OPEN ACCESS
Computer Science & IT Research Journal
P-ISSN: 2709-0043, E-ISSN: 2709-0051
Volume 3, Issue 1, P.36-51, January 2022
DOI: 10.51594/csitrj.v3i1.296
Fair East Publishers
Journal Homepage: www.fepbl.com/index.php/csitrj

A DIAGNOSTIC MODEL FOR THE PREDICTION OF LIVER CIRRHOSIS USING MACHINE LEARNING TECHNIQUES

Ganty, Jamila[1], Gregory Msksha Wajiga[2], Yusuf Musa Malgwi[3] & Abba Hamman Maidabara [4]
[1]No:3 Jiddari Polo near Bus Stop, Maiduguri, Borno State
[2,3]Department of Computer Science, Modibbo Adama University, Yola, P.M.B. 2076 Yola, Adamawa State, Nigeria.
[4] No: 45, Old G.R.A, Polo Ground Maidugu Street, Maiduguri, Borno State

*Corresponding Author: Ganty, Jamila
Corresponding Author Email: jamilaganty@gmail.com
Article Received: 29-12-21 **Accepted:** 10-01-22 **Published:** 25-01-22

ABSTRACT

Liver cirrhosis is the most common type of chronic liver disease in the globe. The ability to forecast the onset of liver cirrhosis sickness is critical for successful treatment and the prevention of catastrophic health implications. As a result, the researchers created a prediction model using machine learning techniques. This study was based on a dataset from the Federal Medical Centre, Yola, which included 583 patient instances and 11 attributes. The proposed model for the prediction of liver cirrhosis sickness employed Nave Bayes, Classification and Regression Tree (CART), and Support Vector Machine (SVM) with 10-fold cross-validation. Accuracy, precision, recall, and F1 Score were used to evaluate the model's performance. Among all the strategies used in this study, the Support Vector Machine (SVM) technique produces the best results, with accuracy of 73%, precision of 73%, recall of 100%, and F1 Score of 84%. Based on medical data from FMC, Yola, this study shows that machine learning methods, specifically the Support Vector Machine, provide a more accurate prediction for liver cirrhosis sickness. This approach can be used to help doctors make better clinical decisions.

INTRODUCTION

Background of the Study

The job of the liver fundamentally is to filter the blood coming from the digestive tract, before passing it to the rest of the body. As it does so, the liver emits bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions.

Liver cirrhosis is an important type of liver damage. It usually occurs as a result of long term damage of liver caused by many forms of liver diseases and circumstances, such as hepatitis and chronic alcoholism or through genetics. Each time the liver is injured it tries to repair itself fibrous scar tissue can be deposited in place of the missing cells which forms the cirrhosis. As cirrhosis progresses, more and more scar tissue forms, hence making it difficult for the liver to function. Advanced cirrhosis is life-threatening. The liver damage done by cirrhosis generally can't be undone. But if liver cirrhosis is diagnosed early and the cause is treated, further damage can be limited and, rarely, reversed.

In addition to fibrosis, the complications of cirrhosis include portal hypertension, ascites, hepatorenal syndrome and hepatic encephalopathy. A poor correlation exists between histologic findings of cirrhosis and the clinical picture. Some patients with cirrhosis are completely asymptomatic and have a reasonably normal life expectancy while some individuals have severe symptoms of end-stage liver disease and limited chance for survival. Common signs and symptoms may arise from decreased hepatic synthetic function (coagulopathy), decreased detoxification capabilities of the liver (hepatic encephalopathy) or portal hypertension (variceal bleeding) (Wolf & Katz, 2013).

ICT has been globally credited for changing the course of history and adding value to human lives in various ways. Of all the technologies that add value and enhance human life, the introduction of telemedicine which perhaps go down in history as the most defining and has the potential to impact positively on humans, especially those living in the rural areas (Ezeorah, Ayatalumo & Ibe-Enwo, 2009).

The machine learning techniques are much popular in medical diagnosis and predicting of diseases. The machine learning techniques have been developed and applied to practices such as diagnosis processes, treatment, personalized medicine, and patient monitoring and care.

Statement of the Problem

In this world, people with liver cirrhosis suffer from lengthy waiting times to be diagnosed as a result of limited medical resources and long processes of diagnosis. Liver cirrhosis is now regarded as the utmost common cause of death among people. Liver cirrhosis progresses slowly and, if diagnosed initially, there is possibility of prolongation of survival which is considerably increased. Reducing diagnostic delays improves early detection and makes treatment outcomes in liver cirrhosis. Thus, reducing these delays improves early detection and makes treatment cost-effective. Doctors or medical practitioners can take prompt action. Again, liver cirrhosis is at times misdiagnosed due to lack of proper tools and longer processes of diagnosis. Long waiting time to diagnose liver cirrhosis may increase the possibility of the disease spreading.

The motivation behind this study is that liver cirrhosis has become a common disease around the world. The death rate due to the disease is becoming alarming. Early detection of the disease may reduce the complication of the disease misfortune on patients. The ease of use of inventive technologies such as the one anticipated in this research may help in alleviating the troubles of holdup in the uncovering and treatment of liver cirrhosis. Also, the data mining tools can be used to assist physicians in predicting and diagnosing the disease to enhance necessary treatment. One more significant drive behind this study is to advance on the works of previous researchers who make their own contributions in this particular field of study.

