International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 3 Issue 10, October – 2019, Pages: 1-11

Predicting Liver Patients using Artificial Neural Network

Musleh M. Musleh[2], Eman Alajrami[1], Ahmed J. Khalil[2], Bassem S. Abu-Nasser[2], Alaa M. Barhoom[2], Samy S. Abu-Naser[2]

[1]Faculty of Information Technology, University of Palestine, Gaza, Palestine
[2]Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine

Abstract: *Liver diagnosis at an early stage is essential for enhanced handling. Precise classification is required for automatic recognition of disease from data samples (utilizing data mining for classification of liver patients from healthy ones). In this study, an artificial neural network model was designed and developed using JustNN Tool for predicting weather a person is a liver patient or not based on a dataset for liver patients. The main factors for input variables are: Age, Gender, Total Bilirubin, Direct Bilirubin, Alkphos Alkaline Phosphotase. Sgpt Alamine Aminotransferase, Sgot Aspartate Aminotransferase, Total Protiens, Albumin, Albumin and Globulin Ratio, and the output variable: Status. The dataset used for training are the data published in the literature for various 583 liver patients. The model was trained and validated, most important factors affecting Status of liver patient identified, and the accuracy for the validation was 99.00%.*

Keywords - Liver disease, ANN, Artificial Neural network, JustNN

1. Introduction

Liver is located in the right upper quadrant of the abdomen, below the diaphragm. Its other roles in metabolism include the regulation of glycogen storage, decomposition of red blood cells and the production of hormones [1].

The liver is an accessory digestive organ that produces bile, an alkaline compound which helps the breakdown of fat. Bile aids in digestion via the emulsification of lipids. The gall bladder, a small pouch that sits just under the liver, stores bile produced by the liver which is afterwards moved to the small intestine to complete digestion [1]. The liver's highly specialized tissue consisting of mostly hepatocytes regulates a wide variety of high-volume biochemical reactions, including the synthesis and breakdown of small and complex molecules, many of which are necessary for normal vital functions [2]. Estimates regarding the organ's total number of functions vary, but textbooks generally cite it being around 500[2].

The liver is a vital organ and supports almost every other organ in the body. Because of its strategic location and multidimensional functions, the liver is also prone to many diseases like [3]:

- **Hepatitis** is a common condition of inflammation of the liver. The most usual cause of this is viral, and the most common of these infections are hepatitis A, B, C, D, and E.
- **Hepatic encephalopathy** is caused by an accumulation of toxins in the bloodstream that are normally removed by the liver. This condition can result in coma and can prove fatal.
- **Budd–Chiari syndrome** is a condition caused by blockage of the hepatic veins (including thrombosis) that drain the liver. It presents with the classical triad of abdominal pain, ascites and liver enlargement[4].
- **Primary biliary cholangitis** is an autoimmune disease of the liver. It is marked by slow progressive destruction of the small bile ducts of the liver, with the intralobular ducts (Canals of Hering) affected early in the disease[5].
- There are also many pediatric liver diseases, including: **biliary atresia, alpha-1 antitrypsin deficiency, alagille syndrome, progressive familial intrahepatic cholestasis, Langerhans cell histiocytosis and hepatic hemangioma a benign tumor** the most common type of liver tumor, thought to be congenital [6].

Classification is a data mining technique comprising of a dual process flow. In the first step the classifier is trained using the training dataset, while the classifier is being tested for its prediction capacity in the second phase using different samples of the test set [8]. Feature selection is the preliminary step to be performed prior to application of classification algorithms for any dataset.

Classification algorithms can be either supervised or unsupervised based on the learning mechanism. Supervised learning is implemented by set of labels defined prior in the training set. The function is mapped for new unseen data to predict the labels. Few examples are Discriminative learning, Artificial Neural Network, Bagging, Boosting, Naïve Bayes, Kernel-based classifiers, Nearest Neighbor algorithm, Decision Trees, Random Forest,

and other ensemble of classifiers. Whereas unsupervised learning identifies the missing or hidden patterns in unlabeled data without any labels. They are commonly used for dimensionality reduction of feature space. The unsupervised ensembles include clustering approaches, self-organization maps, hidden Markov models and adaptive resonance theory [9]. In this study, we developed an ANN model for classification of liver patients using JustNN Tool[10].

2. Literature Review

A study conducted by [11] was used for classification followed by induction of rule set using Learn by Example algorithm. It was followed by execution of fuzzy rules to identify different liver disease types which achieved an accuracy of 96%.

The study of [12] used four algorithms: NB, DT, MLP and k-NN, they evaluated the results on the basis of 4 criteria, which are accuracy, precision, sensitivity, specificity. They used ranking algorithm for feature selection available in WEKA and ordered them by priority on the class. The averages of accuracy, precision, sensitivity and specificity of them are 96.55, 93.69, 0.92 and 0.98 with 12 features, respectively. But when we tried to output prediction results only with default parameters and no filters, the results were very lower than the previous study.

