International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1, 1027-1050 (Mar-2022)
https://dx.doi.org/10.12785/ijcds/110183

Intelligent Identification of Liver Diseases Based on Incremental Hidden Layer Neurons ANN Model

Panduranga Terlapu[1], Ram Prasad Sadi[2], Ramkishor Pondreti[1] and Chalapathi Tippana[1]

[1]*Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, Srikakulam, A.P, India*
[2]*Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, A.P, India*

Received 5 Feb. 2021, Revised 23 Jan. 2022, Accepted 9 Mar. 2022, Published 31 Mar. 2022

Abstract: The liver is a crucial and big organ in the human body, impacts the digestion system. Due to Liver diseases (LDs), so many deaths are occurred in worldwide that nearly 2 million deaths per year. The main LD complications are cirrhosis that 11th position in universal deaths, and others hepatocellular carcinoma and viral hepatitis that 16th leading position for global deaths. Fortunately, 3.5% of deaths are occurred due to LD. The capability of an ML approach for controlling LD can be identified through their factors, co-factors as well as complications respectively. In this research, we gather the personal and clinical information about1460 individuals with 17 LD feature attributes include diagnosis class attribute from 2018 to 2020 with good questionnaire from north coastal districts of A.P., India hospitals, and reputed clinical centers. We apply machine learning (ML) models like Logistic Regression (LR), SVM with RBF kernel, Naive Bayes (NB), KNN, and Decision Tree (DT or Tree). As per the ML model's analysis, the DT model presents the superior classification accuracy that value is 0.9712 (97.12%) than other experimental ML models for the collected LD dataset. Our proposal model incremental hidden layer (HL) neurons ANN (Artificial Neural Network) solutes LD detection with the highest classification and testing accuracy that the value is 0.999 (99.9%) at the 30 HL neurons.

Keywords: Liver Disease, Machine Learning, ANN, Neural Networks

1. Introduction

The weight of the liver is nearer 1 to 1.5 Kg. It occupies 1.5% to 2.5% of body mass. So, it is the elephant or the largest organ of the human body. It evolves with two types of cells that are non-parenchymal and parenchymal. The parenchymal are designated hepatocytes. The non-parenchymal cells are four distinct types including Liver Macrophages or Kupffer cells, Pit cells or killer cells, fat storing or stellate cells, and Sinusoid lining endothelial cells. Clinically, the liver disorders (LDs) distinguish obstructive (cholestatic), hepatocellular, and compounds of both or mixed. The hepatocellular LD is related to necrosis, viral hepatitis, alcoholic LD, liver injury predominantly, and so on. The cholestatic LD leads to cholestatic LDs, gallstone, alcoholic LDs, inhibition of bile flow, and so on. The mixed pattern LDs are related to the viral hepatitis cholestatic forms, drug-induced LDs, and injury of both hepatocellular, and cholestatic [1]. The main essential functionality of the liver is releasing the toxic elements and systematically digesting food. Most cases of Fatty Liver Disease(FLD) causes are alcohol abuse and viruses. So many LDs are there, but some of the LD cases like cirrhosis is the main cause of LD deaths. 20% to 40% of the population suffered from NAFLD (Non-alcoholic fatty liver disease) in developed and developing countries cause of hepatocellular carcinoma. The clinical and epidemiological studies with electronic records in medicine are very crucial to further studies [2]. Hepatitis A, B and C are the Liver Disorders (LDs) cause of viral infections [3]. In this, Hepatitis A is not dangerous than other hepatitis viral infections [4]. Hepatitis B and C are transmitted one to other cause of viral infections. It infects infected persons to health individuals in several ways that are blood transformation, sexual interactions, body fluids, sharing of reused medical equipment and so on [5], [6]. Due to a lack of proper treatment, more than 1 million people are dying every year from liver hepatitis C virus (HCV) diseases. The shape of the liver is differed as per the liver disease. Kohara et al. (2010) [7] researched on normal and abnormal shapes of the livers utilizing statistical and coefficients models. In this experiment, they choose 9 cirrhosis and 9 normal liver shapes hidden valued data and analyzed with Principal Component Analysis (PCA) model. They classified the liver shape components of first and second with feature vectors. As per analysis, identification of the cirrhosis liver by utilizing liver shape model. LD Progression happens in four stages that are fatty liver (FL), hepatitis, cirrhosis, and liver cancerous or carcinoma. The table 1 shows the details about each stage of LD, causes, and symptoms in detail. A medical and clinical symptomatic cycle attempts

E-mail address: vital2927@gmail.com,reddysadi@gmail.com, ramkishor.pondreti@gmail.com, chalapathit520@gmail.com

TABLE I. Description of Liver Disease (LD) Stages and Symptoms

LD Stage	LD Progression Type	LD Causes	Symptoms
First	Fatty Liver (FL)	Overweight, Insulin resistance, High blood sugar or hyperglycemia, type2 diabetes and High levels of fats	loss of appetite, weight loss, weakness, fatigue, nosebleeds, itchy skin, yellow skin and eyes, web-like clusters of blood vessels, under our skin, abdominal pain abdominal swelling, swelling of your legs breast enlargement in men and confusion [8]
Second	Hepatitis	Hepatitis C virus, Hepatitis B virus, Fatty liver Alcohol-related liver disease, Autoimmune hepatitis	Fatigue, Abdominal discomfort Yellowing of the skin and whites of the eyes, jaundice, An enlarged liver, spider angiomas, Skin rashes, Joint pains and Loss of menstrual periods [9]
Third	Cirrhosis	Chronic alcohol abuse, Chronic viral hepatitis, Fat accumulating in the liver, hemochromatosis, Cystic fibrosis, Wilson's disease, biliary atresia, Alpha-1 antitrypsin deficiency, galactosemia, Genetic digestive disorder, autoimmune hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, Infection, brucellosis, Medications	Fatigue, bruising, Loss of appetite, Nausea, swelling in your legs, edema, Weight loss, Itchy skin, jaundice, ascites, Spiderlike blood vessels on your skin, Redness in the palms of the hands, for women, absent or loss of periods not related to menopause for men, loss of sex drive, breast enlargement, Confusion, drowsiness and slurred speech [10]
Fourth	Liver Cancerous or Carcinoma	Liver cells develop changes (mutations) in their DNA, A cell's DNA material, DNA mutations cause changes in these instructions, a mass of cancerous cells, chronic hepatitis infections	Losing weight without trying, Loss of appetite, Upper abdominal pain, Nausea and vomiting, General weakness and fatigue, Abdominal swelling, jaundice, White, chalky stools [11]

to discover the connection between known hidden patterns among disease records having a place with various classes of clinical information separated from actual assessment, past records, and furthermore clinical tests [12] [13] . Intelligent ML (machine learning) and NN (neural networks) models have assumed a crucial job in LD diagnosis. The main intention of all these algorithms has been to analyze LD data and predict the disease. Datasets have been played a vital role in the analysis and identification of the disease in the medical field. Sometimes, the diagnosis of the disease is complex due to the huge and complex data for analysts. Several earlier studies found that ML techniques offered a broad range of tools, methods, and challenges, etc. to address health care problems. This paper has been focused on the identification and classification of Liver Disease using intelligent statistical ML methods including the Incremental Hidden Layer Neurons ANNs model. Novel Andhra Pradesh Liver Disease (APLD) dataset is worked well with novel proposal methodology for detection of liver diseases. In this, increasing the accuracy relatively increasing neurons in hidden layer of ANN until 30 neurons without over fit problem. It is effective and efficient than other experimental traditional ML models and other mentioned past works. The highlights of this research work were mentioned as follows:

- The Data set is collected with a good questionnaire that personal and clinical values of 1460 individuals from reputed hospitals and clinical centers of North coastal districts of Andhra Pradesh, India.
- Five ML algorithms and Incremental HL (hidden layer) neurons ANN (artificial neural network) models have been considered for the analysis of LD. As per the comparative analysis, the proposal model performance is superior to all other experimental models, and researched relative LD works.

The rest of the sections of this paper areas

- Section 2 describes the background of the work. In this literature survey, we reviewed 122 reputed journal papers related to liver diseases, causes of LD, LD detection, LD related to ML, and NN. Most of the

authors express their research on LD with different datasets and models. In this section, we focused on and described some of the popular research works related to this research work.

- Section 3 represents the proposal model working. In this, we present our proposal model incremental HL neurons of ANN and mathematical description of working ANN's loss function value, Gradient descent, learning capabilities of ANN and performance parameters, and so on.

- Section 4 describes experimental setup in detail. In this, we present experimental setup of dataset and materials and methods.

- Section 5 projects result analysis with Five ML algorithms and Incremental HL neurons ANN models. In this section, we have to cover the experiments for detecting LD with several ML models including the proposal system. We compare the models all performance results in one to others in a systematic way and discuss with other LD related works.

2. Literature Survey

The liver is significant and the biggest inward organ of the human body that performs essential capacities, for example, detoxification of medications, chemicals, protein creation, and blood filtration. ML on LD is a novel topic in this decade. In this section, we have referred and described various research works gathered from reputed journals related to this work. In this, we focused on diagnosis liver diseases using different ML, NN (neural networks) and DL (Deep Learning) models for different LD Datasets. Atabaki et al. (2020) [14] researched on Non-alcoholic FLD with various omics and clinical data. In this study, they analyzed 3,029 adult individuals that were 795 individuals T2D diagnosis and 2,234 individual's multi-omics clinical data. They applied the least absolute (LA) shrinkage and selection (SSO) operator for feature selections and RF models for the classification. As per their observations and comparisons, they find an AUC (Area under the ROC Curve) value of 0.84 for all clinical and omics variables and the AUC value of 0.82 for the clinically accessible variables. They conclude that the combination variables of clinical and omics performance were superior to other experimental variable models. Gatos et al. (2017) [15] analyzed chronic LD (CLD) ultrasound SWE (Shear wave elastography) images on 126 (CLD -70 and control - 56) individuals using the SVM model. The model performed the highest CA of controlled instances to CLD with 87.3%. The specificity and sensitivity values are respectively 93.5% and 81.2%, and the AUC value is 0.87. Yip et al. (2017) [16] predicted NAFLD using ML models on 23 parameterized NAFLD clinical data set. In this analysis, they used 922 screening subjects and four ML models that were LR, RR (ridge regression), decision tree (DT), and AdaBoost. The data is splinted into 70% for training and 30% for validation, and they choose the predictor attributes triglyceride, high-density lipoprotein cholesterol (HDLC), alanine aminotransferase (AATF), HA (hemoglobin A1c), hypertension, and count of a white blood cell. They get 0.87 training AUC and 0.88 validation of AUC values. They concluded, "NAFLD ridge score is a simple and robust reference comparable to existing NAFLD scores to exclude NAFLD patients in epidemiological studies." Rahman et al. (2019) [17] researched on Indian Liver Data Set from UCI ML repository with 9 filed. They applied and compared 6 ML models like NB, KNN, SVM, DT, RF, and LR. In this, they found accuracy values 0.53, 0.62, 0.64, 0.69, 0.74 and 0.75 respectively. They conclude the LR model is superior to other experimental ML algorithms. Khusial et al. (2019) [18] analyzed on 2 to 25 age of NAFLD individuals. In this research, they choose a total of individuals 559 (222 NAFLD and 337 non- NAFLD) diagnosed by MRI or liver biopsy. They also assed Liver enzymes, blood lipids, anthropometrics, and glucose and insulin metabolism were also assessed. RF ML approach was applied to the clinical and metabolomics data sets. The data is split into test and training and applied feature selection and dimension reduction. They concluded that "The highest performing classification model was the random forest, which had an area under the receiver operating characteristic curve (AUROC) of 0.94, the sensitivity of 73%, and specificity of 97% for detecting NAFLD cases." The liver is most used for digestion structure where it is Exocrine Gland impacts on fats and normalized pH values of food using alkaline nature. Some abnormal LDs as hyperbilirubinemia identification is difficult in the early stage. One of the specific ways to diagnose LD is a liver function test. Muruganantham et al. (2020) [19] analyzed LD utilizing (BC) Binary classification that the individual suffered from LD is one set and without LD is the second set. They analyzed using an ensemble-based approach to find accuracy. Liver cancer or carcinoma is diagnosed by CAD (computer aided design) for accurate detection, where cancer [20] tissues are not recognized manually. There are a few factors that cause liver malignancy, for example, liquor, smoking, weight, and so on. Finding liver malignancy isn't simple at the beginning stage. Das et al. (2018) [21] researched liver cancer images, for this they used 225 liver cancer CT images and processed them with the model watershed Gaussian-based deep learning (WGDL) model. They got 99.38% accuracy at 200 epochs with DNN classifier. This work is very useful for the analysts for the diagnosis of LD cancer with CT images. Gogi et al. (2020) [22] reviewed so many papers related to LD, especially Liver cancer or Hepatocellular carcinoma(HCC) predictions and scenarios. In this, they analyzed various papers related to HCC that were clinical trial, tumor grading, laboratory and imaging studies in various research works. Pruthvi et al., (2017) [23] reviewed liver cancer images with ML model research works. In this research, they reviewed different methodologies and models of ML with liver cancer CT scan and MRI images. Moreover, explained problems with medical diagnosis systems and solutions with different ML

