

Hyper-Parameter Optimized Models for Liver Disorder Prediction

Vaishnavi Nampally^{1*}, Poojitha Nampally² and Bhattacharyulu N. Ch¹

¹Department of Statistics, Osmania University, Hyderabad, Telangana, India

²Department of Oral Medicine and Radiology, Government Dental College and Hospital, KNRUHS, India

Citation: Nampally V, Nampally P, Bhattacharyulu N. Ch. Hyper-Parameter Optimized Models for Liver Disorder Prediction. *J Artif Intell Mach Learn & Data Sci* 2026 9(1), 3308-3312. DOI: doi.org/10.51219/JAIMLD/vaishnavi-nampally/665

Received: 23 February, 2026; **Accepted:** 26 February, 2026; **Published:** 27 February, 2026

***Corresponding author:** Vaishnavi Nampally, Department of Statistics, Osmania University, Hyderabad, Telangana, India, E-mail: nampallyvaishnavi@gmail.com

Copyright: © 2026 Nampally V, et al., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Public health is crucial study for any of the state & country to make precautionary measures to caution the public. In many situations, patients may require the blood to transfuse in case of surgery, trauma, chronic illness, blood disorders, etc. in such cases it is necessary to examine and to ensure that the donor's blood, whether donor's liver is functioning properly, free from transmissible infections such as Hepatitis-C, Fibrosis, Cirrhosis and health status etc., This paper presents the prediction of liver issues in blood donors based on the sample data set collected from medical records of 615 patients on thirteen attributes using various machine and deep learning techniques. To enhance the predictive performance of these models, applied Random search and Bayesian optimization hyper-parameter optimization techniques and showed significantly improvement in accuracy for identifying individuals with various liver health diseases.

Keywords: Logistic, LSTM, SVM, Random Forest, Boosting, Random Search, Bayesian Optimization.

1. Introduction

Blood donation is a voluntary act of donating the blood, for the purpose of transfusion to individuals in need due to surgery, trauma, chronic illness, blood disorders etc. screening the donated blood is to ensure that the donor's liver is functioning properly, free from transmissible infections such as Hepatitis-C, including stages like 'Fibrosis' and 'Cirrhosis' and health status etc. Liver is the primary organ involved in filtering and detoxifying blood and any dysfunction can compromise blood quality and pose a risk to recipients.

Liver health dataset contains the information about 615 medical records with thirteen attributes under study is collected from 'UCI Machine learning repository'. The features are: (a)

Age, (b) Gender (Female, Male), (c) Blood donor type (i) no suspect; (ii) Suspect blood donor; (iii) Hepatitis-C (inflammation of liver); (iv) Fibrosis (scar tissue starts forming); (v) Cirrhosis (advanced scarring, irreversible damage); (d) Albumin (ALB); (e) Alkaline Phosphatase (ALP); (f) Alanine Aminotransferase (ALT); (g) Aspartate Aminotransferase (AST), (h) Bilirubin (BIL), (i) Cholinesterase (CHE), (j) Cholesterol (CHE), (k) Creatinine (CREA), (l) Gamma-Glutamyl Transferase (GGT) and (m) Total Protein (PROT). Imputed the missing values with median in the variables.

The objective of this study is to identify an optimal machine learning model to classify the blood donors into the five categories based on the attributes under study of liver health

status using biochemical markers, enhancing public health screening and blood donation safety.

2. Exploratory Analysis

In this section, an attempt is made to calculate the descriptive statistics to gain a better understanding of the data and also if the data follows key assumptions such as normality, which is essential for further analysis. Additionally, both categorical and continuous variables were explored using visualizations like histograms and Q-Q plots to identify patterns and assess the distribution of the data in evaluating the data's characteristics and verifying statistical assumptions.

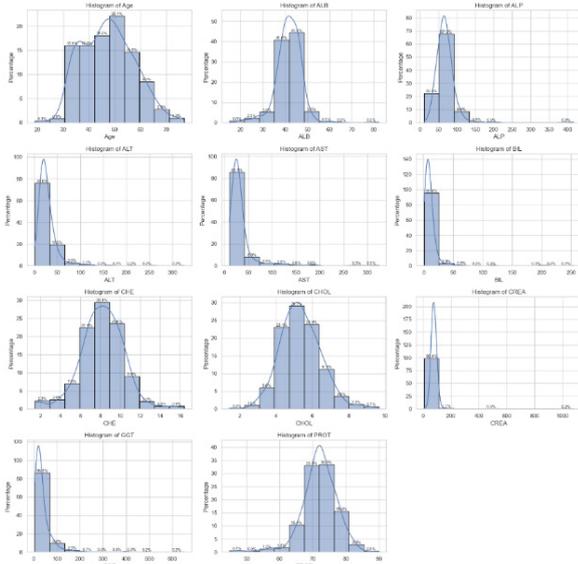


Figure 1: Histograms of (a),(c) to (m) variables respectively.

