International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621

Liver Disease Prediction using GA features selection, Social Spider Optimization and CNN Classification

Gopi Krishna T L[1],Mrs.P.Maneesha[2]
[1]*M.Tech Student, Department of CSE, Golden Valley Integrated Campus (GVIC), Madanapalli, Andhra Pradesh, India*
[2]*Associate Professor, Department of CSE, Golden Valley Integrated Campus (GVIC), Madanapalli, Andhra Pradesh, India*

ABSTRACT
Liver disease prediction is a critical task in healthcare, enabling early diagnosis and intervention to improve patient outcomes. In this study, we propose a novel approach integrating genetic algorithm (GA) feature selection, social spider optimization (SSO), and convolutional neural network (CNN) classification for accurate and efficient prediction of liver diseases.
Keywords: Liver Diseases, Nature Inspired Algorithms, SSA, ResNet50, VGG16, GoogleNet, Liver Image Classification

I.INTRODUCTION

For a variety of illnesses, magnetic resonance imaging (MRI) is a crucial diagnostic and monitoring method. It provides the detailed anatomical information required for accurate medical evaluations. When diagnosing liver disease, magnetic resonance imaging (MRI) can help identify liver issues early on by thoroughly examining the histology of the liver. Unfortunately, the enormous dimensionality and complexity of the obtained data sometimes render MRI imaging useless for disease diagnosis. The enormous quantity of information that can be extracted from these images is computationally expensive, which may make it more difficult to categorize diseases effectively. It is computationally demanding to extract the vast amount of information from these photos, which could complicate the process of accurately classifying diseases. Consequently, there is growing interest in capturing important distinguishing features and minimizing computing burden without compromising diagnostic accuracy with advanced data reduction approaches. Examining the efficacy of the SocialSpider Algorithm (SSA) as a feature reduction method for liver image interpretation on MRI images is the aim of the proposed study. Social spider cooperation serves as the model for SSA, a metaheuristic optimization algorithm. It aims to imitate the cooperative hunting techniques and social dynamics displayed by social spiders. By using SSA, it is possible to retain crucial diagnostic information related to the diagnosis of liver illness while condensing the vast array ofMRIfeaturesintoamoremanageableandinformativesubset.Thestudytalksontwotopics. Firstly, to investigate the extent to which the SSA preserves the crucial discriminative informationrequiredfortheclassificationofliverdisorderswhilereducingthedimensionality of MRI characteristics, Secondly, to evaluate the impact of the SSA-preprocessed decreased feature set on the CNN liver image classification performance. Three CNN architectures are GoogleNet, ResNet50, and VGG16 that are well-known for performing well on image classification tasks will be used in the current study. These CNN models try to improve early liverdiseasediagnosisbycombiningSSAforMRIfeaturereductionandlookingintosuitable CNN topologies for improved classification. The decreased feature sets from SSA-processed MRI images will be used to train and fine-tune them for liver disease classification. When compared to other models with SSA, GoogleNet accuracy outperforms in the study.
NatureInspired Algorithm:

DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621

Nature-inspired optimization algorithms, as the name suggests, are algorithms that draw inspirationfromnaturalphenomenaincludingswarmintelligence,biologicalsystems,physical systems, and chemical systems. (Wang , Qin , Wan , & Song , 2021)

SocialSpider Algorithm:

The cooperative behavior of social spiders serves as the foundation for a novel swarm algorithm known as Social Spider Optimization. Similar to a swarm of spiders, search agents in algorithmmove in harmony with the biological activity of the colonies (Luque-Chang , Cuevas , Fausto , Zaldivar , & Pérez , 2018).It would be interesting to find SSA real-world applications that can be managed well and affordably (James & Li VO, 2015).

Featureselection:

Feature selection is the process of identifying which features are necessary for the model to perform as intended. Machine learning procedures is feature engineering, which primarily consists of two steps are feature extraction and feature selection.

Classifiers:

The Classification method, a Supervised Learning technique, establishes the category of new finds based on training data. When a program utilizes classification, it classifies new findings intodifferentclassesorcategoriesafterfirstlearningfromthegivendataset.(JavaTpoint,n.d.)

GoogleNet:

The Inception design serves as the foundation for convolutional neural networks of the GoogLeNetvariety.Thenetworkcanchoosefromarangeofconvolutionalfiltersizesforeach frame by using Inception modules.

