International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Liver Disease Prediction Using EnsembleTechnique

Mrs.Ancy Y Das
Assistant Professor Computer Science and Engineering
BGS Institute of Technology
Adhichunchanagiri University
ancyydas@bgsit.ac.in

Venkatesh L
20CSE088
Computer Science and Engineering
BGS Institute of Technology
Adhichunchanagiri University
venkatesh.l8402@gmail.com

ABSTRACT- Liver disease is a very serious illness. It occurs in the body, most commonly in the liver. The liver's main job is to remove waste produced by the body. It also stores essential vitamins needed by the body so that they are not wasted. The liver also helps digest food. This disease is serious. The first step is to reduce the chance of contracting it. Being able to detect it early saves lives. Around 3.5 percent of people in the world have liver disease. Machine learning models have been employed in various new methods to predict diseases like KNN, random forest, SVM, and logistic regression. Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) are some of the deep learning approaches that are used for solving this problem. There is no doubt that these approaches will greatly extend the lives of the people suffering from this condition and prevent them from having chronic liver diseases (CLD). The use of automated transactions, bar codes on products alongside better data collection systems has led to collection of huge volumes of information. The proposed system uses ensemble methods like random forest, ADA boost, and gradient boost combined to get betteraccuracy.

I. INTRODUCTION

The liver is a big part of the body so we must all be careful about how we take care of it. It helps make food into things the body can use. It also gets rid of bad stuff. Viruses and drinking too much alcohol can hurt the liver bad. This can make people very sick or even die. There are many liver sicknesses like hepatitis, cirrhosis, tum ours, cancer, and more. The liver is a very importantbody part that needs to be kept healthy. Liver disease is a big health issue in India. One kind is fatty liver disease. It happens when fat builds up in the liver. Every year, there are over million cases in India. No bad effects occur initially, so testing is required to detect it. When any problem affects the proper functioning of the liver and results in illness, it is called liver disease.

The liver does many important jobs in the body. If it gets sick or hurt, those jobs can't happen right. The body can be seriously injured by this. Liver disease is also known as hepatic disease. The term "liver disease" means all the problems that can make the liver not work right. Liver failure occurs when 75% or more of liver tissue has been lost. The term "artificial intelligence" designates a kind of computer thought which permits computer programs to assimilate information, acquire knowledge, and apply it in various domains.AI is used in almost every field now, and it has parts like Deep Learning and Machine Learning. AI helps doctors find liver disease early. It looks at medical tests and scans to see if the liver is not working right. AI can also help pick the best treatment for liver disease, like medicine or surgery. Some common liver diseases are hepatitis, cirrhosis, and liver cancer. Hepatitis is when the liver is swollen and damaged. Cirrhosis is scarring of the liver that makes it hard for the liver to work. Liver cancer is when bad cells grow out of control in the liver. Taking good care of your liver is important. You can do this by eating healthy foods, exercising, and not drinking too muchalcohol. Many things can hurt the liver. Eating too much food, drinking alcohol, breathing bad air, and taking too many pills are problems. This study looks at using machine learning to know when someone has a sick liver. This may help doctors have less work. The liver is a big, important organ. We need to keep it healthy. Machines can spottrouble in the liver by looking at data. Doctors can thenhelp people with liver issues. Machine learning can make finding liver disease easier. It uses math to learn what a sick liver looks like. This tech may make healthcare better for many people.

These are the expected primary outcomes of this project:

1) Prediction of liver disease.
2) Predicting early stage and to prevent other causalities.
3) Decline in deaths associated with liver disease.
4) Better identification of liver disease.

II. MOTIVATION

This paper talks about using machines to predict liver disease. This is a very big deal for doctors and sick people. The research already done has some holes. This paper wants to fix those holes with new ideas. The new ideas let machines predict liver disease very fast, with great accuracy, and for less money. This is better than what is done now. These new ideas can help doctors and sick people in big ways. They can also help people study machines and medicine.

III. LITERATURE SURVEY

Some smart people made this paper. It usescomputer learning to guess if someone has liverdisease. They looked at medical facts about people. They tried four smart ways to guess: Logistic Regression, When it comes to solving classification problems, there are techniques that are commonly used. Support Vector Machine was right most of the time. In fact, it was right 75 percent of the time if we are to be more specific. To be exact, 75 cases were solved correctly by support vector machines out of 100 cases.

Some people made a new way to look at liver sickness. They used a mix of three math thingscalled Random Forest, XGBoost, and Gradient Boosting. They say their way works better than old ways. It can handle lots of info. But, it might make the math model fit the info too much. That can be a problem sometimes.

They proposed this paper that applies machine learning techniques to liver diseaseprediction using non-invasive and readily available data, such as demographic, clinical, and laboratory variables. It reports that Decision Stump has the best accuracy of 91.7%.

