

Liver Disease Classification using Deep Learning Algorithm

L. Anand, V. Neelanarayanan



Abstract- Data Mining is one of the prevalent elucidating portions of programmed request and distinguishing proof. It involves data mining counts and strategies to examine helpful data. Of late, liver dissents have disproportionately expanded and liver infections are complimenting one of the most human pains in different countries. Early assurance of Liver Disorder is essential for the welfare of human culture. This complaint should be considered sincerely by setting up watchful structures for the early break down and expectation of Liver contaminations. The robotized gathering system suffers with non attendance of precision results when differentiated and cautious biopsy. We propose another model for liver issue request for separating the patient's helpful, data using ANN algorithm. The remedial records are organized whether there is a believability of essence of disorder or not. This proposed methodology uses extracted features using M-PSO and ANN for classifying the features. The ANN methodology improves the accuracy when appeared differently in relation to existing request computations. This paper focuses classification of selected features for classification.

I. INTRODUCTION

Information retrieval is a lot of basic methodology in deciding emitted information highlights, propensities and structures in immense datasets, and convey acknowledged recognitions inspired by the reason or condition of certain settings. Through uncovering the example and developments in the dataset, information mining in addition makes the conjecture of future procedures possible. The probability forecast may propose the individuals to carry on with a quality life comparing with their bioinformatics information. Information mining conveys its commitment in therapeutic area for the early finding of illnesses, chance factor investigation, Decision making, Treatment and medicine of medications. The concealed information record of patient is utilized for the forecast of illness in Classification period of information mining. This sickness can be grouped utilizing different morphology factors. The exactness of this kind of model give preferred outcomes over typical conventional strategies. Information mining can be connected in the field of bioinformatics in different applications such like [1] quality finding, [2] protein recognition, [3]function theme identification, [4]protein work derivation, [5]disease conclusion, [6]disease visualization, [7] infection prescription improvement, [8]

protein and quality participation arrange recovery , information laxative, and [9]protein sub-cell position gauge. Current evolvement in information mining study has coordinated to the redesigning of a few efficacious and available methodologies for uncovering interesting examples in gigantic therapeutic databases. Current evolvement in science, restorative science, and DNA innovation has coordinated to the growth of heavenly Volumes of bio - therapeutic information that nerves in the interest of complete examination. Information mining and bioinformatics joins to accomplish solid mining of biomedical information. The information mining procedure could be utilized to discover the probability of ailment risks by utilizing arrangement calculations. The mining procedure uncovers a few discharged and bizarre examples, which may not engaged previously. Various information mining methodologies had been proposed so far for the guess and distinguishing proof of liver maladies. By and large the Liver Disorder could be ends with liver cirrohis. The exactness of anticipating liver infections with therapeutic information mining mostly remains on the component determination. The test is to create and approve conventions connected to clinically significant result measures. Generally Liver maladies are analyzed in the clinical by investigating the degrees of chemicals in the blood. Be that as it may, Information concentrated measurable research on Liver maladies empowers to predict the survival investigation of the illness. It is a challengeable errand to anticipate the episode with the utilization of information mining approach. Enormous volume of gathered restorative dataset pulls in the mining expert to discover the sickness designs and their connections. This measurable mining approach need to investigate enormous number of variable components and medicinal authentic information. The powerful programmed Computer Aided Diagnosis framework ought to group the liver sickness dataset that was gathered from different hepatologist from different nations. The honesty of the grouping strategies endures with the nature of information. The exceptions and imbalanced bioinformatics dataset aggravates the immovability and execution of the expectation models. This expectation information mining calculations needs to confront the accompanying difficulties. Selecting appropriate ascribes to build the extrapolation model.

- Identification and disposal of anomalies with the bioinformatics information.
- Handling the unreasonable information in grouping.
- Differentiating typical hereditary highlights and ailment morphological components.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

L. Anand*, Assistant Professor, School of computing science and Engineering, Vellore Institute of Technology , Chennai, (Tamil Nadu), India. E-mail: anand.l2014@vit.ac.in.