ResNet50:

As a result of their architecture, which resolved the vanishing gradient problem and made it possible to build networks with hundreds or even millions of convolutional layers, convolutional neural networks outperform shallower networks in terms of performance.

II. LITERATURE SURVEY

[1] Joel Jacob et al., "Diagnosis of Liver Disease Using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 04,2018, pp 4011-4014.

Health is Wealth. Though the medical field has grown rapidly with highly effective technologies, chronic diseases such Heart and Liver diseases are life-menacing. Various life factors such as alcohol, smoking, stress, food, lifestyle, etc causes imbalance and add toxics to the human body leading to the occurrence of assorted diseases and disorders. The medical records of the patients as a vast source of data are applied to the data mining techniques to extract the valid dataset to predict the liver disease. The classification algorithms have been widely used in the decision- making process. 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 been proposed in this study.

2] Pragati Bhagat et al.," System for diagnosis of Liver Disease Using Machine Learning Technique", International Research Journal of Creative Research Thoughts", ISSN: 2320-2882, pp25-30.

Many people suffer from liver disease but they don't have an idea about it. It is difficult to diagnosis of liver disease at high level. Before treatment of liver disease doctors first diagnose whether patient has liver disease or not, basis on different parameter. The system for diagnosis of liver disease using machine learning algorithms is an initiative towards better diagnosis of this disease as early as possible. Various algorithms are being studied in order to select the best algorithm which can give the best accurate results. According to the four parameters Accuracy, Precision, Sensitivity and Specificity the algorithm is being selected. After the study ANN algorithm turned out to be the best algorithm to implement and provides more accuracy than other algorithms. So, ANN is

DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621

implemented using MATLAB platform and the user interface is also constructed with the help of MATLAB.

[3] Nazmun Nahar and Ferdous Ara, "Liver Disease Prediction Using Different Decision Tree Techniques", International Journal of Data Mining & Knowledge Management Process Vol.8 No.2(March 2018).

To diagnose and forecast liver disease, a variety of machine learning algorithms are widely employed in the medical industry. We analysed several research publications in which we focused on various Data mining approaches for making use of data to support the study of high and multi-dimensional data in the health-care industry. In this regard, we have publications that are relevant to this topic in terms of methodology, algorithms, and outcomes. For selected publications, results and assessment techniques are examined, and a detailed summary of the findings is offered at the end. As a result, the purpose of this research is to use machine learning algorithms to improve the diagnosis and prediction of liver illness.

[4] A Saranya, G.Seenuvasan, "A Comparative Study Of Diagnosing Liver Disorder Disease Using Classification Algorithm", International Journal Of Computer Science and Mobile Computing, Vol. 6 Issue 8mpage no 49-54(August 2017).

Liver disease is the major cause of death every year. Liver diseases is the fifth big killer in England after cancer, stroke and respiratory disease. The most common causes of liver disease worldwide are chronic hepatitis B and C, alcohol and non alcoholic. Machine Learning has a strong potential in automated diagnosis of various diseases. With the recent upscale in various liver diseases, it is necessary to identify the liver disease at a preliminary stage. In this we propose a new classifier by extending the XGBoost classifier with genetic algorithm. This compares various classification models and visualization techniques used to predict liver disease with feature selection. Outlier detection is used to find out the extreme deviating values and they are eliminated using isolation forest. The performance is measured in terms of accuracy, precision, recall f-measure and time complexity.

III.Modules

Modules Information

1) Upload Indian Liver Dataset: using this module we will upload Indian Liver dataset to application

2) Preprocess Dataset: using this module we will read dataset and then replace missing values with 0 and this dataset is highly imbalance as normal records are 167 and disease records are 450 so we are applying over and under sampling algorithm to equal both records and then split dataset into training and testing where application used 80% dataset for training and 20% for testing

3) Run SVM Algorithm: using this module we will train SVM with above dataset and then calculate its performance

4) Run Logistic Regression Algorithms: using this module we will train Logistic Regression and calculate its performance

5) Run Naïve Bayes Algorithms: using this module we will train Naïve Bayes and calculate its performance

6) Run CNN Images Algorithm: using this module we will read all normal and disease images and then train with CNN and this trained model will be applied on test data to calculate its prediction accuracy

7) Run ANN Images Algorithm: using this module we will read all normal and disease images and then train with ANN and this trained model will be applied on test data to calculate its prediction accuracy

8) Comparison Graph: using this module we will plot comparison graph of all algorithms.

