Research Article

Relative contribution of liver disease risk factors in the Prediction of liver diseases using artificial intelligence

Abstract

The liver is a key organ that filters blood from the digestive tract before it is passed on to the rest of the body. Liver illnesses are diverse, and liver function tests, such as alanine transaminase (ALT), can be used to diagnose them. The study's major goals were to employ neural network analysis to forecast liver disease and to determine the relative contribution of different liver disease predictors. A Kaggle dataset of Indian liver patients was used to do the analysis for liver disease prediction. There were 583 people in the study, and 71.4 percent of them had liver disease. Age, gender, ALT, aspartate aminotransferase (AST), bilirubin, albumin, total protein, albumin/globulin ratio, and alkaline phosphatase were all included as predictors in the study. The prediction model was successful in predicting liver illness in 79.6% of cases. ALT was the most important predictor, while alkaline phosphatase was the least important. Overall, neural network analysis is effective in predicting liver illness from one perspective, and it may be enhanced to provide more accurate results from the other perspective.

Keywords: liver disease, neural network analysis, predictors, dataset, Kaggle

Ahed J Alkhatib[1,2]

[1]Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science and Technology, Jordan
[2]International Mariinskaya Academy, department of medicine and critical care, department of philosophy, Academician secretary of department of Sociology

Correspondence: Ahed J Alkhatib, 1Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science and Technology, Jordan, Tel 00962795905145, Email ajalkhatib@just.edu.jo

Introduction

Obesity and liver disease are interacting health problems. Obesity is a huge global problem that has risen dramatically in recent decades. As a result, obesity and related illnesses are now a severe threat to the current and future health of all human communities. According to the World Health Organization (WHO), more than 1 billion persons worldwide are overweight, with 300 million of them being clinically obese, defined as having a body mass index (BMI) of 30 kg/m2 or higher.[1] The equally alarming rise in childhood obesity,[2] is particularly concerning. Obesity is linked to a slew of other health issues, including insulin resistance (IR), type 2 diabetes, nonalcoholic fatty liver disease (NAFLD), atherosclerosis, degenerative diseases including dementia, respiratory diseases, and even some malignancies.[3]

The liver is a large, meaty organ located on the stomach's right side. The liver is a reddish brown color with a rubbery texture and weighs roughly 3 pounds. The right and left lobes of the liver are separated into two sections. The gallbladder is located beneath the liver, as are sections of the pancreas and intestines. The liver and these organs work together to digest, absorb, and assimilate food. The major function of the liver is to filter blood from the digestive tract before it is passed on to the rest of the body. The liver detoxifies chemicals and metabolizes medications. The liver does this by accumulating bile, which is then released into the intestines. The liver also produces proteins that are needed for blood coagulation and other functions.[4] The term "liver disease" refers to any abnormality with the liver's function that results in illness. The liver is in charge of various dangerous duties in the body, and if it becomes ill or wounded, these functions may be lost, resulting in major bodily harm. The word "hepatic disease" refers to liver disease. Liver disease is a broad term that refers to any condition that causes the liver to fail to execute its functions. More than 75% of the liver tissue, or three quarters of the liver, must be impaired before function declines.[4]

Reena et al.,[5] proposed a liver disease-based data classification system. The training dataset is made up of 345 instances from the UCI repository, each with seven different properties. This paper discusses the findings of Nave Bayes algorithms in the field of data classification. When FT Tree and KStar algorithms were tested on liver disease datasets, they were shown to be faster than other algorithms in terms of running the data for results, with an accuracy of 97.1 percent. The classification accuracy of the FT Tree algorithm is superior than that of other algorithms, according to the results of the studies. Automated disease diagnosis and prediction rely heavily on data mining. It encompasses medical data analysis algorithms and procedures. Liver illnesses have become one of the most deadly diseases in various countries during the last decade.[6]

