IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Liver Disease Prediction using Machine Learning

H Mangalam[1], Arunkumar G[2], Dhivyaprakash K[3], Karthick V[4]
Professor, Department of ECE[1]
Students, Department of ECE[2,3,4]
Sri Ramakrishna Engineering College Coimbatore, India

Abstract: *Machine learning has become an important tool in knowledge and data engineering, particularly in predicting outcomes based on existing data. One of the most commonly used machine learning techniques is classification, which involves learning patterns from an existing dataset and then applying them to a new dataset to make predictions. However, some classification algorithms have limited accuracy. This study proposes a new method called Supervised Learning Technique, which achieves higher accuracy than existing algorithms like Logistic Regression, SVM, and bagging tree classifier. The study also uses a deep learning algorithm called CNN to predict disease levels based on CT scan images. The researchers created a user-friendly web application to allow users to input data and get results. The study used an Indian liver disease dataset to demonstrate how the regression technique can improve prediction accuracy for liver disease. The goal of the study was not only to increase accuracy but also to demonstrate the usefulness of the algorithm for predicting diseases at an early stage.*

Keywords: Dataset**,** Prediction, Accuracy, classification, Input Dataset, Algorithm

I. INTRODUCTION

Liver is one of the largest organs that is present in the upper right part of abdominal cavity, and it is also the second largest organ after skin. Its is wedge shape, and it is also the largest gland of the body which secretes chemical substances called hormones. Liver performs more than 500 functions in human body and also supports most of the organ which is vital for our survival. In adults,it is observed that the liver weighs about 2% of body weight, in Males the liver weighs about 1.4 – 1.8 kgs, in females the liver weighs about 1.2 – 1.4kgs andin new born it weighs 150 g.
The Liver performs the following functions:

- It secretes bile and glycogen.
- Its synthesis serum protein lipids.
- It detoxifies blood room endogenous and exogenous substances such astoxins, drugs, alcohols.
- It stores vitamin D, A, K, E and B1.

II LITERATURE REVIEW

Hartatik et al, (2021) have examined to conclude; based on the findingsof utilising the python application to test the Naive Bayes and KNN algorithms to solve predicting issues for patients with liver illness The Indian Liver Patient Dataset was obtained from the UCI Machine Learning Repository (ILPD). The results reveal that by employing six variables in the prediction model, the NaiveBayes algorithm produces a better value than the KNN, resulting in an increase in accuracy when compared to the results of earlier studies.
Abhishek Chowdhur et al, (2022) has designed different classification techniques, such as Logistic Regression, Support Vector Machine, and K- Nearest Neighbour, in their paper to predict liver disease. All of these algorithmswere compared based on classification accuracy, which was determined using aconfusion matrix. Logistic Regression and K-Nearest Neighbour have the highest accuracy, but logistic regression has the highest sensitivity, according tothe experiment. As a result, we can conclude that Logistic Regression is a good way to predict liver illness
Latha.C.M (2022) proposed an approach, based several associated features and KNN technology to enhance liver disease prediction, and applies a machine learning technique that was highly promising for studies with regard to healthcare and health. To recognize the causes and the identification phases are more important. For this, we applied a machine learning technique that was highly promising for studies with regard to healthcare and health

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Taher M Ghazal et al, (2022) proposed an intelligent model to predict liver disease using machine learning technique, which is more effective and comprehensive in terms of performance, and 0.116 miss-rate. As a result, the purpose of this research is to assess the efficacy of various Machine Learning (ML) algorithms to lower the high cost of liver disease diagnosis through prediction. With the current rise in numerous liver disorders, it's more importantthan ever to detect liver disease early on.

Dr. R. Vijayabhanu (2020) RNN being a text classifier of deep learning technique with the advantage of processing in multiple loops in a sequential manner to obtain best performances measured by the factor of accuracy has beenproposed in this study.

Golmei Shaheamlung et al, (2020) proposed a Liver disease prediction has various levels of steps involved, pre-processing, feature extraction, and classification. In this s research work, a hybrid classification method is proposed for liver disease prediction. And Datasets are collected from the Kaggle database of Indian liver patient records. The proposed model achieved an accuracy of 77.58 .

