

# Feature Selection for Liver Disease using Particle Swarm Optimization Algorithm



L.Anand, V'Neelananarayanan

**Abstract-** Information Mining is one of the preeminent descriptive parts of automatic order and identification. It embroils information mining calculations and methods to inspect restorative information. Lately, liver protests have lopsidedly increased and liver diseases are complimenting one of the most human afflictions in various nations. Early determination of Liver Disorder is basic for the welfare of human culture. This grievance ought to be considered earnestly by setting up canny frameworks for the early analyze and anticipation of Liver infections. The robotized grouping framework endures with absence of exactness results when contrasted and careful biopsy. We propose another model for liver disorder order for breaking down the patient's restorative information utilizing cross breed counterfeit neural system. The therapeutic records are arranged whether there is a plausibility of presence of sickness or not. This proposed strategy utilizes M-PSO for highlight choice of information factors and M-ANN calculation for sickness order. The exhibited crossover approach improves the precision when contrasted with existing order calculations. This paper concentrates on feature selection on using PSO algorithm.

**Index Terms**—Liver, Pso, Spark, Classification.

## I. INTRODUCTION

Information mining is a basic methodology in deciding emitted information highlights, inclinations and structures in tremendous datasets, and convey acknowledged observations intrigued by the reason or condition of sure settings. Through uncovering the example and developments in the dataset, information mining additionally makes the figure of future procedures possible. The likelihood expectation may recommend the individuals to carry on with a quality life comparing with their bioinformatics information. By and by information mining conveys its commitment in medicinal space for the early conclusion of illnesses, hazard factor examination, Decision making, Treatment and remedy of medications. The concealed information record of patient is utilized for the forecast of infection in Classification period of information mining. This malady can be ordered utilizing different morphology factors. The precision of this kind of model give preferable outcomes over typical customary systems. Information mining can be connected in the field of bioinformatics in different applications such like [1] quality

finding, [2] protein discovery, [3]function theme identification, [4]protein work induction, [5]disease conclusion, [6]disease visualization, [7] illness medicine streamlining, [8] protein and quality participation organize recovery, information laxative, and [9]protein sub-cell position figure.

Current evolvement in information mining study has coordinated to the updating of a few strong and available methodologies for unearthing captivating examples in colossal medicinal databases. Current evolvement in science, restorative science, and DNA innovation has coordinated to the gradual addition of wonderful Volumes of bio - therapeutic information that nerves in the interest of complete examination. Information mining and bioinformatics consolidates to accomplish adequate mining of biomedical information.

The information mining procedure could be utilized to discover the likelihood of illness risks by grouping calculations. The mining procedure uncovers a few emitted and unusual examples, which may not engaged previously. Various information mining methodologies had been proposed so far for the forecast and distinguishing proof of liver infections. By and large the Liver Disorder could be as fatty liver, hepatitis, cirrhosis and Liver cancer. The precision of foreseeing liver ailments with therapeutic information mining for the most part remains on the component choice. The test is to create and approve conventions connected to clinically important result measures. Normally Liver illnesses are analyzed in the clinical by breaking down the degrees of catalysts in the blood. Be that as it may, Information concentrated measurable research on Liver ailments empowers to anticipate the survival examination of the illness. It is a challengeable undertaking to anticipate the episode with the utilization of information mining approach. Enormous volume of gathered therapeutic dataset draws in the mining investigator to discover the malady designs and their connections. This factual mining approach need to dissect enormous number of variable components and restorative authentic information. The compelling programmed Computer Aided Diagnosis framework ought to order the liver malady dataset that was gathered from different hepatologist from different nations. The honesty of the arrangement systems endures with the nature of information. The anomalies and imbalanced bioinformatics dataset bothers the solidness and execution of the forecast models. This expectation information mining calculations needs to confront the accompanying difficulties.

