International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 26(1), March, 2026, pp 53-64
DOI: https://doi.org/10.3329/ijss.v26i1.88826

Prediction of Liver Disease using Deep Learning Methods

Md. Abdullah Al Rahat[1], Mohammad Abdul Halim[1] and Md. Tofazzal Hossain[1,*]

[1]Department of Statistics, Faculty of Science, Gopalganj Science and Technology University, Gopalganj-8105, Bangladesh

[*]Correspondence should be addressed to Md. Tofazzal Hossain (Email: tofazzal@gstu.edu.bd)

[Received August 23, 2025; Accepted January 20, 2025]

Abstract

The liver is one of the largest organs in the human body and it performs more than 500 functions in the human body. It also supports most of the organs, which are vital for our survival. Liver disease is a major health challenge nationally and globally. Generally, liver diseases are detected too late and raise the complexity of treatment. Early diagnosis is essential for preventing serious damage to the liver and for decreasing healthcare costs. However, the accuracy of the conventional diagnostic techniques for early prediction of liver disease is currently not at a satisfactory level. Due to the unavailability of enough specialist doctors and the high healthcare cost, people in third-world countries like Bangladesh remain away from regular health checkups, and therefore, like other serious diseases, liver disease is not detected early. Predictive models can serve a promising role in detecting liver disease, primarily before going to a doctor. In recent years, machine learning (ML)/deep learning (DL) models have become a promising technique for improving the diagnosis of human diseases, including liver disease. Therefore, we developed a deep learning-based method for early prediction of liver disease using demographic and clinical data downloaded from Kaggle.com. Several DL models including convolutional neural networks-long short-term memory (CNN-LSTM), long short-term memory (LSTM), gated recurrent units (GRU), categorical boosting (CatBoost), and deep neural network (DNN) were trained and among these models LSTM was selected as the best model based on the evaluation metrics like accuracy, precision, recall (sensitivity), specificity, F1-Score, receiver operating characteristic (ROC) curve, and area under curve (AUC). The prediction accuracy of LSTM was 99% with excellent discrimination (ROC-AUC 0.9991) and feature interpretability. The highest contribution in liver disease was confirmed by the key predictors Alkaline Phosphotase, SGPT Alamine Aminotransferase, and SGOT Aspartate Aminotransferase.

Keywords: Liver disease, Deep learning, Prediction of liver disease, Key predictors of liver disease.

AMS Classification: 62P10, 68T09.

1. Introduction

Liver disease is a condition affecting liver function, including hepatitis, fatty liver, cirrhosis, and liver cancer. It is a major health challenge worldwide, with about two million deaths annually and

4% of all deaths (1 out of every 25 deaths worldwide) (Bray et al., 2018). In Asia, the prevalence of non-alcoholic fatty liver disease (NAFLD) has risen to 33.9% in 2012-17 (Wong et al., 2021). With no exception in Bangladesh, hepatitis, fatty liver disease, cirrhosis, and liver cancer have all grown more widespread in the country. The general people of Bangladesh are highly unaware of liver disease, whereas about 45 million (33.86%) people are affected by NAFLD (Alam et al., 2023). One major challenge to deal with liver disease is that it is often asymptomatic in the early stages. Generally, this disease is often diagnosed at advanced stages, making its treatment more complex. Therefore, it is crucial to detect liver disease at an early stage to increase the survival rate of the patients.

Poor awareness, high cost of diagnosis, and poor access to regular health checkups also complicate early diagnosis of liver disease in Bangladesh and other low-resource nations. Moreover, risk factors, including alcohol consumption, age, sex, and metabolic disorders, are well-established and are likely to lead to the progression and development of liver disease (Becker et al., 1996). Clinical findings, such as bilirubin analysis, liver enzymes, albumin, total proteins, and the A/G ratio, are typically used to make a diagnosis. However, early-stage abnormalities may be mild and can easily go undetected during regular screening, thereby missing the opportunity to be diagnosed on time. All these difficulties highlight the necessity to develop better strategies that could help to diagnose high-risk people in the initial stages of the disease before it enters an acute phase.

Besides the presence of a larger population with a higher incidence of liver disease, there exist several healthcare hurdles to early detection in Bangladesh and other low-resource countries. Lack of top-notch doctors, high diagnostic expenses, and poor routine physical examination are factors that leave several patients without any diagnosis in severe cases. Moreover, the prevalence of liver disease has increased due to common risk factors of unhealthy diet, obesity, metabolic syndrome, alcohol abuse, viral hepatitis, and poor awareness among the population. Bilirubin, transaminases, albumin, and A/G ratio are commonly used as clinical indicators that can be used to diagnose the disease, although at the early stages, they are mild and can be easily missed during the screening process. The above gaps indicate the necessity of creating more reliable means that could be used in supporting early clinical decision-making and allow the identification of high-risk people before the disease develops.

