International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Exploring Deep Learning Techniques for Precision Detection of Liver Cancer

Ms. J.LAHARI, Mr. SIVA KUMAR DEPURU

MCA student, Department of Computer Applications, Mohan Babu University, Tirupathi, Andhra Pradesh, India, LAHARI6303@gmail.com
Assistant Professor, Department of Computer Applications, Mohan Babu University, Tirupathi, Andhra Pradesh, India.

Abstract:
A rising worldwide health problem is liver disease, especially in areas with poor access to early diagnostic tools. Using the Indian Liver Patient Dataset (ILPD) as training data, this work introduces a machine learning-based method for the early prediction of liver disease using logistic regression. The model classifies individuals as probable or unlikely to develop liver disease based on important clinical characteristics such bilirubin levels, enzyme concentrations, protein ratios, and demographic information. Numerous preparation procedures were used, such as normalizing inputs, addressing missing values, encoding categorical variables, eliminating duplicates, and displaying data distributions. After achieving high predictive performance, the trained model was included into a web application built on Flask, which allowed users to enter medical information and obtain predictions in real time via an easy-to-use interface. The trained model achieved strong predictive performance and was integrated into a Flask-based web application, enabling users to input medical details and receive real-time predictions through an intuitive interface. The entire system—combining data analysis, predictive modeling, and web deployment—demonstrates how classical machine learning and modern web technologies can be combined to build accessible, efficient, and scalable healthcare tools. This prototype offers a practical solution for early liver disease screening, particularly in low-resource settings, and lays the groundwork for further advancements in digital diagnostics.

Keywords: Machine Learning, Liver Disease Prediction, Logistic Classifier, Web Application, Flask, Medical Data, Early Detection, Dataset Preprocessing, Model Deployment, Accuracy

, Precision, Healthcare Technology, Disease Classification, Data Preprocessing, Model Training, Binary Classification, Predictive Modeling.

I. Introduction:

Every year, millions of people worldwide are impacted by liver disease, a serious public health concern. One of the most important organs in the human body, the liver carries out a variety of key tasks, including protein synthesis, detoxification, and the manufacture of biochemicals required for digestion. The liver's capacity to carry out these tasks is weakened when it is ill or injured, which can result in serious health issues or even death. Unfortunately, because liver disease's early symptoms are

generally vague or mild, it frequently goes undiagnosed until it has progressed to an advanced state. Increased mortality and medical expenses are mostly caused by this diagnostic delay.

In recent years, the integration of machine learning into the medical domain has opened new doors for automated disease prediction and diagnostics. Machine learning algorithms, particularly classification models, are capable of identifying complex patterns within large medical datasets that might not be evident to the human eye. Among these, Logistic Regression is a widely used algorithm due to its simplicity, interpretability, and efficiency, making it an ideal candidate for clinical decision support systems. This project aims to build a predictive model using Logistic Regression to detect liver disease in patients based on a set of biomedical and demographic features. The model is trained on the Indian Liver Patient Dataset (ILPD), a publicly available dataset containing attributes such as Age, Gender, Bilirubin levels, Liver enzymes, Proteins, and the Albumin-Globulin Ratio—all of which are relevant indicators of liver health.

To ensure the accuracy and robustness of the model, several preprocessing steps are performed. These include handling missing values using median imputation, removing duplicate entries, encoding categorical variables like gender, and scaling the features using standard normalization techniques. Exploratory Data Analysis (EDA) is carried out using boxplots and correlation matrices to understand the distribution of each feature and the relationships among them. Once the data is preprocessed, it is split into training and testing subsets to validate the model's performance. The Logistic Regression model is then trained and evaluated using metrics such as accuracy and precision, which provide insight into the model's ability to correctly identify patients with liver disease.

Beyond just building a machine learning model, this project focuses on usability and real-world application. To make the system accessible and user-friendly, a web application is developed using the Flask framework. This web app serves as the front end of the prediction system, where users—such as doctors, medical staff, or even patients—can input test values through a simple form. Once the input is submitted, the system processes the data, applies the trained machine learning model, and provides an immediate prediction result indicating whether liver disease is likely or not. The model and the scaler are serialized using Python's pickle module, which allows them to be reused during prediction without retraining, thereby increasing the efficiency of the system.