Authors in this paper [13] used C4.5, Random Forest, CART, Random Tree and REP tree classification method and get better accuracy to detect liver disease. They achieved accuracy 79.22% in Random Forest using 80-20% training-testing data partition.

Based on the review of literature, it was portrayed that the past research studies implemented different techniques for classification of liver dataset. ANN model using JustNN tool can be adapted to further increase the prediction accuracy of liver disease.

3. Methodology

3.1 Data Collection

The Indian Liver Patient Dataset (ILPD) was selected from UCI Machine learning repository for this study [7]. It is a sample of the entire Indian population collected from Andhra Pradesh region. The dataset consist of 583 instances based on ten different biological parameters. The Status value was reported based on these parameters as either Liver patient (416 cases) or not liver patient (167 cases) to represent the liver infection.

3.2 Pre-processing and Feature selection

Pre-processing techniques was applied to normalize the missing values. The missing values along with their instances were replaced by null value. It was followed by feature selection to identify relevant attribute for classification. Feature selection was performed using most significant factor method in JustNN tool.

3.3 Randomization and splitting of dataset

The features selected in the preceding step were approved to develop classification models. Initially the dataset was randomized to obtain an arbitrary permutated sample. It was followed by splitting of the dataset into training (83% of the dataset) and test (17%) sets. Training set comprised of 483 instances and test set included the remaining 100 instances.

3.4 Classification algorithms

We used JustNN Tool algorithm for classification of liver patients.

3.5 Dataset description

The Indian Liver Patient Dataset consists of 10 different attributes of 583 patients. The patients were described as either has liver disease (1) or do not have liver disease (2). The detailed description of the dataset is shown in Table 1.

International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 3 Issue 10, October – 2019, Pages: 1-11

Table 1: Description of Liver patient dataset

Sl. No	Attribute name	Attribute Type	Attribute Description
0	Age	Numeric	Age of the patient
1	Gender	Nominal	Gender of the patient
2	Total Bilirubin	Numeric	Quantity of total bilirubin in patient
3	Direct Bilirubin	Numeric	Quantity of direct bilirubin in patient
4	Alkphos Alkaline Phosphotase	Numeric	Amount of ALP enzyme in patient
5	Sgpt Alamine Aminotransferase	Numeric	Amount of SGPT in patient
6	Sgot Aspartate Aminotransferase	Numeric	Amount of SGOT in patient
7	Total Protiens	Numeric	Protein content in patient
8	Albumin	Numeric	Amount of albumin in patient
9	Albumin and Globulin Ratio	Numeric	Fraction of albumin and globulin in patient
10	Status	Numeric {1, 2}	Status of liver disease in patient

3.6 Data Analysis

We used the utilities of the JustNN tool to analyze the input variable with the output (status) variable. Figures 1-10 shows snap shots for the analysis from the age to Status.