Aim and Objectives of the Study

The main aim of this study is to design and implement a diagnostic system that will predict the presence of liver cirrhosis in patients using symptom and risk factors liver cirrhosis. The objectives is to:

i. Develop a fact-gathering tool that will capture the symptom associated with liver cirrhosis disease using Python programming language and a database management tool (MongoDB).
ii. Develop a Classification and Regression Tree (CART), Naïve Bayes and Support Vector Machine (SVM) algorithms using Python and MongoDB that will extract and classify the symptom associated with liver cirrhosis.
iii. Develop a diagnostic model using Python for the modeling and simulation.
iv. Analyze and predict the accuracy score of the model and carry out a comparative analysis on the three machine learning techniques (CART, Naïve Bayes and SVM).

Significance of the Study

The significant of the study are numerous to mention, but few are:

i. It will assist the paramedical personnel's in controlling liver cirrhosis disease
ii. Using the computer will be a consistent solution. Computers do not suffer some of the weaknesses of humans, such as getting tired or ill and they are more easily access than human experts.
iii. It can also be useful where there is no expertise on.

Scope of the Study

The scope of this study is limited to Liver Cirrhosis, use of Python programming language, Naïve Bayes, SVM algorithms and MongoDB in solving the challenges at hand. The study will involve the medical records department of the University of Maiduguri Teaching Hospital and the patients' folders records for the collection and collation of records and the data.

LITERATURE REVIEW

This chapter describes some of the previous work starting with the early diagnosis of liver cirrhosis, Literature survey and finally the research gaps discovered.

Early Diagnosis of Liver Cirrhosis

According to Benvegnù, Fattovich and, Noventa, (1994) the initial diagnosis of liver cirrhosis was initiated in 1986 as a prospective study aimed at investigating the natural history of cirrhosis and obtaining an early diagnosis of HCC. Criteria for inclusion in our follow up programmed were:

(a) Histological or clinical diagnosis of cirrhosis (presence of irregular margins on ultrasound, portal hypertension with laboratory evidence of chronic liver disease);
(b) Presence of compensated disease (stage A or B according to the Child-Pugh classification); and
(c) Absence of clinical and ultrasonography evidence of liver cancer at entry, with α-fetoprotein levels <200 ng/ml. In December 1997, using these criteria 573 patients with compensated cirrhosis were seen for the first time at our institute and 442 were followed up for at least 24 months.

The natural history of initially compensated cirrhosis due to hepatitis B (HBV) or hepatitis C (HCV) virus was only partially defined. Investigations show that, morbidity and mortality rates and the hierarchy of complications in compensated viral cirrhosis over a long follow up period.

Chronic infection with hepatotropic viruses is the main cause of chronic liver disease and cirrhosis worldwide, with a predominant role of the hepatitis B virus (HBV) in the Far East and of the hepatitis C virus (HCV) in Western countries.

Many of these patients die as a consequence of end stage liver disease. The natural history of initially compensated cirrhosis due to HBV or HCV has been only partially defined. The disease often remains asymptomatic for many years, allowing a normal quality of life. Previous studies, often conducted in rather heterogeneous cohorts of patients, indicated an annual risk of developing hepatocellular carcinoma (HCC) of 1–6%, and similar or higher risk of decompensation of liver function (Benvegnù, Fattovich and Noventa, 1994).

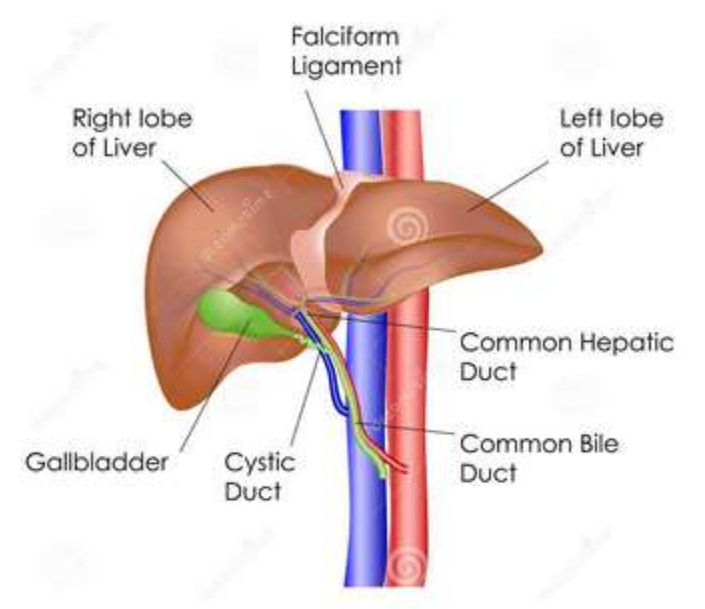

Figure1: Structure of liver

FUNCTION OF LIVER

The liver regulates most chemical levels in the blood and excretes a product called bile. This helps carry away waste products from the liver. All the blood leaving the stomach and intestines passes through the liver. The liver processes this blood and breaks down, balances, and creates the nutrients and also metabolizes drugs into forms that are easier to use for the rest of the body or that are nontoxic. More than 500 vital functions have been identified with the liver. Some of the more well-known functions include the following:

1. Production of bile, which helps carry away waste and break down fats in the small intestine during digestion
2. Production of certain proteins for blood plasma
3. Production of cholesterol and special proteins to help carry fats through the body
4. Conversion of excess glucose into glycogen for storage (glycogen can later be converted back to glucose for energy) and to balance and make glucose as needed
5. Regulation of blood levels of amino acids, which form the building blocks of proteins
6. Processing of hemoglobin for use of its iron content (the liver stores iron)
7. Conversion of poisonous ammonia to urea (urea is an end product of protein metabolism and is excreted in the urine)
8. Clearing the blood of drugs and other poisonous substances
9. Regulating blood clotting
10. Resisting infections by making immune factors and removing bacteria from the bloodstream
11. Clearance of bilirubin, also from red blood cells. If there is an accumulation of bilirubin, the skin and eyes turn yellow.