V. Neelanarayanan, School of computing science and Engineering, Vellore Institute of Technology , Chennai, (Tamil Nadu), India. E-mail: Neelanarayanan.v@vit.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



The principle thought process of this examination work is to give a powerful CADs to analysis and guess of Liver illness. It ought to be viable so that gives dataset Pre-preparing, quality choice and exact classifier. The classifier includes a vivacious job in characterizing the liver issue into predefined classes of ailments.

For effective classifying of liver ailment, bio-enlivened calculations are utilized for straightforward grouping. Bio-motivated methods are the meta-heuristics approach that attempt to resemble the nature innovativeness for the benefit of settling advancement difficulties. The determined wellness worth arrange the kind of infection that could be feasible for the pertinent medicinal data. The principle commitments of this examination are as per the following:

- A new crossover system is intended to discover the odds of happening liver infections dependent on Evolutionary Algorithms.
- A Hybridized Artificial Neural Network grouping is utilized for the forecast of liver malady, for example, typical, greasy, cirrhosis liver, hepatic metastases and heterogeneous liver.

The rest of the areas of this paper are sorted out as different investigates on liver infection order based certain strategies and its confinements has been explored in segment exhibits the nitty gritty depiction of proposed technique incorporates and M-ANN demonstrates the test results and finishes up this work and gives future improvement.

II. RELATED WORK

A couple of peril components are settled for the explanation behind Liver Diseases. Risk parts can be portrayed that raise the upbeat of accomplishing liver sicknesses. In spite of the way that each and every human has the chance to be affected by the Liver issue, these factors sway more. A bit of these danger components can be modifiable and some can't be changed. [10] Liver limit tests are used to measure unequivocal mixes and proteins in your blood. The delayed consequences of the level of synthetic concoctions or proteins that reveals in the test (Whether higher or lower) could demonstrate the condition of the liver. The showing levels of these mixes or proteins are used as attributes in the dataset for executing data mining computations. Abdar [11] reviewed a couple of data mining approach execution and their appraisal results with essential liver ailment dataset. This work also communicated various sorts of liver ailments among 100 different kinds of liver diseases. This work dissected C4.5 estimation, C5.0 count, SVM computation, KNN figuring, CHAID and Neural Network computation to anticipate the liver ailment using Indian Liver disease dataset. It considered 11 attributes for data mining instrument execution. It analyzed diverse quality and impediments of these computations and contemplated that the precision of neural framework stays in 70% to 80%. This work saw the sufficiency of neural framework approach in liver ailment desire and early finding. [12] proposed troublesome affirmation of liver affliction over and completed with a MLPNN procedure concentrated on endless decision tree procedures like C5.0, HAID and CART with boosting technique. It used the estimations for evaluation metrics. The hybridization of B-C5.0 and MLPNN methodologies provided most extraordinary area pace of liver issue related to advance structures. [13] used decision tree based figuring such like Boosted C5.0 and

