35

by Desai n

Submission date: 02-Mar-2023 06:20AM (UTC-0500)
Submission ID: 2026929790
File name: project_Batch35_research_paper_2_-3.doc (1,013K)
Word count: 1895
Character count: 10713

THE DETECTION OF LIVER DISEASE USING MACHINE LEARNING TECHNIQUES

V.R.B.Rohini
Department of Information Technology
S.R.K.R. Engineering College(A)
SRKR Marg, Bhimavaram.
rohini.vegesna@gmail.com

Attili Rishitha
Department of Information Technology
S.R.K.R. Engineering College(A)
SRKR Marg, Bhimavaram.
attilirishitha@gmail.com

Hanumalasetti Sathkeerthi
Department of Information Technology
S.R.K.R. Engineering College(A)
SRKR Marg, Bhimavaram.
sathkeerthi2004@gmail.com

Bapatla Venkata Madhu Soumya
Department of Information Technology
S.R.K.R. Engineering College(A)
SRKR Marg, Bhimavaram.
soumyabapatla@gmail.com

Gadhiraju Hemalatha
Department of Information Technology
S.R.K.R. Engineering College(A)
SRKR Marg, Bhimavaram.
hemalathagadhiraju@gmail.com

***Abstract*— One of the deadliest and most devastating illnesses in the world is liver disease. Increased alcohol consumption, stale air, drug use, and tainted food are all contributing factors to the rapid rise in liver disease cases. Early detection of liver disease plays a key role and helps speed recovery. In the existing system, they used the KNN algorithm for the detection of liver disease, which has low accuracy and is time-consuming. We design a model that uses different types of machine learning algorithms, like random forest, gaussian nave bayes, logistic regression, and ANN, for the detection of liver disease. Our proposed system will overcome the existing system's problem and give better accuracy in a faster way. Generally, there are four stages of the liver. They are healthy but have fibrosis, cirrhosis, and cancer. In our proposed system, we consider a dataset based on those features and detect the stages of liver disease by comparing them with those of a healthy liver in the algorithm. As a result, we can determine whether the liver is healthy or if it has fibrosis, cirrhosis, or cancer.**

Keywords*—*Liver disease, detection, Machine learning, algorithms.

I. INTRODUCTION

One of the most fatal and terrible diseases in the world is liver disease. The causes of liver disease include excessive alcohol consumption, contaminated food, contaminated air, and drug usage. Digestion issues, abdominal pain, dry mouth, constipation, internal bleeding, yellowed skin, spider veins, redness on the feet, brain and memory issues, numbness, and fainting are just a few of the signs of liver disease.

The only solution if the liver is entirely damaged is a liver transplant. A quick recovery is greatly aided by early liver disease identification. Even when liver tissue has suffered only mild damage, it is very difficult to diagnose liver disease in its early stages; in this situation, the majority of medical specialists are helpless. Expert systems struggle to detect the illness. This results in treatment and drug failures. Early identification of liver illness is essential to prevent this. It will help in delivering the right care and preserving the patient's life. So, it is crucial to identify liver illness early on in order to facilitate treatment and assure a quick recovery.

The current approach has a medium level of accuracy for diagnosing liver disease conventionally. Using several machine learning methods, we will create a model that will enhance the precision and quality of the diagnosis of liver illness.

II. LITERATURE SURVEY (RELATED WORK)

For the evaluation of all the various selected algorithms, Bendi et al. [1] used two different input datasets. KNN, backward propagation, and SVM produce superior outcomes based on how well they perform in classification. discovered that SVM, KNN, backward propagation, and C4.5 all had more than 60% accuracy.

Bendi et al.'s modified rotation forest article used the US liver dataset and the Indian liver dataset as inputs. As a consequence, the MLP method with random subset outperforms the CFS algorithm for the liver dataset, achieving a higher accuracy of 64.78% compared to 63.07% for the Indian liver dataset.

The north-east region of the Andhra Pradesh (India) liver dataset was used in a publication by Yugal Kuma and G. Sahoo's [3] proposal, which used a different categorization technique. The Decision Tree (DT) algorithm provides an accuracy of 66.46%, and the findings demonstrate that it is superior to other algorithms.

Using the WEKA (Waikato Environment for Knowledge and Analysis) dataset, S.Dhamodharan [4] offered research

based on two classification methods: naive Bayes and FT trees. We came to the conclusion that Nave Bayes is a superior algorithm to other algorithms since it has an accuracy rate of 65.54% compared to FT Tree's accuracy rate of 52.6624%.