This paper talks about a way to find out if someone has a liver disease. It uses text mining to look at how Chinese medicine talks about this disease. The method uses a machine learning model called an autoencoder to get features from the text. Then, it uses a classifier to predict if someone has the disease or not. The paper is more about mining text than machine learning. The accuracy might not be high, and the predictions could be wrong sometimes. What if the method could not tell if youwere sick or not? That would be bad. We needways to find health problems early.

Maybe mixing this idea with other tests could help make better predictions. Doctors know a lot, so listening to them is wise too. They said that diagnosing diseases is very costly and hard. Therefore, they were assessing the various machine learning algorithms to determine the most efficient ones in saving money used in predicting chronic liver diseases.

Dr. S Vijayarani and Mr. S Dhayanand did a study. The goal was to guess if people have liverproblems. They compared how good they were at guessing right. They looked at how accurate the guesses were and how fast the guesses happened. The results showed that SVM was better at guessing liver problems. But, the way they guessed had some issues. It needed a lot of data to learn from. And it might have learned too much from the data. So it might not guess right for new people.

IV. PROPOSED SYSTEM

In the system we have here, we use an Ensemble way to see if someone has liver issues. This helps the doctor by using computer ways to look at how the person feels. Ensemble ways make many models and then put them all together to get good results. Most times, using many models together works better than just one model. In a nutshell, this technique blends multiple algorithms to make one model. Our system uses Gradient Boosting Classifier + AdaBoost Classifier (Ensemble Technique). The aim is to predict if a person has liver disease based on datasets of people with symptoms. The data helps the system learn patterns and make predictions accurately. Ensemble techniques improve model performance by combining strengths of different algorithms. We use Ensemble Technique to predict disease. This system is very helpful for hospitals and doctors to easily find out what sickness a person has. Regular people can also use this system to check if they might have a sickness. The system looks at many factors to figure out the right sickness. It takes information and makes a good guess about what disease someone might have. This system willchange the way and can be early as possible as it will lead to

save the person's life. This whole work is focused on how we can predict the disease by given datasets so that will help in preventing and curing the disease of the patients.

The datasets follows up through a path that is been indicated in the following diagram:

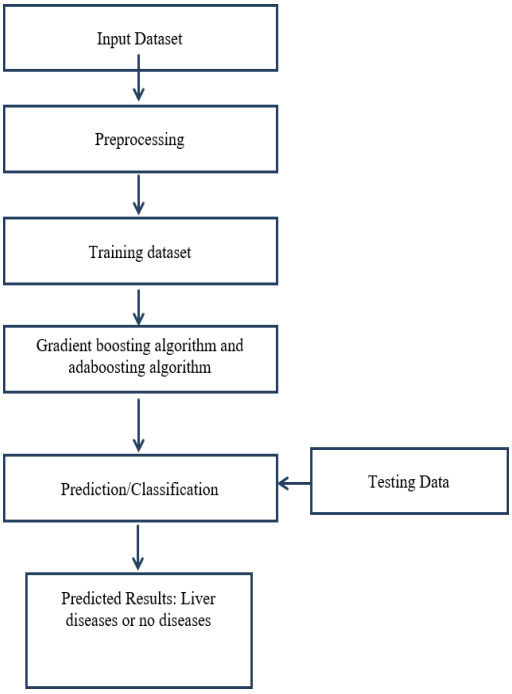

Fig. Data flow Diagram

V. SYSTEM DESIGN

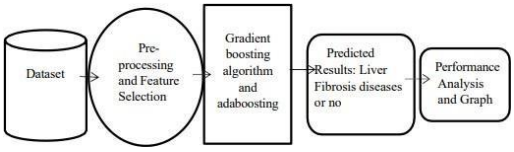

Fig. System Architecture

[1] Collect Data- This is the first big step in making a machine learning model. Getting data is very important. The better the data, the better the model will work. There are many ways to get data, like web scraping and manual work. We will predict liver disease using Ensemble Technique.

[2] Dataset - The dataset has 590 different data points. It has 11 columns. It contains data that is needed to predict liver disease. Only some data is taken from medical test reports.

	age	gender	tot_bilirubin	direct_bilirubin	tot_proteins	albumin	ag_ratio	sgpt	sgot	alkphos	is_patient
0	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	1
1	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
2	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	1
3	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	1
4	72	Male	3.9	2.0	195	27	59	7.3	2.4	0.40	1

Fig. Datasets and required data.

[3]Pre-Processing and Feature Selection - To get the right value, we cleaned the dataset. We took outdata that was not needed, incomplete, or wrong. Weused many python libraries to clean the raw data. At the same time, we also changed data from words to numbers so the data would work with the classifiers.