CHAID counts to shape rules from liver infection dataset. It built up that C5.0 could be especially grasped with multipurpose classifier nearby boosting limit. At any rate these methods are again and again influenced toward separating that attributes have endless levels. This work began combining neural framework and C5.0 computations both with booting approach. [14] associated Support Vector Machine figuring to ILPD and BUPA Liver issue. This work read straight piece work for feature decision. The precision achieved in this recently referenced effort is less. [15] prescribed another idea that combined both Case-Based Reasoning (CBR) theory strengthened by Genetic Algorithm (GA) to streamline the desire results of liver ailment. The heaps were assigned to the features according to their belongings for causing liver illness. The GA registers wellbeing regard that process each example of liver disorder research focus components to foresee whether the individual encounters liver contamination or not. This crossbreed procedure achieved 94.19% of most critical precision. [16] used SVM and Naïve Bayes Algorithms for the estimate. This work communicated that using Naïve classifier limit the required features of the patient's remedial record. [17] offered another fire fly count (FA) and winged creature mating streamlining (BMO) based MLP system for the disease perception. FA incorporates into taking a gander at decided yield and the ordinary respect find the mix-up rate, for instance, mean square screw up and aggregate of squared misstep. The UCI respiratory liver disease dataset is surveyed using this strategy and MLP – BMO evaluated as better. [18] associated adaptable Neuro-Fuzzy reasoning structure for liver issue portrayal. Five layers are used for confirmation and back spread learning standard is associated with change the parameters in the covered layer. The sound and unfortunate liver data were disconnected reasonably. [19] considered the critical components that overwhelmingly quick to the Liver issue. This work pondered the support of different data mining counts such like Naïve Bayes, Decision Tree, Multi-Layer Perceptron and k-NN. It gave a sensible course map and surveying criteria to portray the Liver sicknesses. [20] Fuzzy-cruel case decision remove the clatter of data and works brisk for small planning sets. By then steady quality based trademark gathering is associated. It dealt with the defenselessness data extraordinary and assisted with noisy data. It suffers with huge getting ready data. [21] offered Hybrid strategy of unpredictable woods classifier and feature assurance practice describes the infirmity data. In the event that Good attribute decision approach yet more diminutive get-togethers are Favored over greater social affairs. [22] GA is used for expelling from pleasing and imperative features. Discrete individual and different classifier coordination were used to make exact strategy for sickness recording. [23] Rich game plan of continuous standards are used for the portrayal figuring. Least help check is used to organize the progressive models. The huge hindrances of this philosophy is clashing and lacking records on account of unusual visits of patients impacts continuous procedure. It also suffers with Imbalance and Post-assessment issue. Models reliant on feature assurance and game plan, yet it raised some various issues such like

- More Time multifaceted nature of estimation.
- Some philosophies are fitting only for little data.
- Some blend of classifier over fit with enlightening accumulation and some will be under fit with educational list.
- The dependence of events and credited are uncategorized.
- Some philosophies are not adoptable for ceaseless social occasion of database use.

To decide the issues that were pointed out on this investigation, this paper endeavors to develop another neural network approach (ANN) by executing Bio-breathed life into formative computation and fake neural Network for less than ideal distinctive verification of Liver Disorder robotized system to the welfare of human culture.

A. Modified Feed forward based ANN classifier

In feed-forward systems, a yield sign of a neuron has no more effect on its information – the sign are proliferated uniquely forward. Our proposed is ANN be comprised of at least two layers, each layer covers various neurons. The neurons are joined by coordinated curves associations. The neurons and circular segments build up the system topology. Each circular segment has numerical weight that speaks to the impact among two neurons. Positive loads point towards fortification and the negative loads indicates hindrance. System for the arrangement of Liver malady dataset The work process of ANN execution in the interest of the grouping of dataset is as appeared in figure 3 which conveys brief clarification of significant stages that must be watch out for to utilize ANNs for the benefit of the forecast of Liver illness dataset.

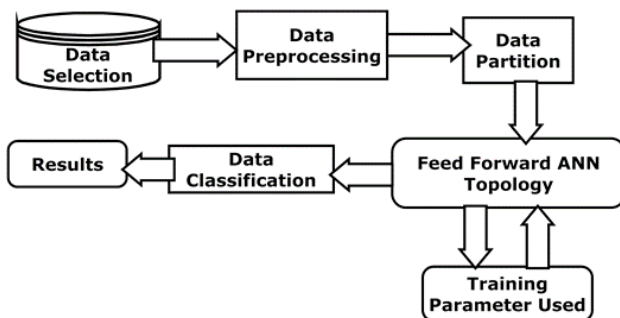


Figure 3. Workflow of M-ANN

Input: Selected Features, labels of the features (L)

