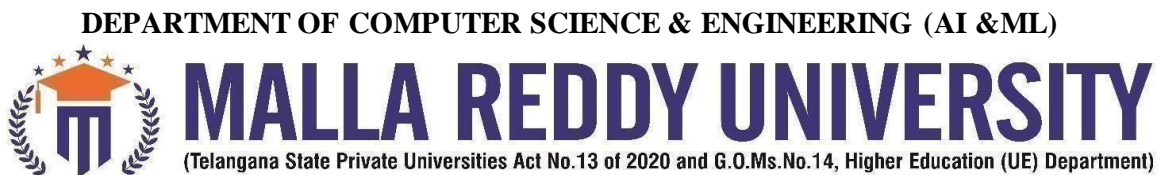# LIVER DISEASE PREDICTION USING DEEP LEARNING.

A project report submitted
to
MALLA REDDY
UNIVERSITY
in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE & ENGINEERING (AI & ML)

Submitted by

A. SRI CHARAN	:	2011CS020019
A. SURYA PAVAN	:	2011CS020020
A. V. S. PAVAN	:	2011CS020021
A. VIJAY	:	2011CS020022

Under the Guidance of
Prof.Veena Rani
(Asst.Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (AI &ML)

MALLA REDDY UNIVERSITY
(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

2023

MALLA REDDY UNIVERSITY

(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

<u>COLLEGE CERTIFICATE</u>

This is to certify that this is the bonafide record of the application development entitled, **"LIVER DISEASE PREDICTION USING DEEP LEARNING**" submitted by **A.SRI CHARAN (2011CS020019), A.SURYA PAVAN(2011CS020020), A.V.S. PAVAN (2011CS020021), A.VIJAY(2011CS020022)** of B.Tech III year II semester, Department of CSE (AI&ML) during the year 2022- 23.The results embodied in the report have not been submitted to any other university or institute for the award of any degree or diploma

INTERNAL GUIDE   **HEAD OF THE DEPARTMENT**

Prof. Veena Rani   **Dr. Thayyaba Khatoon**
(Asst. Professor)   **(Professor)**

ACKNOWLEDGEMENT

We would like to express our gratitude to all those who extended their support and suggestions to come up with this application. Special Thanks to our mentor Asst.Prof. Veena Rani whose help and stimulating suggestions and encouragement helped us all time in the due course of project development.

We sincerely thank our HOD Dr. Thayyaba Khatoon for her constant support and motivation all the time. A special acknowledgement goes to a friend who enthused us from the back stage. Last but not the least our sincere appreciation goes to our family who has been tolerant understanding our moods, and extending timely support.

ABSTRACT

Liver, a crucial interior organ of the human body whose principal tasks are to eliminate generated waste produced by our organism, digest food, and preserve vitamins and energy materials. The liver disorder can cause various fatal diseases, including liver cancer. Early diagnosis, and treating the patients are compulsory to reduce the risk of those lethal diseases. There are various diseases associated with the human liver, some of which are hard to detect using just the information exchanged between a patient and a doctor. Motivated by the vast potential of AI in medicine, in this study, we attempted to find a model which can predict the occurrence of liver disease in a given patient with the highest accuracy, based on different input factors.

A dataset was chosen to train and test this model; Indian Liver Patient Dataset obtained . We implemented different deep learning algorithms (Multi-Layer Perceptron, Stochastic Gradient Descent, Restricted Boltzmann) and filtered out the DL-based MLP (Multi-Layer Perceptron) model as the one providing the highest Accuracy, which was compared for each model along with the Precision, Recall and f1 scores. This research aims to impart insight additional to the current state-of-the-art discoveries by focusing on a comparative analysis of some of the best ML/DL techniques which haven't been scrutinized altogether yet.

