ISSN (O): 2349-5332
CODEN: IRJPHY

PREDICTION OF LIVER DISEASES USING NEURAL NETWORK ANALYSIS

Ahed J Alkhatib[1, 2],Khalid Saud Alharbi[3], Ahmad Mohammed Alsawi[3],Anas Mohammed Aljabal[3], Rayed Fahad Alhammad[3], Saif Saleh Alsaif[3], Abdulrahman Saeed Alshehri[3], Omar Mohamed Alqadi[3]

[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

[3]Ministry of Health, King Saudi Arabia

Submitted on: 18.11.2021; **Revised on:** 07.12.2021; **Accepted on:** 18.12.2021

ABSTRACT: Liver is a vital organ in the body and works to filter blood from the digestive tract before passing it on to the rest of the body. Liver diseases are varied and may be assessed by liver function tests including ALT. The main objectives of this study were to use neural network analysis to predict liver disease, and to identify the relative contribution of liver disease predictors. A dataset of Indian liver patients posted on Kaggle was used to be analyzed for liver disease prediction. The dataset included 583 subjects among whom 71.4% had liver disease. Study predictors included age, gender, ALT, AST, bilirubin, albumin, total protein, albumin/globulin ratio, and alkaline phosphatase. The prediction model was effective in 79.6% predicting the liver disease. The most important predictor was ALT, and the least important predictor was alkaline phosphatase. Taken together, using neural network analysis is effective in predicting liver disease from one side and from another side, it can be improved to give more accurate results.

KEYWORDS: Liver disease, neural network analysis, predictors, dataset, Kaggle

Corresponding Author: Ahed J Alkhatib
E-mail: ajalkhatib@just.edu.jo
Mobile: 00962795905145

Indian Research Journal of Pharmacy and Science; 30(2021)2590-2600;
Journal Home Page: https://www.irjps.in
DOI: 10.21276/irjps.2021.8.4.5

INTRODUCTION

The liver is a big, meaty organ found on the right side of the stomach. The liver is reddishbrown in color and rubbery to the touch, weighing around 3 pounds. The liver is divided into two main portions, known as the right and left lobes. The gallbladder, as well as sections of the pancreas and intestines, are located beneath the liver. To digest, absorb, and process food, the liver and these organs work together. The liver's primary function is to filter blood from the digestive tract before passing it on to the rest of the body. Chemicals are detoxified and medicines are metabolized by the liver. The liver does this by storing bile that eventually returns to the intestines. Proteins required for blood coagulation and other activities are also produced by the liver (medicinenet, 2021). Any problem with the liver's function that produces illness is referred to as liver disease. The liver is in charge ofseveral dangerous tasks in the body, and if it becomes diseased or injured, the loss of such functions can result in serious injury to the body. Hepatic disease is a term used to describe liver disease. Liver disease is a broad phrase that encompasses all potential issues that cause the liver to fail to perform its intended activities. Before a decline in function occurs, more than 75 percent of the liver tissue, or three quarters of the liver, must be compromised (medicinenet, 2021).

Reena et al (2010) suggested a data classification system based on liver disease. The training dataset is made up of 345 instances with seven different attributes that were gathered from the UCI repository. The findings of Nave Bayes algorithms in the realm of data classification are discussed in this work. When FT Tree algorithms and KStar algorithms were evaluated on liver disease datasets, the time taken to run the data for results was rapid when compared to other algorithms, with an accuracy of 97.1%. According to the findings of the experiments, the classification accuracy of the FT Tree algorithm is superior to that of other algorithms.

Data mining is an important aspect of automated disease diagnosis and prediction. It includes algorithms and techniques to analyze medical data. Liver disorders have increased excessively in the past decade, and in several countries, liver disease has become one of the most fatal diseases (Priya et al., 2018).