Jeyalakshmi & Rangaraj[7] did research to develop a deep learning-based method for accurately and reliably predicting liver illness. For the accurate prediction of liver disease outcomes, the Modified Convolutional Neural Network based Liver Disease Prediction System (MCNN-LDPS) was used. In the proposed study, dimensionality reduction was achieved using Modified Principal Component Analysis. The top features are identified using the Score-based Artificial Fish Swarm Algorithm (SAFSA). The SAFSA approach, which gave proper findings, used information gain and entropy values as input variables. A database of Indian liver patients was used to test this research technique. The examination of the research work revealed that the proposed method, MCNN-LDPS, generated better outcomes in terms of accuracy and precision. According to the comparison's findings, MCNN-LDPS improved accuracy by 4.05 percent, had a 21.23 percent F-measure, 4.22 percent precision, and a 34.26 percent recall. This research approach was compared to the existing Multi-layer Perceptron Neural Network for performance evaluation (MLPNN). CNN's inability to encode Orientational and relative spatial relationships, as well as view angle, was a major flaw. Prediction of liver disease is the most concentrated research subject in many medical organizations and industries. Hepatic diseases must be predicted as soon as feasible to ensure early therapy. Predicting the presence of liver disease in an automated and faster manner, on the other hand, is a more difficult task, especially with limited patient

data. The study by Rajeswari and Reena[5] classified the data depending on liver illness. The training dataset is made up of 345 instances from the UCI repository, each with seven different properties. In this study, the results of data classification are discussed. The results were obtained using Nave Bayes algorithms. When the FT Tree approach was tested on liver illness datasets, it took less time to process the data for results than other algorithms, and it had a 97.10 percent accuracy. The classification accuracy of the FT Tree algorithm is superior than that of other algorithms, according to the results of the studies. This method, on the other hand, does not perform well on high-scale data with more noisy characteristics. The algorithms Decision Tree, Naive Bayes, and NB Tree are used. Alfisahrin and Mantoro[8] advocated using the 10 major features of liver disease to assess if patients had liver disease. According to the findings, the NB Tree technique has the highest accuracy, but the Nave Bayes algorithm has the quickest calculating time. Future research will improve on the accuracy of the NB Tree algorithm, with the goal of discovering the most important factor in diagnosing liver disease patients. Traditional methods for predicting liver illness were used in this study, however they did not perform well when dealing with high-dimensional data. Dhamodharan[9] compared the accuracy of the Nave Bayes and FT tree algorithms and found that the Nave Bayes algorithm outperformed both. This research method, on the other hand, has a higher computational cost and does not focus on risk factors. Over the course of a year, Seker et al.,[10] applied data mining techniques such as KNN, SVM, MLP, and decision trees on a single dataset including 16,380 analytical results. This study can be beneficial for reducing the number of analyses since the prediction can be correlated and the correlation can be used to detect the anomaly on the analysis. This study methodology, on the other hand, has a lower accuracy value when processing insufficient patient data.

Study goals

The goal of this study was to use neural network analysis (artificial intelligence) to find the predictors of risk variables for liver disease and their relative relevance.

Methodology

A dataset regarding Indian liver patients that was put to Kaggle[11] was analyzed. There were 583 cases in the dataset. The case processing summary is shown in Table 1. There were 397 examples in the training section (68.6%), and 182 cases in the testing section (31.4 percent). Four cases were ruled out.

Table 1 Case processing summary

		N	Percent
Sample	Training	397	68.60%
	Testing	182	31.40%
Valid		579	100.00%
Excluded		4	
Total		583	

Network information

Network information included input layer, hidden layers, and output layer, as shown in Table 2. Age, total bilirubin, direct bilirubin, alkaline phosphatase, AST, ALT, total protein, albumin, and albumin/globulin ratio were all included in the input layer. The mechanism for rescaling covariates was standardized. There was one hidden layer with seven units on it. The activation function was tangent hyperbolic hyperbolic hyperbolic hyperbolic hyperbolic hyperbolic hyperbolic

hyperbol One variable, the dependent variable, was included in the output layer, as well as a dataset with two units. The error function was cross-entropy, and the activation function was softmax.

Table 2 Network information

Input layer	Covariates	1	Age
		2	Total bilirubin
		3	Direct bilirubin
		4	Alkaline phosphatase
		5	AST
		6	ALT
		7	Total protein
		8	Albumin
		9	Albumin/globulin ratio
	Number of Unitsa	9	
	Rescaling Method for Covariates	Standardized	
Hidden Layer(s)	Number of Hidden Layers	1	
	Number of Units in Hidden Layer 1a	7	
	Activation Function	Hyperbolic tangent	
Output Layer	Dependent Variables	1	Dataset
	Number of Units	2	
	Activation Function	Softmax	
	Error Function	Cross-entropy	

a. Excluding the bias unit

Results

Creating the study's architecture model

Figure 1 shows how the study's variables interacted with hidden layers to predict illness. The main colors were represented as gray and blue lines. Variable intensities and sizes of ea.ch color represent the computed influence of the interaction variable.