III METHODOLOGY

The chapter deals with the three machine learning algorithms and one deep learning algorithm that are used to classify liver disease-based on numerical dataset and image dataset.

Machine learning algorithms:

- Logistic Regression
- Support Vector Machines
- Bagged Tree ClassifierDeep
- Learning Algorithm:
- Convolutional Neural Networks

IV WORKING PRINCIPLE

The system being proposed here uses concept of machine learning and deep learning, and the models are first trained, then tested. Finally, the most accurate model will predict the result. At first, the system asks you to select whether numerical inputs or image inputs which are going to provide by the usersas inputs.

If it was numerical inputs then the system ask the user to enter details including age, gender, total Bilirubin, directt Bilirubin, Alkaline_phosphetase, Alamine_Aminotransferase, Aspartate_Aminotransferase A total proteins, albumin, and Albumin and Globul in ratio. Values of last ten parameters mentioned here, can be known by blood test report of the user. If the user selectimage inputs as their selection then it ask user to upload CT scan image as theirinputs. After taking these inputs from the user, the system compares the data input with the training dataset of most accurate model and then predicts the result accordingly as risk or no risk.No medical expertise required: You don't need to have any knowledge ofmedical science and liver diseases to predict the liver disease using this application. All you need to do is enter the details being asked, which are already present in the blood test report (some like age, gender are already known) and then you will get the results of prediction.

- High accuracy: The system predicts the results with 100 % accuracy for the dataset that we have used while creating this application. While theaccuracymight be different in some cases, it will still be high enough to be trustworthy at a large scale.
- Immediate results: The results here are predicted within seconds of entering the details. You don't need to wait for a doctor to come, unlike in traditional method.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

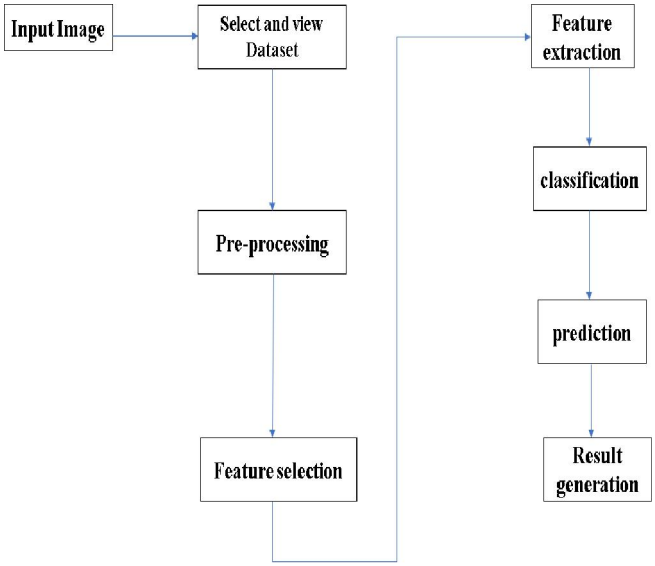

Figure 1.Process Diagram for Liver Disease Classification using image.