Jeyalakshmi andRangaraj (2021) conducted a study to offer a technique that uses deep learning to assure accurate and reliable liver disease prediction. The Modified Convolutional Neural Network based Liver Disease Prediction System (MCNN-LDPS) was utilized for the reliable prediction of liver disease outcomes. Dimensionality reduction was accomplished using Modified Principal Component Analysis in the suggested study. The Score-based Artificial Fish Swarm Algorithm is used to identify the best features (SAFSA). Information gain and entropy values were used as input variables in the SAFSA method, which produced correct results. This study strategy was tested on a database of Indian liver patients. The research work's analysis showed that the proposed approach MCNN-LDPS produced better results in terms of higher accuracy and precision. According to

the results of the comparison, MCNN-LDPS had a 4.05% increase in accuracy, a 21.23% F-measure, 4.22% precision, and a 34.26% recall. For performance study, this research approach was compared to the existing Multi-layer Perceptron Neural Network (MLPNN). The inability of CNN to encode Orientational and relative spatial relationships, as well as view angle, was a severe shortcoming. The most focused research problem in numerous medical organizations and industry is liver disease prediction. To ensure early treatment, hepatic disorders must be predicted as soon as possible. However, predicting the presence of liver disease in an automated and speedier manner is a more difficult issue, especially with inadequate patient data. The data classification in the study of Rajeswari and Reena (2010) is based on liver disease. The training dataset is made up of 345 instances with seven different attributes that were gathered from the UCI repository. The outcomes of data classification are discussed in this study. Nave Bayes methods were used to obtain the results. When FT Tree method was evaluated on liver disease datasets, the time taken to run the data for results was rapid when compared to other algorithms, with an accuracy of 97.10 percent.According to the findings of the experiments, the classification accuracy of the FT Tree algorithm is superior to that of other algorithms. On high-scale data with more noisy features, however, this approach does not perform well. Using a Decision Tree, Naive Bayes, and NB Tree algorithms.Alfisahrin andMantoro (2013) proposed to determine if patients had liver illness based on the 10 key features of liver disease. The NB Tree approach has the highest accuracy, but the Nave Bayes algorithm has the quickest computing time, according to the results. The accuracy of the NB Tree algorithm will be the focus of future research, with the goal of determining the most relevant factor in detecting liver disease patients. This study solely uses traditional algorithms for predicting liver disease, which did not perform well when dealing with high-dimensional data.Dhamodharan (2014) examined the accuracy of the Nave Bayes and FT tree algorithms and determined that the Nave Bayes algorithm is far superior to the other methods. This research style, on the other hand, has a higher computing overhead and does not focus on risk variables. Seker et al (2014)used data mining techniques such as KNN, SVM, MLP, and decision trees on a single dataset comprised of 16,380 analytical results over the course of a year. Because the prediction can be correlated and the correlation can be used to discover the abnormality on the analysis, this study can be effective for minimizing the number of analyses. However, while processing inadequate patient data, this study methodology has a lower accuracy value.

Study objectives:

The objectives of this study were to determine the predictors of risk factors of liver disease and their relative importance using neural network analysis.

METHODOLOGY:

A dataset posted on Kaggle (Kaggle, 2021) about Indian liver patients. The datasetconsisted of 583 cases. As seen in

table (1), case processing summary is given. Training part included 397 cases (68.6%), and testing part included 182 cases (31.4%). Four cases were excluded.

Table 1: Case Processing Summary

		N	Percent
Sample	**Training**	397	68.6%
	Testing	182	31.4%
Valid		579	100.0%
Excluded		4	
Total		583	

Network Information

As illustrated in table (2), network information included input layer, hidden layers, and output layer. Input layer included 9 covariates: age, total bilirubin, direct bilirubin, alkaline phosphatase, AST, ALT, total protein, albumin, and albumin/globulin ratio. Rescaling method for covariates was standardized. There was one hidden layer including 7 units. The activation function was hyperbolic tangent. The output layer included one variable, the dependent variable, the dataset with 2 units. The activation function was softmax, and the error function was cross-entropy.

Table 2: Network Information

Input Layer	**Covariates**	**1**	**age**
		2	**Totalbilirubin**
		3	**Directbilirubin**
		4	**Alkalinephosphatase**
		5	**AST**
		6	**ALT**
		7	**Totalprotein**
		8	**Albumin**
		9	**Albumin/globulinratio**
	Number of Units[a]		**9**
	Rescaling Method for Covariates		**Standardized**
Hidden Layer(s)	**Number of Hidden Layers**		**1**
	Number of Units in Hidden Layer 1[a]		**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

Building the Architecture Model of the Study

The study's covariates interacted with hidden layers to provide a disease prediction, as shown in figure (1). The gray and blue lines were drawn as the major colors. The computed influence of the interacting variable is shown by varied intensities and sizes of each color.