Output: Classified Output

Procedure

Let N_F be the number of features

Let N_R be the number of Records

$T_R=70\%$, $T_S=1-T_R$

Split the data to Training and testing set

$Train_R=N_R(T_R)$

$Test_R=N_R(T_S)$

Initialize N_F, N_C, N_H, N_I

Where N_F is the number of input layers

N_C is the number of class labels and the output layers

N_H is the number of Hidden layers

N_I is the number of Iterations

Set $lrate=50\%$ and $cycles=10$

Initialize random Weight (W_x, W_y), velocity variation (V_x, V_y)

Probable weight and Velocity are

(PW_x, PW_y, PV_x, PV_y)

For $l=1$: cycles

Initialize $lrate$

For $i=1$: N_I

Compute $Nv1=Train_R(i)*W_x -V_x$

Compute $Nv2=Train_R(i)*W_y -V_y$

Compute $H_n(i) = (e^{Nv1} - e^{-Nv1}) / (e^{Nv1} + e^{-Nv1})$

Compute $O_n(i) = (e^{Nv2} - e^{-Nv2}) / (e^{Nv2} + e^{-Nv2})$

$Er = \sqrt{\sum_{i=1}^N (Train_R - O_n)^2 / O_n}$

Normalize Er measure $\theta O_n = (1 + O_n) * (1 - O_n) * ((Train_R - O_n)$

$\theta W_y = O_n * \theta O_n$

$\theta(Nv1) = (1 + O_n) * (1 - O_n) * (\theta O_n * \theta W)$

$\theta W_x = O_n * \theta O_n * (Train_R * \theta(Nv1))$

//Update weight

$W_y = W_y + \theta W_y + PW_y$

$W_x = W_x + \theta W_x + PW_x$

Update Probable weight $PW_y = \theta W_y * Er$

$PW_x = \theta W_x * Er$

Update velocity

$V_y = Er * V_y * W_y$

$V_x = Er * V_x * W_x$

End for i

B. Information Collection

Database ought to contain enough number of various cases to be prearranged as a contribution to the preparation organize. The gathered medicinal information data should comprise of adequate pool of various sorts of liver maladies that could aid the assessment of proposed framework. Each arrangement of information need to assess a measure of occasions, an amount of qualities and various classes.

C. Pre-preparing

Pre-preparing strategy which is utilized to expel the irregularities of the information in the liver Dataset before the assessment of ANN algorithm. The deficient information debase the exhibition of the proposed calculation.

Liver Disease Classification using Deep Learning Algorithm

Information Partition Data parcel that isolates the information into 2 sections; preparing and testing. The exchange work, concealed hubs and the amount of info and yield hubs ought to be named in this segment.

D. Counterfeit Neural Network Training

The artificial neural network is figured with Probable weight and Velocity learning criteria. The framework includes 11 neurons in transit of 11 attributes in the dataset. The framework is organized with two classes addressed as matched structure to such a degree, that 0 exhibits as normal individual and 1 shows Abnormal. The yield layer includes of two neurons to suggest these two classes. The time multifaceted nature of the Modified ANN is O(Reduced Feature set C) addressed in cycles and getting ready dataset size Ts.

So the unpredictability is =O(C+Ts)

In general framework unpredictability is =O($\sum_{i=1}^n I_{[J \times K]}^+$) O(C+Ts)

III. PERFORMANCE ANALYSIS

A. Data Set Description

We assessed our calculation in Indian Liver Patient (ILPD) Dataset. This informational index covers 416 liver patient records and 167 non liver patient narratives. The informational index was gathered from north east of Andhra Pradesh express that lives in the south piece of India. Selector is a class mark which is utilized to arrange the gatherings into (liver patient or not). This informational collection contains 441 male patient records and 142 female patient records. It has 11 qualities which impact in causing Liver issue. This properties were chosen dependent on LFT test outcomes.