People in third-world nations like Bangladesh avoid routine health examinations due to a lack of specialist physicians and the high expense of healthcare; as a result, liver disease is not detected early, like other major illnesses. Machine learning (ML) or deep learning (DL) techniques can offer an effective approach to identify early-stage liver disease by reducing the dependency on costly medical devices, and can be a reliable support to local people (Ganie et al., 2024). Several previous studies used ML methods to identify risk factors of liver disease and to predict the disease, and ML methods achieved significant success in liver disease diagnosis and prediction (Acharya et al., 2012; Nahar and Ara, 2018; Naik and Samant, 2016). Techniques like Random Forest, Support Vector Machines (SVM), and k-nearest neighborhood (KNN) are particularly effective, often reaching high accuracy in predicting liver disease using clinical data, and can achieve up to (90 – 95)% accuracy in distinguishing liver disease cases based on indicators like liver enzymes, age, and BMI (Arbain and Balakrishnan, 2019; Hashem et al., 2020). Despite recent advancements, there is a significant gap in understanding the interplay of various risk factors and their contribution to liver disease progression.

Several studies use traditional ML models like Adaboost, LogitBoost, Bagging, Random Forest, and KNN to predict and investigate the risk factors of liver disease. The study (Nahar et al., 2019)

scored a height accuracy of 71.5% using the LogitBoost model. Another study explored the effectiveness of fibrosis markers in diagnosing chronic alcoholic liver disease. Using primary data, one study (Wong et al., 2021) investigated alcoholic hepatitis. However, this study claimed that there may be bias due to the study being done in a secondary care center. Similar to this study, several researchers investigated the link between current alcohol intake and the future risk of alcohol-related liver disease in both men and women. The participation was from a hospital that may introduce bias. The Poisson regression model (Naveau et al., 2005) was used to analyze the risk factors. This study might use ML/DL models to evaluate risk factors more precisely. Another study tried to use models like RF, SVM, KNN, and Multilayer Perceptron (MLP) (Becker et al., 1996) using the Indian Liver Patient Dataset. However, the results were promising so far, but were not quite satisfactory. Advanced ML/DL models can be used to investigate the risk factors of liver disease precisely.

Another study (Vijayarani and Dhayanand, 2015) used NB and SVM to predict liver disease and evaluated the model performance using accuracy and execution time. The highest accuracy obtained from the SVM was 79.66%. However, it required more execution cost. The study lacked an in-depth analysis of feature selection, data variability, and preprocessing effects. Moreover, it did not show the lower execution cost with higher accuracy scores. DL models can be used to get higher accuracy with a minimum execution cost.

Using the ILPD data, another study (Nahar and Ara, 2018) used J48, LMT, RF, and other traditional ML models and got a height accuracy of 70.67%. This study had limitations, including analysis of feature selection, minor class performance, and long-term disease progression. The study (Kotronen et al., 2009) aimed to predict NAFLD using multivariate regression models and to uncover the risk factors related to NAFLD. New techniques, like DL models, can be used to assess risk factors and predict liver disease more robustly.

Using several traditional models, another study (Dritsas and Trigka, 2023) used techniques like SMOTE and feature ranking to predict liver disease. The voting classifier achieved the highest accuracy of 80.1% and an AUC of 88.4%. The study (Mohaimenul Islam et al., 2018) used Taipei Medical University Hospital patients' data to identify prognostic factors for fatty liver disease (FLD). This study used logistic regression to achieve 76.3% accuracy, and significant clinical differences were found between FLD and non-FLD patients. The LASSO feature selection method was used to identify important attributes for liver disease prediction using various machine learning models, including LR, DT, RF, SVM, KNN, AdaBoost, LDA, and Gradient Boosting (Afrin et al., 2021). This study found DT as the best model with an accuracy of 94.30%.

Artificial intelligence, particularly DL methods, is used in hepatology for the diagnosis and management of various liver diseases. The study (Nam et al., 2022) highlighted the potential application of AI to improve diagnostic accuracy and predict disease progression. It achieved promising results, such as an AUC of 0.98 for classifying hepatocyte ballooning and C-indices up to 0.73 for predicting NASH progression. However, this study did not use a traditional clinical test dataset for early diagnosis of liver disease. Early diagnosis using a clinical dataset helps to identify the severity of liver disease as well as to predict its risk factors. Therefore, more studies employing robust ML/DL methods are required to develop reliable predictive models for early detection of liver disease and to implement them in clinical settings.