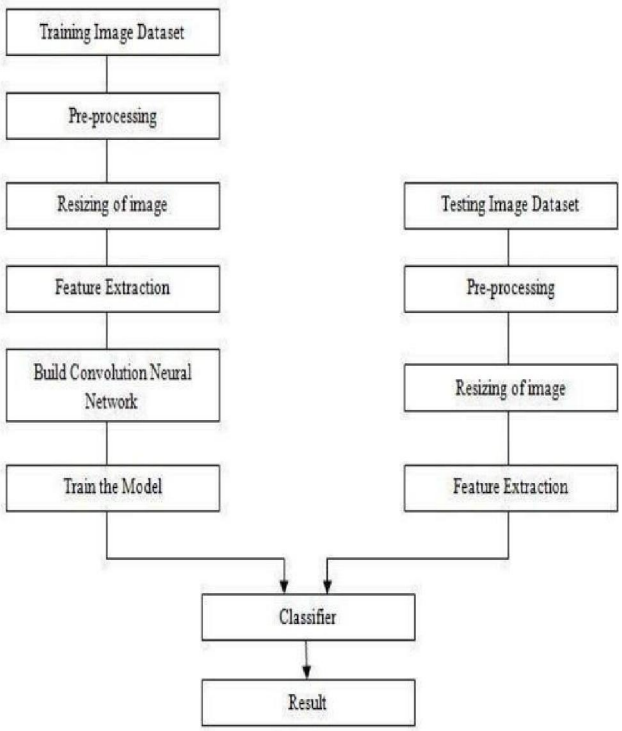

Figure 2 Process Flow Diagram for Liver Disease Classification

Figure 1 The Liver Image Dataset, sourced from a dataset repository, is being used as input for this project. The input images are in .png and .jpg formats and are subjected to pre-processing. The pre-processing step involves resizing the images and converting them to grayscale. In the next step, feature extraction is carried out using Local Binary Pattern (LBP) on the pre-processed images. The images are then split into a train set and a test set for decision-making. The train image set is used for evaluating the model while the test image set is used for predicting the model. Classification is then carried out using a deep learning classification algorithm such as Convolutional Neural Network (CNN). The output of the classification algorithm determines whether the input image is indicative of liver cancer or not based on the input symptoms. In the next step, performance estimation is carried out by analyzing performance metrics such as accuracy and error rate. Finally, the project is implemented as a web application.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

V SOFTWARE IMPLEMENTATION

A. INTRODUCTION

This chapter deals with machine learning algorithms of Logistic Regression, SVM and Bagged tree classifier and deep learning algorithm calledCNN. Also, overall flowchart for the above algorithms. For classifying python is used which is a high-level script language and it is implemented in Jupyter note book and Spyder using anaconda navigator. For liver disease prediction classification, the algorithm was trained with a dataset of around 400 datasets.

B. Algorithm for Logistic Regression:

- Preparing the dataset.
- Pre-Processing of dataset.
- Separating dataset into train model and test model.
- Removing the outliers.
- Dropping the duplicate dataset.
- Invoking the Algorithm function.
- Displaying the confusion matrix and Accuracy results.

C. Algorithm for CNN:

- Load image dataset:
- Pre-process the dataset:
- Define the CNN architecture:
- Train the CNN:
- Evaluate the CNN:

D. FLOW DIAGRAM FOR PROCESSING IMAGE BASED DATASET

- Collect the dataset: Gather a collection of images that you want to process.
- Pre-processing: Perform pre-processing steps to clean and normalize the dataset. This might include resizing images, cropping images, and converting images to grayscale.
- Feature Extraction: Extract relevant features from the images. This could include identifying edges, shapes, colors, and textures in the images.
- Training/Test Set Split: Split the dataset into a training set and a test set. The training set will be used to train the model, while the test set will be used to evaluate the model's performance.
- Model Selection: Choose a suitable machine learning model to process thedataset. This could include convolutional neural networks, decision trees, or support vector machines.
- Model Training: Train the model on the training set. This involves feeding the extracted features from each image into the model and adjusting the model's parameters to minimize the difference between the predicted output and the actualoutput.
- Model Evaluation: Evaluate the model's performance on the test set. This will givean indication of how well the model is able to generalize to new images.
- Model Deployment: Deploy the trained model to perform tasks such as image classification, object detection, or image segmentation.
- Monitoring: Monitor the performance of the deployed model and update it as necessary to improve its accuracy and efficiency.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