Summary of the model

The following is a model summary for the training and testing sections, as shown in Table 3. Cross entropy error 182.239, percent inaccurate prediction 26.4 percent for the training portion. 0:00:00.19 was the training time. The cross-entropy error for the testing section was 78.481, and the percent wrong prediction was 20.3 percent.

Output layer classification by model

The training component included 280 instances, of which 37 were predicted as diseased with an accuracy of 86.8%, as shown in Table 4 & Figure 2. The training portion also contained 117 sick instances, of which 68 were normal with a correction rate of 41.9 percent. 73.6 percent was the overall percentage. The testing portion contained 134 normal instances, with 11 cases correctly predicted as sick (91.8 percent). There were 48 diseased instances in all, with 26 of them being predicted as normal with a 45.8% percent correction. The total percentage adjustment was 79.7%.
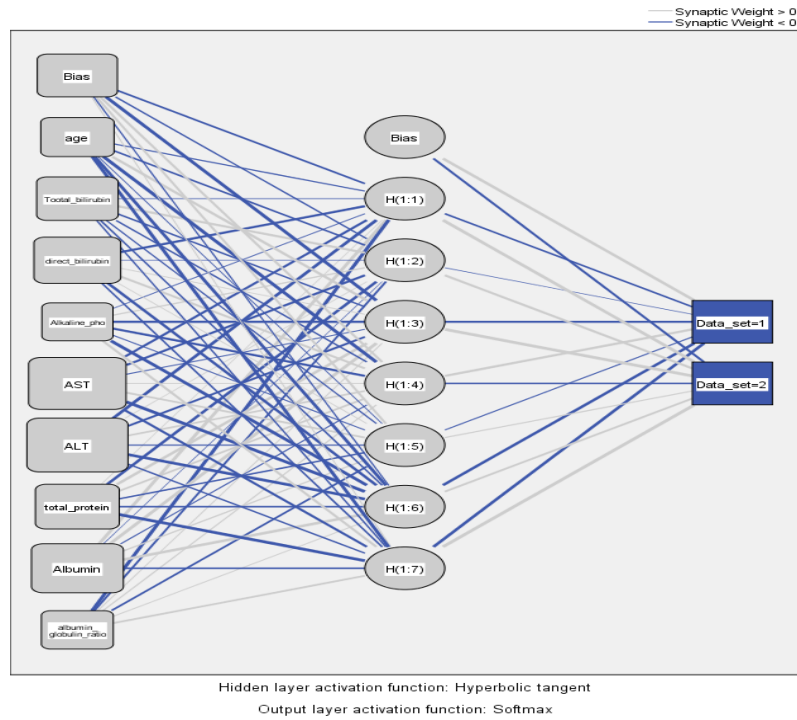
Figure 1 Schematic diagram showing the interactions of liver disease predicting model.

Table 3 Model summary

Training	Cross entropy error	182.239
	Percent Incorrect Predictions	26.40%
	Stopping Rule Used	1 consecutive step(s) with no decrease in errora
	Training Time	00:00.2
Testing	Cross Entropy Error	78.481
	Percent Incorrect Predictions	20.30%

Dependent Variable: Dataset

a. Error computations are based on the testing sample.