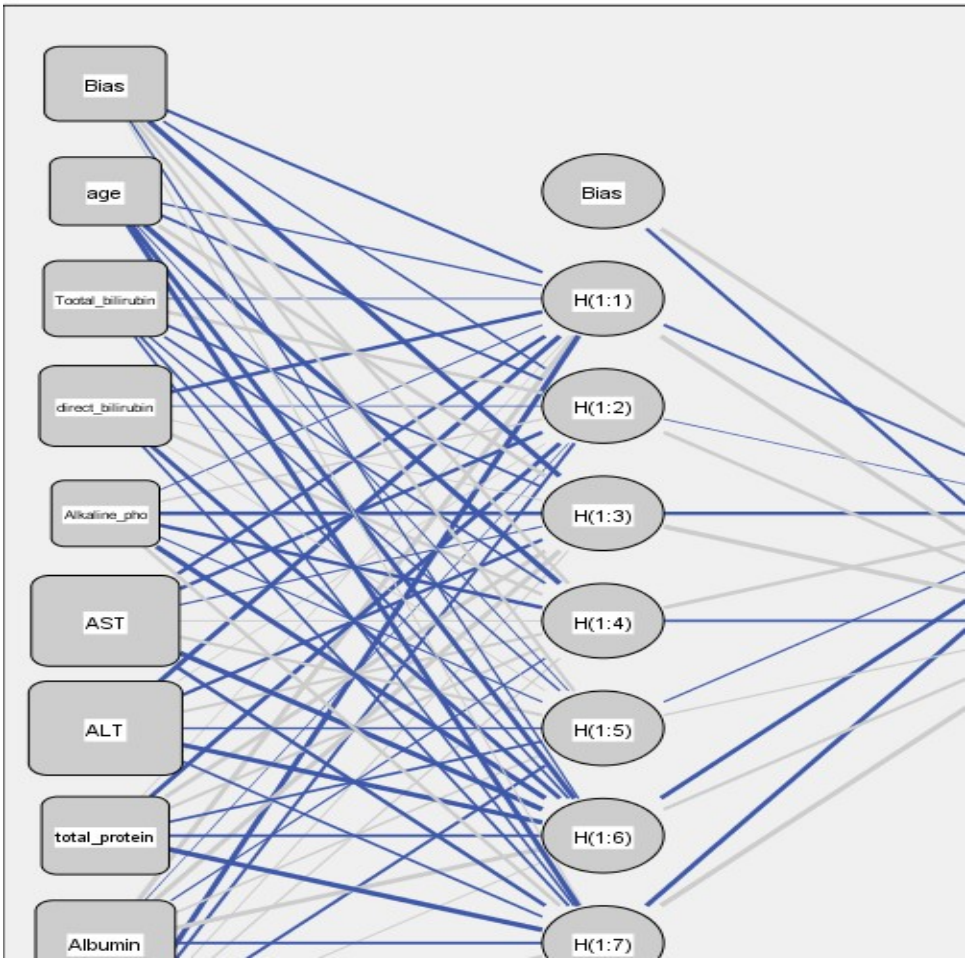

Figure 1: Schematic diagram showing the interactions of liver disease predicting model

Model Summary

As demonstrated in table (3), model summary for training and testing parts was given as follows. for training part, cross entropy error 182.239, percent incorrect prediction 26.4%. training time was 0:00:00.19. For testing part, cross entropy error was 78.481, percent incorrect prediction was 20.3%

Table 3: Model summary

Training	Cross Entropy Error	182.239
	Percent Incorrect Predictions	26.4%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.19
Testing	Cross Entropy Error	78.481
	Percent Incorrect Predictions	20.3%
Dependent Variable: Dataset		
a. Error computations are based on the testing sample.		