and NN algorithms, and compared every analysis and model with different works. Ksia et al. (2018) [24] studied liver cancer with the HCC dataset. In this study, they used 165 patients' data with 49 feature attributes, moreover, they focused on life and death categories of the HCC dataset that are 102 live patients and 63 dead patients that cause liver cancer. They used 10 ML models with/without feature selection on the HCC dataset. The GA algorithm was coupled with 5-fold cross-validation method was performed two times. The GA was used in parallel with the feature selection algorithm and classifier parameter optimization. The proposed model achieved the best accuracy and F1-Score values of 0.8849 and 0.8762 respectively. Naeem et al. (2020) [25] studied MRI and CT images of liver cancer using hybrid feature analysis ML models. In this study, they analyzed 200 (MRI-100 and CT-scan-100) 512 X 512 sized liver cancer images with 10 optimized features that selected by feature selection algorithms. Furthermore, applied 4 ML models like MLP, SVM, RF and J48 utilizing 10-fold validation. In this, they achieved in MLP more accuracy values that were in MRI images 95.78% and in CT images 97.44%. Rajeswari et al. (2010) [26] analyzed LD utilizing DM algorithms on UCI repository LD dataset that contain 7 attributes and 345 instances. In this research, they described causes of LD, symptoms of LD, and types of LD and more over LD with ML analysis. The experiment was computed with ML algorithms like K-star, Naïve Bayes and FT Tree using WEKA tool. As per comparison and findings, NB, FT Tree, and K-Star models' accuracy and time value are 96.52% in 0 sec, 97.10% in 0.2 sec, 83.47% in 0 sec respectively. Akyol et al. (2017) [27] researched on attribute importance of LD datasets, and balanced and unbalanced liver datasets acquired from UCI repository that are ILDP and BUPA. The study showed that the balanced dataset was very accurate than the unbalanced dataset. The accuracy values of BUPA and ILPD unbalanced 5 sub-datasets average values were 71.59% and 71.9%. The accuracy values of BUPA and ILPD balanced 5 sub-datasets average values were 77.24% and 74.85%. Khan et al. (2019) [28] reviewed and analyzed a strategic analysis on LD predictions using various classification algorithms and they found the RF algorithm gave a good accuracy value in so many reviewed researches works on LD. LD infections one of the significant illnesses in the world, Liver is one of the gigantic strong organs in the human body; and is additionally viewed as an organ in light of the fact that among its numerous capacities, it makes and secretes bile. Kefelegn et al. (2018) [29] reviewed on analysis and predictions of LDs utilizing DM techniques. In this systematic review research, they reviewed huge research works related to ML with LD. As per review analysis, the working of different classifiers mainly K-NN, SVM, C4.5, NBC and RF techniques on different LD datasets and performances were explained. The back-propagation ANN is a multi-layered NN organization method discovered by Rumelhart and McClelland. It works by randomizing loads of weights to the different layers relating to the input. The loss function is described as error values within an output and calculate this with gradient loss values. Bahramirad et al. (2013) [30] reviewed several investigations on different UCI LD datasets using DM models in deeper ways. In this study, they had implemented 11 DM models to the various LD datasets and compared the accuracy, recall, and precision values to each other. The table 2 describes about various research scenarios, datasets, models and results represented by different researchers in different years. The detailed analysis shown in the table 2.

3. Proposal Model and Methodologies

In this section, we describe about proposal model work flow in detailed and ANN with back propagation working model and network design. In other hand, we also describe about confusion matrix, performance parameters and loss function values and gradient decent values.

A. *Proposal Model*

In the Figure 1that explains the proposal model of the liver disease (LD) prediction using incremental hidden layer neurons of the back propagation ANN Algorithm. For this, the LD and non-LD data is collected from the North coastal districts (Vizainagaram, Visakhapatnam and Srikakulam) of state A.P., India and stored as *.csv and *.mat formats. The gathering data is spited in two parts of each patient that are personal and clinical records. The preprocessed dataset is input to the back propagation ANN model. The ANN model is trained by the data set in two stages that initially the algorithm decides the network structure of the system and in second section decides the network weights and smoothing parameters. The main aim of the experiment is to classify LD and non-LD. The number of neurons is decided in the input layer by utilizing the feature input vector dimensions. In this problem, there are 16 feature dimensions and one class attribute involved in feature vector that are gender, age, smoke, drink, of the patient, age and remaining dimensions are LD clinical parameters like TB, DB, TP, ALKP, ALAT, ASPAT, Albumin, and AG-Ratio.

B. *ANN (Back Propagation) Model*

The figure 2 shows the ANN model for LD analysis with back propagation. The inputs of the network are X1, X2 ...X16 is the features for LD detections with a target class (Yes or No). The neural network (NN) is composed of three layers that are Input, Hidden, and Output, and each of these layers is made of neurons. The neural training set is established with input and output-based pairs using feature values. Especially, NNs performs this mapping by processing the input through a set of transformations. As per our experiment, the hidden layer neurons are increased 5 at a time in each step, and it can continue until the peak performance goal was reached. In this process, the input or feature or evolutionary values are transformed through the HL then output is predicted at the output layer.

These transformations are depended on the weight (W) and bias (B) values. In mean training time, the network learns and needs to change the weights for minimizing the

TABLE II. Descriptions of Different Research Works with different data sets and models on Liver Diseases (LDs)

Ref. No.	Author	Contribution and Area	Results	Year
[31]	Sontakke et al.,	Diagnosing Chronic Liver Disease (HCC with HCV-related) using ML models dataset contains 4423 CHC patient's clinical values.	Accuracy, Precision Sensitivity, Specificity Values of SVM: 71%, 64.1%, 71.5%, 88.3%, andANN (MLP): 73.2%, 65.7, 73.3, 87.7 respectively.	2017
[32]	Xu et al.	LD identification utilizing LMBP neural network, rough set theory (RS) and hybrid model RS-LMBPNN	Predicting accuracy of LMBPNN-90% and RS-LMBPNN- 96.67%	2016
[33]	Hassan et al.	Diagnosis of Focal LDs SoftMax layer classi[U+FB01]er for Ultrasound Images compared with SVM, KNN and NB	Accuracy of resulting values of Multi-SVM 96.5%±0.019 KNN- 93.6%±0.022Naïve Bayes 95.2%±0.016SoftMax layer classi[U+FB01]er 97.2%±0.023	2017
[34]	Abdar et al.,	Diagnosis of Liver Disease Using MLP NN and Boosted DTs and used UCI Dataset (ILPD)	Accuracy of MLPNNB-C5.0 is 94.12 MLPNNB-CHAID is 79.34 MLPNNB-CART is 79.69	2017
[35]	Özyurtet al.,	The study uses CT images of 41 benign and 34 malign samples Hash-Based CNN is superior than ANN SVM KNN classification	Classification LD ANN SVM KNN Hash-Based CNN 89.3%, 83.9%, 83.9% and98.2%	2018
[36]	Singh et al.,	LD prediction with ML models like K-NN, LR and SVM and comparative analysis	K-NN model Accuracy -73.97%. Sensitivity 0.904 and specificity 0.317LR Model Accuracy - 73.97%. Sensitivity 0.952 and specificity 0.195 SVM Model Accuracy - 71.97% Sensitivity-0.952 and specificity-0.195.	2018
[37]	Auxilia et al.,	LD prediction using ML on ILPD UCI repository Dataset, furthermore apply ML models like DT, NB, RF, SVM and ANN.	Accuracy of the algorithms DT, NB, RF, SVM and ANN values are 81%, 37%, 77%, 77% and 71% respectively.	2018
[38]	Reddy et al.,	Predicted Fatty LD using Ultrasound Imaging dataset with Deep Learning like CNN, VGG16 + Transfer Learning 87.5% VGG16 Transfer Learning + Fine Tuning 90.6%	Classifier Accuracy in (%) CNN 84.3% VGG16 + Transfer Learning 87.5% VGG16 Transfer Learning + Fine Tuning 90.6%	2018
[39]	Srivenkatesh et al.,	LD detection with ML models like K-NN, Support Vector Machines, Logistic Regression, Naive Bayes, Random Forest with MSE value	MSE values of ML models K-NN - 0.55 SVM-0.53 LR-0.48 NB - 0.70 RF- 0.50.	2019
[40]	Ramaiah et al.,	Diagnosing Chronic Liver Disease using 1583 instances Dataset that collected from UCI ML repository	J48 64.37% NB 57.23% 42.77 REP Tree 66.27% and RF 100%.	2019
[41]	Singh et al.,	LD prediction using ML. the feature selections attributes were collected from ILPD UCI repository Dataset and moreover applied ML models like NB, IBK, and RF on optimized datasets, and discussed comparative analysis among the ML techniques.	Accuracy values of NB- 55.74 SMO- 71.35 IBK- 64.15 J48-68.78 RF- 71.53	2020
[42]	Mai et al.,	Liver Cirrhosis Diagnosis using ANN model and analyzed 1152 HBV-related HCC patients AUC values compared with LR model	Performance ANN and LR AUC values are 0.757 and 0.721 respectively	2020
[43]	Kuzhippallil et al.,	Comparative Analysis of ML Models on ILD Data set collected from UCI ML repository	Accuracies of MLP - 0.71, KNN-0.72, LR-0.74, RF-0.74 GB-0.66 AdaBoost 0.68 XG-Boost-0.70 Light GBM-0.70 Stacking Estimator-0.83	2020

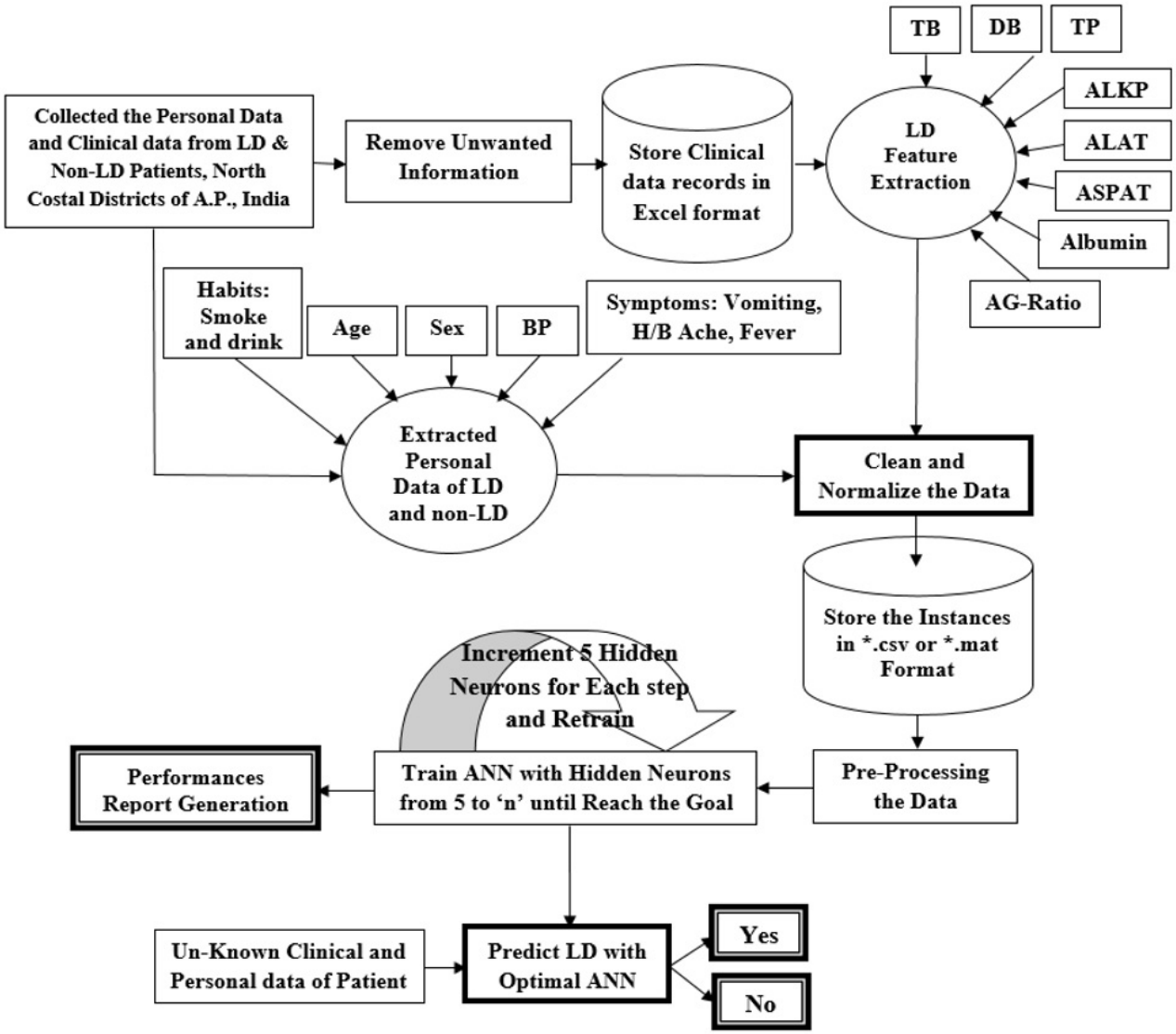

Figure 1. Proposed model for Liver Disease Detection using Incremental hidden layer neurons of the ANN

loss function value (L) or error value between the output and target values. The weights are updated using gradient descent (GD) optimization function at each epoch.

$$W^{(n)} = W^{(n-1)} + \varepsilon \frac{\partial L}{\partial W} or W^{(n+1)} = W^{(n)} + \varepsilon \frac{\partial L}{\partial W} \quad (1)$$

Where W represents the weight value, n indicates the nth weighted value, epsilon (ε) indicates the learning rate, and L is the loss function value (or Error). (∂L) / (∂W) is gradient that measured weight to loss. If this value is larger than it indicates that the weighted value is updated more and more during gradient decent iteration. The general format of activation of a neuron with one feature is calculated using the Eq 2

$$A = X * W + B \quad (2)$$

That the previous layer output value X is multiplied by W (weigh), and added with bias (B). If the networks have more feature values then the neuron activation value is calculated as below Eq 3. This equation is a linear operation.

$$A_i = \sum_{j=1}^{n} X_{ij} W_j + B_i \quad (3)$$

Where, i = 1, 2, 3 ... m. By above linear operation output value is the input of the activation function . The activation function is a sigmoid function for complicated tasks that represented as follow the4.

$$\sigma(x) = \frac{1}{1 + exp(-1)} \quad (4)$$

So, we can write the consequence output layer neurons

computing Yi value as below 5 that

$$Y_i = \sigma(A_i) = \sigma\left(\sum_{j=1}^{n} X_{ij}W_j + B_i\right) \quad (5)$$

By above general equations are implemented to the hidden layer neurons H and output layer neurons Y shown in 6 and 7.