Figure2: Liver with Cirrhosis

SYMPTOMS OF LIVER CIRRHOSIS

The following are the major sign and symptom of liver cirrhosis

I. Fatigue.
II. Easily bleeding or bruising.
III. Loss of appetite.
IV. Nausea.
V. Swelling in your legs, feet or ankles (edema)
VI. Weight loss.
VII. Itchy skin.

VIII. Yellow discoloration in the skin and eyes (jaundice)

RISK FACTORS FOR LIVER CIRRHOSIS

I. Chronic hepatitis B.
II. Chronic hepatitis C.
III. Chronic excessive alcohol intake.
IV. Fatty liver disease (non-alcoholic steatohepatitis)
V. Autoimmune liver disease (autoimmune hepatitis, primary biliary cirrhosis or primary sclerosing cholangitis)

MACHINE LEARNING TECHNIQUE

Machines are by nature not intelligent. Initially, machines were designed to perform specie tasks, such as running on the railway, controlling the trace, digging deep holes, traveling into the space, and shooting at moving objects. Machines do their tasks much faster with a higher level of precision compared to humans. They have made our lives easy and smooth.

Machine learning is a branch of artificial intelligence that aims at enabling machines to perform their jobs skillfully by using intelligent software. The statistical learning methods constitute the backbone of intelligent software that is used to develop machine intelligence. Because machine learning algorithms require data to learn, the discipline must have connection with the discipline of database. Similarly, there are familiar terms such as Knowledge Discovery from Data (KDD), data mining, and pattern recognition. One wonders how to view the big picture in which such connection is illustrated.

Classification and Regression Tree (C ART) Algorithm

Classification and Regression Trees or CART for short is a term introduced by Leo Breiman to refer to Decision Tree algorithms that can be used for classification or regression predictive modeling problems. Classically, this algorithm is referred to as "decision trees",

A CART tree is a binary decision tree that is constructed by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample.

It is a nonparametric method that automatically performs variable selection. It increases the performance by revealing the important relationships of features in dataset and represents them in the form of tree. It can easily handle both categorical and numerical variables. The representation used for CART is a binary tree.

Predictions are made with CART by traversing the binary tree given a new input record. The tree is learned using a greedy algorithm on the training data to pick splits in the tree. It worked in three parts that includes building of maximum tree also known as tree growing, right tree selection also known as tree pruning and classification of testing data using built tree. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. The topmost decision node in a tree which corresponds to the best predictor called root node.

For predicting a class label for a record, we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

The Liver Cirrhosis Prediction Model

The work is designed to predict the presence of liver cirrhosis or not based on the statistical parameters of different Liver cirrhosis symptoms and risk factors. This model will require the detail information of patients (symptoms and risk factors) that she/he is suffering from and other necessary information like Age, Sex, Family History, Presence of Hepatitis B, Presence of Hepatitis C, Excessive alcohol intake, Swelling of the abdomen, Swelling of the ankles and feet, Vomiting of blood or passage of black stools, Yellowing of the eyes and skin (jaundice) associated with dark, tea-coloured urine, given by the medical practitioners through the system's user interface.

LITERATURE

Jankisharan, Jagdeesh and Sanjay (2014) presented computational intelligence techniques for Liver Patient Classification. The effectiveness of the techniques viz. Multiple Linear Regression, Support Vector Machine, Multilayer Feed-Forward Neural Network, J-48, Random Forest and Genetic Programming have been tested on the India Liver Patient Information (ILPD) Dataset. Authors employed under sampling and over sampling for balancing it. The results obtained from experiments indicate that Random Forest over sampling with 200% outperformed all the other techniques.

Omar and Eman (2014) proposed a hybrid classification system for HCV diagnosis, using Modified Particle Swarm Optimization algorithm and Least Squares Support Vector Machine (LS-SVM). Feature vectors are extracted using Standard Component Analysis algorithm. As LS-SVM algorithm is sensitive to the modifications of values of its parameters, Modified-PSO Algorithm was used to search for the optimal values of LS-SVM parameters in less number of iterations. The recommended system was implemented and evaluated on the benchmark HCV data set from UCI repository of machine learning databases. It was associated with another classification system, which developed PCA and LS-SVM. From the experimental results the recommended system achieved maximum classification accuracy than the other systems.

Rosalina and Noraziah (2010) predicted a hepatitis prognosis disease using Support Vector machine (SVM) and Wrapper Method. Before classification process they used binding methods to eliminate the noise features. Firstly SVM carried out feature selection to get greater accuracy. Features selection were applied to minimize noisy or insignificance data. From the experimental results they observed the increased accuracy rate in the clinical lab test cost with minimum execution time. They have reached the target by merging Wrappers Method and SVM techniques.

Karthik, Priyadarishini, Anuradha and Tripathi (2011) applied a soft computing technique for intelligent diagnosis of liver disease. They have implemented classify its type of detection in three phases. In the first phase, they classified liver disease using Artificial Neural Network (ANN) classification algorithm. In the second phase, they generated the classification rules with rough set rule induction using Learn by Example (LEM) algorithm while in the third phase fuzzy rules were applied to identify the types of the liver disease.

Chaitrali and Sulabha (2012) analyzed prediction systems for heart disease using more number of input attributes. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on heart disease database. The performances of these techniques are compared, based on accuracy. The authors' analysis shows that out of these three classification models Neural Networks has predicted the heart disease with highest accuracy.