Model classification of output layer

As shown in table (4) and figure (2), training part included 280 cases among which were 37 cases predicted as diseased with 86.8% percent correction.Training part also included 117 diseased cases of which 68 cases were normal with 41.9% percent correction. Overall percent was 73.6%. Testing part included 134 cases as normal of which 11 cases predicted as diseased with 91.8% percent correction. Diseased cases included 48 cases of which 26 cases were predicted as normal with 45.8% percent correction. The overall percent correction was 79.7%

Table 4: Model classification of output layer

Sample	Observed	Predicted		
		normal	disease	Percent Correct
Training	Normal	243	37	86.8%
	Disease	68	49	41.9%
	Overall Percent	78.3%	21.7%	73.6%
Testing	Normal	123	11	91.8%
	Disease	26	22	45.8%
	Overall Percent	81.9%	18.1%	79.7%
Dependent Variable: Dataset				

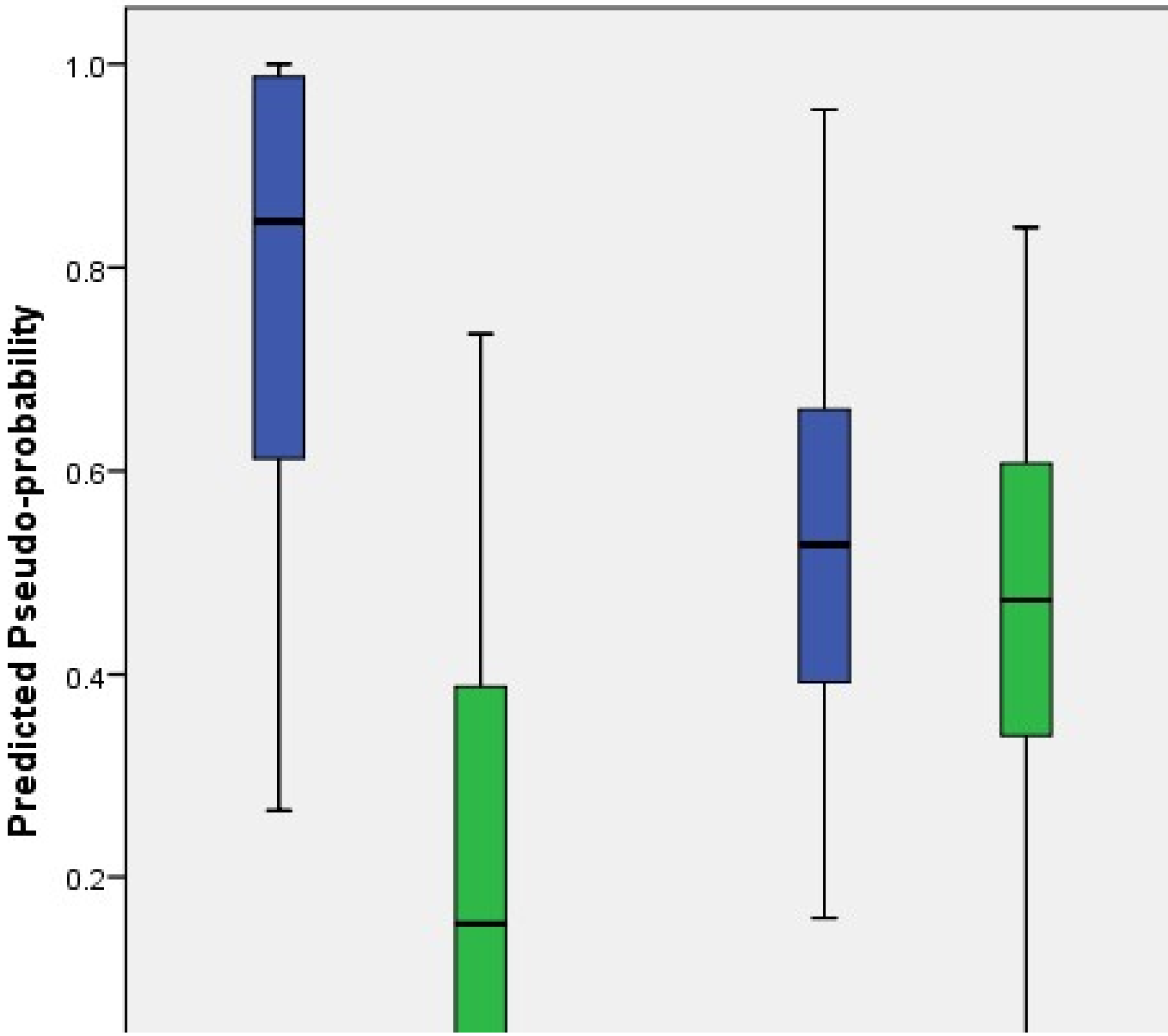

Figure 2: Predicting pseudo-probability of output layer classification of liver disease