$$H = \sigma\left(XW^1 + B_0\right) \quad (6)$$

$$Y = \sigma\left(XW^2 + B_1\right) \quad (7)$$

X is the input vector space, w^1 and w^2 are weighted values of Hidden and Output layer neurons. B0 and B1 are the bias values between 0 and 1 of Hidden and Output layer neurons respectively. We will compute the loss or error value between actual value (target value) T and output value Y. So, the below Eq 8 prompts the error value in mean squared (L).

$$L = \frac{1}{2}(Y - T)^2 \quad (8)$$

The partial differentiation of the Eq **??** with respect to HL weights W2 then we get the solution Eq 9

$$\frac{\sigma L}{\sigma W^2} = \frac{1}{2}\left[\sigma(\frac{[Y-T]^2}{\sigma W^2})\right] = \frac{1}{2}\left[\frac{\sigma(Y^2)}{\sigma(W^2)} - 2T\frac{\sigma(Y)}{\sigma(W^2)}\right] \quad (9)$$

As per equation (7) $Y = \sigma\left(XW^2 + B_1\right)$ substitute in equation (9) then we get the result equation (10) as

$$\frac{\sigma L}{\sigma W^2} = H\left[\frac{exp(-HW^2 - B_1)}{(1 + exp(-HW^2 - B_1))^2}\right]\left[\frac{1}{1 + exp(-HW^2 - B_1))^2}\right] \quad (10)$$

C. Back-Propagation and Gradient Decent (GD) Analysis:

The figure 3 shows the computational NN with one HL that computes the loss function value L. The X represents the inputs, H defines the hidden layer, Y describes the output, and L is loss or error value calculated using equation (8). T describes the target class values or actual values. W1 and W2are weight values propagates the HL and OL. B0 and B1 are bias values propagates the HL and OL respectively. The neurons in this network according to all the computed values for getting the L value. In this process, one variable value is depended on other variable computation that it follows the chain rule of calculus. The bias values and weights are very crucial to compute the gradient of L. In back propagation, we can update the weights and biases values. The equation (11) specifies the partial derivation with respect to W2 weights. It defines the loss value at OL.

$$\frac{\sigma L}{\sigma W^2} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma W^2} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma W^2} \quad (11)$$

The equation (12) specifies the partial derivation with respect to W1 weights. It defines the loss value at HL with chain rule derivation.Like derivation with respect to weights, we can compute the loss function value is computed with respect to bias values. Equations (13) and (14)Instated of squared error, we can apply cross entropy (CE) function. The CE loss function value is computed as following equations(12,13,14,15)

$$\frac{\sigma L}{\sigma W^1} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma W^1} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma W^1}$$

$$= \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma H}\frac{\sigma H}{\sigma W^1}$$

$$= \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma H}\frac{\sigma H}{\sigma A^1}\frac{\sigma A^1}{\sigma W^1} \quad (12)$$

$$\frac{\sigma L}{\sigma B_1} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma B_1} = \frac{\sigma L}{\sigma Y}\frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma B_1} \quad (13)$$

$$\frac{\sigma L}{\sigma Y} = \frac{\sigma Y}{\sigma B_0}\frac{\sigma L}{\sigma Y} = \frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma B_0}\frac{\sigma L}{\sigma Y}$$

$$= \frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma H}\frac{\sigma H}{\sigma B_0}\frac{\sigma L}{\sigma Y} = \frac{\sigma Y}{\sigma A_2}\frac{\sigma A_2}{\sigma H}\frac{\sigma H}{\sigma A_1}\frac{\sigma A_1}{\sigma B_0} \quad (14)$$

$$L(Y,T) = \frac{1}{n\left(\sum_{i=1}^{n} -T^{(i)}logY^{(i)}\right)} - \left(1 - T^{(i)}\right)log\left(1 - Y^{(i)}\right) \quad (15)$$

4. Setup and Performance Analysis

In this section, we describe about Andhra Pradesh Liver Disease (APLD) data-set description in detail. As well as, describe about confusion matrix.

A. Description of Liver Data-set:

The data is collected from north coastal districts of A.P., India from reputed clinical organizations and patients. The table describes each feature and class attribute of the experiment. In this, there are 16 feature attributes (X1 to X16) and class attribute C1 involved with 1460 records (585 Non-LD records and 875 LD records). Table 3 depicts the description of LD data-set in detail.

B. Confusion Matrix:

Confusion Matrix is a performance measurement for classification of ML problems. In this, we represent the confusion matrix for the Liver Disease (LD). The table-IV demonstrates LD confusion matrix with Table 4 distinct combining of Predicted and Actual values and real qualities [44].

C. Performance Parameters:

We need to determine the performance parameters like TPR,-Recall or sensitivity, SPC-Specificity, False Negative Rate (FNR), Miss Rate, FPR, True Negative Rate (TNR), Positive Predictive Value, Precision, ACC-Accuracy, FDR-False Discovery Rate, NPV-Negative Predictive Value, DOR-Diagnostic Odds Ratio, and F1-Score. Some of the performance parameters are describe through the equations

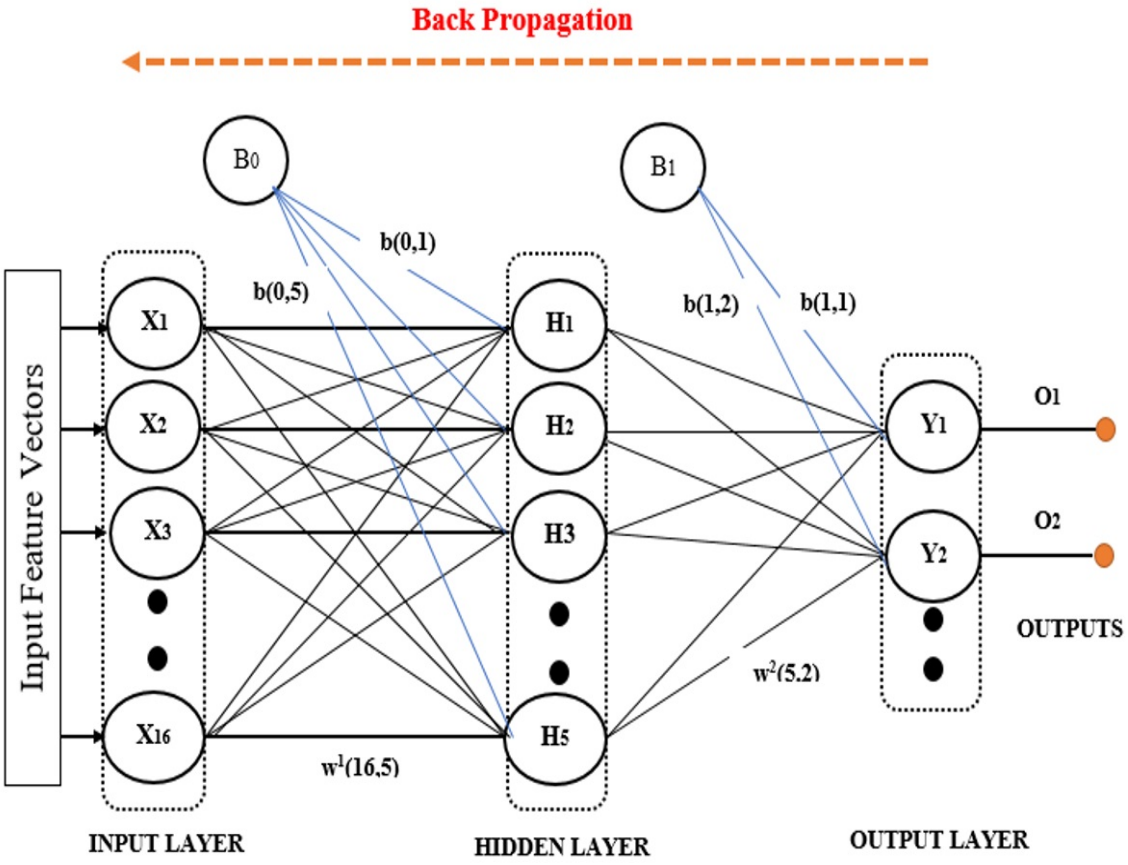

Figure 2. ANN Back Propagation model for Liver Disease Detection

(16,17.18,19,20,21,22,23).

$$Accuracy(ACC) = \frac{\sum TruePositive + \sum TrueNegative}{\sum TotalPopulation} \quad (16)$$

$$TPR = \frac{\sum TruePositive}{\sum ConditionPositive} \quad (17)$$

$$FNR = \frac{\sum TrueNegative}{\sum ConditionPositive} \quad (18)$$

$$FPR = \frac{\sum FalsePositive}{\sum ConditionNegative} \quad (19)$$

$$F_1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (20)$$

$$SPCorTNR = \frac{\sum TrueNegative}{\sum ConditionNegative} \quad (21)$$

$$Prevalence = \frac{\sum ConditionPositive}{\sum TotalPopulation} \quad (22)$$

$$PPVorPRC = \frac{\sum TruePositive}{\sum PredictedConditionPositive} \quad (23)$$

5. Simulation Results and Discussion

Figure 4 shows connection Liver properties are checking out the values between - 1 and +1, and colors (Red and Blue). According to relationship esteems, the heist esteem one determines exceptionally correlated attributes (demonstrated dull red), and fewer qualities are announced as under associated attributes (showed blue tone). Correlated attribute esteems are zero or very closer to 0 (determined tone is white). According to the analysis, self- attribute connection esteems are one that the tone is dull red indicated in figure 4. The X1 and X8 (age and BP) attributes are correlated with the value of 0.64. As well, the clinical feature attributes X9 and X10 (TB and DB) are correlated with the features (AP and AA) X11 and X12 (correlated values are 0.49, 0.51, 0.52, and 0.54). The correlation values are 0.74 for the features X11 and X12, 0.47 for the X14 and X15, and 0.61 for the X15 and X16 attributes.

A. ML Algorithms Analysis:

The figure 5 represents the confusion matrices of ML algorithms. In this, the zero indicates the non-LD and one represents the LD instances. All ML models are efficient that AUC values are greater than 0.95 and classification accuracy (CA) values are greater than 78%. The k-NN model is configured with five neighbors, the distance metric

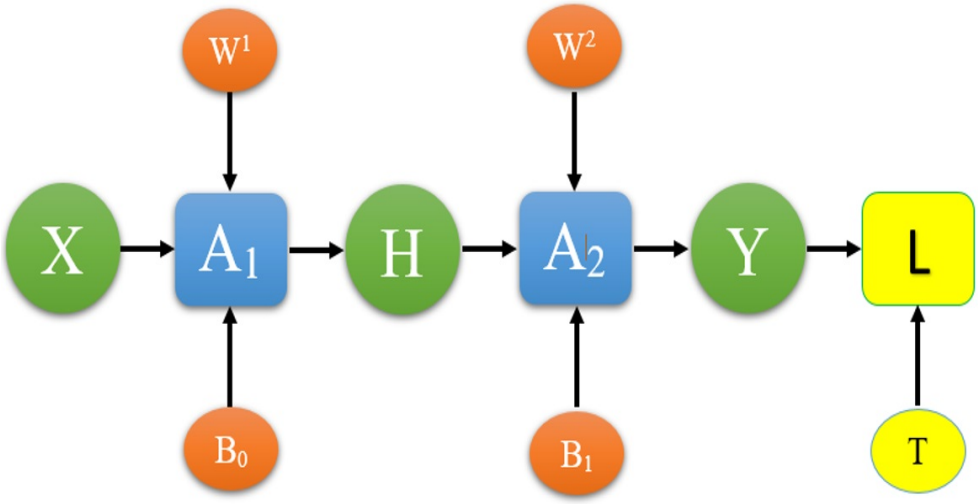

Figure 3. Back Propagation General Architecture for Loss or Error calculations

TABLE III. Descriptions of Different Research Works with different data-sets and models on Liver Diseases (LDs)

Feature and class attributes	Description	Type
Age (X1)	Age 6 to 99	Numeric
Gender (X2)	Patient Gender Male 1Female – 0	Categorical
Smoke (X3)	Has smoking habit or not (YES-1NO-0)	Categorical
Drink (X4)	Has Dirking habit or not, (YES-1NO-0)	Categorical
Vomiting(X5)	Any Symptom of Vomiting, (Present-1Absent-0)	Categorical
Headache/Bone Ache (X6)	Any Symptom of Headache(Present-1Absent-0)	Categorical
Fever (X7)	Any Symptom of Fever, (Present-1Absent-0)	Categorical
BP(X8)	Blood Pressure, Normal-0 Low-1 High-2	Categorical
Total (TB) Bilirubin (X9)	TotalBilirubin (0.4 to 75)	Integral
Direct(DB)Bilirubin (X10)	DirectBilirubin range is 0.1 to 19.7	Integral
Alkaline(AP)Phosphatase (X11)	Alkaline Phosphates range is 10 to 4929	Numeric
Alanine(AAT)Aminotransferase (X12)	Alanine Aminotransferase –range 10 to 2000	Numeric
Aspartate(ASAT)Aminotransferase (X13)	Aspartate Aminotransferase –range 5 to 4929	Numeric
Total-Proteins(X14)	Total Proteins –range is 0.9 to 7.7	Integral
Albumin (X15)	Albumin-range –range is0.9 to 7.7	Integral
A-GRatio (X16)	Albumin andGlobulin Ratio –range is 0.3 to 4.0	Integral
Diagnosis (Class) (C1)	Non-Liver Disease (Class 0) and Liver Disease (Class1)	Categorical Class