This project not only demonstrates the practical utility of classical machine learning techniques in healthcare but also bridges the gap between predictive modeling and user accessibility by providing a functional and deployable application. It showcases how even with limited computational resources and simple models, effective health diagnostic tools can be built and deployed to assist medical professionals in their day-to-day decision-making. In the long term, such systems could play a significant role in improving public health outcomes by enabling faster, cheaper, and more widespread screening for liver disease. Future enhancements may include extending the model to compare multiple classifiers, adding patient history

for more contextual predictions, or integrating the tool into hospital information systems for automated diagnostics.

Related work:

Young persons (those between the ages of 16 and 30) showed a declining tendency, though, with an AAPC of -5.6 (95% CI: -8.1 to -2.9). Applying DL models revealed that, with an accuracy of 0.980 (AUC: 0.886), the convolution neural network (CNN) model performed the best in terms of predicting instances of liver cancer. The study's findings indicate that, between 2002 and 2010, Taiwan's yearly prevalence of hepatitis increased while the number of young people in Taiwan decreased. In a hepatitis cohort, the CNN model can be used to accurately predict liver cancer. [1]. The prediction model for hepatocellular carcinoma (HCC), a type of liver cancer, has been discussed in this study. Since HCC liver cancer is one of the most prevalent forms of the disease and claims the lives of millions of people worldwide, we have suggested a simple method for predicting it using publically accessible data. The dataset undergoes several pre-processing techniques to extract the best information, and our experiment uses a variety of classification models (such as KNN, Decision Tree, etc.) to view the prediction outcomes. We obtained good accuracy and computational time results. **[2].**Based on ten years of use, this study presents further improvement of a prediction score system that was developed in 1980 as a guide to establish a safe limit for hepatectomy. OVERVIEW OF BACKGROUND DATA: The net resection volume or the residual hepatic volume in conjunction with the patient's liver function was formerly used to empirically assess the safety of large resection. However, there were no objectively established standards upon which this conclusion was based. METHODS: This analysis included patients who underwent hepatectomy at some point between 1981 and 1990 and had hepatocellular carcinoma (HCC; n = 376) or metastatic malignancy (n = 58). A multiple classifier equation comprising the patient's age, the indocyanine green retention rate, and the estimated resection rate from the computed tomographic scan was used to calculate the prediction score (PS). A dangerous zone was defined as having a PS above 55, a borderline zone as having a PS between 45 and 55, and a safe zone as having a PS below 45. RESULTS: All patients in the risky zone died from HCC and chronic liver disease, while 33% of patients in the borderline zone and 7.3% of patients in the safe zone perished. All patients in the risky zone died from metastatic cancer with a normal liver, while none of the patients in the safe or borderline zones did. Liver failure mostly occurred following pulmonary infection or abdominal sepsis in the borderline and safe zones, especially for patients with poor prognostic characteristics such impaired glucose tolerance, low platelet counts, and elevated indocyanine green retention rates. CONCLUSION: For individuals with healthy or damaged livers, prediction scores can rule out mortality from excessive resection. Even if the score stays in the safe or borderline range, patients with poor prognostic markers must have cautious surgery and postoperative care to prevent liver failure brought on by intra- or extra-abdominal infection. [3].One of the most important components of automated disease prediction and diagnosis is data