Table 1. Data set of Indian liver patients'

No	Attributes
1	Age
2	Gender
3	TB Total Bilirubin
4	DB Direct Bilirubin
5	Alkphos Alkaline Phosphatase
6	Sgpt Alamine Aminotransferase
7	Sgot Aspartate Aminotransferase
8	TP Total Protiens
9	ALB Albumin
10	A/G Ratio Albumin and Globulin Ratio
11	Selector field used to split the data into two sets

We examine the exhibition of proposed work with regards to Specificity, Accuracy and affectability which is required to quantify the grouping execution and it is contrasted and other existing systems incorporates C5.0 and CHAID. For examining the exhibition of the proposed ANN with Liver Disease Database, the test information is accepted as the contribution to the relating prepared system and the yield of

the net is planned by method for balanced loads. Table 1 speak to the ILPD characteristics taken for the Liver sickness grouping. Here we use SPARK device that adequately support with Big information mining and AI approach. The genuine yield of the previously mentioned net is connected by method for the favored yield to explore the learning capacity of the system for recognizing the Liver malady information.

B. Performance Measures

Accuracy of the liver disease

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

Table 2. Comparison of Classification Algorithms

Methods	Sensitivity	Specificity	Accuracy
SVDD-GSO	95	85.28	83.28
SVDD	87.46	86.85	83.34
ANN	71	77.46	73.22
SVM	73	83.26	79.66
ANN	96.25	95.25	95.409

Table 2 demonstrate different execution proportions of order calculations. The outcome demonstrates that our proposed ANN accomplished better outcomes when contrasting and different methodologies.. The ANN utilizes neurons to successfully order the maladies in two cases. The joined methodology of proposed ANN beats all the current calculations in execution measure. Accuracy