Moloud, Mariam, Resul and I-Hsien (2016) proposed PC helped diagnostic strategy by utilizing novel tree based calculations, which are C5.0 calculation and Chi-square Automatic Interaction Detector (CHAID) calculation for liver infection prediction. They proposed strategy creators utilized C5.0 calculation through boosting method to accomplish the most elevated exactness and the generation of guidelines on liver malady dataset.

Sadiyah, Novita and Teddy (2013) connected three systems, which are Decision Tree, Naive Bayes, and NBTree calculations for determination of liver sickness. They have actualized grouping model and got most astounding precision by utilizing NB Tree calculation. Creators presumed that the Naive Bayes calculation gives the quickest calculation time pursued by Decision Tree and NB Tree calculation and furthermore demonstrated that number of arrangements run of NB Tree calculation is less difficult than the quantity of characterization lead created by Decision Tree calculation.

Bendi, Surendra and Venkateswarlu (2012) proposed five arrangement calculations for liver sickness finding. Creators connected Naive Bayes grouping, C 4.5 Decision Tree, Back Propagation, K-Nearest Neighbor and Bolster Vector Machine calculations on two distinctive sort of liver Informational collections, which are BUPA liver issue information and India Liver Patient Information (ILPD). In the paper, the above calculations are considered for assessing their arrangement execution as far as Accuracy, Precision, Sensitivity and Specificity in grouping liver patient's dataset. At long last they acquired most noteworthy precision by utilizing K-Nearest Neighbor what's more, Support Vector Machine calculations.

Sumedh, Jay and Reshul (2017) proposed two techniques with the end goal to order the eternal liver illness, one strategy is a characteristic way to deal with finding, and second one includes a hereditary way to deal with the finding. Proposed approach is the application of Artificial Neural Networks and Multi-Layer Perceptron to Reduce scale Array Analysis. They utilized these two techniques to enhance the proficiency of two calculations Back Propagation and Support Vector Machine (SVM) to characterize the liver disease.

Gregory (2015) proposed two real liver patient datasets were investigated for building classification models in order to predict liver diagnosis. Eleven data mining classification algorithms were applied to the datasets and the performance of all classifiers are compared against each other in terms of accuracy, precision, and recall. Based on the experimental results the classification accuracy is found to be better using FT Tree algorithm compare to other algorithms., it also shows the enhanced performance according to the attributes and it gives 78.0% of Accuracy, 77.5% of Precision, 86.4% of Sensitivity and 38.2% of Specificity results respectively.

Research Gaps Discovered

While a lot of studies were been done by diverse scholars in the diagnosis of liver cirrhosis, using the various data mining techniques. It is experimental that most of the studies made in the literature were based on radiology testing such as computed tomography (CT), ultrasound or magnetic resonance imaging (MRI) or via a needle biopsy of the liver. Also, a new imaging technique called elastography, which can be performed with ultrasound or MRI is used for diagnosis cirrhosis. No mentioned is the used of the risk factors and symptoms in the diagnosis of liver cirrhosis.

For instance, in Insha and Chiranjit (2018) they worked on liver disease detection due to excessive alcoholism using data mining techniques. Their work proposed identification just as to foresee the nearness of liver sickness utilizing information mining calculations. Subsequent to deciding the principles, they utilize diverse information mining calculations to prepare and test the dataset to distinguish the liver sickness. They concentrated on the information (Dataset) that was collected from UCI repository which is dependent on laboratory results and comprises of 7 unique qualities with 345 occurrences.

Medical personnel need high-performance system that is flexible, efficient and minimizes errors of analysis giving an instant result (Real-time) using the risk factors and symptoms of liver cirrhosis without going to the laboratory. The system when implemented will help the medical practitioners capture fit knowledgeable decisions during the diagnosis within the shortest possible time. Incorporating symptoms and risk factors together with data mining technique will fill in the gap that exists in the reported literature

METHODOLOGY

Introduction

In this section, we will describe the location, population size and the sampling technique, source of data, method of data collection and the proposed methodology. The chapter will also look at the three algorithms to be use (Classification and Regression Tree, Naïve Bayes and Support Vector Machine algorithms), the liver cirrhosis prediction model and finally the description of variables for the dataset.

Location of the Study

The study is to be conducted on the data that will be supplied by the Federal Medical Center, Yola. The reason for the choice is nearness and relationship and have enough medical manpower that handles complicated Liver diseases.

Study Population Size and Sampling Technique

The study population size will be composing of all patients that are diagnosed with Liver Cirrhosis the past three years in FMC Yola. Purposive sampling techniques will be used in selecting FMC Yola as a study center. All the patients from July, 2017 to July, 2020 constituted the sample population size of the data from the hospital.

Purposive sampling technique will be used to select record from the Laboratories and Medical Records Departments for collation and collecting past records in order to form the primary data using interview for the study. The research design is going to be cross-sectional (observational study).

Source of Data

The sources of the data will be from the Medical Records Departments and the Patient folders at FMC Yola.

Method of Data Collection

In collating and collecting necessary data and information needed for the model, the primary source will be collected from the Medical Records Department and the secondary source which include journals, e-books as well as visiting relevant websites, medical dictionaries and other research materials will be use.

The completed data of the diagnosed patients with Liver Cirrhosis between July, 2017 to July, 2020 will be collected from the Medical Records Department and analyze to determine the variables or predictors for the Liver Cirrhosis which will be coded to values.