From (Figures 1 and 2): It is evident that ‘Age’, ‘ALB’, ‘CHE’, ‘CHOL’ and ‘PROT’ approximately normally distributed and rest of the variables are skewed to the right. For normal

distributed variables, ANOVA test is conducted and test statistics are FAge =11.04, FALB, = 45.81, FCHE =40. 68, FCHOL =16.55, FPROT = 29.85. Based on the ANOVA results for the variables, it is evident that all five variables show statistically significant differences across the five categories under study at a 5% significance level ($\alpha = 0.05$) as each variable is greater than the Critical F-value (= 2.3865) at a 5% significance level ($\alpha = 0.05$). This indicates that the means of these variables vary significantly from one category to another. Such findings confirm that these variables are not only statistically significant but also potentially meaningful in distinguishing between groups. Hence, they can be considered important predictors or indicators in further category-wise analysis, whether it be for medical diagnosis, disease staging or classification tasks in a machine learning framework (Table 1).

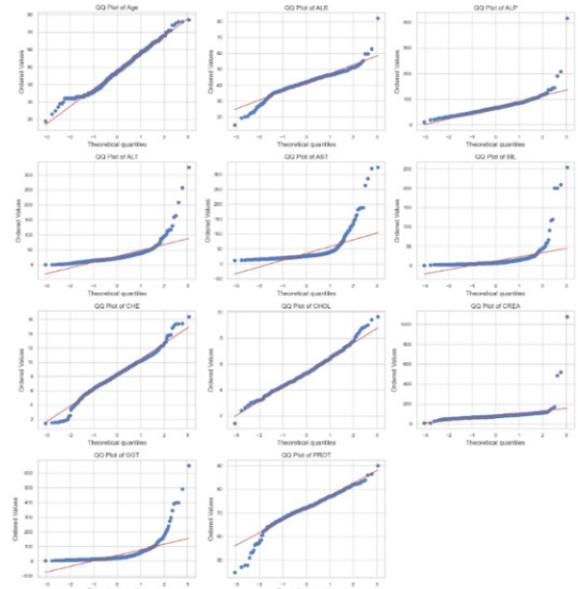


Figure 2: QQ Plots of (a), (c) to (m) variables respectively.

Table 1: Descriptive Statistic w.r.t each category of Blood Donor type.