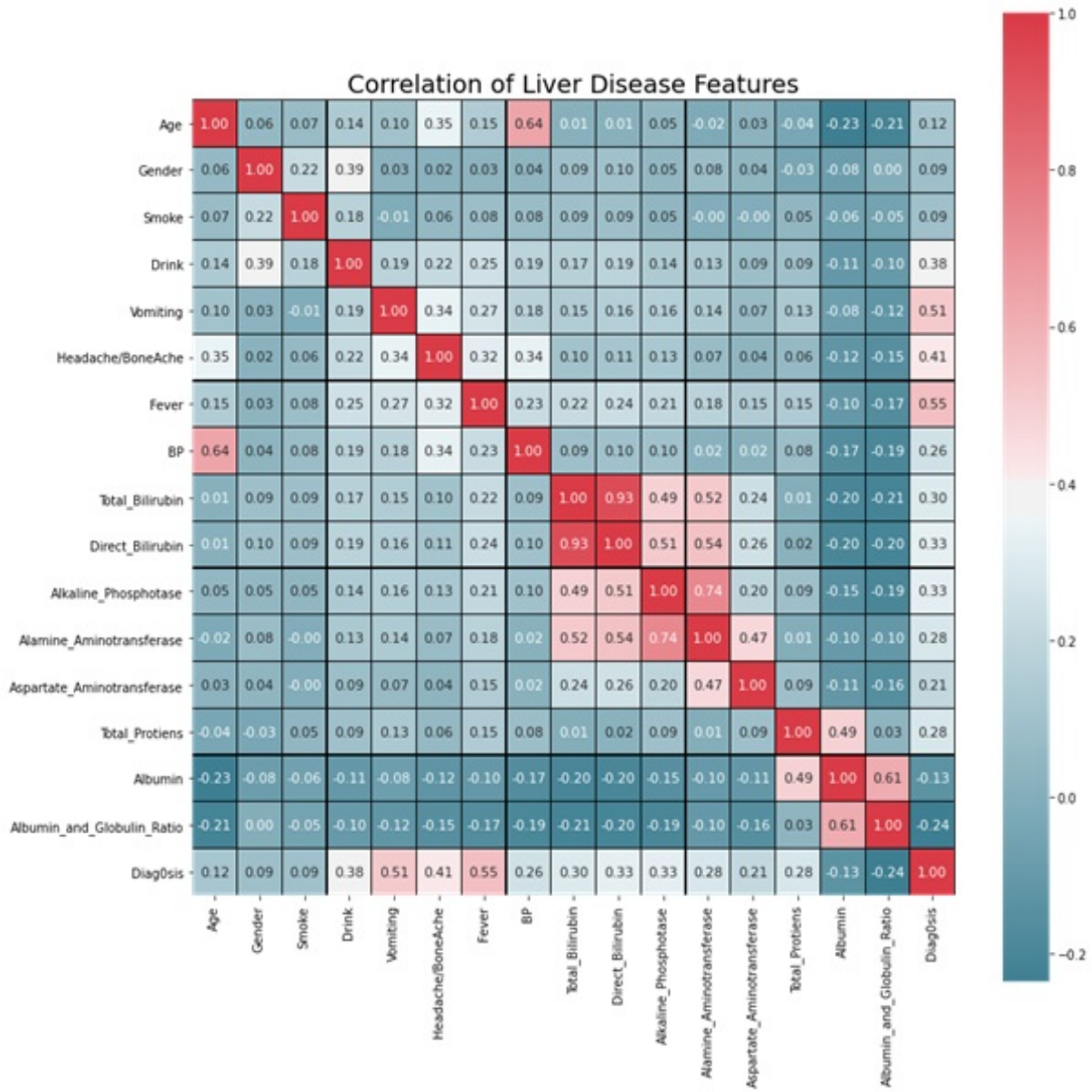

Figure 4. Correlation coefficients for feature attributes of A.P. Liver Data-set (APLD)

is Euclidean and weight measures are in a uniform. Figure 5 (A) is analyzed k-NN model with confusion matrix. As per analysis, the k-NN ML model classifies 480 instances correctly and 106 instances classify incorrectly out of 586 total instances in class 0, as well as 81 instances are classified incorrectly and 793 instances are classified correctly out of 874 instances of class 1(LD). So, 1273 (480 +793)class 0(non-LD) and class 1 (LD)) instances are classified correctly out of 1460 imbalanced classes of instances. The accuracy of the k-NN is 0.8719 (87.2%) and recall values is superior then precision value. The class one (LD) accuracy is superior with 0.8191 (82%) accuracy than the class zero (non-LD) accuracy (0.907322(91%)). The Tree ML model is constructed with parameters that are induce binary tree, minimum number of instances in leaves 2, do not split subsets smaller than 5, and limit the maximal tree depth to 20. The Tree ML model is superior to other experimental ML models with CA (0.9712) and AUC (0.9712) values. Figure 5(B) is analyzed Tree model with confusion matrix. The Tree ML model classifies 566 instances correctly and 20 instances classify incorrectly out of 586 total instances in class 0, as well as 22 instances are classified incorrectly and 852 instances are classified correctly out of 874 instances of class 1 (LD). So, 1418 (566 +852 class 0(non-LD) and class 1 (LD)) instances areclassified correctly out of 1460 imbalanced classes of

TABLE IV. Analysis of Confusion Matrix Structure

Confusion Matrix

		Predicted values	
Actual Values	Classes	Non-Liver Disease (0)	Liver Disease (1)
	Non-Liver Disease (0)	(0, 0)	(0,1)
	Liver Disease (1)	(1,0)	(1,1)

instances. The SVM model is configured as the cost value is 1.00, regression loss epsilon is 0.10, tolerance value is 0.0010, iteration limit is 100, and kernel is RBF. Gaussian RBF is familiar and efficient kernel used in SVM. The RBF is calculated with $K(a1, a2) = exp(-gamma||a1-a2||)$ where $||a1 - a2||$ is Euclidean distance between $a1$ and $a2$, and gamma value is 0.01. The SVM model, AUC and accuracy values are 0.9530 and 0.7828 respectively. It performs least than other experimental MLs that it classifies correctly only 1143 out of 1460. Figure 5(C) is analyzed SVM model with confusion matrix. The class 1 (LD) classification performance (0.988558352) is higher than class 0 (Non-LD) CA values (0.476109215). So, it is very accurate model for predict only diseased individuals. Figure 5(D) is analyzed NB model with confusion matrix. As per analysis of the NB ML model classifies 517instances correctly and 69 instances classify incorrectly out of 586 total instances in class 0, as well as 95 instances are classified incorrectly and 779 instances are classified correctly out of 874 instances of class 1 (LD). So, 1296 (517 + 779) class 0(non-LD) and class 1 (LD)) instances are classified correctly out of 1460 imbalanced classes of instances. The accuracy of the NB is 0.0.8876 (88.8%) and recall values is superior then precision value. The class one (LD) accuracy is superior with 0.8868 (89%) accuracy than the class zero (non-LD) accuracy (0.8817 (88%)). The Logistic Regression (LR) model, AUC and accuracy values are 0.9675 and 0.9130 respectively. It performs moderately compare to other experimental MLs that it classifies correctly 1296 out of 1460 total instances. Figure 5(E) is analyzed LR model with confusion matrix. The class 1 (LD) classification performance (0.92791762) is higher than class 0 (Non-LD) CA values (0.890784983). The table 5shows the performance parameters AUC, CA, F1, Precision and Recall values. In this analysis the Tree model is superior to other experimental ML algorithms with all performance parameters.

B. ROC Analysis of ML models:

The ROC curves of ML models are constructed in two dimensional plots that x-axis specifies the FP – Rate or specificity and y-axis indicates the sensitivity or TP Rate. The ML models k-NN, Tree, SVM, Naive Bayes, and Logistic Regression ROC curves are specified with individual colour shown in Figure 6. The k-NN represented with green colour and AUC values is 0.9524. The ROC curve of Tree is specified with violet colour and AUC value is 0.9887. The SVM represented with orange colour and AUC value is 0.9524. The ROC curve of NB model is specified with brown and AUC value is 0.9596. The LR model AUC value is 0.9675 and the ROC is represented with green colour. The Figure 6 shows the comparative analysis of experimental ML models. The blue colour bars represented the AUC values and the brown colour bars represented the classification accuracy values. As per analysis, the superior values of AUC and CA are 0.9887 and 0.9712 in the model Tree. The minor values of AUC and CA are 0.9524 and 0.7828 that are in models k-NN and SVM respectively.

C. ML Comparative Analysis using AUC and CA Values:

The Figure 7 shows the combative analysis of experimental ML models. In this the decision tree model performs well with 0.9887 and 0.9712 of AUC and CA values respectively. All models' performances are good with the AUC values are above 0.95, and CA values are above 0.78.

D. Incremental HL neurons ANN Model evolutions

We analyze the incremental hidden neurons of the hidden layer (HL) ANN model. We start with Five neurons hidden layer to analyzing the ANN model and increment the five neurons in each iteration step until getting the peak performance without the over fitting the problem of the ANN model. In this process, the ANN set up is that the data division is random, use the Liebenberg-Marquardt training algorithm, and the performance is computed using mean squared error.

1) Confusion Matrix Analysis

The Figure 8 shows the confusion matrices of 5 to 30 hidden layer (HL) neurons ANN models. The confusion matrix is built using TP, TN, FP, and FN values as per classifications of target classes and output classes (LD (class 1) and Non-LD (class 2)). Class 1 specifies the Liver Disease instances, and class 2 defines the Non- Liver Disease instances. Figure 8(A) shows the 5 HL neurons confusion matrix analysis that the total accuracy is 92.1%, class 1 classifies 829 instances out of 874 with 94.9% accuracy, and class 2 classifies 515 instances out of 586 with 87.9% accuracy. As per the analysis, the recall value (0.920377) is superior to the precision value (0.913676). Figure 8(F) shows the 30 HL neurons confusion matrix analysis that the total accuracy is 99.9%, the class 1 accuracy is 100% that class, and the accuracy of class 2 is 99.8% classifies 585 instances out of 586. The recall and precision values are 0.999147 and 0.992009 relatively.

The table 6 shows performance parameters like AUC, CA, F1, and the precision, and the recall analyzing values. As per observations, the performance values are increased propositionally increasing of HL neurons of ANN. At the position 30 HL neuron, the accuracy value is in the pea that the value is 0.999315. As per observations, above thirty neurons of HL ANN were performed with over fitting problems. So, we stopped at 30 HL neurons of ANN for

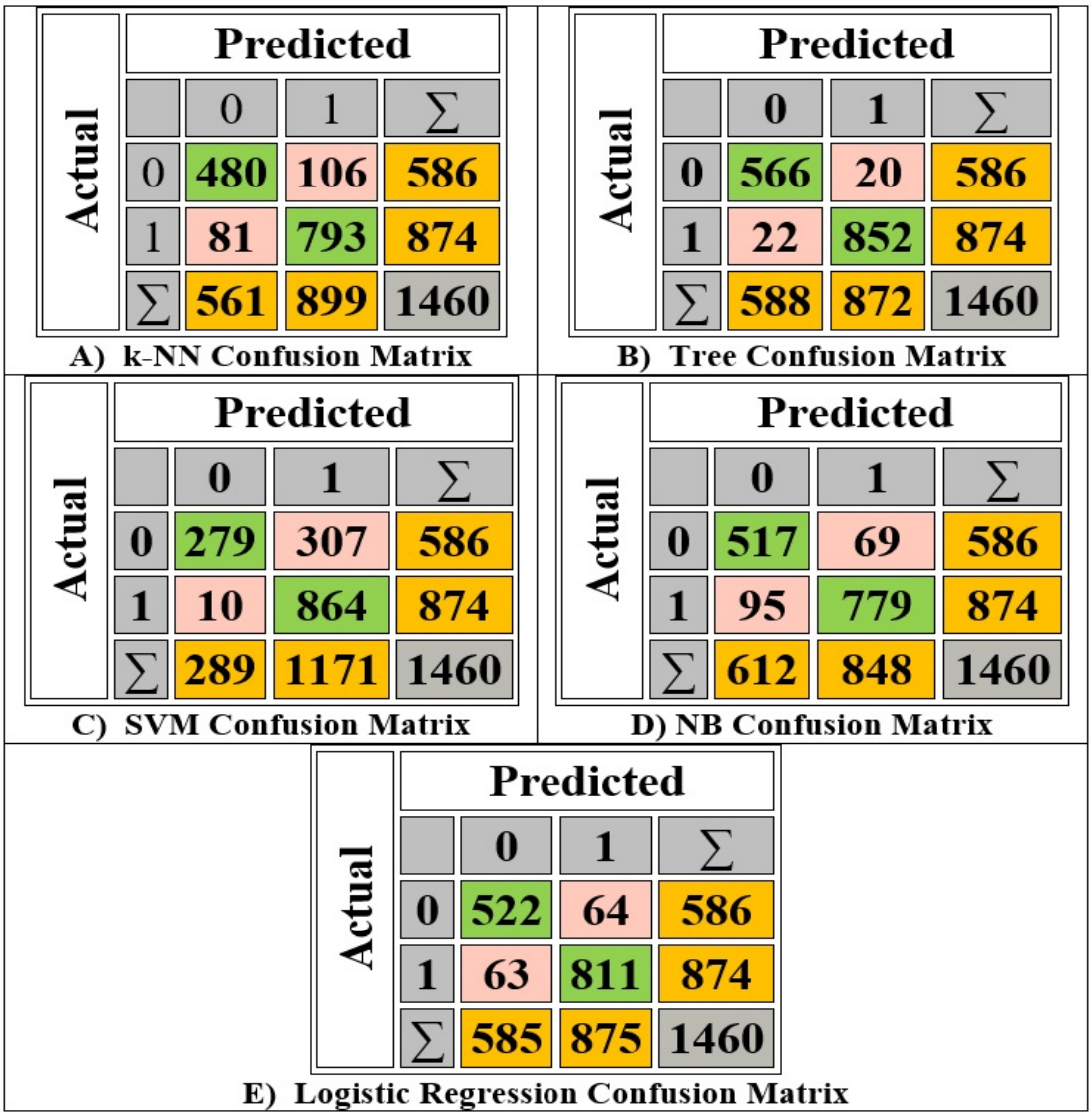

Figure 5. Confusion matrices of Experimental ML models

predicting Liver Diseases. The highlighted figures in the table show the high-performance values.

