the model on the webpage.

**4.2 RESULT**

In the system diabetes disease prediction model used knn algorithm, heart disease uses the xgboost algorithm and liver uses the random forest algorithm as these gave the best accuracy accordingly. There when the patient adds the parameter according to the disease it will show whether the patient has a disease or not according to the disease selected. The parameters will show the range of the values needed and if the value is not between the range or is not valid or is empty it will show the warning sign that add a correct value.

## CHAPTER 5

**CONCLUSION**

The main objective of this project was to create a system that would predict more than one disease and do so with high accuracy. Because of this project the user doesn’t need to traverse different websites which saves time as well. Diseases if predicted early can increase your life expectancy as well as save you from financial troubles. For this purpose, we have used various machine learning algorithms like Random Forest, XGBoost, and K nearest neighbor (KNN) to achieve maximum accuracy.

## CHAPTER 6

**FUTURE SCOPE**

* In the future we can add more diseases in the existing API.
* We can try to improve the accuracy of prediction in order to decrease the mortality rate.
* Try to make the system user-friendly and provide a chatbot for normal queries.

**REFERENCES**

[1] Priyanka Sonar, Prof. K. JayaMalini,” DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES”, 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC) .

[2] Archana Singh ,Rakesh Kumar, “Heart Disease Prediction Using Machine Learning Algorithms”, 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3) .

[3] A.Sivasangari, Baddigam Jaya Krishna Reddy,Annamareddy Kiran, P.Ajitha,” Diagnosis of Liver Disease using Machine Learning Models” 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).