	Overall		0		1		2		3		4	
	Mean ± SD	(Min, Max)	Mean ± SD	(Min, Max)	Mean ± SD	(Min, Max)	Mean ± SD	(Min, Max)	Mean ± SD	(Min, Max)	Mean ± SD	(Min, Max)
Sample size	615		533		7		24		21		30	
Age	47.41 ± 10.06	(19, 77)	47.61 ± 9.78	(19, 76)	57.57 ± 6.96	(48, 67)	38.71 ± 12.83	(21, 70)	46.76 ± 13.14	(29, 75)	47.8 ± 9.45	(38, 74)
ALB	41.62 ± 5.78	(14.9, 82.2)	42.13 ± 4.88	(27.4, 60.1)	24.4 ± 6.32	(25.9, 55.6)	43.83 ± 5.82	(26.6, 52.8)	41.13 ± 9.72	(29.3, 52.9)	38.47 ± 9.74	(27.9, 47.8)
ALP	68.22 ± 25.65	(11.3, 416.6)	65.48 ± 22.28	(11.3, 154.5)	107.3 ± 79.68	(22.3, 416.6)	45.43 ± 10.84	(30.2, 152.5)	49.67 ± 11.27	(30.6, 192.9)	63.61 ± 29.17	(30.2, 161.3)
ALT	28.44 ± 25.45	(0.9, 325.3)	26.12 ± 18.98	(2.2, 142.5)	102.1 ± 112.7	(4.7, 325.3)	25.59 ± 11.63	(3.7, 112.6)	38.9 ± 32.59	(5.3, 105.7)	42.13 ± 37.17	(3.1, 135.2)
AST	34.79 ± 33.09	(10.6, 324)	26.55 ± 11.17	(10.6, 324.0)	69.06 ± 74.41	(13.3, 214.4)	59.96 ± 54.34	(13.8, 141.5)	81.17 ± 72.96	(16.1, 158.1)	69.02 ± 47.41	(17.3, 136.2)
BIL	11.4 ± 19.67	(0.8, 254)	9.92 ± 18.46	(1.0, 254.0)	4.69 ± 2.82	(1.5, 59.2)	15.62 ± 14.55	(0.8, 53.9)	13.43 ± 10.17	(1.2, 34.5)	9.99 ± 5.51	(1.6, 19.3)
CHE	8.2 ± 2.21	(1.42, 16.41)	8.28 ± 2.05	(3.9, 16.41)	7.48 ± 3.77	(3.55, 12.61)	9.28 ± 1.98	(4.26, 11.21)	8.62 ± 1.44	(4.46, 11.27)	7.58 ± 2.13	(5.07, 9.43)
CHOL	5.37 ± 1.12	(1.43, 9.67)	5.49 ± 1.06	(2.86, 9.67)	4.45 ± 1.42	(2.48, 7.58)	5.2 ± 0.81	(41.0, 187.0)	5.05 ± 1.06	(41.0, 150.0)	4.92 ± 0.99	(50.0, 164.0)
CREA	81.29 ± 49.76	(8, 1079.1)	81.54 ± 23.15	(35.0, 204.0)	95.81 ± 53.84	(43.0, 295.0)	89.77 ± 19.41	(41.0, 187.0)	85.52 ± 21.77	(41.0, 150.0)	112.1 ± 66.13	(50.0, 164.0)
GGT	39.53 ± 54.66	(4.5, 650.9)	34.7 ± 26.6	(5.4, 204.6)	107.13 ± 109.64	(26.7, 377.9)	40.29 ± 27.53	(10.1, 99.1)	54.45 ± 37.89	(19.5, 150.1)	55.9 ± 62.71	(11.4, 274.7)
PROT	72.04 ± 5.4	(44.8, 90)	73.69 ± 9.25	(60.0, 85.5)	52.93 ± 11.15	(59.6, 79.6)	75.59 ± 11.1	(60.8, 79.3)	69.52 ± 12.27	(61.5, 79.2)	70.25 ± 12.74	(64.4, 77.3)

Further, Significant group comparisons across various variables based on Tukey's Honest Significant Difference (HSD) test was conducted. The test was applied to evaluate the mean differences between different groups and statistically significant results are Age: $t_{02} = -8.42$, $t_{12} = -18.86^*$, $t_{23} = 13.63$, $t_{24} = 14.76$; ALB: $t_{01} = -17.84$, $t_{04} = -9.44$, $t_{12} = 19.43^*$, $t_{13} = 17.36$, $t_{14} = 8.40$, $t_{24} = -11.04$, $t_{34} = -8.96$; CHE: $t_{04} = -4.58$, $t_{14} = -3.66$, $t_{24} = -5.47^*$, $t_{34} = -4.52$; CHOL: $t_{03} = -0.86$, $t_{04} = -1.4^*$, $t_{03} = -1.00$; PROT: $t_{03} = 3.99$, $t_{12} = 20.79$, $t_{13} = 22.19^*$, $t_{14} = 16.21$, $t_{24} = -4.58$, $t_{34} = -5.98$ where 0:No suspect; 1: Suspect; 2:Hepatitis-C; 3:Fibrosis; 4:Cirrhosis. These results underscore the importance of these variables in differentiating between the groups, making them crucial for further analysis and interpretation in the clinical context.

Kruskal-Wallis tests were conducted on various non-normally distributed variables and their test statistics (significant pairs by Dunn's test with p-values) are HALP = 51.65 ($p_{02} = 9.29 \times 10^{-7}$, $p_{12} = 0.00003$, $p_{24} = 0.000014$), HALT = 24.74 ($p_{04} = 0.00016$, $p_{14} = 0.0065$), HAST = 161.65 ($p_{04} = 1.47 \times 10^{-16}$), HBIL = 98.66 ($p_{02} = 0.000332$, $p_{03} = 0.00005$, $p_{04} = 3.12 \times 10^{-14}$), HCREA = 15.83 (no $p < 0.01$ and $p < 0.05$), HGGT = 112.38 ($p_{02} = 4.5 \times 10^{-6}$, $p_{03} = 9.13 \times 10^{-8}$, $p_{04} = 7.52 \times 10^{-13}$). The Kruskal-Wallis test statistics and their corresponding p-values of significant pairs revealed significant differences between groups in all variables except for CREA. The observed insignificance of CREA (Creatinine) in liver health analysis aligns with established clinical practices that it is primarily used for assessing kidney function and does not directly correlate with liver function or liver-related conditions. Consequently, it can be concluded that creatinine is not a significant biochemical marker for evaluating liver health. As a result, this variable was removed from the analysis to enhance the accuracy and comprehensiveness of liver function prediction, with a focus on liver-specific biomarkers.