The importance of independent variables

As shown in table (5) and figure (3), the independent variables were arranged in the following pattern according to their importance: 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 (28.9%).

Table 5: The importance of independent variables

	Importance	Normalized Importance
Age	.063	33.9%
Totalbilirubin	.099	53.6%
Directbilirubin	.121	65.7%
Alkalinephosphatase	.053	28.9%
AST	.169	91.6%
ALT	.185	100.0%
Totalprotein	.109	58.9%
Albumin	.144	78.1%
Albumin/globulinratio	.056	30.4%

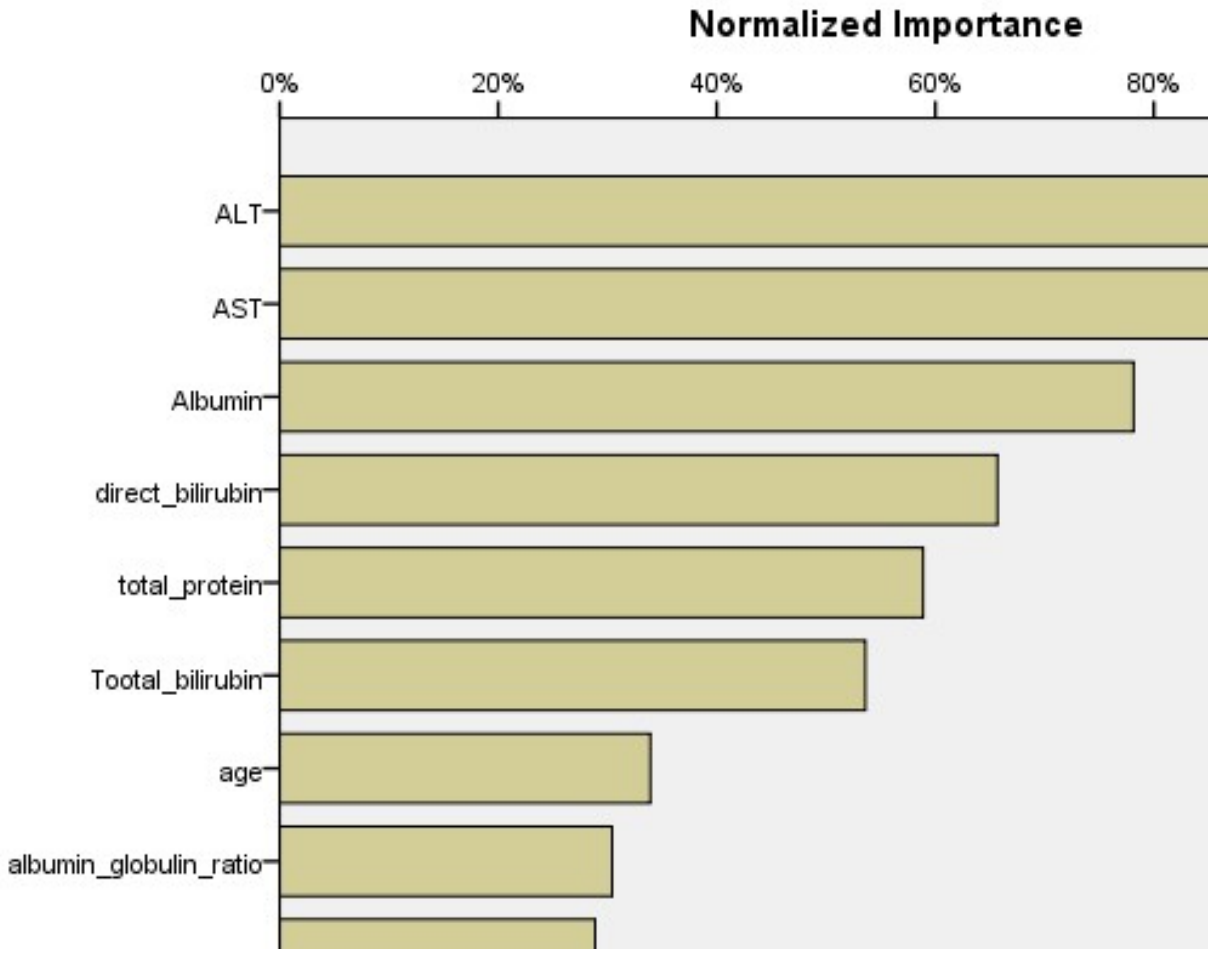

Figure 3: Normalized importance of covariates on liver disease

DISCUSSION

The present study showed that liver disease can be predicted accurately using neural net work analysis. In the testing part, the overall percent of prediction was 79.6%.Several studies have reported the successfulness of liver prediction using neural network analysis (Rajeswariet al., 2010; Dhamodharan, 2014).

The results of this study showed that ALT and AST were the most important predictors of liver disease. This is in line with previous studies that showed hepatocellular disease is defined by elevated ALT and AST levels (Ribeiro et al., 2019).

The results showed that albumin was the third predictor of liver disease. Albumin is a very important liver function test because it is responsible for the transfer of chemicals such as unconjugated bilirubin and certain hormones, accounting for 65% of TSP in the blood. It is responsible for maintaining the blood's 80% colloid osmotic pressure and is utilized as a long-term indicator of malnutrition, resulting in nutrition-related chronic deficiencies diagnosis (Tian et al., 2014).

The results showed that direct bilirubin was the fourth predictor of liver disease. However, higher than normal levels of direct bilirubin may indicate that liver isn't clearing bilirubin properly (mayoclinic, 2021).

The results showed that total protein was the fifth predictor of liver disease. A total protein is a biochemical test for determining the total amount of protein in serum (webmd, 2021).Total serum proteins (TSP) are evaluated in the body to diagnose nutritional issues including protein energy waste (PEW), which is a condition in which the body's protein and energy stores are depleted. This is caused by a lack of protein and energy-rich foods, and it happens when people are malnourished (Sabatino et al., 2017).