2) ROC and AUC Curves Analysis

The Figure 9 shows the ROC curves of 5 to 30 hidden layer (HL) neurons ANN models. The ROC curve is created between false positive (FP) rate values (zero to one) on the X-axis, and true positive (TP) rate values (zero to one) on the Y-axis as per target classes and output classes (LD (1) and Non-LD (2)). Class 1 determines the Liver Disease instances, and class 2 defines the Non- Liver Disease instances. The blue colour curve specifies the class 1 (LD) and the green colour curve indicates the class 2 (Non-LD). Figure 9(A) shows the 5 HL neurons ROC curves analysis that the total AUC is 0.953676 that the class 1 AUC value is 0.947617, and the class 2 AUC value is 0.959676. As per observations, class 1 AUC is superior to class 2. Figure 9(B) shows the 10 HL neurons ROC curve analysis that the total accuracy is 0.968492 that the class 1 AUC value is 0.969912, and the class 2 AUC value is 0.966492. Figure 9(C) shows the 15 HL neurons ROC curves analysis that the total AUC is 0.974551 that the class 1 AUC value is 0.976998, and the class 2 AUC value is 0.976998. As per the investigation, the class 1 AUC is superior to class 2. Figure 9(D) shows the 20 HL neurons ROC curve analysis that the total accuracy is 0.980369 that the class 1 AUC value is 0.980839, and the class 2 AUC value is 0.979901. Figure 9(E) shows the 25 HL neurons ROC curves analysis that the total AUC is 0.991493 that the class 1 AUC value is 0.992493, and the class 2 AUC value is 0.990493. Figure 9(F) shows the 30 HL neurons ROC curve analysis that the total accuracy is 0.968492 that the class 1 AUC value is

TABLE V. Experimental ML models performance parameters values

Model	AUC	CA	F1	Precision	Recall
k-NN	0.9524	0.8719	0.866	0.8633	0.8688
Tree	**0.9887**	**0.9712**	**0.9701**	**0.9699**	**0.9704**
SVM	0.953	0.7828	0.7874	0.7323	0.7828
Naive Bayes	0.9596	0.8876	0.8842	0.8817	0.8868
Logistic Regression	0.9675	0.913	0.9094	0.9095	0.9094

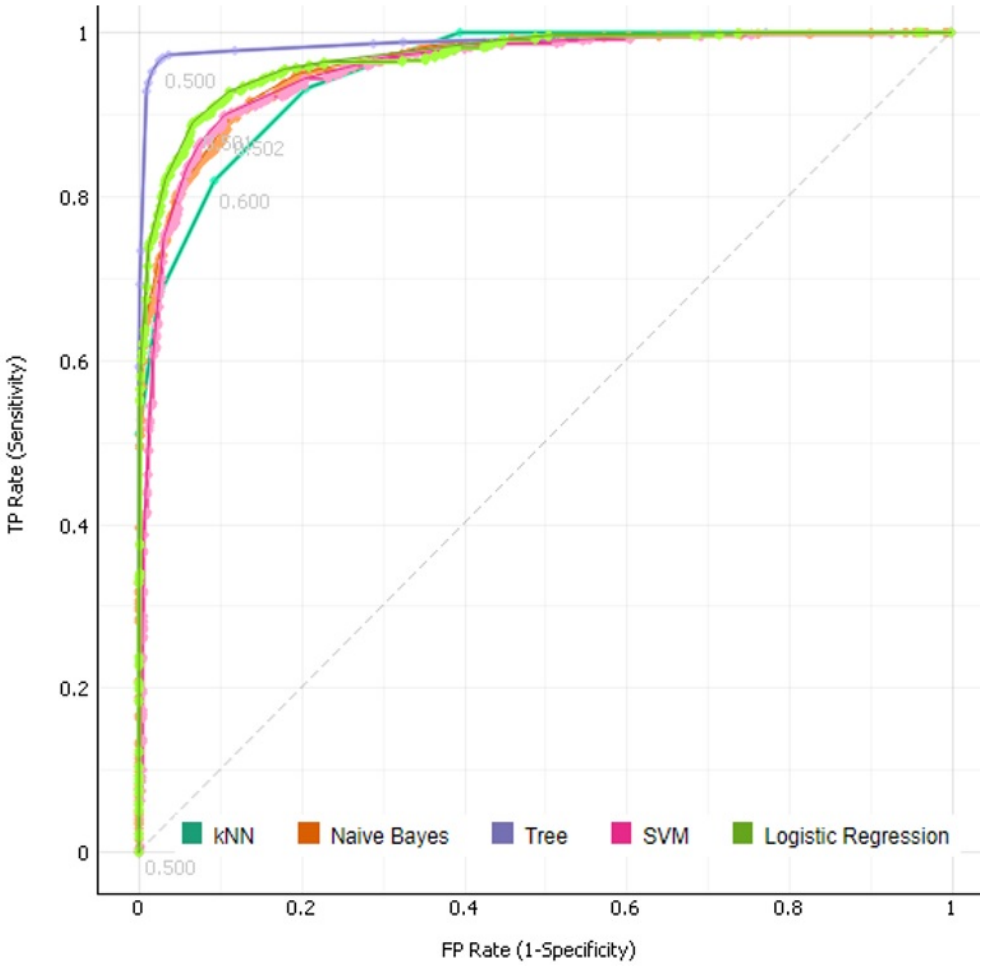

Figure 6. Analysis of Experimental ML models Performances using ROC curves

one, and the class 2 AUC value is also one.

3) Regression (R) Value Analysis:

The Figure 10 shows the Regression (R) analysis values of 5 to 30 hidden layer (HL) neurons ANN models. The training R value is calculated using target values and output values that it describes about data set fitness value. The target values are lie between 0 and 1. The blue colour line indicates the data fit line, the dotted line indicates peak fitted line that the output data values are equal to target values (Y=T). The circle symbols describe classified data points. Figure 10(A) shows the ANN model with 5 HL neurons regression analysis that the total R value is 0.88581. Most of data points class 1 and class 2 (LDs and non-LDs) are fitted according to output. The output is formulated as 0.78*Traget+0.11 in Y-axis. Figure 10(B) shows the 10 HL neurons ANN model regression analysis that the total R value is 0.93913. All the data points class 1 and class 2 (target values 0 and 1) are fitted according to output. The output is formulated as 0.88*Traget+0.059

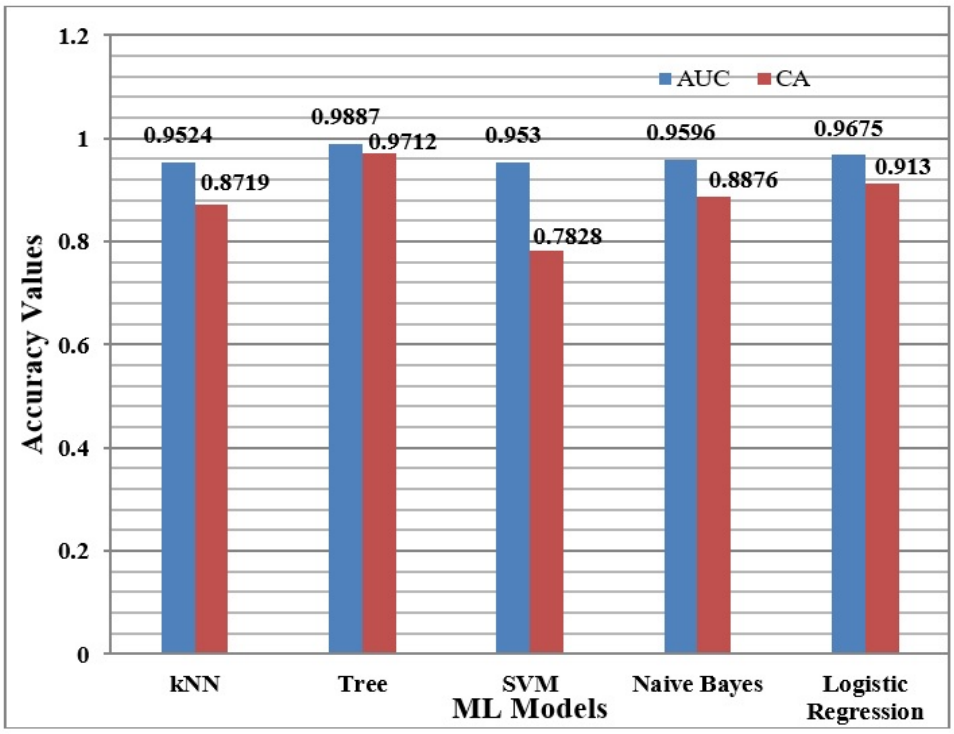

Figure 7. ML Comparative Analysis using AUC and CA Values

in Y-axis. Figure 10(C) shows the ANN model with 15 HL neurons regression analysis that the total R value is 0.95316. Most of data points class 1 and class 2 (LDs and non-LDs) are fitted according to output. The output is formulated as 0.91*Traget+0.046 in Y-axis. Figure 10(D) shows the 20 HL neurons ANN model regression analysis that the total R value is 0.97293. All the data points class 1 and class 2 (target values 0 and 1) are fitted according to output. The output is formulated as 0.95*Traget+0.027 in Y-axis. Figure 10(E) shows the 25 HL neurons ANN model regression analysis that the total R value is 0.98464. All the data points class 1 and class 2 (target values 0 and 1) are fitted according to output. The output is formulated as 0.97*Traget+0.044 in Y-axis. Figure 10(F) shows the ANN model with 30 HL neurons regression analysis that the total R value is 0.99375. Most of data points class 1 and class 2 (LDs and non-LDs) are fitted according to output. The output is formulated as 0.99*Traget+0.018 in Y-axis. As per investigation, the R values are increased proportional to HL neurons of ANN model.

4) Best Training Performance Analysis:

The Figure 11 shows the Best training performances analysis of 5 to 30 hidden layer (HL) neurons ANN models. The performance curve is created by number epochs on X-axis and mean squared error (MSE)on Y-axis as per classified data points. The blue colour curve indicates the training performance line and the dotted line gives the information about best performance point. As per observations, the performance value is in proportional to number of hidden neurons of ANN model that the error value is decreased as per incremental hidden neurons. Figure 11(A) shows the 5 HL neurons best performance analysis that the best training performance value is 0.053836 at the epoch 1000. Figure 11(B) shows the 10 HL neurons best performance analysis that the best training performance value is 0.02951at the epoch 661. Figure 11(C) shows the 15 HL neurons best performance analysis that the best training performance value is 0.017672at the epoch 521. Figure 11(D) shows the 20 HL neurons best performance analysis that the best training performance value is 0.013354 at the epoch 380.Figure 11(E) shows the 25 HL neurons best performance analysis that the best training performance value is 0.0097475 at the epoch 1000.Figure 11(F) shows the 30 HL neurons best performance analysis that the best training performance value is 0.0075644 at the epoch 1000

5) Error Histograms Analysis:

The Figure 12shows the error histograms of 5 to 30 hidden layer (HL) neurons ANN models with 20 bins. The error value of each data point is calculated using Target value minus output value. The error values are allotted to X-axis, and the numbers of instances are allocated to the Y-axis for plotting histograms. The orange colour line indicates the zero-error value. The blue colour strips describes the training data points with error values. Figure 12(A) shows the 5 Neurons ANN model error histogram.

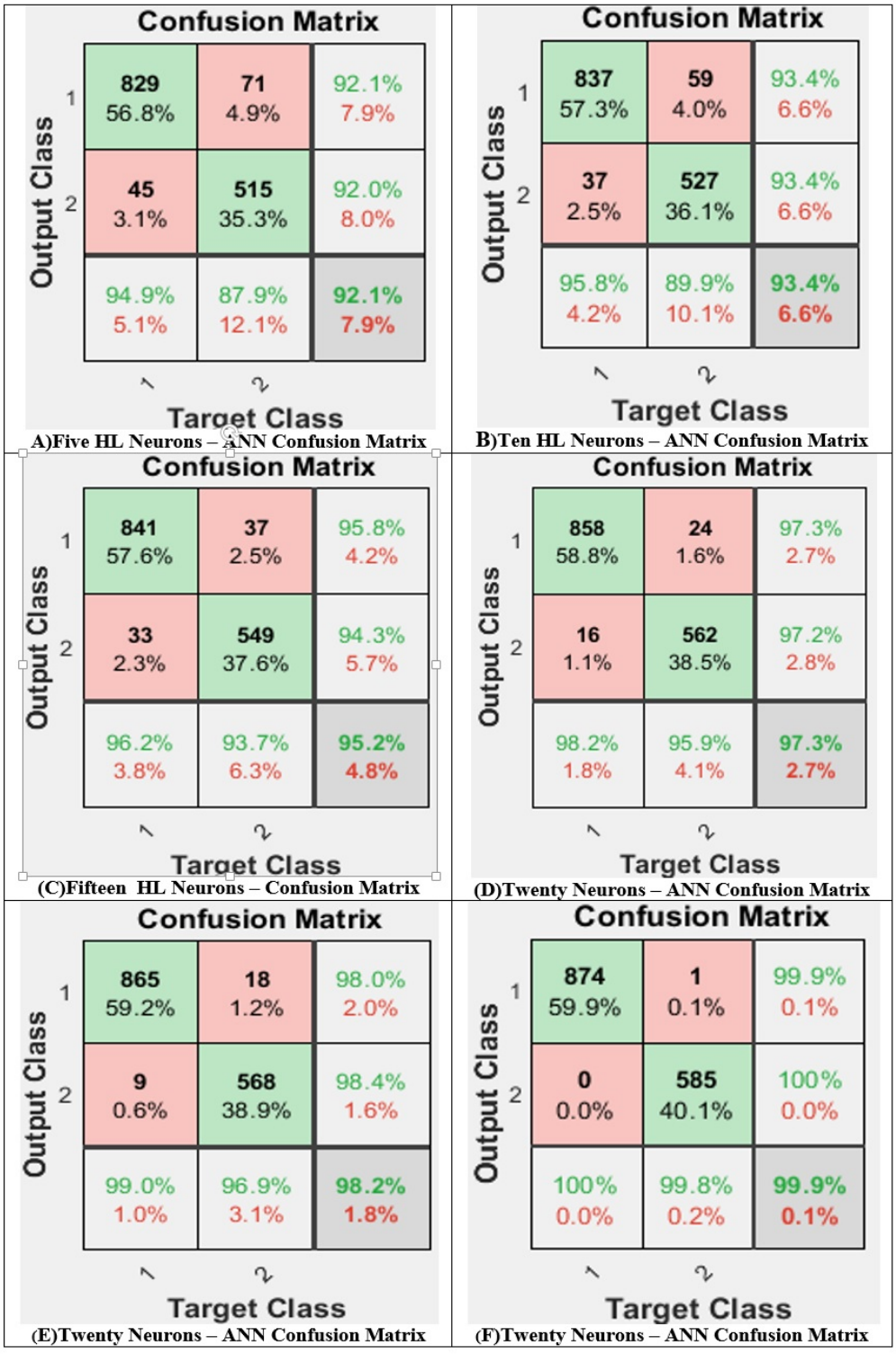

Figure 8. ANN model Confusion Matrices for 5, 10, 15, 20, 25 and 30 Hidden Layer Neurons