Here, we developed an improved DL model for the early detection of liver disease using demographic as well as clinical data downloaded from Kaggle.com. Several DL models, including

convolutional neural networks-long short-term memory (CNN-LSTM), long short-term memory (LSTM), gated recurrent units (GRU), categorical boosting (CatBoost), and deep neural network (DNN), were trained, and among these models, LSTM was selected as the best model with an accuracy of 99%. The key predictors of liver disease were Alkaline Phosphatase, SGPT Alamine Aminotransferase, and SGOT Aspartate Aminotransferase.

2. Methodology

2.1 Data Description

The Indian Liver Disease Patient Dataset was collected for this study, which is available publicly at Kaggle (*Liver Disease Patient Dataset 30K Train Data*, n.d.). The dataset included patients' age, gender, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, SGPT, SGOT, Total Proteins, ALB Albumin, and A/G ratio–a total of 10 variables. Besides, an outcome variable 'Result', which contained the labelled data of liver disease patients and non-liver disease patients. The number of duplicate records detected and eliminated was 14,302. Once these duplicates had been removed, there was a total of 16,389 records that contained unique information out of 30,691.

2.2 Data Processing

Before starting data analysis, several preprocessing steps were carried out, including handling missing values and duplicates. Primarily, the missing values were checked and removed if detected. After that, duplicate cases were also removed before analysis. Then, the dataset was standardized to ensure all features were on the same scale to improve model performance and to reduce the possibility of overfitting.

2.3 Feature Selection

As we developed the DL model, the feature selection was automatic within the DL model. Therefore, the DL models inherently performed the feature selection process.

2.4 Model Training and Validation

The dataset was randomly divided into training and testing subsets to evaluate model performance. A stratified split was applied to maintain class balance, allocating 80% of the data for training and 20% for testing. This approach ensured that the model was trained on the majority of the data, while the testing set was reserved for performance evaluation.

2.5 Deep Learning Models

CNN-LSTM: CNN-LSTM (Yildiz et al., 2023), a hybrid model, is the integration of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to detect both the spatial and time dependencies. CNN derives hierarchical, and LSTM derives sequential patterns, which allows them to perform well overall in a single model.

LSTM: Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) network, which is a type of Recurrent Neural Networks (RNNs), is created to resolve vanishing and exploding gradient issues with conventional RNNs. LSTM retains or discards information using gated cell structures, which are basically selective in capturing long-term dependencies in sequential data.

GRU: The Gated Recurrent Unit (GRU) (Lecun et al., 2015) is a variant of the RNN that is designed to mitigate the vanishing gradient issue. Besides, it enhances the modelling efficiency of long-term relationships in sequential data.

CatBoost: CatBoost (Categorical Boosting) (Dorogush et al., n.d.) is a modern gradient boosting algorithm, which can directly operate on categorical variables without processing them on a large scale. CatBoost is based on oblivious (symmetric) decision trees, and it is more effective at generalizing, minimizing overfitting, and training in parallel.

DNN: A Deep Neural Network (DNN)(Yi et al., 2017) is an artificial neural network that contains several hidden layers in between the input and output layers. With these layers, the model can learn hierarchical data representations, with each neuron in the model doing a linear transformation, then a non-linear activation to recognize more complex relationships than the traditional approaches.

2.6 Evaluation Metrics

Among several DL models, the best one was selected by the following evaluation metrics.

$$Precision = \frac{TP}{TP + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F_1 - measure = \frac{2 * precision * sensitivity}{precision + sensitivity}$$

Here, TP, TN, FP, and FN were the number of true positives, true negatives, false positives, and false negatives, respectively.

ROC curve and AUC were also used to test the model's performance. The ROC curve plotted the true positive rate (sensitivity) against the false positive rate (1 - specificity), and the AUC was the area of the ROC curve. Higher AUC values indicated a better-performing model.

3. Results

3.1 Descriptive Statistics

There was a total of 16,389 patient records in the dataset, with a mean age of 43.77 and a range of 4-90 years. The Total Bilirubin and Direct Bilirubin values were 3.36 $\pm$ 6.21 mg/dL and 1.53 $\pm$2.89 mg/dL, respectively, with a very high maximum (75.0 mg/dL and 19.7 mg/dL), representing extreme cholestatic or hepatocellular damage in some patients. The Alkaline Phosphatase value was 290.83 $\pm$ 240.95 IU/L (max: 2110 IU/L), whereas transaminases (Sgpt and Sgot) demonstrated very high variability (means: 80.15 and 111.37 IU/L; maxima: 2000 and 4929 IU/L, respectively). The average Total Proteins and Albumin were 6.49 g/dL and 3.14 g/dL, respectively, whereas the average A/G Ratio was 0.95 g/dL.