mining. It uses data mining methods and algorithms to examine medical information. The prevalence of liver diseases has skyrocketed in recent years, and in many nations, they are now among the most deadly illnesses. The datasets of liver patients are examined in this thesis in order to develop classification algorithms that can forecast liver illness. In order to increase the prediction accuracy of Indian liver patients, this thesis used a feature model development and comparison analysis in three stages.The initial stage involves applying the min max normalization method to the initial liver patient datasets that were gathered from the UCI repository. In the second stage of liver dataset prediction, a subset (data) of the entire normalized liver patient dataset that only includes significant features is created by the application of PSO feature selection. In the third stage, the data set is subjected to classification methods. The accuracy will be determined in the fourth phase using the root mean square value and the root mean error value.After using PSO feature selection, the J48 method is thought to perform better. Ultimately, accuracy values are used to do the evaluation. With the aid of feature selection, the J48 method outperforms all other classification algorithms with an accuracy of 95.04%, according to the outputs from suggested classification implementations. [4].The body's largest organ, the liver, is vital for both food digestion and the elimination of harmful substances. Alcohol use and infections cause liver damage, which can result in a potentially fatal illness. In addition to cirrhosis, hepatitis, liver tumors, and liver cancer, there are numerous more forms of liver illnesses. Among these, cirrhosis and liver disorders are the leading causes of death.As a result, liver disease is among the world's most serious health issues. Approximately 2 million people worldwide pass away from liver disease each year . According to the Global Burden of Disease (GBD) research, published in BMC Medicine, one million peoples are perished in 2010 because of cirrhosis and million are suffering from liver cancer . In the biomedical field, machine learning has greatly impacted the diagnosis and prognosis of liver diseases . [5]. [6]. Coastal areas are seriously threatened by Liver cancers, hence effective disaster preparedness and response measures are required. Understanding and forecasting Liver cancer features and arrival times is the goal of this research endeavor. Comprehensive data gathering covering historical landslide, Liver cancer, and earthquake data from 1800 to 2023 is part of the first stage. Important details about location, magnitude, and date are obtained from reliable databases, scientific journals, historical archives, and government organizations. The gathered information serves as the basis for further analysis and forecasting. Accurate predictions of Liver cancer wave speed, travel time, and vulnerability are made by utilizing machine learning approaches such as the Logistic classifier Regressor model and K-means clustering.Prediction capabilities are further improved by integrating real-time data and taking climate change into account. Comprehensive visualizations, such as maps showing Liver cancer-prone areas and tables showing the months with the highest rates of damage, are the result of the research. These observations help coastal towns and disaster management authorities create focused plans for catastrophe preparedness and mitigation.

In addition to providing vital information for disaster risk reduction and coastal resilience decision-making, the study's international collaboration and sophisticated predictive models aid in Liver cancer protection for susceptible locations [6].

II. Proposed system:

The proposed system is designed to automate the early detection of liver disease using a machine learning-based predictive model that is both efficient and user-friendly. The core of the system is a Logistic Regression classifier trained on the Indian Liver Patient Dataset, which includes features such as Age, Gender, Total and Direct Bilirubin, Liver enzyme levels (Alkaline Phosphotase, Alamine Aminotransferase, Aspartate Aminotransferase), Total Proteins, Albumin, and the Albumin-Globulin Ratio. These features were selected for their clinical relevance in indicating liver function and potential abnormalities. To prepare the data for modeling, the system performs several preprocessing tasks, including the removal of duplicate entries, handling of missing values using median imputation, and normalization of numerical values using Standard Scaler to ensure consistent input scales. Categorical variables like Gender are converted to numerical form, and a detailed exploratory data analysis (EDA) is conducted using boxplots and correlation matrices to better understand the patterns in the dataset.

Once the model is trained and validated, it is deployed into a Flask-based web application, creating a complete end-to-end solution that is accessible through a web browser. The web app allows users to enter medical data through a simple form and instantly receive a prediction indicating whether the patient is likely to have liver disease. The model and scaler are stored using Python's `pickle` module to enable efficient loading during prediction without retraining. This modular approach ensures the system is scalable, maintainable, and easy to extend in the future. The user interface is clean, responsive, and intuitive, designed to make the diagnostic tool usable even by those with limited technical knowledge. This system thus represents a practical application of machine learning in healthcare, making early liver disease screening more accessible, especially in areas where advanced diagnostic facilities may be lacking. Future upgrades could involve incorporating other machine learning models, adding more patient metadata, or connecting the system to a hospital database for real-time clinical use.

The mathematical model of logistic classifier is based on the logistic function, also known as the sigmoid function. The logistic classifier model predicts the probability that a given input belongs to a particular class.

Let X be the input features, and β be the coefficients associated with each feature.