1. INTRODUCTION

With a growing trend of sedentary and lack of physical activities, diseases related to liver have become a common encounter nowadays. In rural areas the intensity is still manageable, but in urban areas, and especially metropolitan areas the liver disease is a very common sighting nowadays. Liver diseases cause millions of deaths every year. Viral hepatitis alone causes 1.34 million deaths every year. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An early diagnosis of liver problems will increase patient's survival rate. Liver failures are at high rate of risk among Indians. It is expected that by 2025 India may become the World Capital for Liver Diseases. The widespread occurrence of liver infection in India is contributed due to deskbound lifestyle, increased alcohol consumption and smoking. There are about 100 types of liver infections. With such alarming figures, it is necessary to have a concern towards tackling these diseases. Afterall, we cannot expect a developed and prosperous nation, with unhealthy youths. In this project we have taken UCI ILPD Dataset which contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos and contains 415 as liver disease patients and 167 as non liver disease patients.As we got through the next parts of this paper we will explain what proccess as taken place for the selection of best model and building necccessary sytem for the prediction of liver disease.

liver is well-thought-out to be one of the central organs in any living body with fundamental functions such as processing leftover products, generating enzymes, and eliminating exhausted tissues or cells . We can stay alive merely a couple of days if our liver shuts down. Fortunately, the liver can continue its role even when up to 75% of it is contaminated or removed. is due to its astonishing capability to produce new liver tissues from finefettle liver cells that quiet exist . It shows a significant role in several bodily functions such as protein creation and blood clotting to glucose (sugar), cholesterol, and iron metabolism. It has a range of functions, comprising eliminating toxins from the body, and is crucial for survival . harm of these functions can reason to momentous destruction to the body. Once the liver is diseased with a virus, injured by chemicals, or under attack from its immune system, the elementary hazard is similar; that is, the liver will become so spoiled that it can no lengthier retain an individual alive.

According to World Health Organization (WHO) and World Gastroenterology Organization (WGO), 35 million individuals pass away due to chronic diseases, and liver failure is one of the apprehensive diseases stated . It is further stated that more than 50 million grown-ups will be affected with chronic liver disease (CLD), and it requests for instantaneous responsiveness for actions in a conference held in Paris that deliberated the shocking drifts of liver disease worldwide . Moreover, agreeing to the current figures, 25 million US residents are pretentious by the liver or biliary ailment, and out of these, 50% populace have no symptoms. In the United Kingdom, nearly 25% of death due to liver disease is from extreme alcohol drinking.

As soon as the liver becomes diseased, it can ground severe destruction to our health. There can be numerous equipment and health conditions that can naively reason for liver damage . Alcohol. Dense alcohol drinking is the utmost collective reason for liver damage. Once individuals drink alcohol, the liver becomes distracted from its other functions and provides attention mostly on converting alcohol into a smaller amount of toxic form. Obesity. People who are fat have the leftover quantity of body obese which inclines to accrue nearby the liver causing fatty liver disease (FLD). Diabetes. Devising diabetes upturns the hazard of liver disease by 50 percent. Increased level of compelling insulin results in FLD.

Hepatitis. It is an ailment produced by a virus feast due to manure pollution or direct interaction with the septic bloody fluids . Cirrhosis. It is the utmost severe liver disease that happens when normal liver cells are swapped by mutilation tissue as the CLD. Liver Cancer. the danger of consuming liver cancer is higher for individuals who have cirrhosis and another type of hepatitis . In the current era, we have been confronted with a cumulative amount of records kept in several societies such as hospitals, universities, and banks that inspire us to discover an approach to mine information from this huge number of records and to proficiently use them, especially in the healthcare organizations. In the recent era, researchers are focusing on using data from healthcare organisations for early and accurate prediction of syndromes. Nowadays, data mining (DM) and machine learning (ML) become elementary in healthcare due to its approaches, e.g., classification, clustering, and association rule mining, for determining repeated patterns pragmatic for disease extrapolation on medical data .

2.Literature survey

Liver Disease Prediction by using different Decision Tree techniques
Authors: Nazmun Nahar and Ferdous Ara

Early prediction of sickness is incredibly necessary to save lots of human life and take correct steps to regulate the disease call Tree algorithms are with success applied in varied fields, particularly in life science. This analysis work explores the first prediction of disease exploitation varied call tree techniques. The disease dataset that is chosen for this study is consisting of attributes like total animal pigment, direct animal pigment, age, gender, total proteins, simple protein and simple protein magnitude relation. The most purpose of this work is to calculate the performance of varied call tree techniques and compare their performance. The choice tree techniques employed in this study area unit J48, LMT, Random Forest, Random tree, REP Tree, call Stump, and Hoeffding Tree. The analysis proves that call Stump provides the very best accuracy than different techniques.