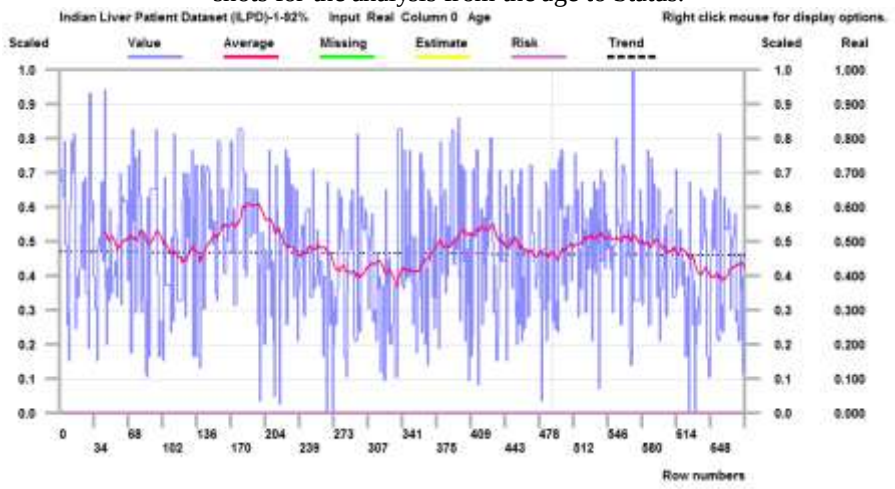

Figure 1: Data analysis with respect to Age

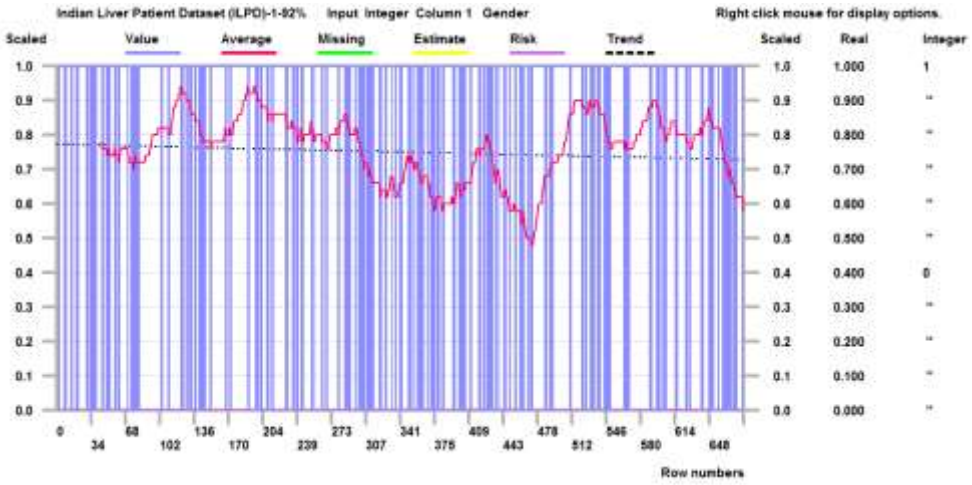

Figure 2: Data analysis with respect to Gender

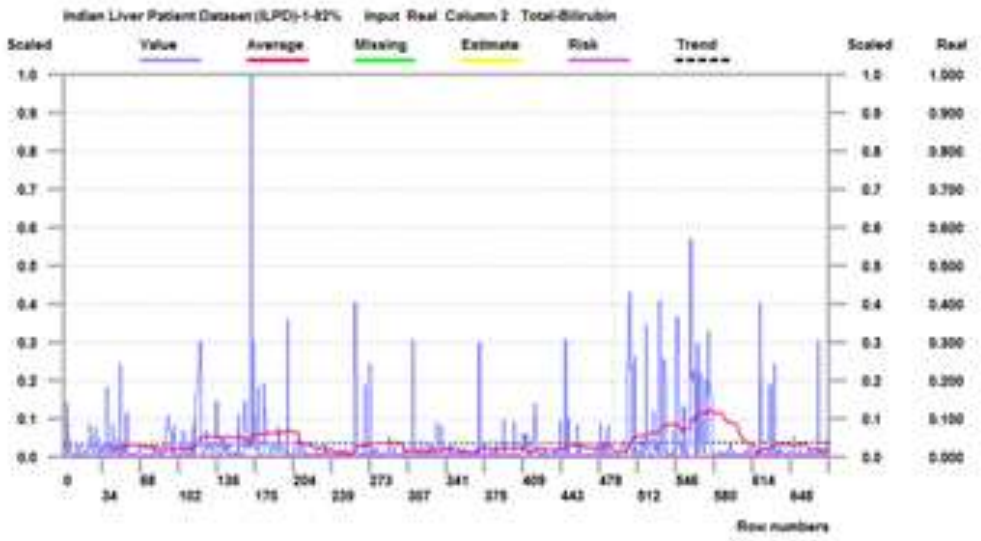

Figure 3: Data analysis with respect to Total Bilirubin

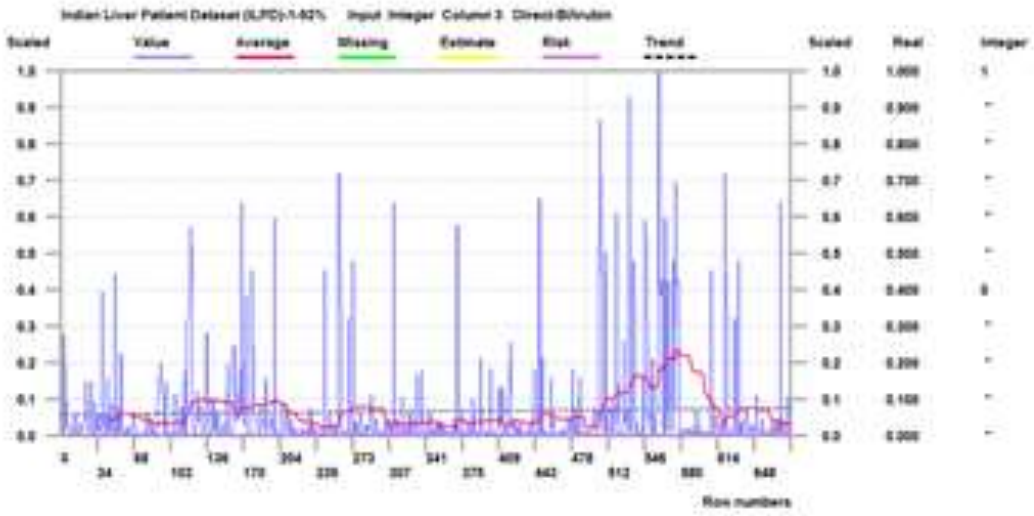

Figure 4: Data analysis with respect to Direct Bilirubin