The In-depth interview will be conducted with the staff of the Medical Records Department for the study. This approach will be taken to incorporate the experiences of the respondents.

Proposed Methodology

The experimental type of research design will be use and being a quantitative research method. Basically, it is a research conducted using a scientific approach, where a set of variables will be kept constant while other set of variables are being measured as the subject of the experiment

Liver Cirrhosis Prediction Model

Figure 2 represent the functional blocks of the liver cirrhosis prediction model. The whole process is initiated from the liver cirrhosis database. The dataset follows the data preprocessing which processing step itself consists of a combination of two different tasks. One of them is data cleaning, and the other ones are the data integration and data transformation. In the data cleaning (data cleansing) phase noise data and the data which is irrelevant to the analysis are removed from the collected data. While in the data integration phase, multiple data sources are combined into a common source. It is worth mentioning here that the data sources are often heterogeneous.

The training dataset is a dataset used for learning, which is to fit the parameters (e.g., weights). Most approaches that search through training data for empirical relationships tend to overfit the data, meaning that they can identify and exploit apparent relationships in the training data that do not hold in general.

The test dataset is a dataset that is independent of the training dataset, but that follows the same probability distribution as the training dataset. If a model fit to the training dataset also fits the test dataset well, minimal over fitting has taken place. A better fitting of the training dataset as opposed to the test dataset usually points to over fitting. A test set is therefore a set of examples used only to assess the performance (i.e. generalization) of a fully specified classifier.

Description of Variables for the Dataset

<u>Patients' Details</u>

i. Excessive Alcohol Consumption

Alcohol is a psychoactive, potentially toxic and addictive substance. It is rapidly absorbed by the body where it is detoxified by the liver. Additional risk factors include gender (women are more susceptible), a genetic predisposition and co-occurring liver disease. Individuals consuming excess alcohol can develop alcoholic liver disease (ALD). Although a dose response effect can be observed, there is no exact level of consumption that can predict the onset of ALD nor is

dependence on alcohol a pre-requisite (Gramenzi *et al* 2006). At a population level, the average population consumption of alcohol is closely related to mortality from CLD (Ramstedt 2003).
Alcohol Intake: (Yes) = 1, (No) = 0.

ii. Blood Borne Viruses

The ***Hepatitis C*** virus is highly infectious and is transmitted through blood borne routes, principally by sharing of injecting drug use equipment, via blood transfusion (prior to instigation of blood donor testing in 1991) and, rarely, by sexual activity and mother to child transmission. One fifth of those infected with the virus recover spontaneously whilst of the remainder who develop chronic Hepatitis C infection, 5-15% will develop cirrhosis over the next 20 years (SIGN 2006). The proportion of liver-related deaths that occurred in HCV-diagnosed individuals increased from 2.8% (1995-1997) to 4.2% (2004-2006) (*McDonald et al* 2010).

The ***Hepatitis B*** virus is highly infectious and is transmitted through blood and bodily fluids by unprotected sex; sharing of injecting drug equipment and from mother to child transmission. Less than 10% develop chronic infection which in turn can result in CLD. It cannot be determined from routine mortality data what proportion of CLD deaths are due to Hepatitis B.
Presence of Hepatitis B/C: Yes = 1 No = 0.

iii. Obesity

Non-alcoholic fatty liver disease (NAFLD) is a disease of the liver characterized by fatty infiltration with or without inflammation (non alcohol steatohepatitis or NASH). Previously thought to be benign, it can progress to fibrosis and cirrhosis in 15-20% of patients. It can also result in liver cancer. Development of NASH and fibrosis is associated with obesity, type 2 diabetes, hypertension and high triglycerides (Roderick *et al* 2004). In European and US studies, NAFLD affects 3-30% of the population, depending on whether blood tests or liver scans are the screening test (Clarke *et al* 2002). Deaths from NAFLD have risen in Scotland from 3 in 1979 to 40 in 2007 though this figure is likely to be an underestimate (Scottish Government 2008).
Presence of Obesity: Yes = 1, No = 0

iv. Swelling of the abdomen

Presence of swelling of the abdomen due to accumulation of fluid (ascites) is an associated symptom for liver cirrhosis.
Presence of swelling of the abdomen: Yes = 1, No = 0

v. Swollen of ankles

Presence of Swelling of the ankles and feet (pedal oedema) is an indication for the presence of liver cirrhosis.
Presence of swollen ankles: Yes=1 No = 0

iv. Vomiting of blood or passage of black stool

Vomiting of blood or passage of black stools due to bleeding from ruptured varices (variceal bleeding) is a strong indication for the presence of Liver cirrhosis
Vomiting of blood or passage of black stool: Yes = 1, No = 0

vii. Yellowing of eye, skin and tea coloured urine

A presence of yellowing of the eyes and skin (jaundice) associated with dark, tea-coloured urine is a strong indication for the presence of liver cirrhosis.

Yellowing of eye, skin and tea coloured urine: Yes = 1, No = 0

viii. Drowsiness or Confusion

The presence of drowsiness or confusion due to inability of the liver to break down toxins in the blood (hepatic encephalopathy) is an indication for the possibility of liver cirrhosis in the body.

Presence of drowsiness or confusion: Yes = 1, No = 0

RESULTD AND DISCUSSION

The Diagnostic System

The system interfaces, as well as their source codes, are a result the researcher realized at the deployment phases in the realization during the implementation of this task.

The Sign-Up Page

This page allows system users to register their credentials by creating an account, which will be used throughout the authentication phase of the system.