Decision Tree (DT): Decision Tree calculation has a place with the supervised learning algorithms In contrast to other supervised learning algorithms, a decision tree algorithm can be utilized for taking care of regression and classification issues as well. The general thought process of utilizing Decision Tree is to make a training model that can use to predict class or estimation of objective factors by taking in choice standards derived from earlier data (training data).

Liver Disease Prediction using SVM and Naïve Bayes Algorithms
Authors: Dr. S. Vijayarani1 , Mr.S.Dhayanand

Data mining has become an easy use for disease prediction in recent years in health care sectors. Data mining is the dredging process of information from massive datasets or warehouses or other repositories. Predicting the diseases from the voluminous medical databases is a very challenging task for researchers. The researchers use data mining techniques such as classification, clustering, rules of association and so on to over come this issue. The main objective of this research work is to use classification algorithms to predict liver diseases. Naïve Bayes and support vector machine (SVM) are the algorithms used in this work. The algorithms are classify the factors on performance. These classifier algorithms are compared on the basis of performance factors i.e. accuracy of classification and time of execution. It is observed from the experimental results that the `SVM is a better classifier for liver disease prediction.

SVM Algorithm :

SVM algorithm tries to give out hyper planes and split the data into different categories. The scikit-learn package in python is employed for implementing SVM. The pre-processed information is split into check information and coaching set that is of twenty fifth and seventy fifth of the entire dataset severally. A SVM technique builds hyper planes in an exceedingly dimensional area. a decent separation is achieved by the hyper plane that has the most important distance to the closest coaching information of any category (so-called purposeful margin), since generally the larger the margin the lower the generalization error of the classifier.

Naïve Bayes:

Naïve Bayes is one of the basic probabilistic classifiers which classifies the specific class with the given tuple. It is categorised by hypothesising that every attribute has a solitary effect on the class attribute by not depending on other attribute values.

Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques
Authors : Insha Arshad , Chiranjit Dutta.

Many people are consuming alcohol in the current period. Nowadays alcohol consumption is directly associated with threatening liver disease called cirrhosis. Early detection of liver disease caused by excessive alcohol consumption would help many people save lives It can be diagnosed in time by detecting liver disease at its early stage and can lead to full recovery in some patients. This paper proposes that the presence of liver disease be detected and predicted using data mining algorithms. We will make a dataset decision tree and then generate the rules. To coach and check the dataset.

3.PROPOSED METHODOLOGY

3.1 Introduction:

Liver is well-thought-out to be one of the central organs in any living body with fundamental functions such as processing leftover products, generating enzymes, and eliminating exhausted tissues or cells. We can stay alive merely a couple of days if our liver shuts down. Fortunately, the liver can continue its role even when up to 75% of it is contaminated or removed. Is due to its astonishing capability to produce new liver tissues from fine fettle liver cells that quiet exist . It shows a significant role in several bodily functions such as protein creation and blood clotting to glucose (sugar), cholesterol, and iron metabolism. It has a range of functions, comprising eliminating toxins from the body, and is crucial for survival. Harm of these functions can reason to momentous destruction to the body. According to World Health Organization (WHO) and World Gastroenterology Organization (WGO), 35 million individuals pass away due to chronic diseases, and liver failure is one of the apprehensive diseases stated. It is further stated that more than 50 million grown-ups will be affected with chronic liver disease (CLD), and it requests for instantaneous responsiveness for actions in a conference held in Paris that deliberated the shocking drifts of liver disease worldwide. Moreover, agreeing to the current figures, 25 million US residents are pretentious by the liver or biliary ailment, and out of these, 50% populace have no symptoms. In the United Kingdom, nearly 25% of death due to liver disease is from extreme alcohol drinking.all the machine learning techniques are unable to provide high accuracy because of large amount of data .

3.2 Existing System:

- The liver disease prediction is performed by various machine learning techniques.
- Algorithms like,
- **support vector machine**: SVM algorithm tries to give out hyper planes and split the data into different categories. The scikit-learn package in python is employed for implementing SVM. The pre-processed information is split into check information and coaching set that is of twenty fifth and seventy fifth of the entire dataset severally. A SVM technique builds hyper planes in an exceedingly dimensional area. a decent separation is achieved by the hyper plane that has the most important distance to the closest coaching information of any category (so-called purposeful margin), since generally the larger the margin the lower the generalization error of the classifier.