Using a decision tree technique and the Cario University data set, Heba Ayeldeen et al. [5] offer a paper for the prediction of liver fibrosis phases. The outcome demonstrates that the decision tree classifier is effective.

A study on the classification of liver disease is reviewed by D. Sindhuja and R. Jemina Priyadarsini [6]. In this study, various data mining classification strategies are examined and proven to produce superior outcomes to other algorithms.

III. System Implementation (Methodology)

- Proposed approach

We create a model that detects liver disease using various machine learning algorithms such as Random Forest, Gaussian Nave Bayes, Logistic Regression, and Artificial Neural Networks. We consider a dataset with various liver-related features. We will detect liver disease based on these characteristics. Our proposed system will solve the current system's problem by providing greater accuracy in a shorter period of time.

Machine learning is understandably one of the most widely used paradigms of big data management, where a significantly large set of distinct raw data can be effectively collated to make appropriate inferences and eventually to come up with a typical collection of contextually useful collection of integrative information. It could be further modified to any extent by.

- System Architecture

Users of the system can register their credentials by setting up an account on the Registration page, which will be used throughout the system's authentication process. Following the registration procedure, the database manages and stores the user-provided data. Users of the system can be recognized using their identification on the login page. The credentials supplied are compared to the information kept in a database for authorized users. The data is loaded once the user has successfully logged in. The loaded data was shown on the screen. In the see data phase, the user is able to see the data.

The model of the method is chosen during the model selection step depending on the dataset that we have collected. The model is built in the model building step whenever a model is chosen. There are three stages. Data for testing, validation, and training We compare our models in Model Score and select the one with the highest accuracy. Our ML Model's scores can be obtained using the Sk Learn library. Hence, to find our output, pick the algorithm with the highest score. To improve our models, it's also a good idea to calculate errors like mean squared and mean absolute errors and aim to minimize them. Finally, the programme will give us the Detected value.

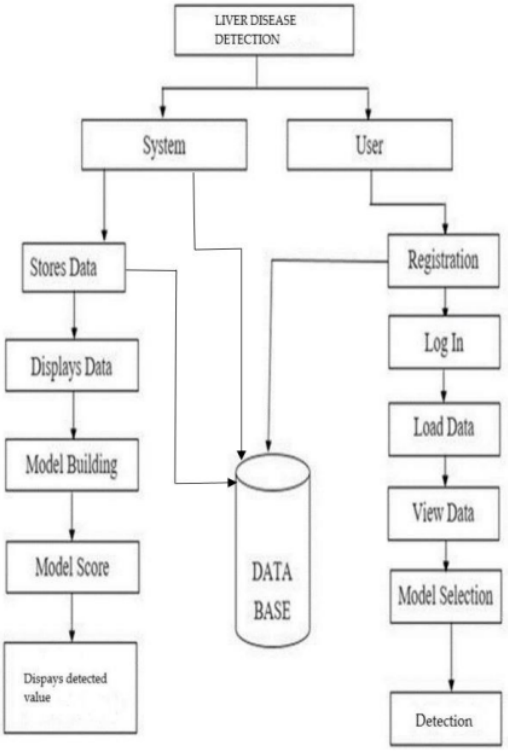


Fig 1: System Architecture

IV. Experiments & Results

Machine learning techniques are used to conduct the experiments.

We employ four different types of algorithms to find liver disease: random forest, gaussian naive bayes, logistic regression, and artificial neural networks.

For classification and regression issues, the Random Forest Algorithm is a supervised machine learning technique.

For binary and multi-class classification issues, the machine learning algorithm Gaussian Naive Bayes is utilized.

A supervised machine learning approach called logistic regression is used to determine whether something is likely to happen or not. This algorithm's output is either 0 or 1.