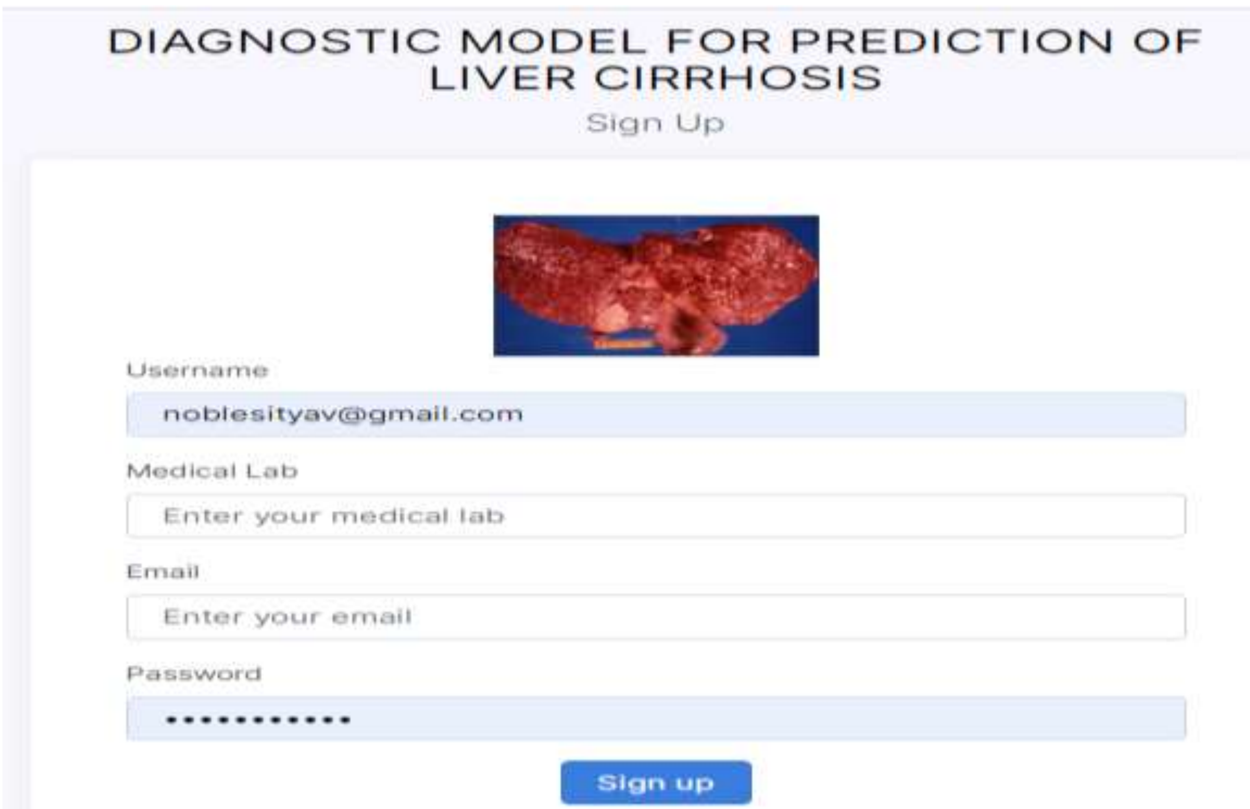

Figure 6: The Sign-Up Page

The Login Page

This page allows system users to be identified based on their identification. The credentials provided are compared to those stored data in a database for authorized users.