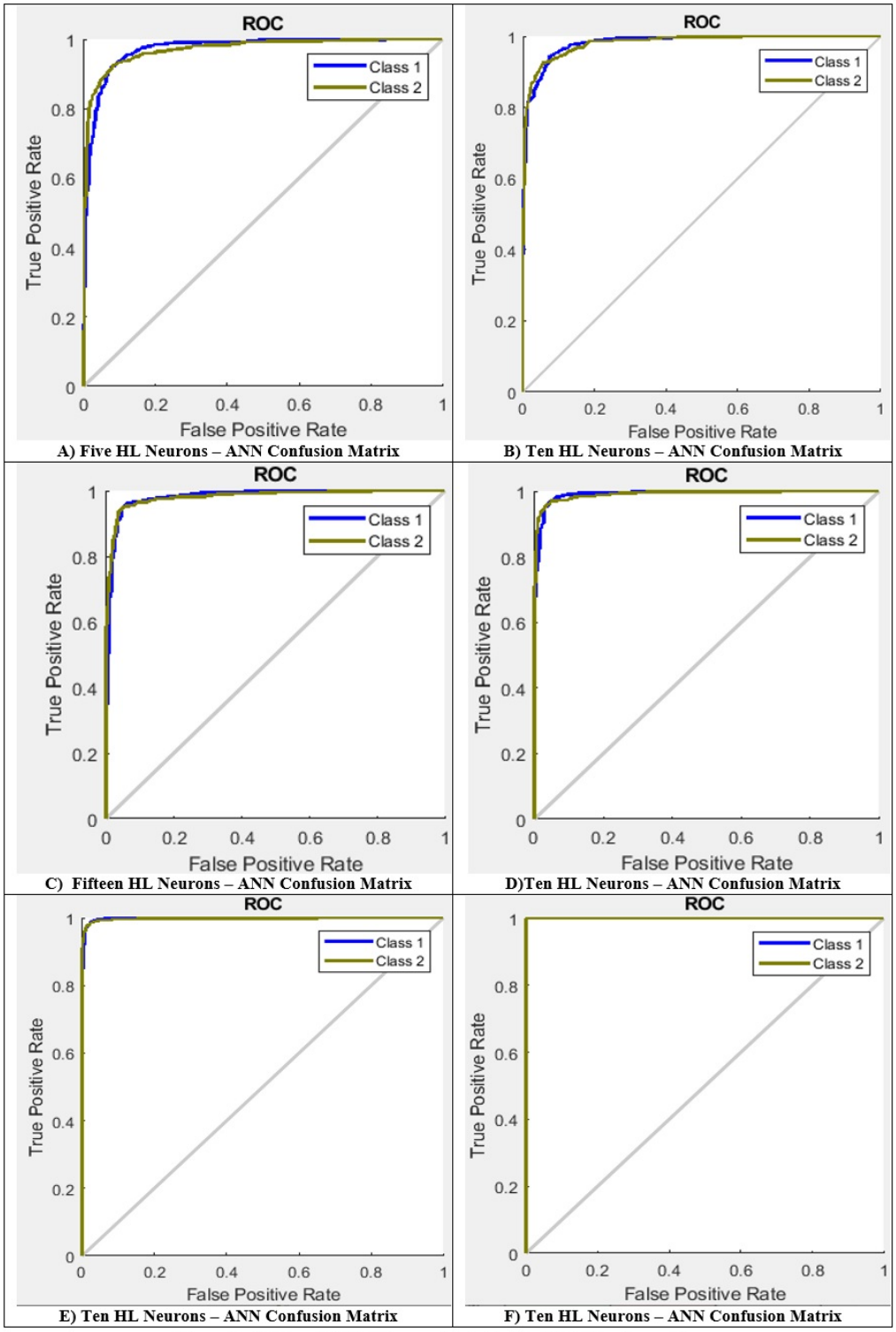

Figure 9. ANN model ROC Curves Analysis for 5, 10, 15, 20, 25 and 30 Hidden Layer Neurons

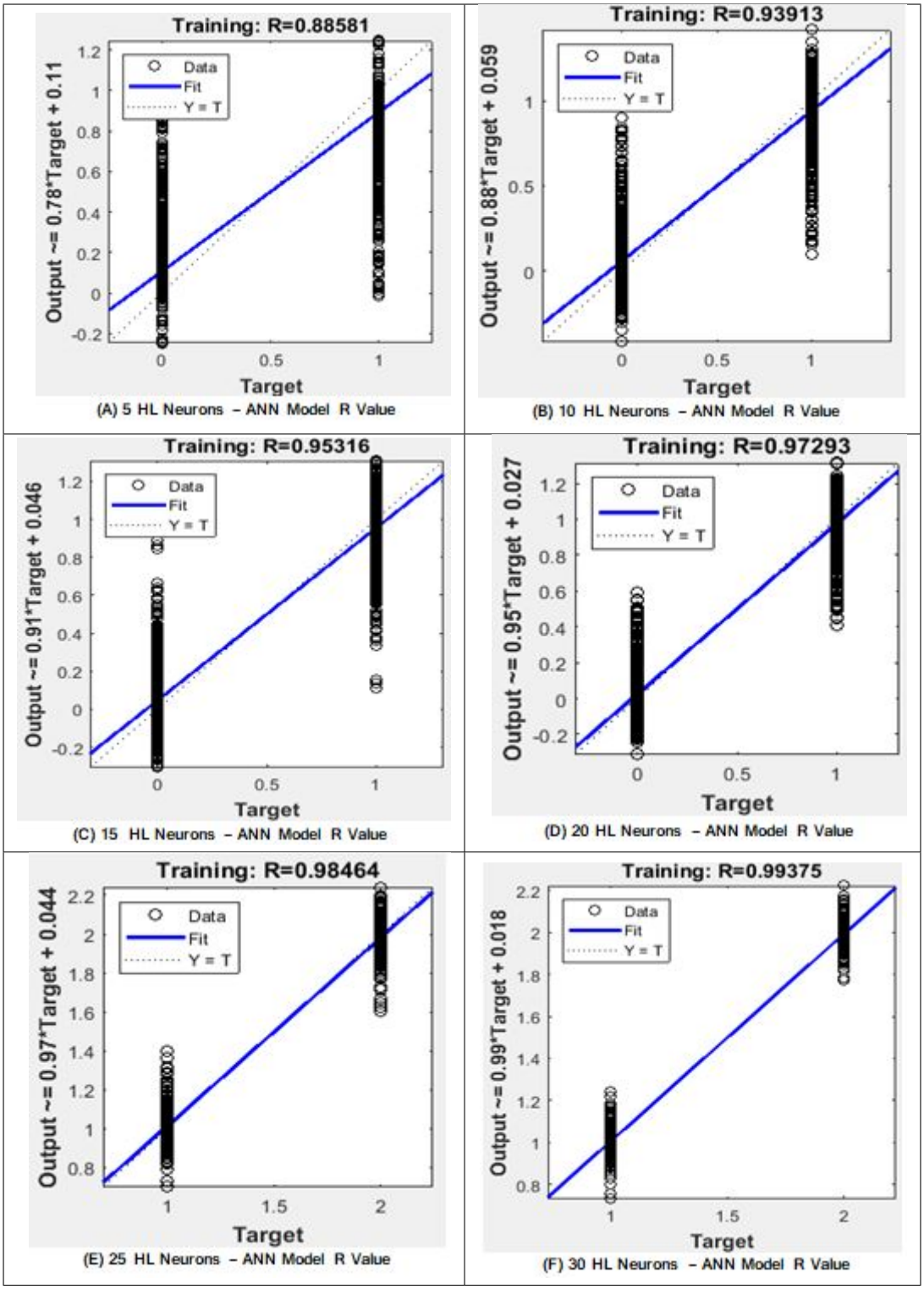

Figure 10. ANN model Regression (R) Values for 5, 10, 15... 30 Hidden Layer Neurons

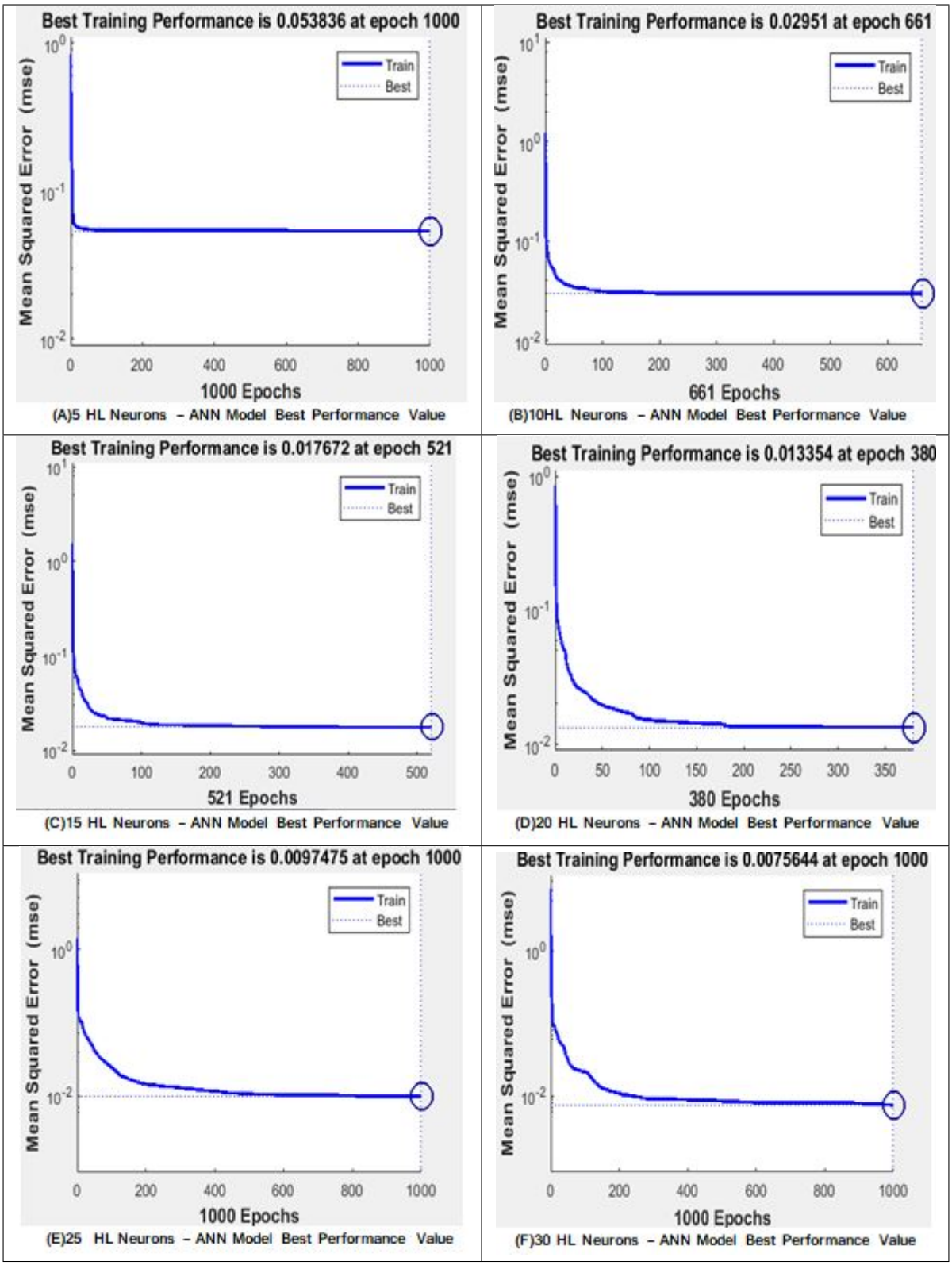

Figure 11. ANN model Best Performance Values for 5, 10, 15... 30 Hidden Layer Neurons

The marginal values of 20 bins are -0.9666 to 0.9666, and most of data points are nearer to zero that the values are between -0.1526 to 0.1526. Figure 12(B) indicates the 10 HL neurons ANN model error histogram. The marginal values of 20 bins are -0.9919 to 0.9919, and most of data points are nearer to zero that the values are between -0.1566 to 0.1566. Figure 12(C) describes the 15 HL neurons ANN model error histogram. The marginal values of 20 bins are -0.7885 to 0.7885, and most of data points are nearer to zero that the values are between -0.1245 to 0.1245. Figure 12(D) presents the 20 HL neurons ANN model error histogram. The marginal values of 20 bins are -0.562 to 0.562, and most of data points are nearer to zero that the values are between -0.08874 to 0.08874. Figure 12(E) shows the 25 HL neurons ANN model error histogram. The marginal values of 20 bins are -0.4519 to 0.4519, and most of data points are nearer to zero error line that the values are between -0.07937 to 0.04481. Figure 12(F) tells the30 HL neurons ANN model error histogram. The marginal values of 20 bins are-0.3759 to 0.3759, and most of data points are nearer to zero that the values are between -0.06047 to 0.05782. As per error histograms analysis, the number of data point's error values are decreased as well as the range of error value is also decreased, and numbers of data points are increased with no error value that they are reached orange line according to the HL neurons of ANN.