The results showed that total bilirubin was the sixth predictor of liver disease. Total bilirubin, which includes both unconjugated and conjugated fractions, is commonly reported by most laboratories. As a result, increases in either percentage will result in an increase in the measured bilirubin concentration. Gilbert's syndrome, a hereditary metabolic condition that results in defective conjugation due to diminished activity of the enzyme glucuronyltransferase, is the most prevalent cause of an isolated high bilirubin levels (Monaghan et al., 1996).

Age was the seventh predictor of liver disease. This result is in line with other studies that showed age is a predicting factor for acute liver disease. Aging is a condition in which a person's ability to maintain homeostasis gradually deteriorates owing to structural changes or dysfunction, leaving them vulnerable to external stress or injury (Kim et al., 2015).

Albumin/globulin ratio was the eighth predictor of liver disease. The results of albumin/globulin ration vary according to the severity of liver disease. It is significantly higher in normal subjects compared with patients who have liver disease (p=0.000). This may indicate that liver patients had less globulins than albumins, this is usually involved in cancer cases (Suh et al., 2014).

Alkaline phosphatase was the least important risk factor predicting liver disease. Alkaline phosphatase (ALP) is mostly produced in the liver (by the biliary epithelium), although it is also abundant in bone and found in smaller amounts in the intestines, kidneys, and white blood cells. Furthermore, cholestasis (elevated ALP levels and/or bilirubin) might be caused by hepatic congestion caused by right-sided heart failure.When ALP is increased in isolation, glutamyltransferase can be used to determine whether the ALP is hepatic or non-hepatic (Posen and Doherty, 1981; Whitehead et al., 1991).