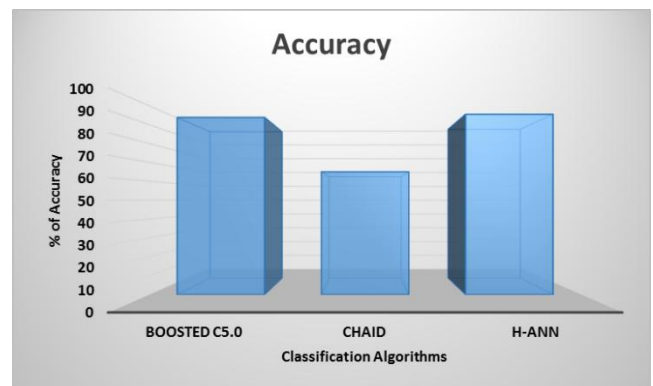


Figure 4. Accuracy Comparison of HANN on Liver Datasets

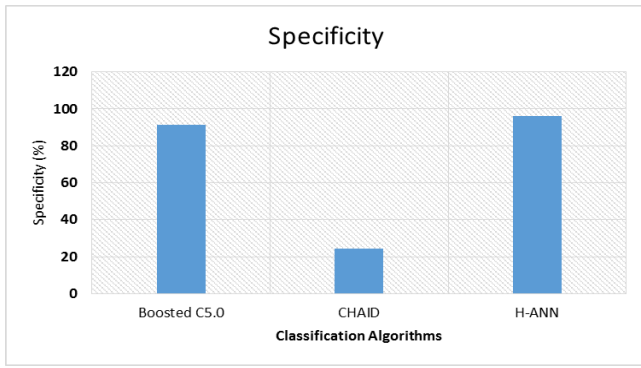


Figure 5. Comparison on specificity for liver datasets

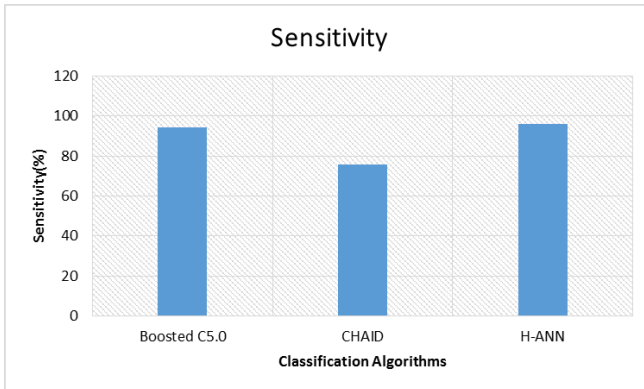


Figure 5. Accuracy Comparison of HANN on Liver Datasets

Figures(4) (5) (6) demonstrates a correlation of execution measurements, for example, precision, particularity and affectability with existing strategies such like Boosted c5.0, CHAID and proposed technique. The proposed strategy demonstrates effective outcome for Liver dataset when contrasted with different strategies. Obviously the proposed strategy in (Figure. 4) for Liver Disease dataset demonstrate the productive consequence of early conclusion precision of accurately ordered liver infections. The proposed HANN accomplishes 95.4% precision which is higher than the current Boosted c5.0, CHAID 93.75%, 65% separately. The precision of the proposed HANN accomplishes 2% higher than the current best Boosted c5.0.

Specificity of calculations are appeared in Figure. 5. The particularity estimations of Boosted c5.0, CHAID and HANN is 91.42%, 24.24% and 96.25% separately. Our proposed calculation accomplished 5% more noteworthy explicitness execution contrasting and Boosted c5.0 calculation.

Sensitivity of calculations are appeared in Figure. 6. The affectability estimations of Boosted c5.0, CHAID and HANN is 94.4%, 75.59% and 96.25% individually. Our proposed HANN approach accomplished 2% more noteworthy affectability execution contrasting and Boosted c5.0 calculation.

False Negative Rate (FNR)

False negative rate is the extent of positives which yield negative test results with the test. By and large it very well may be determined as 1-affectability.

$$FNR = \frac{FN}{TP + FN}$$

False positive Rate (FPR)

It estimates the grouping of the liver illness which are negative, when the condition is false. It is perceive as false - positive rate.

$$Specificity = \frac{TN}{TN + FP}$$

Fig 7 describes the values of the False positive algorithm algorithms. The obtained graph shows the ANN has better performance than other algorithm

Precision and F1 measure Accuracy and specificity are another much of the time utilized arrangement of execution estimates that can be determined. Appraisal classifications are determined by the equations