E. USE CASE DIAGRAM

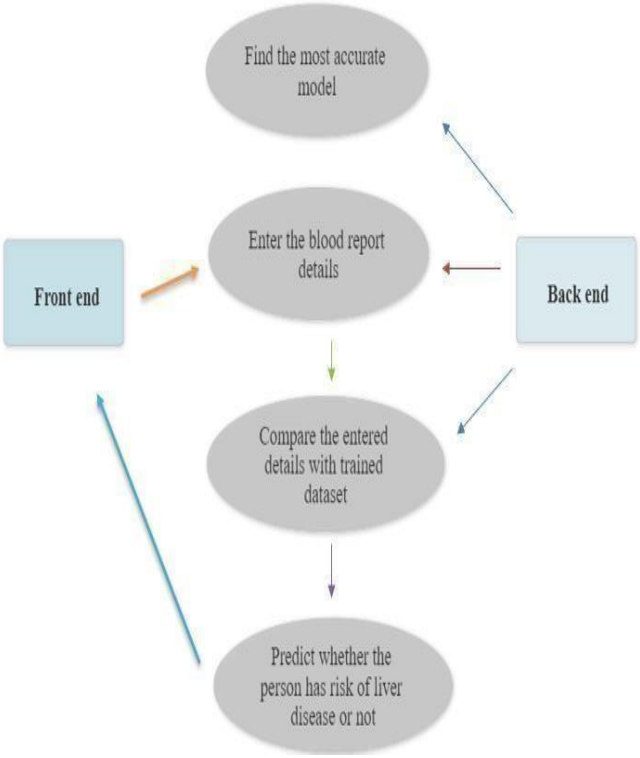

Figure 3 Use Case Diagram for Liver Disease Classification

The Use Case Diagram for Liver Disease Classification in figure 4.2 can help toprovide a visual representation of the interactions between users and the system, and to identify the key functionalities and features of the system. It can also help toensure that the system meets the needs and requirements of its users.

F. SUMMARY

This chapter discusses different machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Bagged tree classifier, and deep learning algorithm called Convolutional Neural Network (CNN) for liver disease prediction classification. Python is used as the programming language for implementing these algorithms using Jupyter notebook and Spyder through the Anaconda Navigator. The algorithms were trained with a dataset of around 400 data points to predict the occurrence of liver disease. A flowchart of the overall process is also mentioned.

VI RESULTS AND DISCUSSION

DATA SET

This data set contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 femalepatient records.

LOGISTIC REGRESSION:

Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Here we used logistic regression as first algorithm, we had given the train dataset values tothe corresponding algorithm to study and predict.

SVM:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

Bagged Tree classifier:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction.

Convolutional Neural Networks

A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data.

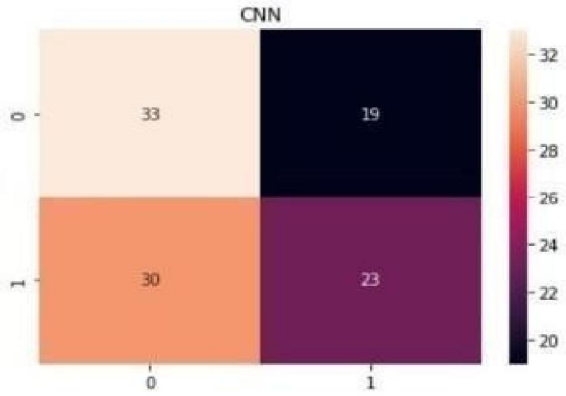

Figure 4 Confusion matrix of CNN

SI.NO	METRICES	ACCURACY IN %
1	Accuracy	99.307%
2	Precision	90.909%
3	F1 Score	100.0%
4	Recall Score	95.238%

Comparing the Metrices of Evaluation for Liver DiseaseClassification Using CNN Algorithm..

VII. FINAL RESULT

Webpage:

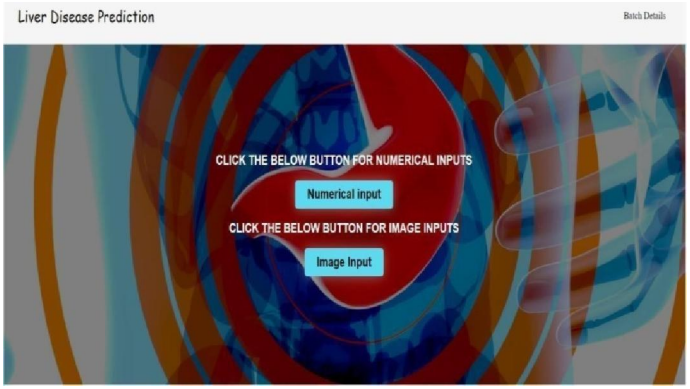

Figure 5 Screenshot of the home page

Figure 5 shows the homepage of the website where the user can able tochoose type of data which is preferrable for the user to provide to the system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