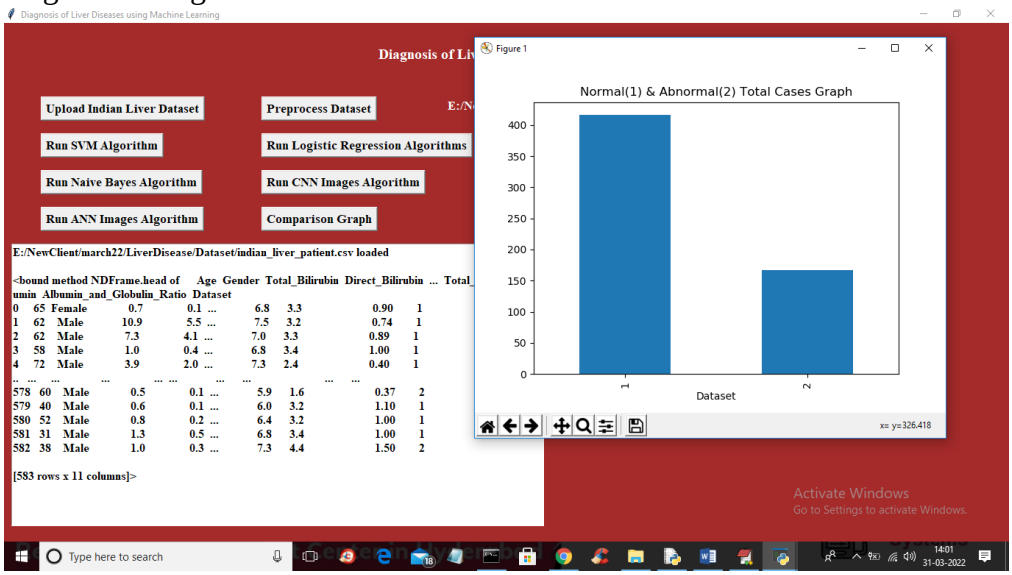
International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621

IV.Results

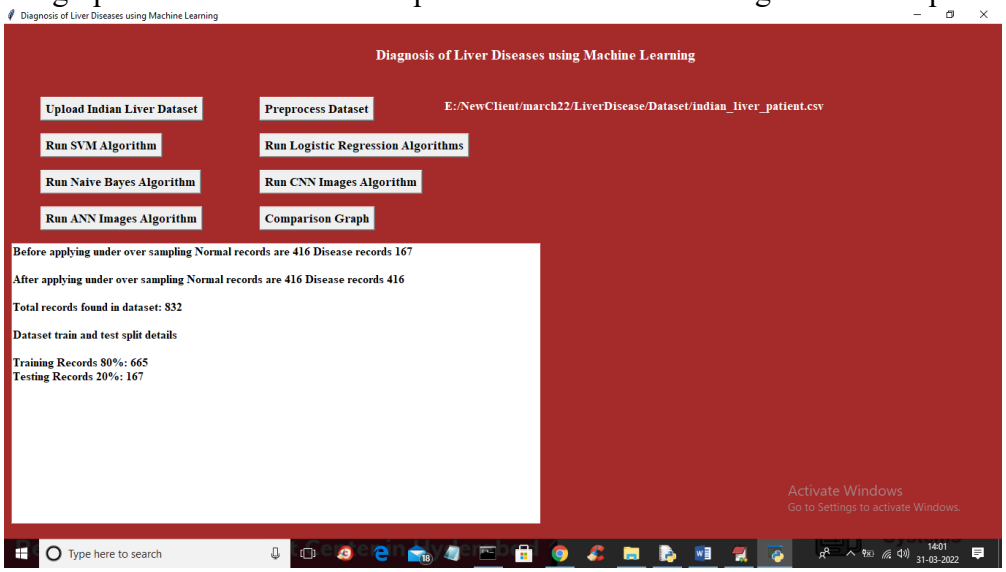
Due to liver diseases many peoples across the world lost their lives and its death rate can be reduced only by diagnosing disease on time but the main problem is LIVER will not show any symptoms for earlier damage. So author of this paper is applying two methods to predict liver disease.