[4] Feature selection- Finding key data is a step that helps solve issues. It finds a smaller set of data from the big data set. This makes the data smallerto work with. Data cleaning and picking key parts Gradient boosting method and adaboosting Outcomes: Liver sickness or no sickness.

[5] Model Selection- We used an important machine learning approach. It mixes two methods - Gradient Boosting and AdaBoost. These two ways work together. We call this an ensemble technique. With this smart approach, we got very high accuracy of 92.1% on the test data. So we used this ensemble algorithm in our final system.

[6] Ensemble Technique- Ensemble methods are ways to make models give better results. They do this by combining many models together instead of using just one model. When you combine many models, the results become much more accurate. This has made ensemble methods very popular for machine learning.

[7] Performance Analysis- We look at different things to check how good this model works. Accuracy is how many times the model guessed right out of all the tries. Classification error is how much the model got wrong. These are important to see if the model works well or not.

V. ALGORITHMS

[1] When faced with a problem of numerical values, we make use of Gradient Boosting Regressor. Whereby a sorting issue arises, we turn to a Gradient Boosting Classifier. They only differ in their respective "Loss function". The objective is to minimize this loss function by adding gradually using gradient de- scent. Being that it is based on loss function; there would be other loss functions for numbers problems such as Mean squared error (MSE).

[2] AdaBoost is a Machine Learning method. Itjoins easy ways to boost their use. The most common way with AdaBoost is trees with one level. These are called Decision Stumps. They have only one question to split data.

VI. CONCLUSION

We did the whole work to get the new output. That included making the data ready, getting it fixed, choosing different models, and putting them together. The result will help the medical field predict liver disease better. This can save lives of people with the disease. The model had high accuracy scores. The training accuracy was 100%. The test accuracy was 92% and above. This work can help the whole medical industry a lot. High accuracy can also be used by developers in real-time web applications. Different ensemble methods can also be used by them too. All machine learning and ensemble techniques can be compared by them as well.

VII. REFERENCES

[1]A. M. K. T. G. K. M. K. B.Muruganantham, ”Liver Disease Prediction Using Classification Algorithms”.

[2] S. E. A. I. E.-D. Engy El-Shafeiy, ”Prediction of Liver Diseases Based on Machine Learning Technique for Big Data”.

[3] Z. T. Joy Roy, ”A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning”.

[4] S. R. D. K. Vasan Durai, ”Liver disease prediction using Machine Learning”.

[5] M. S. A. K. G. S. P. 1Rakshith D B, ”Liver Disease Prediction System using Machine Learning Techniques,” 2021.

[6] J. C. M. J. M. E. I. Joel Jacob, ”Diagnosis of Liver Disease Using Machine Learning Techniques,” 2018.

[7] L. J. T. Banu Priyal, ”Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms”.

[8] R. T. Ribeiro, R. T. Marinho and J. M. Sanches, ”Classification and Staging of Chronic Liver Disease From Multimodal Data,” in IEEE Transactions on Biomedical Engineering, vol. 60, no. 5, pp. 1336-1344, May 2013, doi: 10.1109/TBME.2012.2235438.

[9] J. Lara, Y. Khudyakov, F. Xavier L o´pez-Labrador, F. Gonzalez Candelas and M. Berenguer, ”Hepatitis C virus genetic association to rate of liver fibrosis progression,” 2013 IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences (ICCABS), 2013, pp. 1-1, doi:10.1109/ICCABS.2013.6629225

[10] S. Ambesange, V. A, R. Uppin, S. Patil and V. Patil, ”Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques,” 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2020, pp. 98-102, doi: 10.1109/CCEM50674.2020.00030

[11] F. Destrempes et al., ”Machine learning based on quantitative ultrasound for assessment of chronic liver disease,” 2020 IEEE International Ultrasonics Symposium(IUS),2020,pp.1-3,doi:10.1109/IUS46767.2020.9251512.

[12] H. -q. Zhang et al., ”Using text mining to understand traditional Chinese medicine pathogenesis of nonalcoholic fatty liver disease,” 2013 IEEE International Conference on Bioinformatics and Biomedicine, 2013, pp. 311-314, doi: 10.1109/BIBM.2013.6732702.

[13] Shah and S. Modi,”Comparative Analysis of Psychometric Prediction System,” 2021 Smart Technologies, Communication and Robotics (STCR),2021,pp.1-5,doi:10.1109/STCR51658.2021 .9588950.

[14] Sivasangari, B. J. Krishna Reddy, A. Kiran and P. Ajitha,”Diagnosis of Liver Disease using Machine Learning Models,” 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 627-630,doi:10.1109/ISMAC49090.2020.9243375.