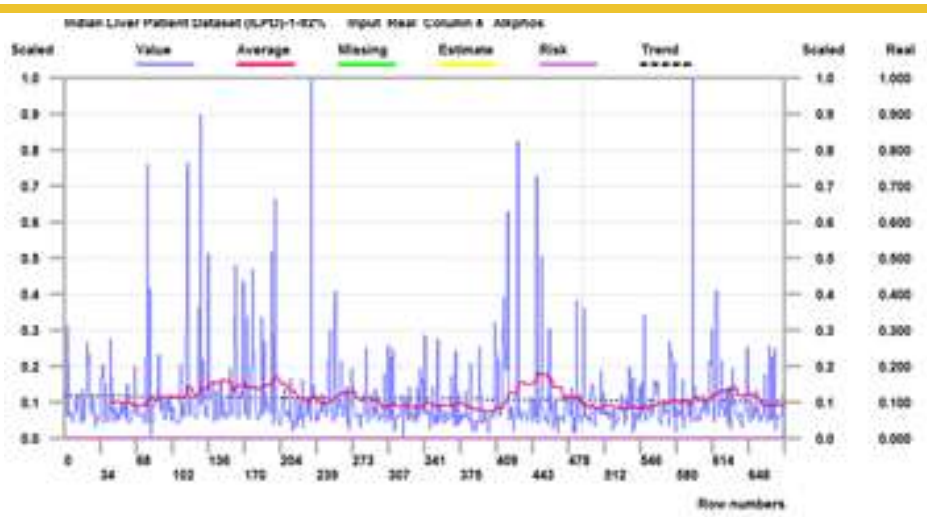

Figure 5: Data analysis with respect to Alkphos Alkaline Phosphotase

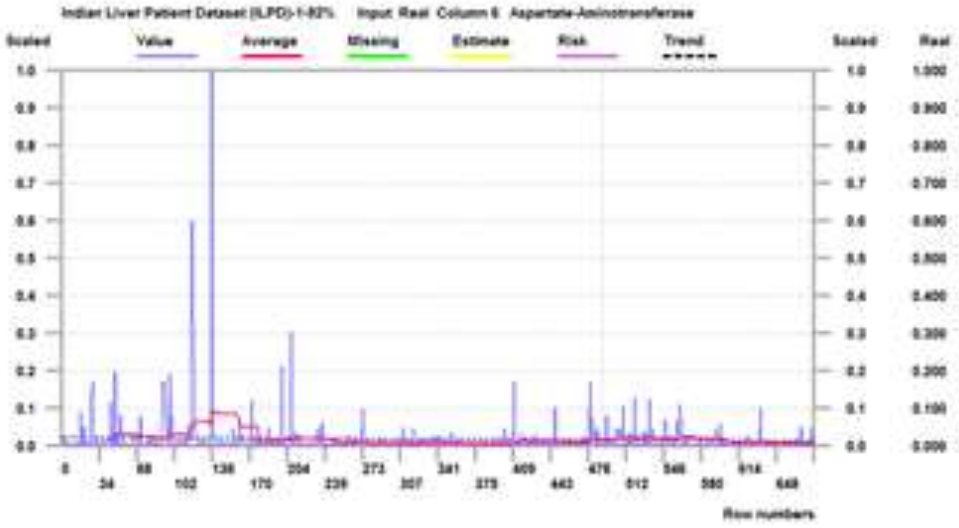

Figure 6: Data analysis with respect to Sgpt Alamine Aminotransferase

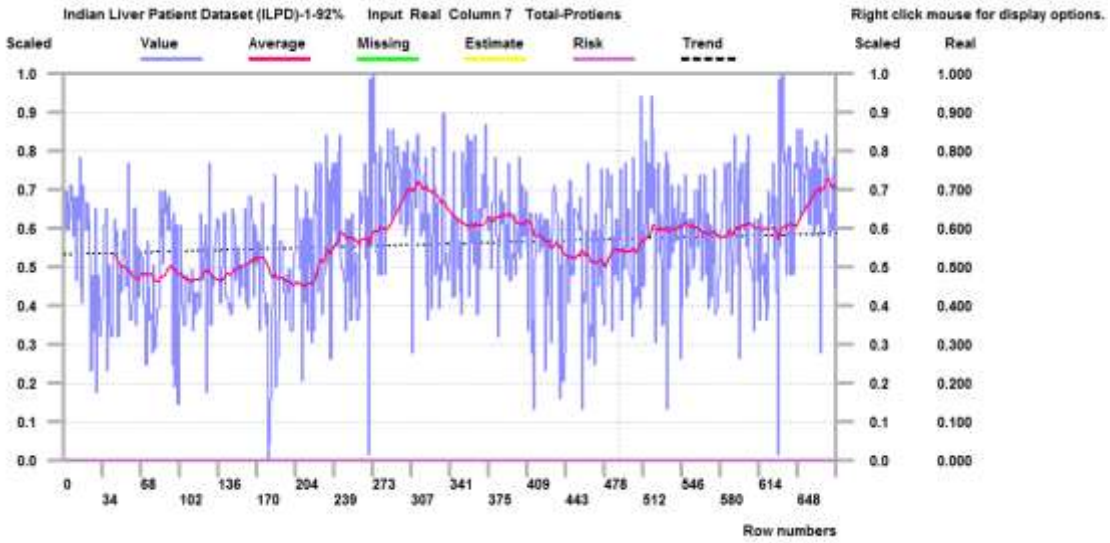

Figure 7: Data analysis with respect to Sgot Aspartate Aminotransferase