Machine learning algorithms called artificial neural networks are based on the idea of biological neural networks seen in the human brain. An effort to mimic the functions of the human brain led to the creation of artificial neural networks. Without assistance from a person, this algorithm is able to process large amounts of data and produce precise results.

```
# create random_forest object

random_forest = RandomForestClassifier(max_depth=3,n_estimators=56,criterion='entropy')
random_forest.fit(X_train, y_train)
```

```
RandomForestClassifier(criterion='entropy', max_depth=3, n_estimators=56)
```

```
#Predict Output

rf_predicted = random_forest.predict(X_test)
```

```
random_forest_score = round(random_forest.score(X_train, y_train) * 100, 2)
random_forest_score_test = round(random_forest.score(X_test, y_test) * 100, 2)
print('Random Forest Score: \n', random_forest_score)
print('Random Forest Test Score: \n', random_forest_score_test)
print('Accuracy: \n', accuracy_score(y_test,rf_predicted))
print(confusion_matrix(y_test,rf_predicted))
print(classification_report(y_test,rf_predicted))
```

```
Random Forest Score:
 75.74
Random Forest Test Score:
 68.57
Accuracy:
 0.6857142857142857
[[118   6]
 [ 49   2]]
              precision    recall  f1-score   support

           1       0.71      0.95      0.81       124
           2       0.25      0.04      0.07        51

    accuracy                           0.69       175
   macro avg       0.48      0.50      0.44       175
weighted avg       0.57      0.69      0.59       175
```

```
# Create gaussian object

gaussian = GaussianNB()
gaussian.fit(X_train, y_train)
```

```
GaussianNB()
```

```
#Predict Output
gauss_predicted = gaussian.predict(X_test)
```

```
gauss_score = round(gaussian.score(X_train, y_train) * 100, 2)
gauss_score_test = round(gaussian.score(X_test, y_test) * 100, 2)
print('Gaussian Score: \n', gauss_score)
print('Gaussian Test Score: \n', gauss_score_test)
print('Accuracy: \n', accuracy_score(y_test, gauss_predicted))
print(confusion_matrix(y_test,gauss_predicted))
print(classification_report(y_test,gauss_predicted))
```

```
Gaussian Score:
 56.13
Gaussian Test Score:
 53.14
Accuracy:
 0.531428571428571
[[44 80]
 [ 2 49]]
              precision    recall  f1-score   support

           1       0.96      0.35      0.52       124
           2       0.38      0.96      0.54        51

    accuracy                           0.53       175
   macro avg       0.67      0.66      0.53       175
weighted avg       0.79      0.53      0.53       175
```

```
#Predict Output

log_predicted= logreg.predict(X_test)
logreg_score = round(logreg.score(X_train, y_train) * 100, 2)
logreg_score_test = round(logreg.score(X_test, y_test) * 100, 2)


#Equation coefficient and Intercept

print('Logistic Regression Training Score: \n', logreg_score)
print('Logistic Regression Test Score: \n', logreg_score_test)
print('Coefficient: \n', logreg.coef_)
print('Intercept: \n', logreg.intercept_)
print('Accuracy: \n', accuracy_score(y_test,log_predicted))
print('Confusion Matrix: \n', confusion_matrix(y_test,log_predicted))
print('Classification Report: \n', classification_report(y_test,log_predicted))
```

```
Logistic Regression Training Score:
 72.06
Logistic Regression Test Score:
 68.0
Coefficient:
 [[ -0.00994993  -0.09851216  -0.30688723  -0.0008293   -0.01078828  -0.00275598
   -0.23099674   0.4020893    0.59475582   0.25335289   0.09115993]]
Intercept:
 [0.361086]
Accuracy:
 0.68
Confusion Matrix:
 [[107  17]
 [ 39  12]]
Classification Report:
              precision    recall  f1-score   support

           1       0.73      0.86      0.79       124
           2       0.41      0.24      0.30        51

    accuracy                           0.68       175
   macro avg       0.57      0.55      0.55       175
weighted avg       0.64      0.68      0.65       175
```

```
Classification Report
              precision    recall  f1-score   support

           0       0.70      1.00      0.82        82
           1       0.00      0.00      0.00        35

    accuracy                           0.70       117
   macro avg       0.35      0.50      0.41       117
weighted avg       0.49      0.70      0.58       117
```

V. EVALUATION METRICS

The performance is analyzed by using the various metrics such as F-beta score, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN)

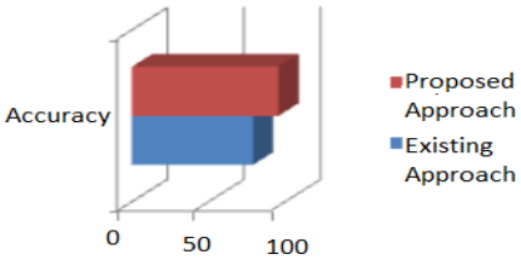
Precision=TP/ (TP+FP)

Recall=TP/ (TP+FN)

Accuracy = (TP+TN)/ (TP+TN+FP+ FN)

F-beta score =

(1+β2)*precision*recall/((β2*precision)+recall)