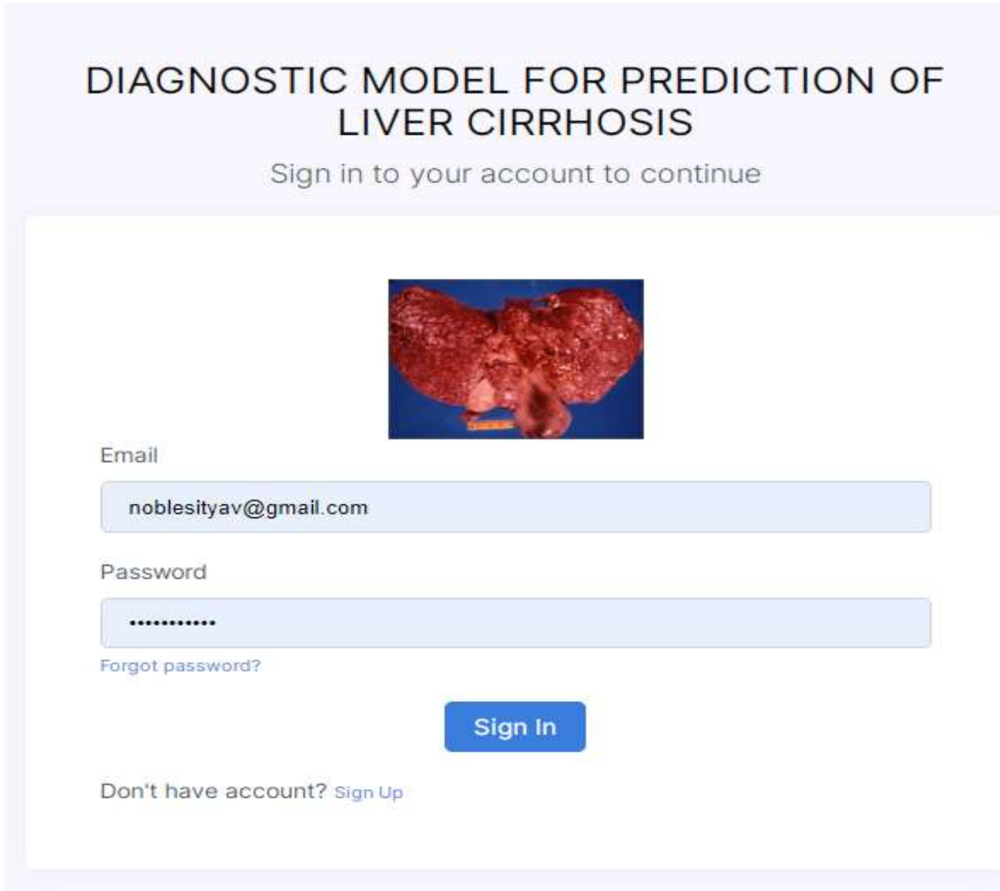

Figure 7: The Login Page

Dashboard Page

In this work, a dashboard is utilized to graphically track, analyze, and show key performance indicators (KPI) and essential data points in order to traverse the system. Figure 8 depicts the dashboard of this study, which comprises four (4) command buttons (navigation interfaces).

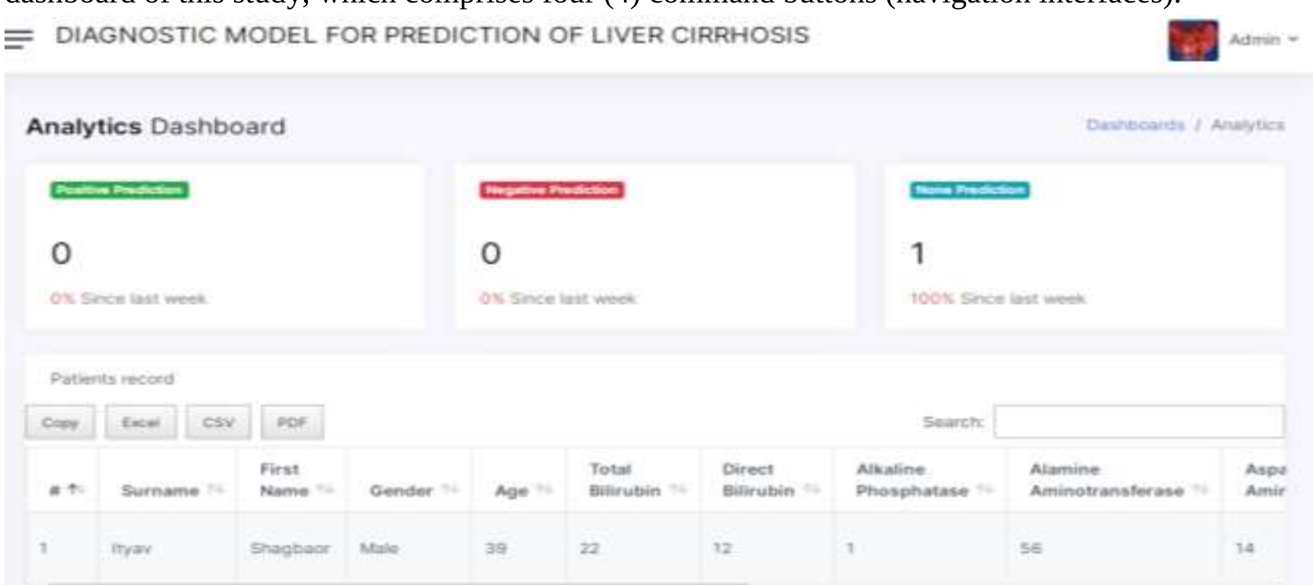

Figure 8: The Dashboard Page

The Patient Information Page

This page allows system users to enter risk factors of patience for the prediction of the occurrence of liver cirrhosis illness.