Table 1: Descriptive statistics of the study variables.

Attributes	Count	Mean	Std	Min	25%	50%	75%	Max
Age of the patient	16389	43.77052	16.52949	4	32	45	55	90
Total Bilirubin	16389	3.360431	6.208708	0.4	0.8	1	2.7	75
Direct Bilirubin	16389	1.530429	2.894558	0.1	0.2	0.3	1.3	19.7
Alkaline Phosphatase	16389	290.8268	240.946	63	175	209	298	2110
SGPT Alamine Aminotransferase	16389	80.14729	180.0102	10	23	35	62	2000
SGOT Aspartate Aminotransferase	16389	111.3676	280.666	10	25	42	88	4929
Total Proteins	16389	6.487705	1.090549	2.7	5.8	6.6	7.2	9.6
ALB Albumin	16389	3.136573	0.794006	0.9	2.6	3.1	3.8	5.5
A/G Ratio: Albumin and Globulin Ratio	16389	0.946612	0.323337	0.3	0.7	0.93	1.1	2.8

3.2 Correlation Analysis

The correlation heatmap (Figure 1) depicted that there was a positive correlation between Total Bilirubin and Direct Bilirubin (r = 0.90), as well as the liver enzymes SGPT and SGOT (r = 0.78). There was also a strong relationship between Total Proteins and ALB Albumin (r = 0.78). The target variable (Result), Direct Bilirubin (r = -0.25), and Total Bilirubin (r = -0.23), had the highest correlation, and the remaining measures (ALB Albumin r = 0.16; A/G Ratio r = 0.16) were observed to have moderate, positive correlations. Negligible correlations were observed between age and gender, and disease status.

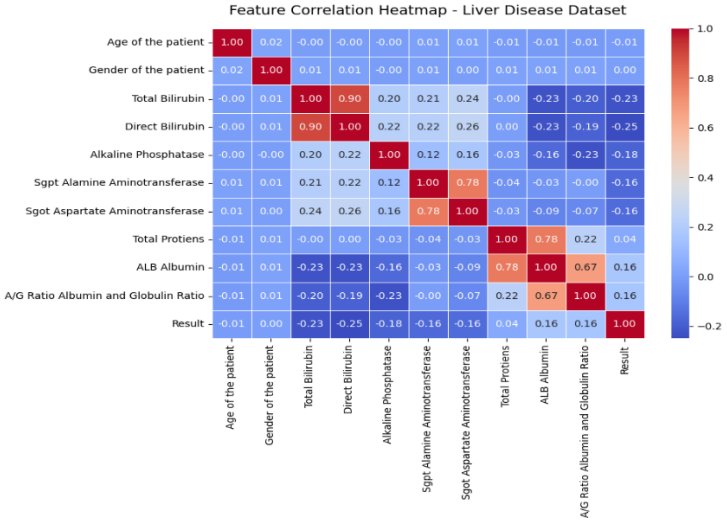

Figure 1: Correlation Analysis among Attributes.

3.3 Model Performance Comparison

The performance of the different models was summarized in Table 2, Figures 2, 3, and 4, both within the non-disease (Class 0) and the disease (Class 1) classes. The CNN-LSTM model performed well and was well-balanced in both classes, with class 1 and 0 having a precision of 0.99 and 0.96, respectively. Along with a recall of 0.98 for both classes (0 and 1), it achieved an overall accuracy of 0.98. It scored 0.9960 in ROC-AUC, which was decent, however, not as much as LSTM and DNN.

Table 2: Performance comparison of different DL models.

Model	**Precision**		**Recall**		**F1-score**		**Accuracy**	**ROC-AUC**
	Class 1	Class 0	Class 1	Class 0	Class 1	Class 0		
CNN–LSTM	0.99	0.96	0.98	0.98	0.98	0.97	0.98	0.9960
LSTM	**0.99**	**0.99**	**1.00**	**0.99**	**1.00**	**0.99**	**0.99**	**0.9991**
GRU	0.94	0.88	0.95	0.85	0.95	0.86	0.92	0.9781
CatBoost	0.93	0.99	1.00	0.81	1.00	0.89	0.94	0.9890
DNN	0.97	0.99	1.00	0.92	1.00	0.95	0.97	0.9972

The LSTM model achieved the best overall accuracy (0.99) and the highest recall for Class 1 (1.00), as well as a higher recall for Class 0 (0.99), compared to the CNN-LSTM. The LSTM model achieved an ROC-AUC score of 0.9991, which was the highest among all models. The GRU model performed comparatively less well, especially in terms of precision for Class 0 (0.88), indicating more mistakes in predicting non-disease classes. However, the recall at Class 1 was good (1.00). The overall accuracy of GRU was 0.92.