3. Analysis Of Liver Health Condition Using Machine & Deep Learning Techniques

The accuracy, precision, recall and F1 score of the machine learning methods at 80-20 train-test splitting are presented in (Table 2). LSTM has shown higher performance in terms of Accuracy, Precision, Recall, F1-Score. The classification report and confusion matrix for LSTM is presented in (Tables 3 and 4).

Table 2: Comparison of classification methods.

S.No.	Classification Method	Accuracy	Precision	Recall	F1-Score
1	Logistic Regression	0.89	0.90	0.89	0.88
2	KNN	0.80	0.73	0.80	0.74
3	Support Vector Machine	0.86	0.85	0.86	0.84
4	Decision Tree	0.85	0.83	0.85	0.84
5	Random Forest	0.86	0.84	0.86	0.83
6	Bagging	0.89	0.85	0.89	0.86
7	Ada-Boost	0.87	0.86	0.67	0.75
8	XG-Boost	0.90	0.90	0.90	0.89
9	Multilayer Perceptron	0.89	0.89	0.89	0.87
10	LSTM*	0.91	0.91	0.91	0.91

Table 3: Confusion matrix for LSTM before optimization 112/123.

Actual \ Predicted	No Suspect	Suspect Blood Donor	Hepatitis C	Fibrosis	Cirrhosis
No Suspect	96	0	0	0	0
Suspect Blood Donor	1	2	0	0	0
Hepatitis C	0	0	6	2	1
Fibrosis	3	0	0	3	0
Cirrhosis	2	0	0	2	5

Table 4: Classification report for LSTM before tuning optimization.

	Precision	Recall	F1-Score
No Suspect	0.94	1.00	0.97
Suspect Blood Donor	1.00	0.67	0.80
Hepatitis C	1.00	0.67	0.80
Fibrosis	0.43	0.50	0.46
Cirrhosis	0.83	0.56	0.67

The LSTM model performs exceptionally well overall, achieving an accuracy, precision, recall and F1-Score of 0.91, which reflects balanced and reliable performance across multiple metrics. However, the "Fibrosis" class presents a challenge, with a low precision of 0.43 and a recall of 0.50, indicating the model's struggle to identify these instances accurately. Additionally, the "Suspect Blood Donor" and "Hepatitis C" classes show a recall of 0.67 and an F1-Score of 0.80, suggesting potential issues with class imbalance or inadequate representation of these categories in the training data. The 'Cirrhosis' class also struggles with a recall of 0.56, highlighting a substantial number of false negatives, which could be improved with better handling of rare cases. These results emphasize the need for further optimization.

4. Hyper-Parameter Optimization

Hyper-parameter tuning is a critical step in optimizing machine learning models as it involves adjusting key parameters that control the learning process and performance of the model. In this study, two popular hyper-parameter tuning techniques are applied: Random Search and Bayesian Optimization to enhance the LSTM model's performance. These methods were chosen for their ability to explore the hyper-parameter space efficiently and help identify the optimal settings that maximize accuracy and minimize classification errors (Tables 5-10).

After hyperparameter tuning, the LSTM model showed substantial performance improvements. Random Search increased accuracy to 0.9467, Bayesian Optimization further enhanced accuracy to 0.9756, improving precision, recall and F1-Score for almost all the classes. While both methods improved handling of imbalanced classes, Bayesian Optimization took significantly longer, requiring 325 minutes compared to Random Search's 138 minutes.

5. Remarks

From Figure 1 and Figure 2, the variables 'Age', 'ALB', 'CHE', 'CHOL' and 'PROT' appear to follow an approximately normal distribution. Therefore, ANOVA was applied to determine whether there were significant differences in means across the different categories of the response variable. The results confirmed the presence of statistically significant differences, leading to the application of Tukey's HSD post hoc test. This analysis revealed that most group pairs showed significant differences for each of these variables.

Table 5: Default and Hyperparameter space for LSTM for optimization.