The logistic function, denoted by $\sigma(z)$, is defined as:

$$\sigma(z) = 1/1 + e^{-z}$$

where z is the linear combination of input features and their corresponding coefficients:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Here,

- β_0 is the intercept,

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

- $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_n x_n$ are the coefficients,

- $x_1 + x_2 + \ldots. + x_n$ are the input features.

The logistic function transforms the linear combination of input features into a value between 0 and 1, representing the probability of belonging to the positive class (typically denoted as class 1).

The logistic classifier model predicts the probability \(p \) that an input belongs to the positive class using the logistic function:

$$p(X) = 1/1 + e^{\beta 0 + \beta 1 x1 + \beta 2 x2 + \ldots. + \beta n x n}$$

The projected probabilities are subjected to a threshold θ in order to determine a binary categorization. Generally, the input is categorized as belonging to the positive class (class 1) if p(X) is larger than or equal to the threshold; if not, it is categorized as belonging to the negative class (class 0).

In order to reduce the error between the actual class labels of the training data and the projected probabilities, the logistic classifier model is trained by maximizing the coefficients β. Usually, methods like gradient descent or maximum likelihood estimation are used for this optimization.

III. FLOW CHART

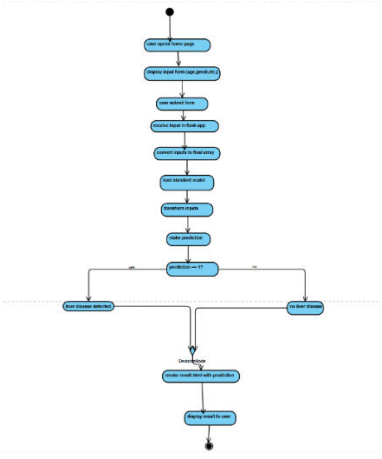

Figure-1: Flow chart of the system

IV. EXPERIMENTING RESULTS:

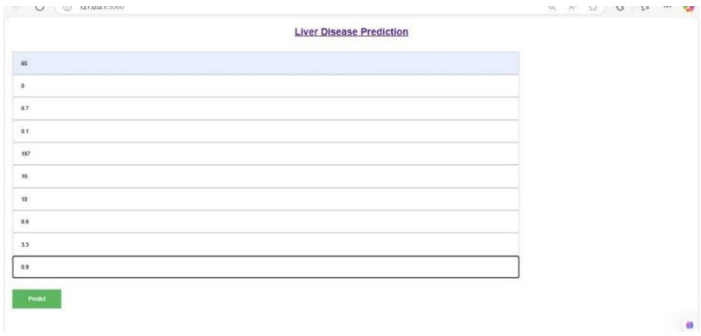

Figure-2: Webpage for the project

Figure-3: Result of the project

Algorithms	Accuracy	Precision
KNN	0.685	0.737
Support vector machine	0.720	0.723
Artificial neural network	0.713	0.781

Table-1 : Comparision table for Algorithms

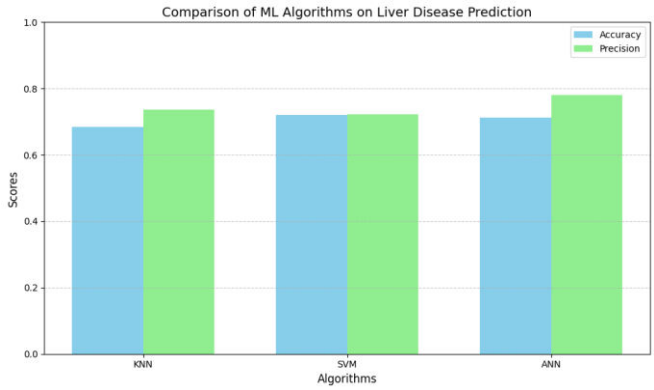

Figure-4: Comparison Graph

V. CONCLUSION

This project successfully demonstrates the application of multiple machine learning algorithms to predict the likelihood of liver

disease in patients. By leveraging Logistic classifer , the model was able to achieve accurate predictions, with precision values ensuring minimal false positives. The implementation of data preprocessing steps like handling missing values and scaling features contributed to the models' effectiveness. The comparison of these algorithms highlights the strengths of each, offering insights into how different techniques can be utilized based on performance metrics such as accuracy and precision. Ultimately, this project serves as a valuable tool for early detection, aiding healthcare professionals in diagnosing liver disease more efficiently.