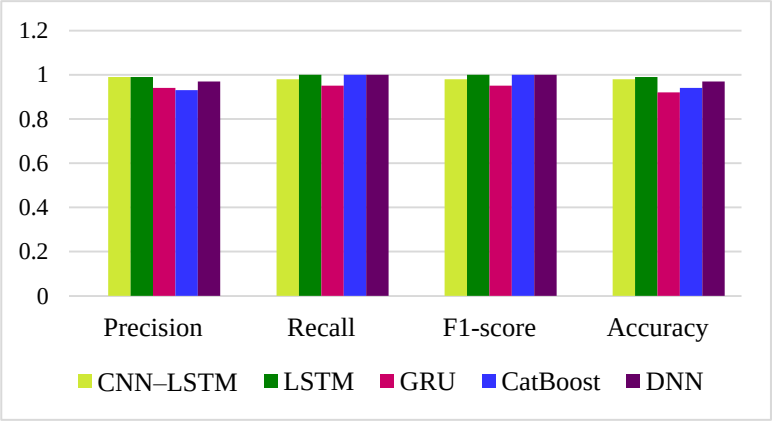

Figure 2: Comparison of different DL models (precision, recall, F1-score for class 1, and the overall accuracy).

The CatBoost algorithm achieved a high precision of Class 0 (0.99), whereas the recall of Class 0 was a little bit lower (0.81). It scored 0.9890 for ROC-AUC, indicating an incredibly good performance in discriminating between cases of the disease and non-disease.

The DNN model produced slightly lower scores than LSTM, especially in the class of disease (precision of 0.97). The overall accuracy of DNN was 0.97 with an ROC-AUC score of 0.9972.

Comparing the evaluation metrics, we concluded LSTM as a better performer among different DL models.

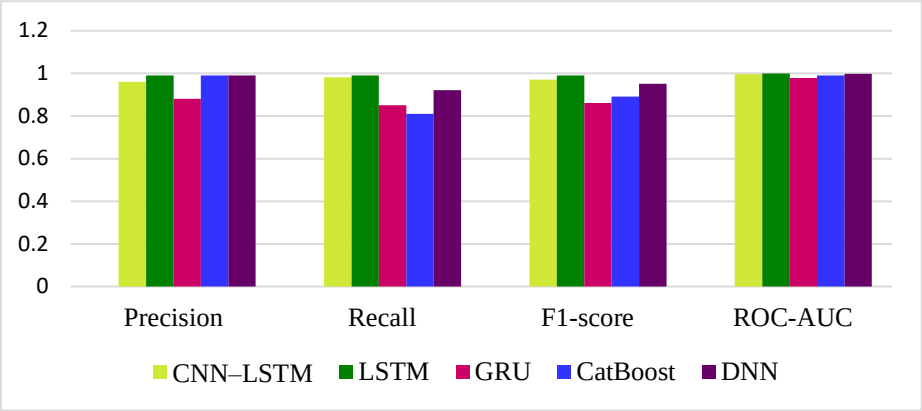

Figure 3: Comparison of different DL models (precision, recall, F1-score for class 0, and the ROC-AUC).

Figure 4 shows the ROC curves for different deep learning models: CNN-LSTM, LSTM, GRU, DNN, and CatBoost. All the models performed very well as their curves were very close to 1. Here, the LSTM model achieved the highest AUC (0.999), followed closely by DNN (0.997) and CNN-LSTM (0.996). GRU (0.978) and CatBoost (0.989) were also good, but gained slightly lower values. Overall, all models were highly precise in predicting liver disease. However, LSTM showed the highest performance among these models.