Method	Default Hyper-parameters	Hyper-parameter Space
Long Short-Term Memory (LSTM)	'units' = 64, 'dropout' = 0.3, 'learning_rate' = 0.0001, 'batch_size' = 32 'epochs' = 20, 'activation' = 'tanh', 'recurrent_activation' = 'sigmoid',	'units': randint(10, 100), 'dropout': uniform(0.1, 0.9), 'learning_rate': loguniform(10 ⁻⁴ , 10 ⁻²), 'batch_size': randint(10, 100) 'epochs': randint(20, 51), 'activation' = categorical(['tanh', 'relu', 'sigmoid', 'softmax']) 'recurrent_dropout' = uniform(0.0, 0.9), 'optimizer': categorical(['adam', 'rmsprop', 'nadam']), 'recurrent_activation' = categorical(['tanh', 'relu', 'sigmoid', 'softmax'])

Table 6: List of best hyper-parameters for the study, accuracy and time complexity.

Hyper-parameter tuning methods	Best hyper-parameters	Accuracy	Time
Random Search	'units' = 64, 'dropout' = 0.3, 'learning_rate' = 0.0001, 'batch_size' = 16 'epochs' = 30, 'activation' = 'tanh', 'recurrent_activation' = 'sigmoid',	0.9467	138m 18s
Bayesian Optimization	'units' = 32, 'dropout' = 0.457, 'learning_rate' = 0.000179, 'batch_size' = 32 'epochs' = 31, 'activation' = 'tanh', 'recurrent_activation' = 'sigmoid',	0.9756	325m 33 s

Table 7: Confusion matrix for LSTM after optimization using Random Search 116/123.

Actual \ Predicted	No Suspect	Suspect Blood Donor	Hepatitis C	Fibrosis	Cirrhosis
No Suspect	102	1	1	2	0
Suspect Blood Donor	1	2	0	0	0
Hepatitis C	0	0	4	1	0
Fibrosis	0	0	0	4	0
Cirrhosis	0	0	0	2	4

Table 8: Confusion matrix for LSTM after optimization using Bayesian Optimization: 120/123 = 0.975.

Actual \ Predicted	No Suspect	Suspect Blood Donor	Hepatitis C	Fibrosis	Cirrhosis
No Suspect	105	0	0	1	0
Suspect Blood Donor	1	2	0	0	0
Hepatitis C	0	0	4	1	0
Fibrosis	0	0	0	4	0
Cirrhosis	0	0	0	1	5

Table 9: Classification report for LSTM after optimization using Random Search.

	Precision	Recall	F1-Score
No Suspect	1.00	0.96	0.98
Suspect Blood Donor	0.67	1.00	0.80
Hepatitis C	0.80	0.80	0.80
Fibrosis	0.57	1.00	0.73
Cirrhosis	1.00	0.67	0.80

Table 10: Classification report for LSTM after optimization using Bayesian Optimization.

	Precision	Recall	F1-Score
No Suspect	0.98	0.99	0.99
Suspect Blood Donor	0.99	0.97	1.00
Hepatitis C	0.80	0.80	0.80
Fibrosis	0.80	1.00	0.89
Cirrhosis	1.00	0.83	0.91

For variables that did not follow a normal distribution, the Kruskal-Wallis test was employed. This test indicated significant differences between the groups of the response variable. However, post hoc analysis showed that the variable 'CREA' did not exhibit significant differences among the groups, suggesting that it is less important from a biochemical standpoint and hence was excluded from further analysis.

Among ten supervised algorithms for classification health issues, Long Short-Term Memory (LSTM) model performed best across all metrics—accuracy (0.91), precision (0.91), recall (0.91) and F1-score (0.91) indicating excellent prediction capabilities.

Hyperparameter optimization significantly improved accuracy, precision, recall and F1-score, proving its effectiveness. Random Search increased accuracy from 0.91 to 0.9467 and Bayesian Optimization to 0.9756. From Table 7 and Table 9, The LSTM confusion matrix showed a significant decrease in false negatives, indicating better capture of liver disease cases post-optimization.

6. Acknowledgements

The first author is thankful to Department of Science and Technology, Ministry of Science and Technology, Government of India for providing fellowship under INSPIRE to pursue this work.

7. Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. References

1. Hutter F, Kotthoff L, Vanschoren J. Automated Machine Learning: Methods, Systems, Challenges. Springer Nature, 2019.
2. Schiff ER, Maddrey WC, Reddy KR. Schiff's Diseases of the Liver. Wiley-Blackwell, 2017.
3. Mitchell TM. Machine Learning. New York: McGraw-hill, 1997.