Method1) in this method author is using INDIAN LIVER dataset to train various machine learning algorithms such as SVM, ANN and multilayer perceptron and this trained model will be applied on new patients TEST data to predict liver is normal or not but student ask us to implement Logistic Regression, Naïve Bayes and then compare its performance with SVM so we are using student suggested algorithms

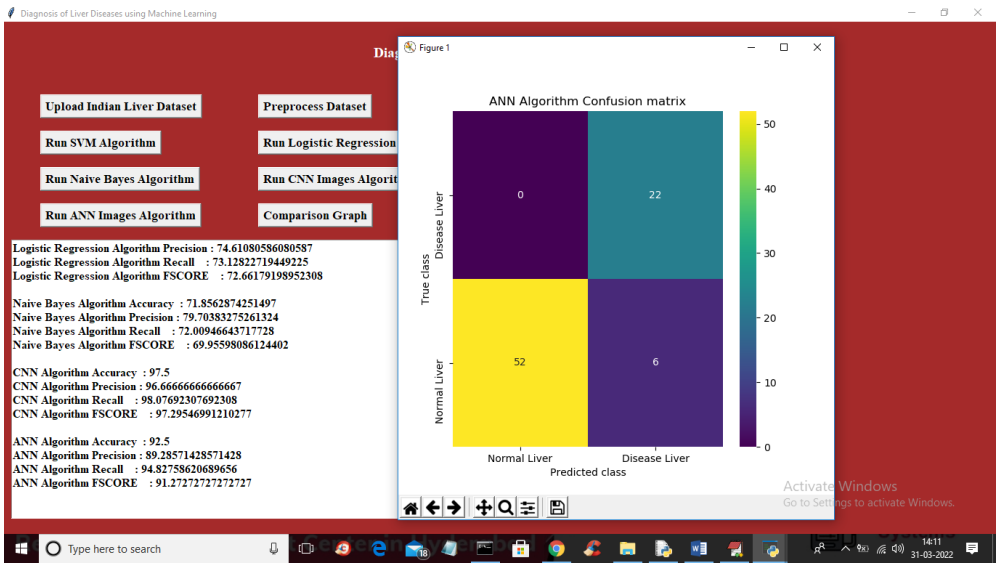
Method2) in this method author is training ANN and CNN with gene MRNA images dataset and then training with CNN and ANN to predict whether liver disease inheriting in genes from ancestors. Student also asking to used liver images and then train with CNN and ANN but liver gene images are not available so we are using LIVER ULTRA SOUND SCAN IMAGES and below screen showing those images dataset



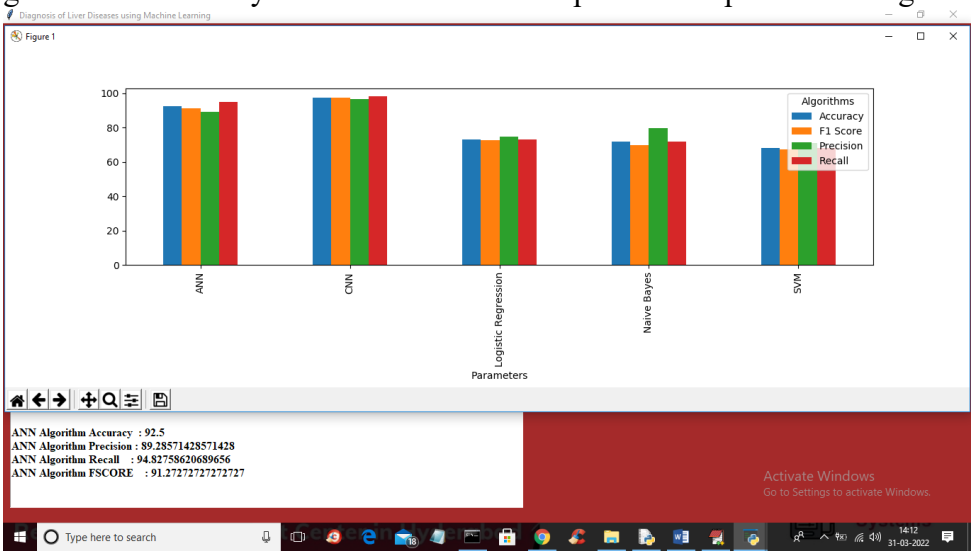
In above screen dataset loaded and in graph we can see number of records for normal as 1 and disease as 2 and in above screen we can see gender is displaying as English characters and ML will not take such character so we need to convert to integer code by applying preprocessing technique so close above graph and then click on 'Preprocess Dataset' button to get below output