VI. REFERENCES:

[1]. Nguyen, V.C., Phan, D.V., Chan, C.L., Li, A.H.A., and Chien, T.Y., 2020. A deep learning method for predicting liver cancer in a cohort with viral hepatitis. 147(10), International Journal of Cancer, pp. 2871–2878.
[2]. Choudhury, N.A., Moulik, S., and Rajesh, S., 2020, December. Machine learning techniques for the prediction of hepatocellular carcinoma (HCC) liver cancer. India Council International Conference (INDICON), IEEE 17th edition (pp. 1–5). IEEE.
[3] In 1994, Yamanaka, N., Okamoto, E., Furukawa, K., Fujimoto, J., Tanaka, T., tomoda, F., and Furukawa, K. A predictive grading method for choosing liver cancer surgery. more improvement after ten years of usage. p. 342 in Annals of Surgery, 219(4).

[4] Shiani, A., Pena, L., Friedman, M., and Narayanan, S. (2017). The prognosis of primary liver cancer depends on the identification and management of underlying liver disease. 24(3), p. 1073274817729240; Cancer Control.
[5]. Juliet, P.L., Tamilselvi, P.R., and Priya, M.B. (2018). evaluation of the effectiveness of machine learning algorithms for the prediction of liver disease. 206-211 in Int. Res. J. Eng. Technol, 5(1).
[6] Rahman, A.S., Hossain, S.A., Roy, J., Tasnim, Z., and Shamrat, F.J.M., 2019. A comparative analysis of supervised machine learning methods for liver disease prediction. 8(11), pp. 419–422, International Journal of Scientific & Technology Research.

[7]. Zhang, X. and Chen, X. (2025) research on the associated molecular mechanism and liver cancer prediction model based on chronic liver disease. Hepatology Annals, 30(2), p. 101572.
[8]. Tarazi, H., Hatamleh, W.A., Shukla, P.K., Zakariah, M., and Tiwari, B. (2022). A new AI-driven method for detecting and forecasting liver cancer that uses a cascaded fully convolutional neural network. p. 4277436, Journal of Healthcare Engineering, 2022(1).
[9] Liu, Y., Zheng, J., Hao, J., Wang, R.R., Liu, X., Gu, P., Yu, H., Yu, Y., Wu, C., Ou, B., and Peng, Z., 2022. According to the 2019 worldwide Burden of Disease Study, the worldwide burden of primary liver cancer by five etiologies is projected to reach 2035. pp. 1310–1323 in Cancer Medicine, 11(5).

[10]. Fujii, L.L., Murad, M.H., Wang, Z., Asrani, S.K., Ehman, R.L., Kamath, P.S., and Talwalkar, J.A. (2013), A comprehensive review and meta-

analysis found that liver stiffness is linked to the risk of decompensation, liver malignancy, and mortality in individuals with chronic liver disorders. Hepatology and Clinical Gastroenterology, 11(12), pp. 1573-1584.

[11] Cai, T., Zhu, E., Cong, H., Li, J., and Tao, Y., 2022. use biased epidemiological data to forecast liver cancers. Medical Artificial Intelligence, 124, p. 102234.

[12]. Sohn, S.Y., Yoon, C.N., and Kim, Y.S. (2003). analysis of screening test data for a growth curve-based liver disease prediction model. Pharmacotherapy & Biomedicine, 57(10), pp. 482-488.

[13]. Radha, P. and Divya, R. (2019). An APO algorithm-based multiple time series clinical liver cancer dataset for optimized HCC recurrence prediction. Medical Systems Journal, 43(7), pp. 1–12.

[14]. Khan, H., Williamson, M., Hasan, E., and Mostafa, F., 2021. Predicting liver illness using statistical machine learning techniques. 1(4), Livers, pp. 294-312

[15]. Laversanne, M., Clark, P.J., Petrick, J.L., McGlynn, K.A., Bray, F., and Valery, P.C. (2018). estimates of primary liver cancer in 30 nations globally by 2030. 600–611 in Hepatology, 67(2).