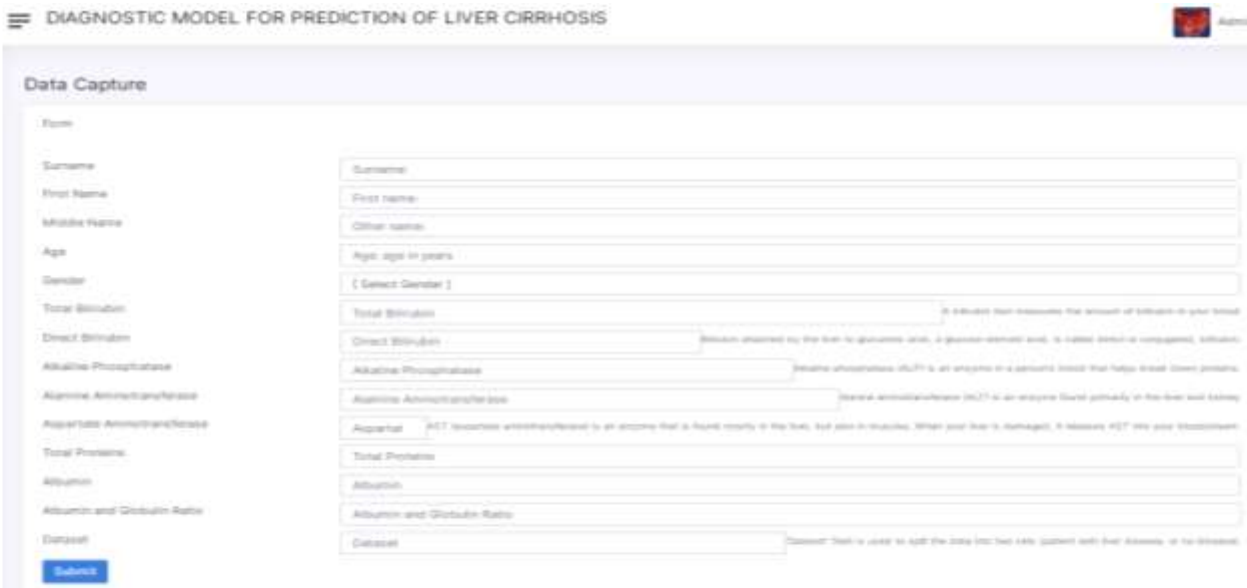

Figure 9: The Patient Information Pages

RESULTS

In this study, the researchers used Nave Bayes, Classification and Regression Tree, and Support Vector Machine as machine learning algorithms to generate a set of outputs for the prediction of liver cirrhosis cases. The implementations of this model were used to train and evaluate our model for predicting liver cirrhosis illness on a dataset from FMC, Yola. The dataset contains 583 patient cases diagnosed with liver cirrhosis, including 11 raw attributes for each clinical instance of patient.

Performance Evaluation of the Model

This is the crucial stage of the machine learning process. The researcher consider Accuracy, Precision, Sensitivity (Recall), and F1_Score as the measures use in evaluating the model considering its confusion matrix.

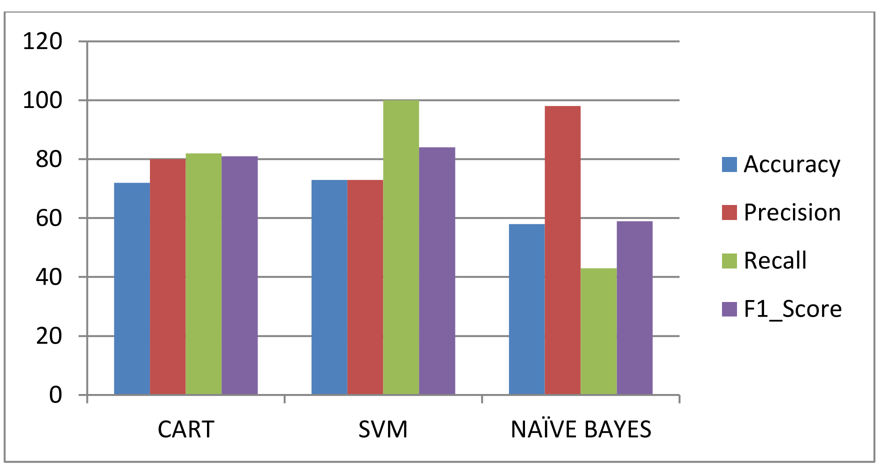

Figure 10: Comparative Chart for Classification Performance

CONCLUSION

Machine Learning method is the most promising technique for the prediction of liver cirrhosis illness, which will aid in the appropriate decrease in the fertility rate and the construction of large-scale alerts. Using Python and MangoDB, the suggested model for liver cirrhosis disease prediction using machine learning method mentioned in this study was effectively designed, developed and simulated. The model captures indicators and factors associated to liver cirrhosis illness in order to identify a patient with liver cirrhosis and present them in a Graphical User Interface (GUI) form where the user enters the attribute and the system diagnoses the patient using the training dataset utilized in Support Vector Machines.

Consequently, the study determined that a liver cirrhosis prediction system delivers good predictions and obtains greater results based on the results of this study.

References

Bendi, V. R., Surendra, P. B., & Venkateswarlu, N.B. (2012). A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis. *International Journal of Database Management Systems (IJDMS), 3*(2), 101-111.

Benvegnù, L., Fattovich, G. & Noventa, F. (1994). *Concurrent hepatitis B and C virus infection and risk of hepatocellular carcinoma in cirrhosis*. A prospective study. PP.74-76.

Chaitrali, S. D., & Sulabha, S. A. (2012). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications, 47*(10), 44-48.

Ezeorah E.U., Ayatalumo, J.C., & Ibe-Enwo, O. (2009). Application of Telemedicine in Primary Healthcare Delivery: A focus on Nigeria. *Nigerian Computer Society National Conference, 2*(4), 181-186.

Gregory, O. (2015). Prediction of Liver Disease (Biliary Cirrhosis) Using data Mining Technique. *International Journal of Emerging Technology & Research, 10*(2), 37- 42

Insha, A., & Chiranjit D. (2018). Liver disease detection due to excessive alcoholism using data mining techniques. *International Conference on Advances in Computing and Communication Engineering* (ICACCE). IEEE. PP.76-80.

Jankishran, P., Jagdeesh, M., & Sanjay, P. (2014). Liver patient classification using intelligence techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(2), 295-299.

Karthik, S., Priyadarishini, A., Anuradha, J. & Tripathi. B. K. (2011). Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types. *Advances in Applied Science Research, 2*(3), 334-345.

Omar, S. S., & Eman, A. E. (2014). Classification of Hepatitis C Virus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine. *International Journal of Scientific & Engineering Research,5*(3), 122-126.

Rosalina, A. H., & Noraziah, A. (2010). Prediction of Hepatitis Prognosis Using Support Vector Machine and Wrapper Method, IEEE, PP.20-22.

Sadiyah, N., Novita, A., & Teddy M. (2013). Data Mining Techniques for optimization of liver disease classification. *International Conference on Advanced Computer Science Applications and Technologies, 4*(5), 33-39.

Sanjay, K., & Sarthak, K. (2014). Effective Analysis and Diagnosis of Liver Disorder by Data Mining. *Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA).* PP.44-52.

Sumedh, S., Jay, L. & Reshul, D. (2017). Diagnosis of liver diseases using machine learning. *International Conference on Emerging Trends & Innovation in ICT (ICEI)*. PP.129-133