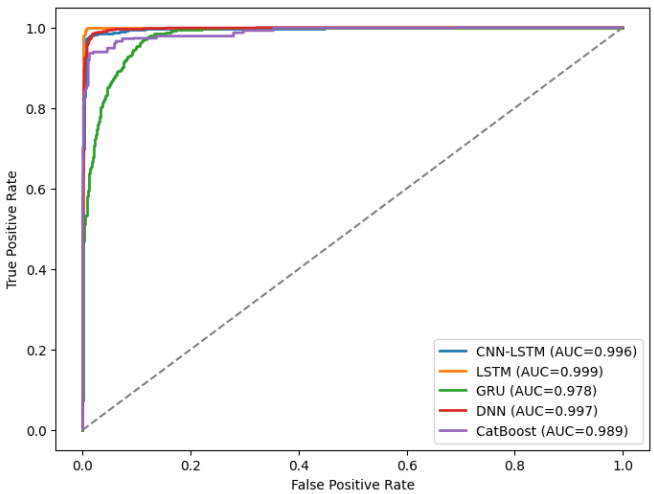

Figure 4: ROC Curve for different DL models.

3.4 Importance of Features

The feature importance produced by the LSTM (Figure 5) indicated the significance of each predictor in classifying liver disease. The LSTM model maintained a baseline accuracy (indicated by the dashed orange line) of approximately 0.85 before permutating the feature importance. Features such as SGOT Aspartate Aminotransferase, SGPT (Alamine Aminotransferase), and Alkaline Phosphatase led to a more noticeable drop in accuracy when permuted. This highlighted their stronger influence on predictions. Therefore, these three features were the most contributing factors to liver disease. Besides, the A/G ratio, ALB (Albumin), Total Proteins, Direct Bilirubin, and Total Bilirubin also led to a moderate drop in accuracy when permutated. This suggested that these features were also contributors to liver disease. Demographic factors, such as gender, did not contribute significantly. However, the Age of the patients showed overlap in baseline, which implied dropping the feature age, increasing accuracy. Overall, these results suggested that the model relies more heavily on clinical and biochemical indicators than on demographic characteristics for accurate classification.

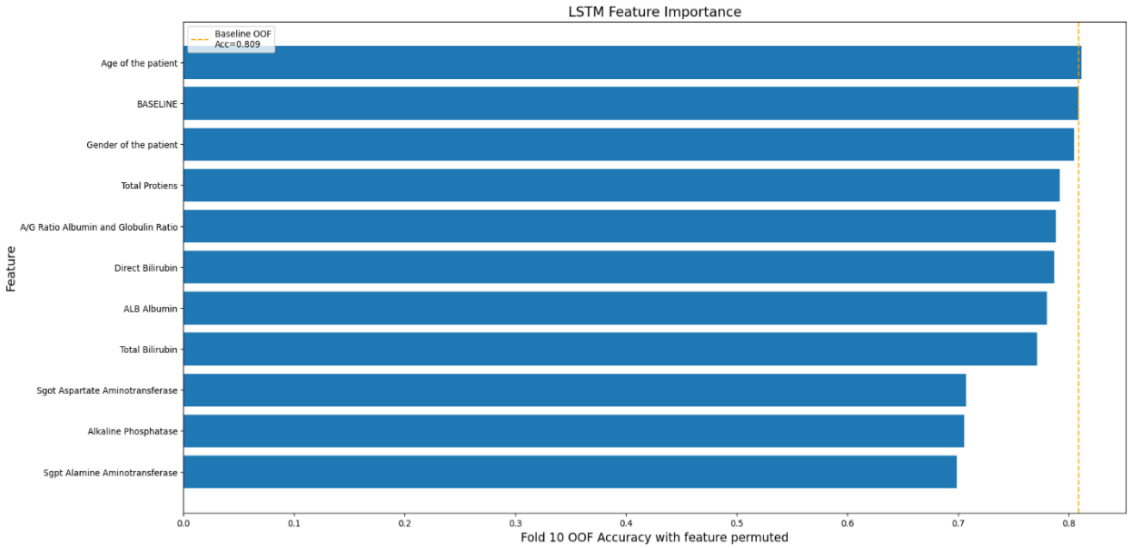

Figure 5: Important factors (features) for predicting liver disease.

4. Discussion

This study showed that DL models can predict liver disease with much higher accuracy than most traditional ML methods used in earlier research. Several studies used traditional methods like RF, KNN, SVM, or Logistic Regression, with an accuracy between 70-80% (Cho et al., 2014; Dorogush et al., n.d.; Nahar and Ara, 2018). This study achieved 99% accuracy with LSTM, which consists of a balanced recall for both disease and non-disease cases. The CNN–LSTM model also performed strongly (98% accuracy), outperforming the 80.1% from voting classifiers (Acharya et al., 2012). These results suggested that deep learning can better capture complex clinical patterns. With a ROC–AUC of 0.9991, LSTM was able to match the top-performing conventional models, such as Decision Trees, with LASSO feature selection, without sacrificing class balance (Becker et al., 1996).