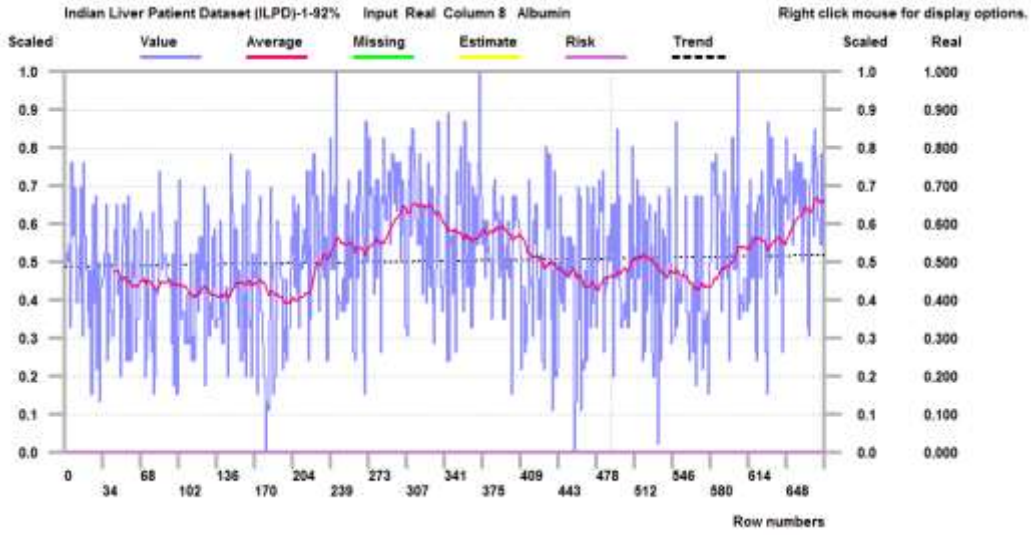

Figure 8: Data analysis with respect to Albumin

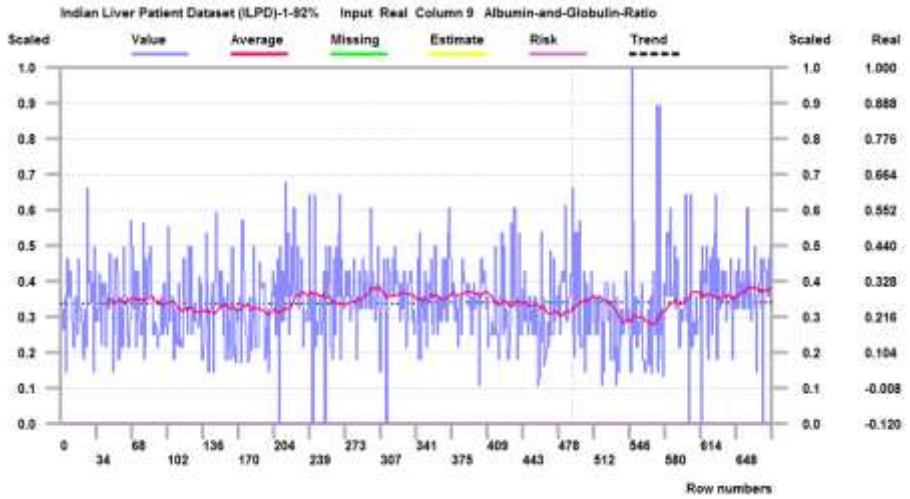

Figure 9: Data analysis with respect to Albumin and Globulin Ratio

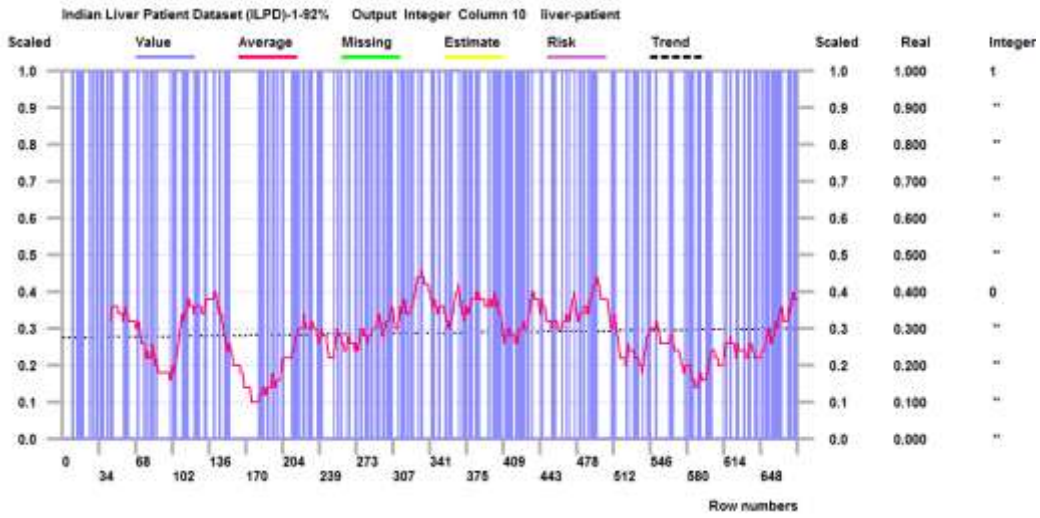

Figure 10: Data analysis with respect to Status

3.7 ANN Model

The resulted predictive ANN model is shown in Figure 11 and Figure 14.