- **decision trees:** Decision Tree calculation has a place with the supervised learning algorithms In contrast to other supervised learning algorithms, a decision tree algorithm can be utilized for taking care of regression and classification issues as well. The general thought process of utilizing Decision Tree is to make a training model that can use to predict class or estimation of objective factors by taking in choice standards derived from earlier data (training data).

- **naïve bayes :** Naïve Bayes is one of the basic probabilistic classifiers which classifies the specific class with the given tuple. It is categorised by hypothesising that every attribute has a solitary effect on the class attribute by not depending on other attribute values.

  These algorithms are performed to predict liver disease prediction in existing systems.
- But due to the large amount of data obtaining accuracy was difficult.

The accuracies provided by the existing system algorithms are shown below.

Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	TP Rate	Precision	F Measure
Naïve Bayes	61.28	38.72	0.612	0.558	0.251
SVM	79.66	20.34	0.796	0.766	0.331

The flow chart of existing system are:

III. METHODOLOGY
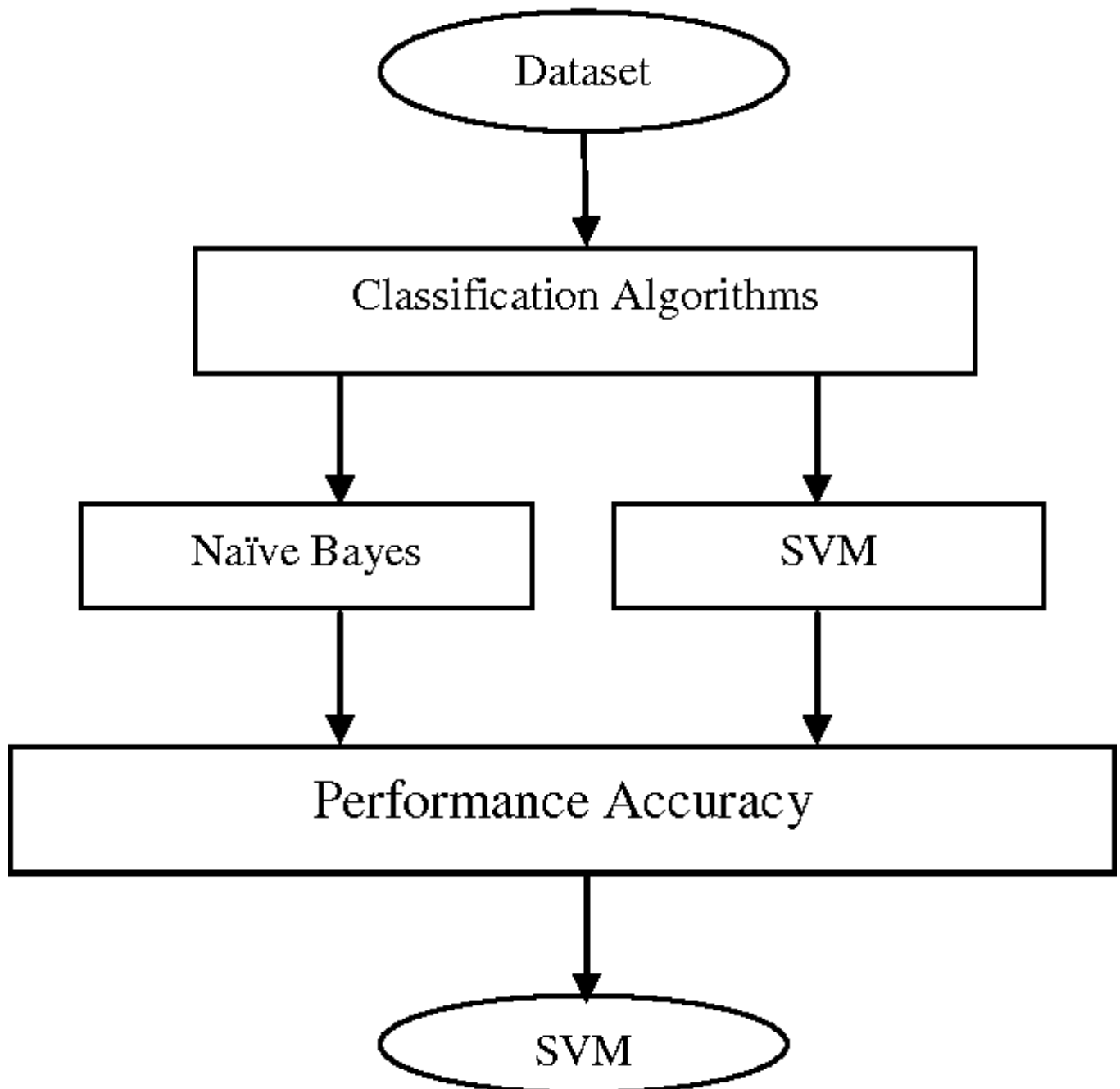
3.3 Proposed System:

- For obtaining high accuracy we used deep learning techniques for liver disease prediction.
- Deep learning techniques like multilayer perceptron, restricted Boltzmann and gradient descent are performed on the dataset.
- Among the algorithms multilayer perceptron algorithm gives highest accuracy for liver disease prediction.

MULTILAYER PERCEPTRON:

The Multilayer Perceptron was developed to tackle this limitation. It is a neural network where the mapping between inputs and output is non-linear. A Multilayer Perceptron has input and output layers, and one or more hidden layers with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.

3.4 Software Requirement Specification:

3.4.1 Software Requirements:

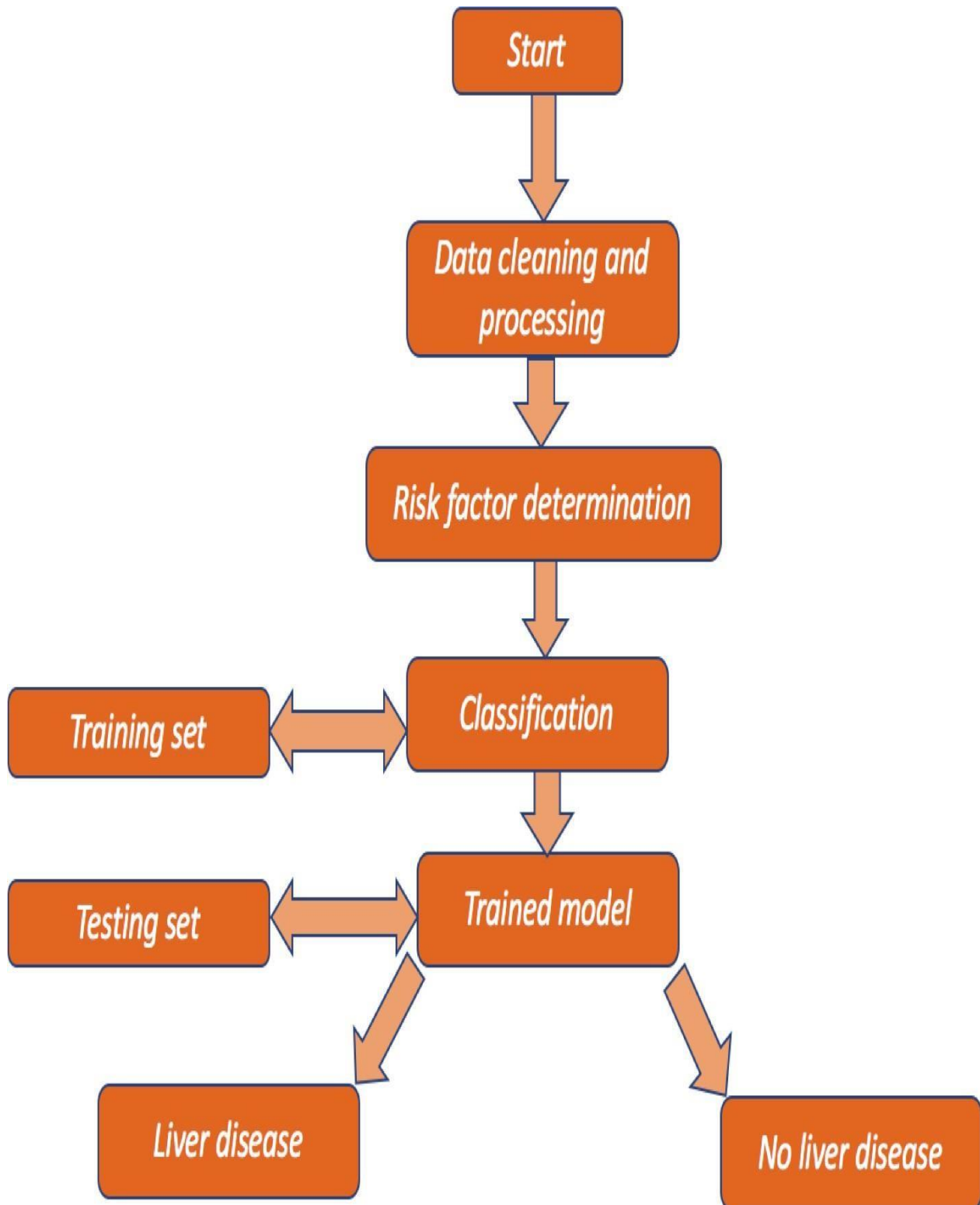During this project, we have used some software tools to perform the tasks of the modules.