The key predictors, according to feature importance analysis, were Alkaline Phosphatase, SGPT Alamine Aminotransferase, and SGOT Aspartate Aminotransferase. In addition to being in line with clinical knowledge, this offered a clear ranking of predictors, something that was frequently absent from previous studies (Afrin et al., 2021; Dritsas and Trigka, 2023; Straw and Wu, 2022).

These findings implied that, from a clinical standpoint, liver disease screening programs might give priority to a limited number of highly predictive biochemical tests, which would save costs without sacrificing accuracy. However, this analysis was restricted to a single dataset and might not account for variability at the population level. Future research should validate these models on multiple datasets, explore their integration into hospital decision-support systems, and assess the impact on early intervention outcomes.

Despite the dataset's imbalance, the high accuracy (~99%) is supported by additional metrics such as Precision, Recall, F1-score, and ROC-AUC, confirming that the performance is not merely due to class dominance. All models achieved balanced Precision–Recall results, with the LSTM attaining the highest ROC-AUC (0.9991), followed closely by DNN and CNN–LSTM, indicating strong discriminative ability. Cross-validation, 80% train and 20% test data, was used to minimize overfitting and data leakage. The consistently high F1-scores (0.95–1.00) across models further demonstrate that the proposed DL models, particularly LSTM and CNN–LSTM, generalize well for early liver disease prediction.

5. Conclusion

Liver disease represents a major global health concern, with its prevalence steadily increasing worldwide. Early prediction of liver disease is crucial, as it can decrease the serious damage to the liver as well as increase the survival rate. In this study, several deep learning models were used to assess the risk factors and to predict liver disease. Among different models, LSTM, CNN–LSTM, and DNN demonstrated better performance in liver disease classification. Among all, LSTM achieved the highest accuracy of 99% with excellent classification performance (ROC-AUC 0.9991) and feature interpretability. Key predictors included Alkaline Phosphatase, SGPT Alamine Aminotransferase, and SGOT Aspartate Aminotransferase, confirming the highest contribution in liver disease.

Funding: The authors receive fund from Grant ID: BSMRSTU-RC-25-032 for this research work.

Acknowledgement: The authors express their gratitude to the reviewer for his kind comments and helpful recommendations.

References

[1] Acharya, U. R., Sree, S. V., Ribeiro, R., Krishnamurthi, G., Marinho, R. T., Sanches, J., and Suri, J. S. (2012). Data mining framework for fatty liver disease classification in ultrasound: A hybrid feature extraction paradigm. Medical Physics, 39(7), 4255–4264. https://doi.org/10.1118/1.4725759,

[2] Afrin, S., Javed Mehedi Shamrat, F. M., Nibir, T. I., Muntasim, M. F., Moharram, M. S., Imran, M. M., and Abdulla, M. (2021). Supervised machine learning based liver disease prediction approach with LASSO feature selection. Bulletin of Electrical Engineering and Informatics, 10(6), 3369–3376. https://doi.org/10.11591/EEI.V10I6.3242

[3] Alam, S., Islam Alin, M. S., Begum, F., Fahim, S. M., Tasnim, Z., and Alam, M. M. (2023).

Estimating the cost of illness of non-alcoholic fatty liver disease in Bangladesh. JGH Open, 7(9), 629–635. https://doi.org/10.1002/JGH3.12960,

[4] Arbain, A. N., and Balakrishnan, B. Y. P. (2019). A Comparison of Data Mining Algorithms for Liver Disease Prediction on Imbalanced Data. International Journal of Data Science and Advanced Analytics, 1(1), 1–11. https://doi.org/10.69511/IJDSAA.V1I1.2

[5] Becker, U., Deis, A., Sørensen, T. I. A., Grønbæk, M., Borch-Johnsen, K., Müller, C. F., Schnohr, P., and Jensen, G. (1996). Prediction of risk of liver disease by alcohol intake, sex, and age: A prospective population study. Hepatology, 23(5), 1025–1029. https://doi.org/10.1053/jhep.1996.v23.pm0008621128

[6] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394–424. https://doi.org/10.3322/caac.21492

[7] Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. 103–111.