$$Precision = TP / (TP + FP)$$

$$F1 = 2TP / (2TP + FP + FN)$$

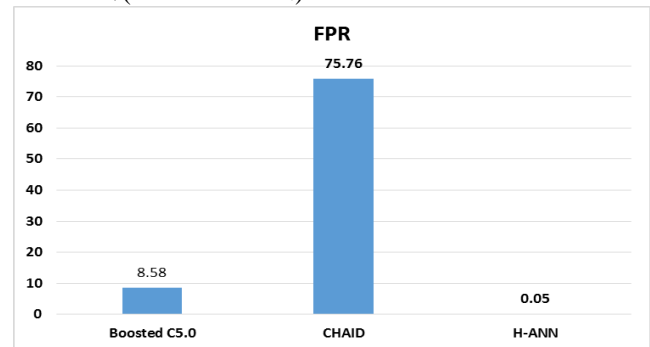


Figure 7. Comparison of FNR Predictive Measure

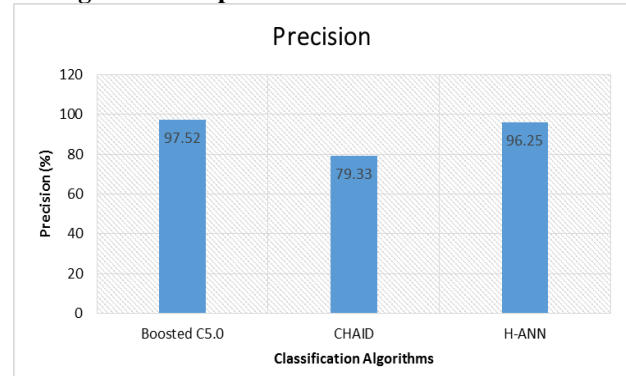


Figure 8. Comparison of Precision Predictive Measure

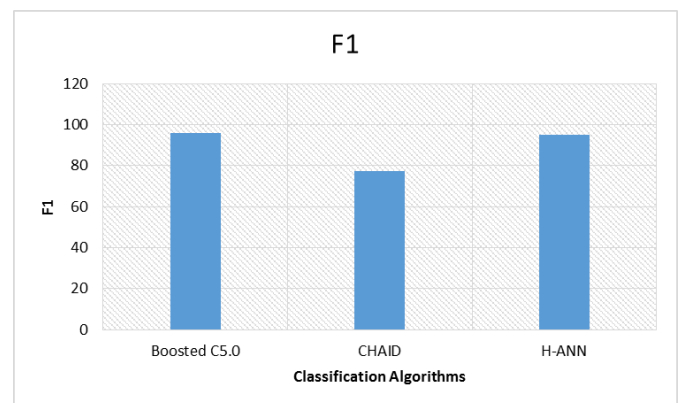


Figure 9. Comparison of F1 Predictive Measure

The figure (8) (9) shows the evaluation estimations of Precision and F1 Measures.



The helped c5.0 accomplished 97.52 % and our proposed calculation accomplished 96.25%. in the estimating purpose of Precision. The F1 measure consequences of our proposed calculation is 96.16% and supported C5.0 evaluated as 95.93%. According to our exploratory outcomes and just as other existing examinations on early forecast of liver malady, proposed ANN had adequate execution.

IV. CONCLUSION AND FUTURE WORK

The early conclusion of liver issue is a critical true restorative issue in which characterization of liver ailment assumes a significant job and can be utilized in the medicinal field for right grouping. This exploration work presents computational insight systems for early expectation of Liver issue dependent on ANN algorithm and Evolutionary calculations. It actualizes the similar investigation of proposed and existing strategy for improving expectation exactness of liver illness patients in which preprocessing method is utilized to expel the clamor from the given Liver malady dataset. These chose highlights are additionally utilized for arrangement of liver ailments which are accomplished by utilizing the characterization procedure of proposed ANN. The proposed technique is assessed by persuaded execution metric, for example, precision, affectability and particularity to quantify the order execution. The trial results demonstrate that proposed strategy indicates proficient arrangement of liver sicknesses and they show the early forecast of liver illness when contrasted with other existing techniques. In future the exhibition of characterization of liver based ailments will additionally improved by utilizing advancement systems and diagnosing the various sorts of liver sicknesses.

REFERENCES

1. T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Brucocoleri, et al., "antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters," *Nucleic acids research*, vol. 43, pp. W237-W243, 2015.
2. J. Ji, A. Zhang, C. Liu, X. Quan, and Z. Liu, "Survey: Functional module detection from protein-protein interaction networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 261-277, 2014.
3. J. D. Watson, R. A. Laskowski, and J. M. Thornton, "Predicting protein function from sequence and structural data," *Current opinion in structural biology*, vol. 15, pp. 275-284, 2005.
4. M. Deng, K. Zhang, S. Mehta, T. Chen, and F. Sun, "Prediction of protein function using protein-protein interaction data," *Journal of computational biology*, vol. 10, pp. 947-960, 2003.
5. J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *International Journal of Computer Applications*, vol. 17, pp. 43-48, 2011.
6. M.-J. Huang, M.-Y. Chen, and S.-C. Lee, "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis," *Expert systems with applications*, vol. 32, pp. 856-867, 2007.
7. M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. P. Kotsia, K. V. Vakalis, K. K. Naka, et al., "Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, pp. 447-458, 2008.
8. J. Xu and Y. Li, "Discovering disease-genes by topological features in human protein-protein interaction network," *Bioinformatics*, vol. 22, pp. 2800-2805, 2006.
9. B. Stapley, L. A. Kelley, and M. J. Sternberg, "Predicting the sub-cellular location of proteins from text using support vector machines," in *Biocomputing 2002*, ed: World Scientific, 2001, pp. 374-385.
10. H. Vilstrup, P. Amodio, J. Bajaj, J. Cordoba, P. Ferenci, K. D. Mullen, et al., "Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of