Table 4 Model classification of output layer

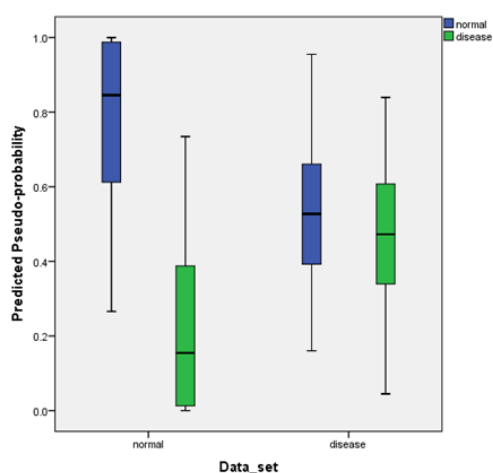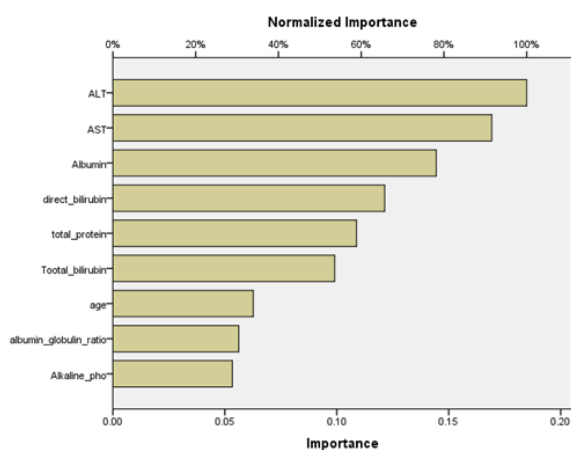| Sample	Observed	Predicted		
		normal	disease	Percent Correct
Training	Normal	243	37	86.80%
	Disease	68	49	41.90%
	Overall Percent	78.30%	21.70%	73.60%
Testing	Normal	123	11	91.80%
	Disease	26	22	45.80%
	Overall Percent	81.90%	18.10%	79.70%

Dependent variable: Dataset

Relative importance of independent variables

The independent variables were ordered in the following pattern according to their relevance, as shown in Table 5 & Figure 3: ALT (100%), AST (91.6%), albumin (78.6%), direct bilirubin (65.7%), total proteins (58.9%), total bilirubin (53.6%), age (33.9%), albumin/globulin ratio (30.4%), and alkaline phosphatase (30.4%) were all tested (28.9 percent).

Citation: Alkhatib AJ. Relative contribution of liver disease risk factors in the Prediction of liver diseases using artificial intelligence. *Adv Obes Weight Manag Control.* 2022;12(1):1–5. DOI: 10.15406/aowmc.2022.12.00356

Table 5 The importance of independent variables

	Importance	Normalized Importance
Age	0.063	33.90%
Total bilirubin	0.099	53.60%
Direct bilirubin	0.121	65.70%
Alkaline phosphatase	0.053	28.90%
AST	0.169	91.60%
ALT	0.185	100.00%
Total protein	0.109	58.90%
Albumin	0.144	78.10%
Albumin/globulin ratio	0.056	30.40%

Figure 2 Predicting pseudo-probability of output layer classification of liver disease.

Figure 3 Normalized importance of covariates on liver disease.

Discussion

The current study found that neural network analysis can properly predict liver illness. The overall percent of prediction in the testing phase was 79.6 percent. The success of liver prediction utilizing neural network analysis has been demonstrated in several research.[5,9] ALT and AST were found to be the most important predictors of liver disease in this investigation. Hepatocellular disease is defined by high ALT and AST levels, according to earlier research.[12]

Albumin was found to be the third best predictor of liver disease. Albumin, which accounts for 65 percent of TSP in the blood, is a significant liver function test because it is responsible for the transport of substances such as unconjugated bilirubin and some hormones. It maintains the blood's 80% colloid osmotic pressure and is used as a long-term biomarker of malnutrition, leading in the identification of nutrition-related chronic deficiencies.[13] Direct bilirubin was found to be the fourth best predictor of liver disease. Higher than usual amounts of direct bilirubin, on the other hand, may suggest that the liver isn't properly eliminating bilirubin.[14]

Total protein was found to be the fifth best predictor of liver disease. A total protein test determines the total quantity of protein in serum using a biochemical method.[15] Protein energy waste (PEW) is a condition in which the body's protein and energy supplies are depleted. Total serum proteins (TSP) are examined in the body to determine nutritional disorders such as PEW. When people are malnourished, this is caused by a lack of protein and energy-rich diets.[16]

Total bilirubin was found to be the sixth best predictor of liver disease. Most laboratories will report total bilirubin, which includes both unconjugated and conjugated fractions. As a result, increases in either percentage will result in an increase in the bilirubin concentration being measured. Gilbert's syndrome is the most common cause of isolated high bilirubin levels. It is a genetic metabolic disorder that results in faulty conjugation due to decreased activity of the enzyme glucuronyltransferase.[17]

The eighth predictor of liver disease was age. This finding is consistent with prior research that has found that age is a predictor of acute liver illness. Aging is a condition in which a person's ability to maintain homeostasis deteriorates over time due to structural changes or dysfunction, leaving them vulnerable to stress or injury from the outside world.[18]