[8] Dorogush, A. V., Ershov, V., and Yandex, A. G. (n.d.). CatBoost: gradient boosting with categorical features support. Retrieved August 21, 2025, from https://github.com/Microsoft/LightGBM

[9] Dritsas, E., and Trigka, M. (2023). Supervised Machine Learning Models for Liver Disease Risk Prediction. Computers 2023, Vol. 12, Page 19, 12(1), 19. https://doi.org/10.3390/COMPUTERS12010019

[10] Ganie, S. M., Dutta Pramanik, P. K., and Zhao, Z. (2024). Improved liver disease prediction from clinical data through an evaluation of ensemble learning approaches. BMC Medical Informatics and Decision Making, 24(1), 1–24. https://doi.org/10.1186/S12911-024-02550-Y/TABLES/6

[11] Hashem, S., ElHefnawi, M., Habashy, S., El-Adawy, M., Esmat, G., Elakel, W., Abdelazziz, A. O., Nabeel, M. M., Abdelmaksoud, A. H., Elbaz, T. M., and Shousha, H. I. (2020). Machine Learning Prediction Models for Diagnosing Hepatocellular Carcinoma with HCV-related Chronic Liver Disease. Computer Methods and Programs in Biomedicine, 196, 105551. https://doi.org/10.1016/J.CMPB.2020.105551

[12] Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735,

[13] Kotronen, A., Peltonen, M., Hakkarainen, A., Sevastianova, K., Bergholm, R., Johansson, L. M., Lundbom, N., Rissanen, A., Ridderstråle, M., Groop, L., Orho-Melander, M., and Yki-Järvinen, H. (2009). Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. Gastroenterology, 137(3), 865–872. https://doi.org/10.1053/J.GASTRO.2009.06.005

[14] Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/NATURE14539;SUBJMETA=117,639,705;KWRD=COMPUTER+SCIENCE,MATHEMATICS+AND+COMPUTING

[15] Liver Disease Patient Dataset 30K train data. (n.d.). Retrieved August 21, 2025, from https://www.kaggle.com/datasets/abhi8923shriv/liver-disease-patient-dataset/

[16] Mohaimenul Islam, M., Wu, C. C., Poly, T. N., Yang, H. C., and Li, Y. C. (2018). Applications of Machine Learning in Fatty Live Disease Prediction. Studies in Health Technology and Informatics, 247, 166–170. https://doi.org/10.3233/978-1-61499-852-5-166

[17] Nahar, N., and Ara, F. (2018). Liver Disease Prediction by using Different Decision Tree

Techniques. International Journal of Data Mining and Knowledge Management Process (IJDKP), 8(2). https://doi.org/10.5121/ijdkp.2018.8201

[18] Nahar, N., Ara, F., Neloy, M. A. I., Barua, V., Hossain, M. S., and Andersson, K. (2019). A Comparative Analysis of the Ensemble Method for Liver Disease Prediction. ICIET 2019 - 2nd International Conference on Innovation in Engineering and Technology. https://doi.org/10.1109/ICIET48527.2019.9290507

[19] Naik, A., and Samant, L. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 85, 662–668. https://doi.org/10.1016/J.PROCS.2016.05.251

[20] Nam, D., Chapiro, J., Paradis, V., Seraphin, T. P., and Kather, J. N. (2022). Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction. JHEP Reports, 4(4). https://doi.org/10.1016/j.jhepr.2022.100443

[21] Naveau, S., Raynard, B., Ratziu, V., Abella, A., Imbert-Bismut, F., Messous, D., Beuzen, F., Capron, F., Thabut, D., Munteanu, M., Chaput, J. C., and Poynard, T. (2005). Biomarkers for the prediction of liver fibrosis in patients with chronic alcoholic liver disease. Clinical Gastroenterology and Hepatology, 3(2), 167–174. https://doi.org/10.1016/S1542-3565(04)00625-1

[22] Straw, I., and Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health and Care Informatics, 29(1). https://doi.org/10.1136/BMJHCI-2021-100457

[23] Vijayarani, D. S., and Dhayanand, M. S. (2015). Liver Disease Prediction using SVM and Naïve Bayes Algorithms.

[24] Wong, G. L. H., Yuen, P. C., Ma, A. J., Chan, A. W. H., Leung, H. H. W., and Wong, V. W. S. (2021). Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. Journal of Gastroenterology and Hepatology (Australia), 36(3), 543–550. https://doi.org/10.1111/JGH.15385,

[25] Yi, H., Shiyu, S., Duan, X., and Chen, Z. (2017). A study on Deep Neural Networks framework. Proceedings of 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2016, 1519–1522. https://doi.org/10.1109/IMCEC.2016.7867471

[26] Yildiz, E. N., Cengil, E., Yildirim, M., and Bingol, H. (2023). Diagnosis of Chronic Kidney Disease Based on CNN and LSTM. Acadlore Transactions on AI and Machine Learning, 2(2), 66–74. https://doi.org/10.56578/ataiml020202