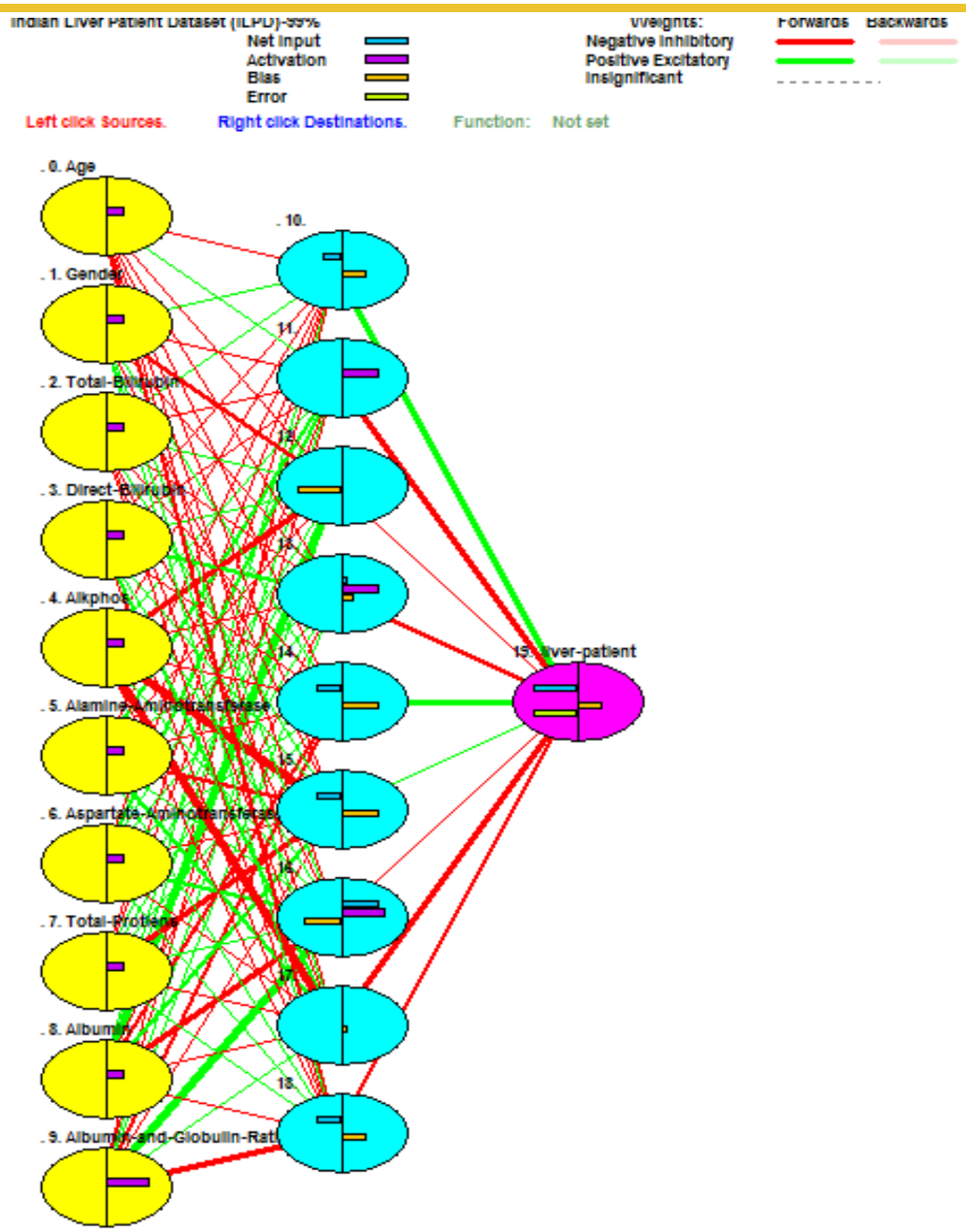

Figure 11: ANN model of our proposed Liver Patient system

3.8 Validation

Our ANN model was able to predict weather a person is a liver patient or not with 99.00% accuracy, with about 0.003 errors as seen in Figure 12. Furthermore, The Model showed that the most effective factor in liver patient is: Alkphos Alkaline Phosphotase, Albumin and Globulin Ratio, Albumin. More details are shown in Figure 13.

International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 3 Issue 10, October – 2019, Pages: 1-11

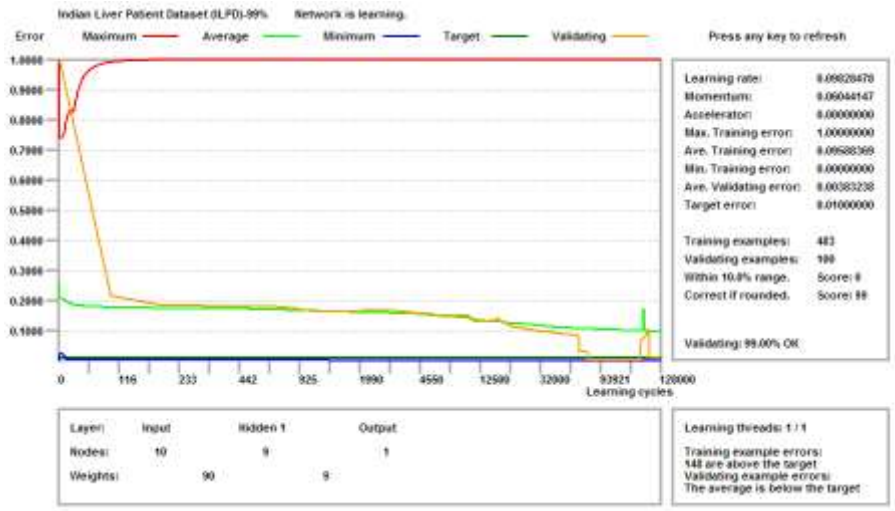

Figure 12: Training and validating our ANN model of our proposed Liver Patient system

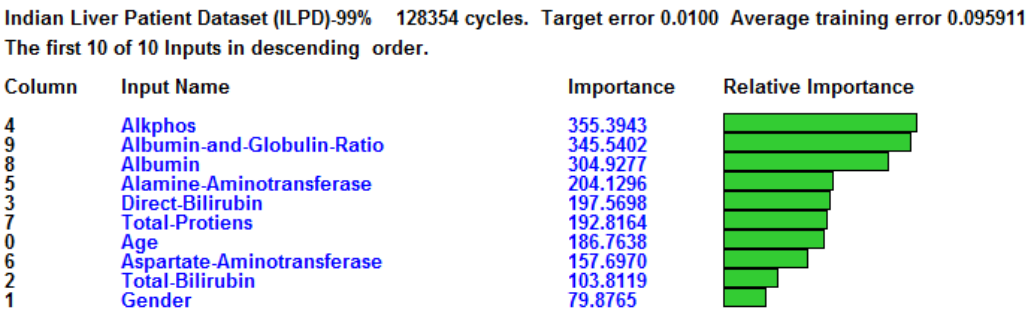