6) Comparative 5 to 30 HL Neurons of ANN model Analysis:

The table 6 shows the comparative analysis of time, epochs, and accuracy like R value, training accuracy, and gradient values all 5 to 30 HL neurons of ANN models. As per table description, the training time is increased according to the number of HL neurons of ANN. The gradient values and number of epochs for training are flickered that is increase and decreasing the values. The detailed figures are shown in the table. Figure 13 shows the detailed comparative analysis of R values. In this, the X-axis represents the number of HL neurons of ANN models, and Y-axis specifies the Regression values between 0 and 1. The blue line indicates R values according number of HL neurons.As per observations, the R values are increased proportional to number of HL neurons of ANN model.This graph is very useful for the analyzing and estimating remain HL neuron's R values like 6 HL neurons,7 6 HL neuronsand so on. The graph indicates that the model performance is in stable (gradual increment) and fair and also it describes about fitness of data set values in each stage. The peak value is getting the value 0.99375 at 30 HL neurons ANN model [45]. Figure 14 shows the detailed comparative analysis of best training performance or MSE (mean squared error) values in graph model. In this, the X-axis represents the number of HL neurons of ANN models, and Y-axis specifies the MSE values between 0.0 and 0.06. The blue line indicates MSE values according number of HL neurons. As per observations, the MSE values are decreased proportional to number of HL neurons of ANN model. This graph is very use full for the analyzing and estimating remain neuron's MSE values and analyzing every stage of ANN training. The highest MSE value is 0.053836 at 5HL ANN and lowest value is 0.007564 at 30HL ANN.

E. Discussions and Comparative Internal and External Analysis

As per this research study, the ML algorithms perform parameters AUC and CA values for k-NN is 0.9524 and 0.8719, Decision Tree values are 0.9887 and 0.9712, SVM with RBF kernel values are 0.9530 and 0.7828, Naive Bayes values are 0.9596 and 0.8876, and Logistic Regression values are 0.9675 and 0.9130 respectively. The incremental HL neurons ANN proposal model performs with AUC and CA values that The Five neurons HL ANN values are 0.953676 and 0.920548, Ten neurons HL ANN 0.968492 and 0.934247, Fifteen neurons HL ANN 0.974551 and 0.952055, Twenty neurons HL ANN 0.980369 and 0.972603 Twenty-Five neurons HL ANN 0.991493 and 0.981507Thirty neurons HL ANN 1 and 0.999315 respectively. As per internal comparative analysis the 30 neurons HL ANN performs superiorly with AUC value 1 than other experimental models. As well as compare to other related works our incremental HL neurons ANN model is the best. Wu et al. (2018) [46] analyzed fatty liver (LD) diseased 577 individual's clinical data that was collected from Banqiao New Taipei City Hospital. In this research, their vital aim was to predict the FLD using ML models like RF, LR, NB, and ANN with threefold, fivefold, and tenfold cross-validations. For the evaluation of performance, they choose and develop the AUC and accuracy of classification for four models that were RF, LR, NB, and ANN. The highly performed values with 10fold were 0.925, 0.888, 0.895, and 0.854 in AUC analysis, and the classification accuracy (CA) analysis values were respectively 87.48, 82.65, 81.85, and 76.96% at the cross-validation of 10fold. In this research, they concluded that the 10flod RF model was superior to other experimental methods with 0.925 (AUC) and 87.48 (CA) values. The table shows the comparative analysis.

6. Conclusions

Early identification of LD is one of the exorbitant and complex assignments. Thus, one of the answers for recognize this disease should be possible through the investigation of clinical and personal data items and records in depth. In this paper, we analyzed esteemed analysis on LD versus non-LD.For this, we use the proposal model "back propagation(BK) ANN with incremental neurons in hidden layer (HL)". We checked the 5 to 30 HL neurons in BK-ANN for better performance, and compare to ML models. The ML models like DT, Naive Bayes (NB), SVM,k-NN, and Logistic Regression (LR). DT and NB were performed 97.12% and 88.76%with accuracy first and second positions respectively. The training performance of BK-ANN is step by step increased with respect to the addition HL neurons of BK-ANN.At the best, thirty neurons HL of BK-ANN performance is very high compare to other BK-ANN and ML models that the accuracy is 99.9% (0.999) and AUC is 1(100%) in test and train, the gradient value is 5.1624e-05,

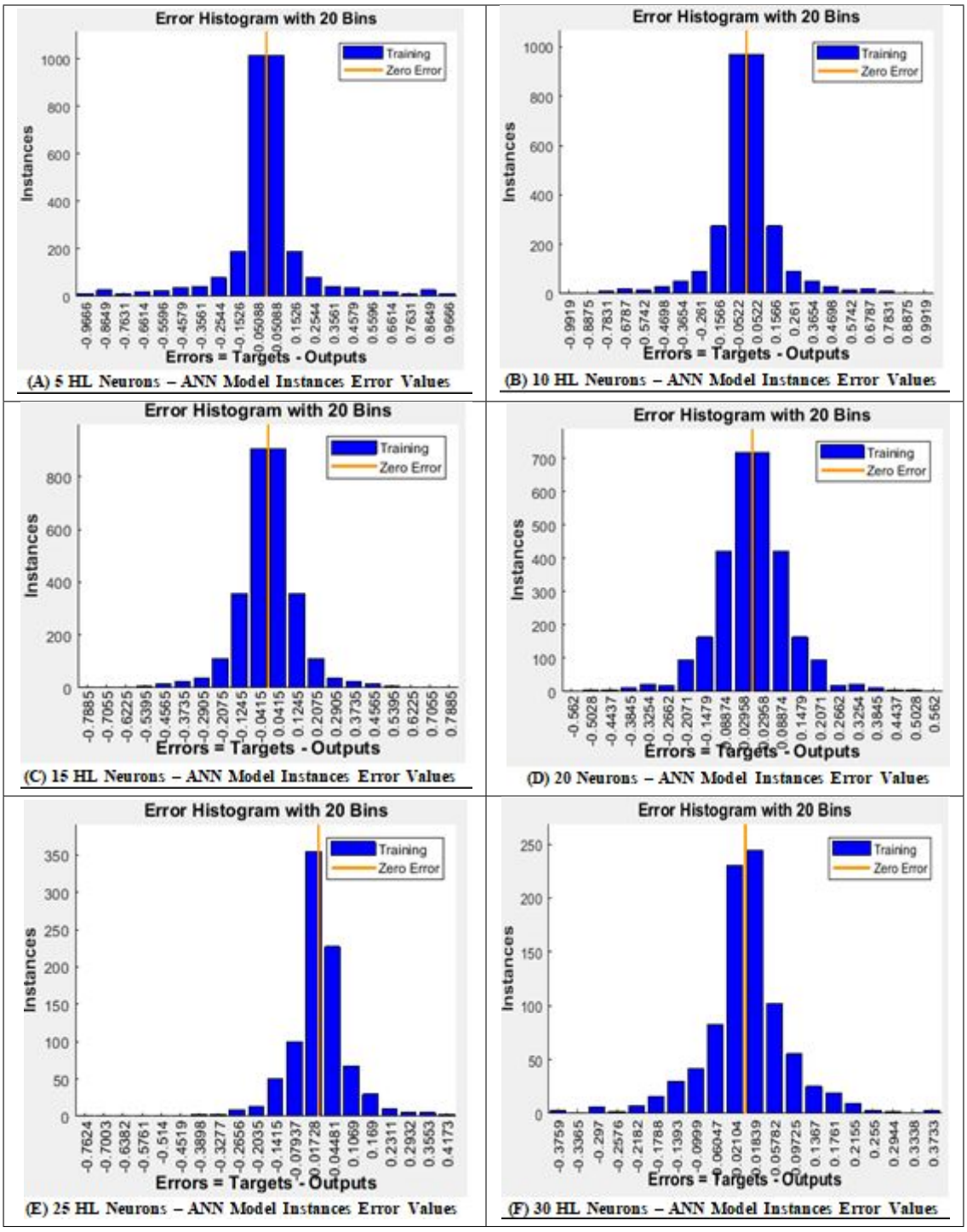

Figure 12. Caption

TABLE VI. Experimental Incremental HL neurons (5 to 30) ANN models performance parameters

ANN HL Neurons	AUC	CA	F1	Precision	Recall
Five	0.953676	0.920548	0.917014	0.913676	0.920377
Ten	0.968492	0.934247	0.931374	0.928492	0.934274
Fifteen	0.974551	0.952055	0.950065	0.949551	0.950579
Twenty	0.980369	0.972603	0.97146	0.970369	0.972554
Twenty-Five	0.991493	0.981507	0.980749	0.979493	0.982009
Thirty	**1**	**0.999315**	**0.999288**	**0.999147**	**0.998861**

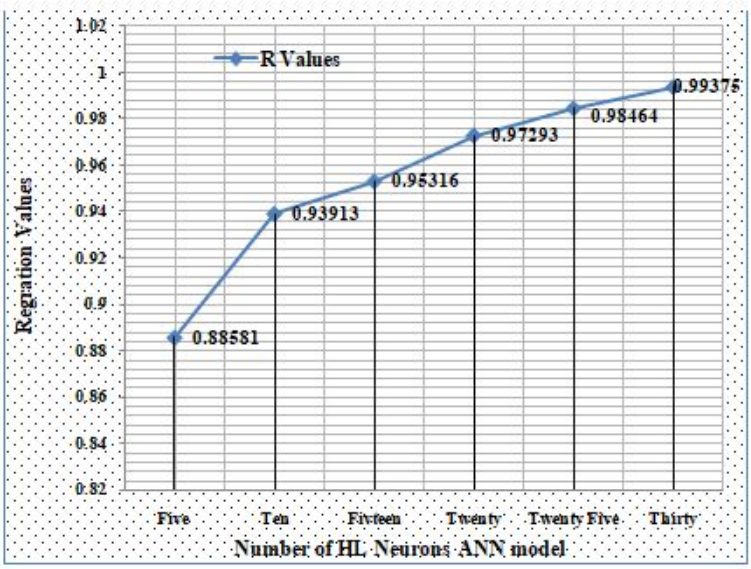

Figure 13. Comparative Analysis of R Values

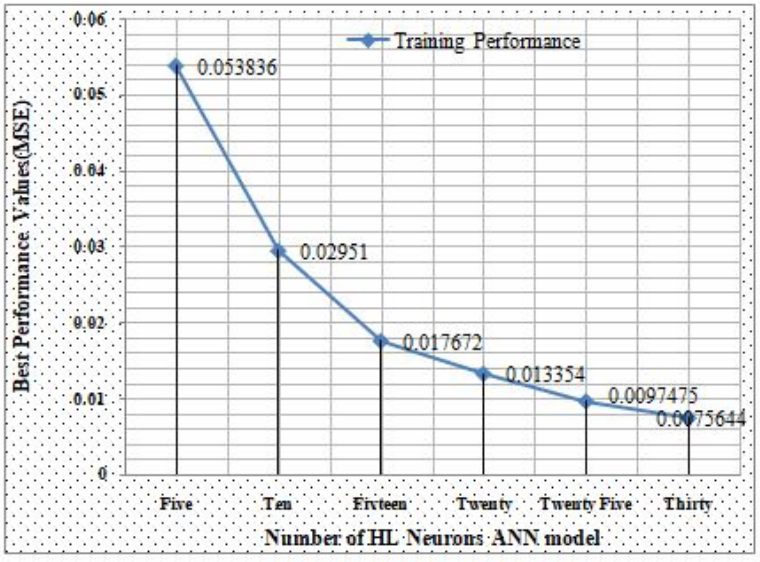

Figure 14. Comparative Analysis of Best Training Performance or MSE

and the performance 0.007564, and it takes 0.52 seconds for 1000 processed iterations. So, 30 HL BK-ANN is very

TABLE VII. Comparative Analysis

Ref. No.	Author	Contribution and Area	Results	Year
[47]	Musleh et al.,	Prediction LD using ML-ANN (Multilayer) and SVM. UCI ML Repository data set with 583 liver patients (416 LD individuals 167 Control Individuals)	Accuracy, Precision, Sensitivity, and Specificity Values of SVM: 71%, 64.1%, 71.5%,88.3%, and ANN (MLP): 73.2%, 65.7%, 73.3% and 87.7% respectively.	2019
[48]	Yao et al.,	Diagnosis LD using LR, RF, DNN, Dense-DNN, the dataset collected from northwestern China 76,914 samples	LR, RF, DNN, and Dense-DNN AUC values are 0.7977, 0.8790, 0.8867, and 0.8919 respectively.	2019
[49]	Chen et al.	Diagnosis LD using ANN the dataset split 70% for training and 30% testing of 14,792 samples UCI Repository	Accuracies of Fatty Liver Index, Hepaticsteatosis Index and ANN are 0.796, 0.802 and 0.9821 respectively.	2020
[50]	Praveen et al.,	Diagnosis of LD using ML models	SVM- 0.864 (86.4%), Random Forest-0.986 (98.7%) Naive Bayes- 0.887(88.8%)	2021
–	**This Study (ML Models)**	**Prediction LD using k-NN, DT, SVM, Naive Bayes and Logistic Regression. The data Collected from North costal Districts of A.P. (Synthetic real-world data)**	**Accuracy values of k-NN - 0.8719, DT-0.9712, SVM-0.7828, Naive Bayes-0.8876 and Logistic Regression-0.9130**	—
–	**This Study (Proposal Model)**	**Prediction LD using incremental HL neurons ANN model. The data Collected from North costal Districts of A.P. (Synthetic real-world data)**	**30 HL neurons ANN – Accuracy (CA) - 0.999315(99.9%) AUC -1(100%)**	—

effective and efficient model to early detection of LD with this dataset. As per comparison of ML algorithms and other research works, the proposal model is efficient for detecting LD in early with low cost.