The albumin/globulin ratio was found to be the sixth most accurate predictor of liver disease. The albumin/globulin ration yields different outcomes depending on the severity of liver disease. When compared to patients with liver disease, it is considerably higher in normal subjects (p=0.000). This could mean that the liver patients had fewer globulins than albumins, which is common in cancer instances.[19]

The least important risk factor for liver disease was alkaline phosphatase. The biliary epithelium of the liver produces alkaline phosphatase (ALP), which is also abundant in bone and found in smaller levels in the intestines, kidneys, and white blood cells. Hepatic congestion caused by right-sided heart failure may also cause cholestasis (elevated ALP levels and/or bilirubin). Glutamyltransferase can be used to detect whether ALP is hepatic or non-hepatic when it is raised in isolation.[20,21]

Conclusion

Neural network analysis can be used to classify liver disorders. With more and more trials, the accuracy of neural network analysis can be improved. The model used in this study was able to predict liver disorders by 79.6% and identify liver disease predictors. The most important predictor of liver disease was ALT.

Acknowledgments

Conflicts of interest

The authors declare that they have no competing interests.

Funding

None.

References

1. WHO The World Health Report 2002 – Reducing Risks, Promoting Healthy Life. 2009.

2. Rocchini AP. Childhood obesity and a diabetes epidemic. *N Engl J Med.* 2002;346:854–855.

3. Semenkovich CF Insulin resistance and atherosclerosis. *J Clin Invest.* 2006;116:1813–1822.

4. https://www.medicinenet.com/liver_disease/articl e.html

5. Rajeswari P, Sophia Reena G. Analysis of Liver Disorder Using Data Mining Algorithm. Global Journal of Computer Science and Technology, 2010;10.

6. Banu Priya M, Laura Juliet P, Tamilselvi PR. Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms. *International Research Journal of Engineering and Technology (IRJET).* 2018;5 (1), 206–211.

7. Jeyalakshmi K, Rangaraj R. Accurate liver disease prediction system using convolutional neural network. *Indian Journal of Science and Technology.* 2021;14(17):1406–1421.

8. Alfisahrin SNN, Mantoro T. Data mining techniques for optimization of liver disease classification. In: and others, editor. International Conference on Advanced Computer Science Applications and Technologies. 2013;379–384.

9. Dhamodharan S. Liver disease prediction using bayesian classification. *COMPUSOFT:An International Journal of Advanced Computer Technology.* 2014;1–3.

10. Seker SE, Unal Y, Erdem Z, et al. Ensembled Correlation Between Liver Analysis Outputs. *International Journal of Biology and Biomedical Engineering.* 2014;8:1–5.

11. https://www.kaggle.com/sanjames/liver–patients–analysis–prediction–accuracy.

12. Ribeiro AJS, Yang X, Patel V, et al. Liver Microphysiological Systems for Predicting and Evaluating Drug Effects. *Clin Pharmacol Ther.* 2019;106(1):139–147.

13. Tian CR, Qian L, Shen X, et al. Distribution of serum total protein in elderly Chinese. *PLoS ONE.* 2014;9:e101242.

14. https://www.mayoclinic.org/tests–procedures/bilirubin/about/pac–20393041

15. https://www.webmd.com/a–to–z–guides/what–is–a–total–serum–protein–test.

16. Sabatino A, Regolisti G, Karupaiah T, et al. Protein–energy wasting and nutritional supplementation in patients with end–stage renal disease on hemodialysis. *Clin Nutr.* 2017;36:663–671.

17. Monaghan G, Ryan M, Seddon R, et al. Genetic variation in bilirubin UPD–glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet.* 1996;347:578–581.

18. Kim IH, Kisseleva T, Brenner DA. Aging and liver disease. *Current opinion in gastroenterology.* 2015;31(3):184–191.

19. Suh B, Park S, Shin DW, et al. Low albumin–to–globulin ratio associated with cancer incidence and mortality in generally healthy adults. *Ann Oncol.* 2014;25(11):2260–2266.

20. Posen S, Doherty E The measurement of serum alkaline phosphatase in clinical medicine. *Adv Clin Chem.* 1981;22:163–245.

21. Whitehead MW, Hawkes ND, Hainsworth I, et al. A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. *Gut.* 45:129–133.