CONCLUSION:

Liver diseases can be classified using neural network analysis. The accuracy of neural network analysis can be increased with more and ore trials. The model in this study was able to predict liver diseases by 79.6%, and to identify predictors of liver diseases. ALT was the most important predicting factor of liver disease.

REFERENCES:

1. https://www.kaggle.com/sanjames/liver-patients-analysis-prediction-accuracy, retrieved in 17/12/2021.
2. https://www.medicinenet.com/liver_ disease/articl e.htm, retrieved in 17/12/2021.
3. P. Rajeswari,G. Sophia Reena (2010), Analysis of Liver Disorder Using Data Mining Algorithm. Global Journal of Computer Science and Technology, 10.
4. M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi (2018). Performance Analysis of Liver Disease Prediction Using MachineLearning Algorithms. International Research Journal of Engineering andTechnology (IRJET), 5 (1), 206-211.
5. Jeyalakshmi K, Rangaraj R. (2021). Accurate liver disease prediction system using convolutional neural network. Indian Journal of Science and Technology. 14(17): 1406-1421. https://doi.org/10.17485/IJST/v14i17.451.
6. Alfisahrin SNN, Mantoro T (2013). Data mining techniques for optimization of liver disease classification. In: and others, editor. 2013 International Conferenceon Advanced Computer Science Applications and Technologies, 379–384. doi:10.1109/ACSAT.2013.81.
7. Dhamodharan S (2014). Liver disease prediction using bayesian classification. COMPUSOFT: An International Journal of Advanced Computer Technology. 1–3. Available from: https://ijact.joae.org/index.php/ijact/article/view/443/378.
8. Seker SE, Unal Y, Erdem Z, Kocer HE (2014). Ensembled Correlation Between Liver Analysis Outputs. International Journal of Biology and Biomedical Engineering.8:1–5. Available from: https://arxiv.org/abs/1401.6597. doi:2014.
9. Ribeiro AJS, Yang X, Patel V, Madabushi R, Strauss DG (2019).

LiveMicrophysiological Systems for Predicting and Evaluating Drug Effects. Clin PharmacolTher., 106(1):139-147. [PMC free article] [PubMed].

10. Tian C.R., Qian L., Shen X.Z., Li J.J., Wen J.T (2014). Distribution of serum total protein in elderly Chinese. PLoSONE,.9:e101242. doi: 10.1371/journal.pone.0101242.
11. https://www.mayoclinic.org/tests-procedures/bilirubin/about/pac-20393041, retrieved in 18/12/2021.
12. https://www.webmd.com/a-to-z-guides/what-is-a-total-serum-protein-test, retrieved in 6/12/2021.
13. Sabatino A., Regolisti G., Karupaiah T., Sahathevan S., Singh B.K.S., Khor B.H., Salhab N., Karavetian M., Cupisti A., Fiaccadori E (2017). Protein-energy wasting and nutritional supplementation in patients with end-stage renal disease on hemodialysis. Clin. Nutr. 36:663–671. doi: 10.1016/j.clnu.2016.06.007.
14. Monaghan G, Ryan M, Seddon R, et al (1996). Genetic variation in bilirubin UPD-glucuronosyl transferase gene promoter and Gilbert's syndrome. Lancet, 347:578–81.doi: 10.1016/S0140-6736(96)91273-8.
15. Kim, I. H., Kisseleva, T., & Brenner, D. A. (2015). Aging and liver disease. Current opinion in gastroenterology, 31(3), 184–191. https://doi.org/10.1097/MOG.0000000000000176.
16. Suh B, Park S, Shin DW, Yun JM, Keam B, Yang HK, Ahn E, Lee H, Park JH, Cho B (2014). Low albumin-to-globulin ratio associated with cancer incidence and mortality in generally healthy adults. Ann Oncol., 25(11):2260-2266.
17. Posen S, Doherty E (1981). The measurement of serum alkaline phosphatase in clinical medicine. Adv Clin Chem, 22:163–245.
18. Whitehead MW, Hawkes ND, Hainsworth I, et al (1999). A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. Gut, 45:129–33.doi:10.1136/gut.45.1.129.

CONFLICT OF INTEREST REPORTED: NIL; SOURCE OF FUNDING: NONE REPORTED