References

[1] R. Manco, I. A. Leclercq, and L.-A. Clerbaux, "Liver regeneration: Different sub-populations of parenchymal cells at play choreographed by an injury-specific microenvironment," *International journal of molecular sciences*, vol. 19, no. 12, p. 4115, 2018.

[2] Y. K. Jung and H. J. Yim, "Reversal of liver cirrhosis: current evidence and expectations," *The Korean journal of internal medicine*, vol. 32, no. 2, p. 213, 2017.

[3] R. A. d. A. Ximenes, L. M. B. Pereira, C. M. T. Martelli, E. Merchán-Hamann, A. T. Stein, G. M. Figueiredo, M. C. Braga, U. R. Montarroyos, L. M. Brasil, M. D. Turchi *et al.*, "Methodology of a nationwide cross-sectional survey of prevalence and epidemiological patterns of hepatitis a, b and c infection in brazil," *Cadernos de saude publica*, vol. 26, pp. 1693–1704, 2010.

[4] S.-H. Jeong and H.-S. Lee, "Hepatitis a: clinical manifestations and management," *Intervirology*, vol. 53, no. 1, pp. 15–19, 2010.

[5] J. Lucifora, S. Arzberger, D. Durantel, L. Belloni, M. Strubin, M. Levrero, F. Zoulim, O. Hantz, and U. Protzer, "Hepatitis b virus x protein is essential to initiate and maintain virus replication after infection," *Journal of hepatology*, vol. 55, no. 5, pp. 996–1003, 2011.

[6] M. P. Manns, M. Buti, E. Gane, J.-M. Pawlotsky, H. Razavi, N. Terrault, and Z. Younossi, "Hepatitis c virus infection," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1–19, 2017.

[7] S. Kohara, T. Tateyama, A. H. Foruzan, A. Furukawa, S. Kanasaki, M. Wakamiya, and Y.-W. Chen, "Application of statistical shape model to diagnosis of liver disease," in *The 2nd International Conference on Software Engineering and Data Mining*. IEEE, 2010, pp. 680–683.

[8] H. Bhatt and R. Smith, "Fatty liver disease in diabetes mellitus. hepatobiliary surg nutr. 2015; 4 (2): 101-8."

[9] D. L. Dumitrascu and M. G. Neuman, "Non-alcoholic fatty liver disease: an update on diagnosis," *Clujul medical*, vol. 91, no. 2, p. 147, 2018.

[10] D. A. Schady and M. J. Finegold, "Contemporary evaluation of the pediatric liver biopsy," *Gastroenterology Clinics*, vol. 46, no. 2, pp. 233–252, 2017.

[11] A. Saini, M. Kumar, S. Bhatt, V. Saini, and A. Malik, "Cancer causes and treatments," *Int J Pharm Sci & Res*, vol. 11, no. 7, pp. 3121–34, 2020.

[12] S. M. Abd El-Salam, M. M. Ezz, S. Hashem, W. Elakel, R. Salama, H. ElMakhzangy, and M. ElHefnawi, "Performance of machine learning approaches on prediction of esophageal varices for egyptian chronic hepatitis c patients," *Informatics in Medicine Unlocked*, vol. 17, p. 100267, 2019.

[13] T. P. Vital, G. S. Murty, K. Y. Rao, and T. Sriram, "Empirical study and statistical performance analysis with ann for parkinson's vowelized data set," in *Computational Intelligence in Data Mining*. Springer, 2020, pp. 767–780.

[14] N. Atabaki-Pasdar, M. Ohlsson, A. Viñuela, F. Frau, H. Pomares-Millan, M. Haid, A. G. Jones, E. L. Thomas, R. W. Koivula, A. Kurbasic *et al.*, "Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the imi direct cohorts," *PLoS medicine*, vol. 17, no. 6, p. e1003149, 2020.

[15] I. Gatos, S. Tsantis, S. Spiliopoulos, D. Karnabatidis, I. Theotokas, P. Zoumpoulis, T. Loupas, J. D. Hazle, and G. C. Kagadis, "A machine-learning algorithm toward color analysis for chronic liver disease classification, employing ultrasound shear wave elastography," *Ultrasound in medicine & biology*, vol. 43, no. 9, pp. 1797–1810, 2017.

[16] T.-F. Yip, A. Ma, V.-S. Wong, Y.-K. Tse, H.-Y. Chan, P.-C. Yuen, and G.-H. Wong, "Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (nafld) in the general population," *Alimentary pharmacology & therapeutics*, vol. 46, no. 4, pp. 447–456, 2017.

[17] A. S. Rahman, F. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *International Journal of Scientific & Technology Research*, vol. 8, no. 11, pp. 419–422, 2019.

[18] R. D. Khusial, C. E. Cioffi, S. A. Caltharp, A. M. Krasinskas, A. Alazraki, J. Knight-Scott, R. Cleeton, E. Castillo-Leon, D. P. Jones, B. Pierpont *et al.*, "Development of a plasma screening panel for pediatric nonalcoholic fatty liver disease using metabolomics," *Hepatology communications*, vol. 3, no. 10, pp. 1311–1321, 2019.

[19] B. Muruganantham, R. Mahapatra, K. Taparia, and M. Kumar, "Liver disease prediction using an ensemble based approach," in *Intelligent Computing and Applications*. Springer, 2021, pp. 507–518.

[20] T. P. Vital, M. M. Krishna, G. Narayana, P. Suneel, and P. Ramarao, "Empirical analysis on cancer dataset with machine learning algorithms," in *Soft Computing in Data Analytics*. Springer, 2019, pp. 789–801.

[21] A. Das, U. R. Acharya, S. S. Panda, and S. Sabut, "Deep learning based liver cancer detection using watershed transform and gaussian mixture model techniques," *Cognitive Systems Research*, vol. 54, pp. 165–175, 2019.

[22] V. J. Gogi and M. Vijayalakshmi, "Review of machine learning methods for the survey on hcc scenario and prediction strategy," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE, 2020, pp. 949–951.

[23] P. Pruthvi, B. Manjuprasad, and B. Parashiva Murthy, "Liver cancer analysis using machine learning techniques-a review," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCICCNDA*, vol. 5, no. 22, 2017.

[24] W. Ksiażek, M. Abdar, U. R. Acharya, and P. Pławiak, "A novel machine learning approach for early detection of hepatocellular carcinoma patients," *Cognitive Systems Research*, vol. 54, pp. 116–127, 2019.

[25] S. Naeem, A. Ali, S. Qadri, W. Khan Mashwani, N. Tairan, H. Shah, M. Fayaz, F. Jamal, C. Chesneau, and S. Anam, "Machine-learning based hybrid-feature analysis for liver cancer classification using fused (mr and ct) images," *Applied Sciences*, vol. 10, no. 9, p. 3134, 2020.

[26] P. Rajeswari and G. S. Reena, "Analysis of liver disorder using data mining algorithm," *Global journal of computer science and technology*, 2010.

[27] K. Akyol and Y. Gültepe, "A study on liver disease diagnosis based on assessing the importance of attributes," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 11, p. 1, 2017.

[28] B. Khan, P. K. Shukla, M. K. Ahirwar, and M. Mishra, "Strategic analysis in prediction of liver disease using different classification algorithms," in *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*. IGI Global, 2021, pp. 437–449.

[29] S. Kefelegn and P. Kamat, "Prediction and analysis of liver disorder diseases by using data mining technique: survey," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 9, pp. 765–770, 2018.

[30] S. Bahramirad, A. Mustapha, and M. Eshraghi, "Classification of liver disease diagnosis: a comparative study," in *2013 Second International Conference on Informatics & Applications (ICIA)*. IEEE, 2013, pp. 42–46.

[31] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," in *2017 International Conference on Emerging Trends & Innovation in ICT (ICEI)*. IEEE, 2017, pp. 129–133.

[32] D. Xu, H. Fu, and W. Jiang, "Research on liver disease diagnosis based on rs_lmbp neural network," in *2016 12th International Conference on Computational Intelligence and Security (CIS)*. IEEE, 2016, pp. 646–649.

[33] T. M. Hassan, M. Elmogy, and E.-S. Sallam, "Diagnosis of focal liver diseases based on deep learning technique for ultrasound images," *Arabian Journal for Science and Engineering*, vol. 42, no. 8, pp. 3127–3140, 2017.

[34] M. Abdar, N. Y. Yen, and J. C.-S. Hung, "Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees," *Journal of Medical and Biological Engineering*, vol. 38, no. 6, pp. 953–965, 2018.

[35] F. Özyurt, T. Tuncer, E. Avci, M. Koç, and İ. Serhatlioğlu, "A novel liver image classification method using perceptual hash-based convolutional neural network," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3173–3182, 2019.

[36] A. S. Singh, M. Irfan, A. Chowdhury *et al.*, "Prediction of liver disease using classification algorithms," in *2018 4th international conference on computing communication and automation (ICCCA)*. IEEE, 2018, pp. 1–3.

[37] L. A. Auxilia, "Accuracy prediction using machine learning techniques for indian patient liver disease," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 45–50.

[38] D. S. Reddy, R. Bharath, and P. Rajalakshmi, "A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2018, pp. 1–5.

[39] "[37] srivenkatesh, m. (2019). performance evolution of different machine learning algorithms for prediction of liver disease volume-9 issue-2, 1115-1122. - google search."

[40] M. Ramaiah, P. Baranwal, S. B. Shastri, M. Vanitha, and C. Vanmathi, "Analytical comparison of machine learning techniques for liver dataset," in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, vol. 1. IEEE, 2019, pp. 1–5.

[41] J. Singh, S. Bagga, and R. Kaur, "Software-based prediction of liver disease with feature selection and classification techniques," *Procedia Computer Science*, vol. 167, pp. 1970–1980, 2020.

[42] R.-y. Mai, H.-z. Lu, T. Bai, R. Liang, Y. Lin, L. Ma, G.-b. Wu, L.-q. Li, J.-z. Ye *et al.*, "Artificial neural network model for preoperative prediction of severe liver failure after hemihepatectomy in patients with hepatocellular carcinoma," *Surgery*, vol. 168, no. 4, pp. 643–652, 2020.

[43] M. A. Kuzhippallil, C. Joseph, and A. Kannan, "Comparative analysis of machine learning techniques for indian liver disease patients," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020, pp. 778–782.

[44] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection. maics, 710, 120-127," *DOI: 10.35940/ijitee. G5451. 059720 Published By: Blue Eyes Intelligence Engineering*, 2011.

[45] N. B. Chaudhuri, D. Chandrika, and D. K. Kumari, "Review on mental health using soft computing and neuro fuzzy techniques," in *the Proceedings of International Conference on Global Trends in*

Engineering Technology and Management (ICGTETM), 2016, pp. 390–394.

[46] C.-C. Wu, W.-C. Yeh, W.-D. Hsu, M. M. Islam, P. A. A. Nguyen, T. N. Poly, Y.-C. Wang, H.-C. Yang, and Y.-C. J. Li, "Prediction of fatty liver disease using machine learning algorithms," *Computer methods and programs in biomedicine*, vol. 170, pp. 23–29, 2019.

[47] M. M. Musleh, E. Alajrami, A. J. Khalil, B. S. Abu-Nasser, A. M. Barhoom, and S. A. Naser, "Predicting liver patients using artificial neural network," *International Journal of Academic Information Systems Research (IJAISR)*, vol. 3, no. 10, 2019.

[48] Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," *Neural Networks*, vol. 123, pp. 299–304, 2020.

[49] Y.-S. Chen, D. Chen, C. Shen, M. Chen, C.-H. Jin, C.-F. Xu, C.-H. Yu, and Y.-M. Li, "A novel model for predicting fatty liver disease by means of an artificial neural network," *Gastroenterology report*, vol. 9, no. 1, pp. 31–37, 2021.

[50] A. D. Praveen, T. P. Vital, D. Jayaram, and L. V. Satyanarayana, "Intelligent liver disease prediction (ildp) system using machine learning models," in *Intelligent Computing in Control and Communication*. Springer, 2021, pp. 609–625.

Dr. Panduranga Vital Terlapu pursed Bachelor of Science in Computer Science from Andhra University of A.P, India in 1995 and Master of computer Application from Andhra University in year 1998. He completed his M. Tech in Computer Science and Engineering from Acharya Nagarjuna University of A.P, India and he received his-PhD in Computer Science and Engineering from GITAM University of A.P, India. He has 20 years of teaching and 13 years of research experience. He is currently working as Associate Professor in Department of Computer Science and Engineering, Aditya Institute of Technology and Management (AITAM), India. He is a member of ACM, Life Time Membership from International Computer Scienceand Engineering Society (ICSES), USA and Life Time Membership from Indian Society for Technical Education (ISTE), New Delhi, India. He has published more than 40 research papers in reputed international journals including SCOPUS and SCI indexed. He is reviewer of reputed journals like Springer, Elsevier and IEEE. His main research work focuses on Machine Learning, Deep Learning and Data Mining, Data and Big Data Analytics, Voice Analysis and Voice Processing, Bioinformatics.

Dr. Ram Prasad Reddy Sadi received his PhD in Computer Science and Engineering from Jawaharlal Nehru Technological University, Kakinada. Currently he is working as Associate Professor of Information Technology in Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam. His research Interests include data privacy, machine learning and distributed systems.

Mr. Ramkishor Pondreti received his M.Tech in Software Engineering from Jawaharlal Nehru Technological University, Kakinada. Currently he is working as Assistant Professor of Aditya Institute of Technology and Management, Tekkali. His research interests include Data Analytics, Machine Learning, Image Processing and Deep Learning.

Mr.Chalapathi Rao Tippana received his M.Tech in Computer Science and Engineering from Jawaharlal Nehru Technological University, Ananthapur. Currently he is working as Sr. Assistant Professor of Aditya Institute of Technology and Management, Tekkali. He has published more than 10 research papers in reputed international journals including SCOPUSHis research interests include Data Analytics, Machine Learning, Image Processing, and IoT.