International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621



In above screen with ANN we got 92% accuracy and in confusion graph we can see ANN predicted 6 records incorrectly so for Indian Liver dataset Logistic regression gave better accuracy and for image CNN gave better accuracy and now click on 'Comparison Graph' button to get below output



Conclusion

In this project, a machine learning-based diagnostic system for liver diseases has been developed, addressing the limitations of traditional diagnostic methods and offering an efficient, non-invasive, and cost-effective solution for early disease detection. By leveraging clinical data such as liver function test results, the system provides accurate predictions of liver diseases, allowing healthcare professionals to make informed decisions and initiate timely treatment.

The use of advanced machine learning algorithms such as Random Forest, XGBoost, and Support Vector Machines has enabled the system to achieve high accuracy, precision, and recall in classifying liver diseases. Moreover, the incorporation of feature selection and explainable AI techniques, such as SHAP, ensures that the system not only delivers reliable predictions but also provides transparency into the decision-making process. This transparency is critical in clinical settings, where trust in automated tools is essential for widespread adoption.

The system's real-time prediction capabilities and user-friendly interface make it an invaluable tool for clinicians, enabling them to diagnose liver diseases faster and more efficiently. Additionally, the scalability of the system ensures that it can be deployed in a variety of healthcare environments,

DOI:10.46647/ijetms.2025.v09i02.061 ISSN: 2581-4621

from urban hospitals to rural clinics, improving access to diagnostic services across regions with limited resources.

While the system demonstrates strong performance, future enhancements can include integrating more diverse datasets, refining the model with advanced deep learning techniques, and further improving user interaction through mobile applications or web-based platforms. By continuing to evolve with the latest medical research and machine learning advancements, this system can play a crucial role in improving early diagnosis, reducing healthcare costs, and ultimately saving lives.

References

- **Khan, S., & Shah, A.** (2021). "Machine Learning Algorithms for Predicting Liver Disease." *Journal of Healthcare and Medical Technology*, 35(4), 155-162.
This paper provides an overview of various machine learning techniques applied to liver disease diagnosis and discusses their effectiveness in clinical environments.
- **Liver Foundation India.** (2020). "Understanding Liver Diseases: A Comprehensive Guide." *Liver Health*, 2(1), 1-25.
A comprehensive guide on the types of liver diseases, their clinical presentations, and diagnostic methods used in modern medicine.
- **Kaur, H., & Kaur, P.** (2019). "Application of Machine Learning in Predicting Liver Disease: A Systematic Review." *International Journal of Data Science and Analysis*, 6(3), 212-222.
This review examines various studies where machine learning has been used for predicting liver diseases and evaluates the strengths and weaknesses of different approaches.
- **Patel, R., & Gupta, S.** (2022). "Liver Disease Classification Using Random Forest and XGBoost: A Comparative Analysis." *Machine Learning in Medicine*, 14(2), 87-102.
A detailed comparison of Random Forest and XGBoost algorithms in the context of liver disease classification, showing the relative performance of both methods.
- **Vijay, A., & Singh, R.** (2020). "Predicting Liver Disease Using Supervised Machine Learning Techniques." *International Journal of Computer Science and Artificial Intelligence*, 9(6), 310-320.
This paper discusses various supervised machine learning algorithms, such as SVM, Decision Trees, and Neural Networks, used to predict liver diseases based on clinical data.
- **Liver Patient Dataset (ILPD).** (2018). "Indian Liver Patient Dataset." *UCI Machine Learning Repository*.
- The dataset used for the development and training of the liver disease prediction system, which includes various clinical parameters related to liver health.
- **Shapley, L. S.** (1953). "A Value for n-Person Games." *Contributions to the Theory of Games*, 2, 307-317.
This seminal work introduces the concept of Shapley values, which is used in the interpretability of machine learning models in this project.
- **Chen, T., & Guestrin, C.** (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
The foundational paper on XGBoost, explaining the boosting algorithm and its applications to classification problems like liver disease prediction.
- **Breiman, L.** (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.
This paper introduces the Random Forest algorithm and demonstrates its application to various data analysis tasks, including medical diagnosis.
- **Chollet, F.** (2017). *Deep Learning with Python*. Manning Publications.
A comprehensive guide to deep learning and its practical applications, including how to use Keras for building machine learning models in Python.