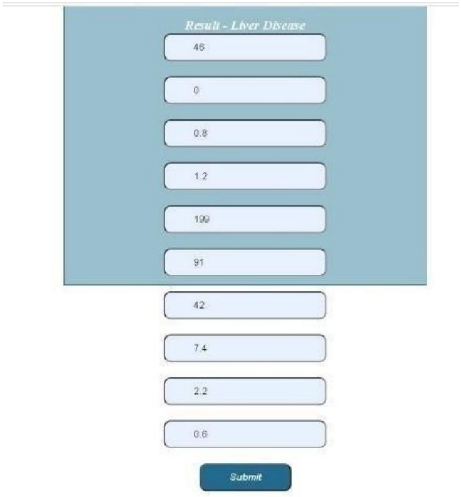

Figure 6 Screenshot of the numerical data input webpage.

Figure 7 Screenshot of CT Scanimage input web page

VIII CONCLUSION

In this research, an intelligent machine-learning based predictive system was proposed for the diagnosis of liver disease. The K-fold cross-validation method was used in the system for validation. In order to check the performance of classifiers different evaluation metrics were also adopted. The logistic regression algorithms select important fields that improve the performance of classifiers in terms of classification accuracy, specificity, and sensitivity. The Bagged classifier showed best accuracy when compared with Logistic Regression and SVM classifier. Due to the good performance of with Bagged classifier, it is a better predictive system in terms of accuracy.

IX. FUTURE SCOPE

Designing a decision support system through machine-learning-based method will be more suitable for diagnosis of liver disease. Additionally, some irrelevant features reduced the performance of the diagnosis system and increased the computation time. So, another innovative dimension of this study was the usage of feature selection algorithms to choose best features that improve the classification accuracy as well as reduce the execution time of the diagnosis system. In the future, we will perform more experiments to increase the performance of these predictive classifiers for liver disease diagnosis by using others feature selection algorithms and optimization technique.

REFERENCES

[1] M. Kumar, R. Kumar, and N. Kaur, "Liver disease prediction using machine learning techniques: A review," International Journal of Computer Applications, vol. 179, no. 26, pp. 27-33, 2021.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

[2] R. K. Singh and A. K. Singh, "A comparative study of machine learning algorithms for liver disease prediction," Journal of Big Data, vol. 8, no. 1, pp. 1- 23, 2021 M. N. U. Khan et al., "A comparative study of machine learning algorithms for liver disease prediction," Computers in Biology and Medicine, vol. 132, pp. 104307, 2021.
[3] Y. Pan, C. Wang, and H. Shen, "Prediction of liver disease using machine learning: A review," Journal of Healthcare Engineering, vol. 2021, pp. 1-17, 2021.
[4] M. N. U. Khan et al., "A comparative study of machine learning algorithms for liver disease prediction," Computers in Biology and Medicine, vol. 132, pp. 104307, 2021.
[5] A. J. Chen et al., "Liver disease diagnosis using machine learning algorithms: A systematic review," Journal of Healthcare Engineering, vol. 2021, pp. 1-19, 2021.
[6] A. Singh et al., "Predictive modeling of liver disease using machine learning algorithms: A comparative study," Journal of Medical Imaging and Health Informatics, vol. 11, no. 5, pp. 1068-1076, 2021.
[7] R. Arora and M. Kaur, "Prediction of liver disease using machine learning algorithms: A systematic review," International Journal of Advanced Research in Computer Science, vol. 11, no. 5, pp. 12-18, 2020.
[8] A. Tiwari et al., "Liver disease prediction using machine learning techniques," Journal of Medical Systems, vol. 43, no. 7, pp. 1-9, 2019.
[9] A. Mishra, S. Kumar, and R. Shukla, "Predictive modeling of liver disease using machine learning algorithms," International Journal of Data Mining & Knowledge Management Process, vol. 9, no. 3, pp. 1-12, 2019.
[10] A. H. Ahmed, "Comparative study of machine learning algorithms for liver disease diagnosis," International Journal of Computer Science and Network Security, vol. 18, no. 5, pp. 66-73, 2018.