Algorithms	Accuracy
Existing Approach	56.76%
Proposed Approach	70.45%

VI . CONCLUSION

Liver disease is a condition that is spreading more and more with time. With the help of machine learning techniques, we proposed a model in this study for determining a patient's stage of liver illness. We employed four machine learning methods: gaussian naive Bayes, random forest, artificial neural networks, and logistic regression. We take into account a set of data that includes features like total bilirubin, direct bilirubin, total proteins, albumin, the A/G ratio, SGPT for Alanine Aminotransferase, SGOT for Aspartate Aminotransferase, and alkaline phosphatase. With the help of the proposed model, we may identify liver illnesses based on those features. In comparison, the proposed model outperforms previous models and has a higher accuracy. Using the F-beta score as a performance indicator, a performance evaluation was conducted.

VII. FUTURE WORK

The major goal of this study is to identify liver disease at an early stage and aid in a quick recovery for those affected. One of the deadliest illnesses in the world is liver disease. The importance of early liver disease identification cannot be overstated. Using machine learning algorithms, we create a model to identify liver illness. On the subject of diagnosing liver illness, there have been numerous relevant works. You can refer to the most recent diagnostic and therapeutic choices for patients with liver diseases in "The Handbook of Liver Disease" by Drs. Lawrence S. Friedman

and Emmet B. Keefe. Researchers refer to the "gestational alloimmune liver disease" (GALD) treatment as one that improves survival rates.

ACKNOWLEDGMENT *(Heading 5)*

We appreciate Bendi, Pascal Vincent, Andrej Karpathy, and Vladimir Vapnik and their helpful comments and recommendations. We also wish to thank Noventa for helping to make this initiative possible. We would like to thank everyone who took part in the user research. Projects PMCID: PMC8832876 and ID-LIVER Consortium, funded by the National Health Service Foundation Trust, Manchester, United Kingdom, have helped to develop this work in part.

REFERENCES

[1]Heimbach JK et al authors performs a research and give guidelines for the treatment of hepatocellular carcinoma. Hepatology. 2018 Jan;67(1):358-380.

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

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

[4] Chaitrali, S. D., & Sulabha, S. A. (2012).has Improved Study for predicting heart disease System using Data Mining Classification Techniques.

[5] Paul R. Harper, A review and comparison of machine learning classification algorithms for medical decision making.

[6] BUPA Liver Disorder Dataset. UCI repository machine learning databases. [6] Prof Christopher N. New Automatic Detection of Liver Status Using Bayesian Classification.

[8] B. Y. Ramana, M. S. P. Babu and N. B. Venkateswarlu. (2011), "A Critical Study of Selected Classification Algorithms for Liver Disease" is very useful for the liver disease detection.

[9] Detection", International Journal of Database Management Systems (IJDMS), vol.3, no.2, pp. 101-114

[10] S. Vijayarani, and S. Dhayanand. (2015) "Liver disease prediction using SVM and Naïve Bayes algorithms." This journal is very useful for the liver disease detection in early stage.

[11] International Journal of Science, Engineering and Technology Research (IJSETR) vol. 4, no. 4, pp. 816-820

[12] Schalk BW, Visser M, Bremmer MA, et al. studied on Change of serum albumin and risk of cardiovascular disease and all-cause mortality: Longitudinal Aging Study Amsterdam. Am J Epidemiol 2006;164:969–77.Abstract/FREE Full Text Google Scholar

published.

35

ORIGINALITY REPORT

15%	9%	13%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Gitam University Student Paper	4%
2	Golmei Shaheamlung, Harshpreet Kaur, Mandeep Kaur. "A Survey on machine learning techniques for the diagnosis of liver disease", 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020 Publication	3%
3	www.warse.org Internet Source	2%
4	Keshali Pathak, Dheeraj Kumar Singh. "Convolutional Neural Networks based Liver Tumor Classification", 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2022 Publication	2%
5	Submitted to University of Moratuwa Student Paper	1%
6	Submitted to Wright State University Student Paper	1%

7	repository.nusamandiri.ac.id Internet Source	1 %
8	centaur.reading.ac.uk Internet Source	1 %
9	scholarsjunction.msstate.edu Internet Source	1 %
10	Moloud Abdar, Mariam Zomorodi-Moghadam, Resul Das, I-Hsien Ting. "Performance analysis of classification algorithms on early detection of liver disease", Expert Systems with Applications, 2017 Publication	<1 %
11	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
12	www.ncbi.nlm.nih.gov Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		