Figure 13: Most relevant factors in our ANN model of our proposed Liver Patient system

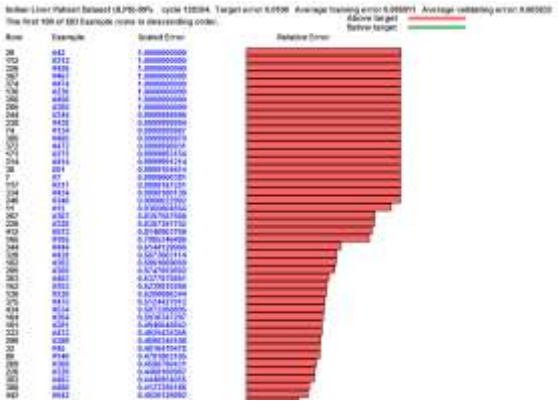

Figure 14: Errors in our ANN model of our proposed Liver Patient system

4. Conclusion

The liver is an essential body organ that forms an important barrier between the gastrointestinal blood which contains large amounts of toxins and antigens in the body. The impairment of this organ is the main reason of illness and death. In this paper, Liver Patient has been investigated using Artificial Neural Network model to predict weather a person is Liver Patient or not and analysis using JustNN Tool was used to determine the effect of input variables based on the data in the literature.