11. M. ABDAR, "A survey and compare the performance of IBM SPSS modeler and rapid miner software for predicting liver disease by using various data mining algorithms," *Cumhuriyet Science Journal*, vol. 36, pp. 3230-3241, 2015.
12. 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*, pp. 1-13, 2017.
13. M. Abdar, M. Zomorodi-Moghadam, R. Das, and I.-H. Ting, "Performance analysis of classification algorithms on early detection of liver disease," *Expert Systems with Applications*, vol. 67, pp. 239-251, 2017.
14. E. M. Hashem and M. S. Mabrouk, "A study of support vector machine algorithm for liver disease diagnosis," *American Journal of Intelligent Systems*, vol. 4, pp. 9-14, 2014.
15. S. Takkar and A. Singh, "Impact of Genetic Optimization on the Prediction Performance of Case-Based Reasoning Algorithm in Liver Disease," *International Journal of Performability Engineering*, vol. 13, p. 383, 2017.
16. S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms," *International Journal of Science, Engineering and Technology Research*, vol. 4, pp. 816-820, 2015.
17. N. Behera, A. Routray, J. Nayak, and H. Behera, "Bird mating optimization based multilayer perceptron for diseases classification," in *Computational Intelligence in Data Mining-Volume 3*, ed: Springer, 2015, pp. 305-315.
18. P. Tavakkoli, D. M. Souran, S. Tavakkoli, M. Hatamian, A. Mehrabian, and V. E. Balas, "Classification of the liver disorders data using Multi-Layer adaptive Neuro-Fuzzy inference system," in *Computing, Communication and Networking Technologies (ICCCNT), 2015 6th International Conference on*, 2015, pp. 1-4.
19. H. Jin, S. Kim, and J. Kim, "Decision factors on effective liver patient data prediction," *International Journal of Bio-Science and Bio-Technology*, vol. 6, pp. 167-178, 2014.
20. A. Onan, "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer," *Expert Systems with Applications*, vol. 42, pp. 6844-6852, 2015.
21. C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *Journal of Biomedical Science and Engineering*, vol. 6, p. 551, 2013.
22. E. Aličković and A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest," *Neural Computing and Applications*, vol. 28, pp. 753-763, 2017.
23. Y.-T. Cheng, Y.-F. Lin, K.-H. Chiang, and V. S. Tseng, "Mining Sequential Risk Patterns From Large-Scale Clinical Databases for Early Assessment of Chronic Diseases: A Case Study on Chronic Obstructive Pulmonary Disease," *IEEE journal of biomedical and health informatics*, vol. 21, pp. 303-311, 2017.

AUTHORS PROFILE



and infectious disease

Anand L received his M.tech from SRM university and currently doing his Ph.D. degree in VIT university in area of Big data Analytics under Dr Neela Narayanan in VIT University, Chennai .He is currently working has Assistant Professor in school of computing sciences at SRM INSTITUTE OF SCIENCE AND TECHNOLOGY His research interests include Big data analytics, Machine Learning



between VIT University and industries such as CDAC and DLink.

Neelanarayanan Venkataraman received his Master of Science in Computer Science from Madurai Kamaraj University, India in 1995 and PhD from IT University of Copenhagen, Denmark in 2012. Currently, he is an Associate Professor at VIT University, Chennai, India. He is heading the Cyber Physical Systems (CPS) department. He was instrumental for initiating joint research collaboration

He has published more than 30 papers in various peer-reviewed international conferences and journals. He has organised various national workshops, international conference and symposium. He was heading the Cloud Computing Research Group at VIT University from 2016 - 19, Chennai campus and six students are pursuing their PhD and one student has completed her PhD under his guidance. Before joining VIT University he has worked as a Scientist at Centre for Advanced Computing (CDAC), India and as a Lecturer in Madurai Kamaraj University, India and its affiliated institutions. He received the research award in VIT University for the year 2015 and 2018 for his achievements, exemplary commitment, dedication and motivation towards research publication