- Jupyter Notebook

- PyCharm

- Google collab

3.4.2 Hardware Requirements:

We performed this project based on some hardware requirements to better performance of the project.

• RAM : 4 GB

• ROM : 20 GB

• PROCESSOR : i5

3.5 ARCHITECTURE

DATA CLEANING AND PROCESSING:

Steps In Data Cleaning and Preprocessing:

1. Gathering the data

2. Import the dataset & Libraries

3. Dealing with Missing Values

4. Divide the dataset into Dependent & Independent variable

5. dealing with Categorical values

6. Split the dataset into training and test set

7. Feature Scaling

RISK FACTOR DETERMINATION:

While dealing with health related projects we need to focus on the risk factors. In general, risk factors can be categorised into the following groups:

- Behavioural (risk factor of liver disease)
- Physiological
- Demographic
- Environmental
- Genetic

Behavioural:

Behavioural risk factors usually relate to 'actions' that the individual has chosen to take. They can therefore be eliminated or reduced through lifestyle or behavioural choices. Examples include:

- smoking tobacco
- drinking too much alcohol
- nutritional choices
- physical inactivity

CLASSIFICATION:

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc. Classes can be called as targets/labels or categories.

TRAINED MODEL:

The model is trained using "indian_liver_patient.csv" dataset obtained from Kaggle. 40% of data is used for training and 60% of data is used for testing. The trained data is labelled as "TRAINING SET" and testing data is labelled as "TESTING SET" in the architecture.

4. RESULTS:

Indian_liver_patient.csv dataset obtained from Kaggle.

The dataset contains following attributes

['Age' , 'Gender' , 'Total_Bilirubin' , 'Direct_Bilirubin' , 'Alkaline_Phosphotase','
Alamine_Aminotransferase' , 'Aspartate_Aminotransferase' , 'Total_Protiens' , 'Albumin' ,
'Albumin_and_Globulin_Ratio ', 'dataset']

 The dataset is imported the head of the data is printed and found the shape of the dataset using
Jupiter notebook platform which is shown in below fig. The sample of data is printed to know about
the data. The data almost contains all the numerical data except gender attribute. The fig 1.1 shows
the output of sample data and fig 1.2 shows the output of shape of the dataset.

```
     Age  Gender  Total_Bilirubin  Direct_Bilirubin  Alkaline_Phosphotase  \
0    65   Female              0.7               0.1                   187
1    62     Male             10.9               5.5                   699
2    62     Male              7.3               4.1                   490
3    58     Male              1.0               0.4                   182
4    72     Male              3.9               2.0                   195

     Alamine_Aminotransferase  Aspartate_Aminotransferase  Total_Protiens  \
0                          16                          18             6.8
1                          64                         100             7.5
2                          60                          68             7.0
3                          14                          20             6.8
4                          27                          59             7.3

     Albumin  Albumin_and_Globulin_Ratio  Dataset
0        3.3                        0.90        1
1        3.2                        0.74        1
2        3.3                        0.89        1
3        3.4                        1.00        1
4        2.4                        0.40        1
```

Fig 1.1

```
the shape of the dataset is:

Out[4]:  (583, 11)
```

Fig 1.2

The pre-processing techniques are applied on the dataset especially the categorical data is converted into numerical data. Here, used dataset contains the numerical data in gender attribute and pre-processing techniques are applied to change it from categorical data to numerical data. Some of the attributes in the dataset is assigned to one variable and some attributes are assigned to another variable and using this variable the splitting, training and testing of the data set will be done and we can see in the below fig 1.3.