- A simple static neural network model gives a very good prediction (99.00%) in comparison with the original data sets of [7].

- The Alkphos Alkaline Phosphotase, Albumin and Globulin Ratio, and Albumin have significant effects on whether a person is Liver Patient or not for the present problem.

References

1. Tortora, Gerard J.; Derrickson, Bryan H. (2008). Principles of Anatomy and Physiology (12th ed.). John Wiley & Sons. p. 945. ISBN 978-0-470-08471-7.
2. Zakim, David; Boyer, Thomas D. (2002). Hepatology: A Textbook of Liver Disease (4th ed.). ISBN 9780721690513.
3. Rajani R, Melin T, Björnsson E, Broomé U, Sangfelt P, Danielsson A, Gustavsson A, Grip O, Svensson H, Lööf L, Wallerstedt S, Almer SH (Feb 2009). "Budd-Chiari syndrome in Sweden: epidemiology, clinical characteristics and survival – an 18-year experience". Liver International. 29 (2): 253–259. doi:10.1111/j.1478-3231.2008.01838.x. PMID 18694401.
4. Hirschfield, GM; Gershwin, ME (Jan 24, 2013). "The immunobiology and pathophysiology of primary biliary cirrhosis". Annual Review of Pathology. 8: 303–330. doi:10.1146/annurev-pathol-020712-164014. PMID 23347352.
5. Dancygier, Henryk (2010). Clinical Hepatology Principles and Practice of. Springer. pp. 895–. ISBN 978-3-642-04509-7. Retrieved 29 June 2010.
6. Saxena, Romil; Theise, Neil (2004). "Canals of Hering: Recent Insights and Current Knowledge". Seminars in Liver Disease. 24 (1): 43–48. doi:10.1055/s-2004-823100. PMID 15085485.
7. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. ILPD (Indian Liver Patient Dataset) Data Set.
8. S. Karthik A, Priyadarishini, J. Anuradha and B. K. Tripathy (2011). Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types. Advances in Applied Science Research. Journal of Intelligent Systems. 4(1): 9-14.
9. B. Venkata Ramana1, M. S. P. Babu and N. B. Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Systems (IJDMS) vol. 3, no. 2, (2011), pp. 101-104.
10. A. Kumar1, N. Sahu, " Categorization of Liver Disease Using Classification Techniques", International Journal for Research in Applied Science & Engineering Technology (IJRASET) vol. 5 no. 5, (2017).
11. Zaqout, I., & Al-Hanjori, M. (2005). An improved technique for face recognition applications. Information and Learning Science, 119 (9/10), 529-544.
12. Zaqout, I. S. (2012). Printed Arabic Characters Classification Using A Statistical Approach. International Journal of Computers & Technology, 3 (1), 1-5.
13. Zaqout, I. (2019). Diagnosis of skin lesions based on dermoscopic images using image processing techniques. Pattern Recognition-Selected Methods and Applications.
14. Zaqout, I., Zainuddin, R., & Baba, S. (2004). Human face detection in color images. Advances in Complex Systems, 7 (03n04), 369-383.
15. Zaqout, I. S. (2005). An integrated approach for detecting human faces in color images. Fakulti Sains Komputer dan Teknologi Maklumat, Universiti Malaya.
16. Zaqout, I. (2011). A Statistical Approach For Latin Handwritten Digit Recognition. IJACSA Editorial.
17. Zaqout, I. S. (2017). An efficient block-based algorithm for hair removal in dermoscopic images. Компьютерная оптика, 41 (4).
18. Zaqout, I., Zainuddin, R., & Baba, S. (2005). Pixel-based skin color detection technique, Machine Graphics and Vision. 14 (1), 61.
19. Musleh, M. M. (2019). Predicting Blood Donation using Artificial Neural Network.
20. Khalil, A. J. (2019). Blood Donation Prediction using Artificial Neural Network.
21. Jacobson, T., & Segerberg, G. (2019). A Machine Learning-Based Statistical Analysis of Predictors for Spinal Cord Stimulation Success.
22. Naz, H., & Ahuja, S. (2019). Deep learning approach for early detection of diabetes in India.

23. Rabiha, S. G. (2019). Analysis of the Indicator's Performance to Predict Indonesian Teacher Engagement Index (ITEI) using Artificial Neural Networks. Procedia Computer Science, 157, 266-273.

24. Maitra, S., Eshrak, S., Bari, M. A., Al-Sakin, A., Munia, R. H., Akter, N., & Haque, Z. Prediction of Academic Performance Applying NNs: A Focus on Statistical Feature-Shedding and Lifestyle.