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase \
0	65	0	0.7	0.1	187
1	62	1	10.9	5.5	699
2	62	1	7.3	4.1	490
3	58	1	1.0	0.4	182
4	72	1	3.9	2.0	195

	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens \
0	16	18	6.8
1	64	100	7.5
2	60	68	7.0
3	14	20	6.8
4	27	59	7.3

	Albumin	Albumin_and_Globulin_Ratio	Dataset
0	3.3	0.90	1
1	3.2	0.74	1
2	3.3	0.89	1
3	3.4	1.00	1
4	2.4	0.40	1

Fig 1.3

Multilayer perceptron is imported from sklearn module and created a mlp-clasifier model and x trained data and y trained data is fitted into the model and tried to reduce loss function by taking number of epochs.

Total no of epochs is : 144
The loss function is nearly reduced to : 0.5429
The total taken hidden layers are :4,5
The initial learning rate is : 0.01
Taken random state is : 5

```
Iteration 1,  loss = 2.39533516
Iteration 2,  loss = 1.45732902
Iteration 3,  loss = 1.09416111
Iteration 4,  loss = 1.03139572
Iteration 5,  loss = 0.98074398
Iteration 6,  loss = 0.92037273
Iteration 7,  loss = 0.85952581
Iteration 8,  loss = 0.81196534
Iteration 9,  loss = 0.78202008
Iteration 10, loss = 0.75763907
Iteration 11, loss = 0.74321128
Iteration 12, loss = 0.73625022
Iteration 13, loss = 0.73336120
Iteration 14, loss = 0.73258469
Iteration 15, loss = 0.73242459
Iteration 16, loss = 0.70244369
Iteration 17, loss = 0.65131436
Iteration 18, loss = 0.63142191
Iteration 19, loss = 0.62978614
Iteration 20, loss = 0.62800616
```
Fig 1.4

```
Iteration 125, loss = 0.54406098
Iteration 126, loss = 0.54364269
Iteration 127, loss = 0.54962569
Iteration 128, loss = 0.54420265
Iteration 129, loss = 0.54246175
Iteration 130, loss = 0.54247451
Iteration 131, loss = 0.54134177
Iteration 132, loss = 0.54063506
Iteration 133, loss = 0.53984179
Iteration 134, loss = 0.54077057
Iteration 135, loss = 0.54331706
Iteration 136, loss = 0.54476847
Iteration 137, loss = 0.54047295
Iteration 138, loss = 0.54052840
Iteration 139, loss = 0.54024175
Iteration 140, loss = 0.54022089
Iteration 141, loss = 0.54023504
Iteration 142, loss = 0.54012461
Iteration 143, loss = 0.54132159
Iteration 144, loss = 0.54297004
```

Fig 1.5

Finally, accuracy score is imported from sklearn module and prediction of accuracy is found usingthis accuracy score. The Obtained accuracy is 0.73.

The accuracy obtained through multi-layer perceptron is
accuracy= 0.7328767123287672

5. CONCLUSION

For liver disease prediction the following algorithms are used :

- Multilayer perceptron
- Restricted Boltzmann

MULTILAYER PERCEPTRON:

The Multilayer Perceptron was developed to tackle this limitation. It is a neural network where the mapping between inputs and output is non-linear. A Multilayer Perceptron has input and output layers, and one or more hidden layers with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.

RESTRICTED BOLTZMANN:

Restricted Boltzmann Machine (RBM) is a type of artificial neural network that is used for unsupervised learning. It is a type of generative model that is capable of learning a probability distribution over a set of input data. It is a type of neural network that consists of two layers of neurons – a visible layer and a hidden layer. The visible layer represents the input data, while the hidden layer represents a set of features that are learned by the network.

The RBM is called "restricted" because the connections between the neurons in the same layer are not allowed. In other words, each neuron in the visible layer is only connected to neurons in the hidden layer, and vice versa. This allows the RBM to learn a compressed representation of the input data by reducing the dimensionality of the input.

The final conclusion obtained is multilayer perceptron provides good accuracy than the other algorithms and it is best from machine learning algorithms and the obtained accuracy is 0.73.