- Selecting relevant features to develop the proposed model.
- Identify and end of exceptions with the bioinformatics information.
- Managing the out of line information in arrangement.

Manuscript published on 30 September 2019

\* Correspondence Author

L.Anand\*, School of computing science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

V'Neelananarayanan, School of computing science and Engineering, SRMIST, Chennai, Tamil Nadu, India

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

• Differentiating ordinary hereditary highlights and malady morphological elements.  
The fundamental thought process of this exploration work is to give a viable CADS to analysis and guess of Liver infection.

It ought to be successful so that gives dataset Pre-preparing, characteristic choice and exact classifier. The classifier includes an energetic job in ordering the liver issue into existing classes of maladies. The effective indexing of the liver ailment, bio-propelled calculations are utilized for straightforward grouping.

Bio-propelled methods are the meta-heuristics approach that attempt to resemble the nature inventiveness for the benefit of settling streamlining inconveniences. The determined wellness worth group the kind of malady that could be feasible for the applicable medicinal data. The principle commitments of this examination are as per the following:

• A new half and half structure is intended to discover the odds of happening liver sicknesses dependent on simple Algorithms.

• MPSO calculation is utilized for the choice of hazard highlights from the restorative dataset.

• HANN characterization is utilized for the forecast of liver infection, for example, typical, greasy, and liver.

Rest of the areas of this paper are sorted out as different explores on liver ailment arrangement based certain systems and its impediments has been assessed in segment 2, segment 3 displays the nitty gritty depiction of proposed approach incorporates M-PSO work and gives future upgrade.

## II. RELATED WORK

A few hazard elements are resolved for the reason for Liver Diseases. Hazard components can be characterized that raise the happy of achieving liver illnesses. Despite the fact that every single human has the opportunity to be influenced by the Liver issue, these variables impact more. A portion of these hazard elements can be modifiable and some can't be changed. [10] Liver capacity tests are utilized to gauge explicit compounds and proteins in your blood. The aftereffects of the degree of chemicals or proteins that uncovers in the test (Whether higher or lower) could show the state of the liver. The demonstrating levels of these compounds or proteins are utilized as traits in the dataset for executing information mining calculations. Abdar [11] overviewed a few information mining approach execution and their assessment results with basic liver malady dataset. This work additionally expressed different sorts of liver sicknesses among 100 divergent types of liver ailments. This work analyzed C4.5 calculation, C5.0 calculation, SVM calculation, KNN calculation, CHAID and Neural Network calculation to foresee the liver sickness utilizing Indian Liver malady dataset. It considered 11 traits for information mining instrument execution. It examined different quality and disadvantages of these calculations and reasoned that the exactness of neural system remains in the middle of 70% to 80%. This work saw the adequacy of neural system approach in liver illness expectation and early finding. [12] proposed untimely assurance of liver ailment over and finished with a MLPNN process focused on incalculable choice tree procedures like C5.0, HAID and CART with boosting method. It utilized the measurements (explicitness, affectability, exactness, false positive rate (FPR), false negative rate (FNR), F1, and precision) for the appraisal of

calculations. The hybridization of B-C5.0 and MLPNN strategies provided most extreme location pace of liver disorder identified with further frameworks. [13] utilized choice tree based calculations such like Boosted C5.0 and CHAID calculations to shape rules from liver sickness dataset. It established that C5.0 could be very much embraced with multipurpose classifier alongside boosting capacity. Anyway these techniques are over and over affected toward parting that characteristics have countless levels. This work started consolidating neural system and C5.0 calculations both with booting approach. [14] connected Support Vector Machine calculation to ILPD and BUPA Liver issue. This work read straight piece work for highlight choice. The exactness accomplished in this previously mentioned exertion is less. [15] recommended another thought that consolidated both Case-Based Reasoning (CBR) philosophy fortified by Genetic Algorithm (GA) to streamline the expectation consequences of liver sickness. The loads were allocated to the highlights as per their effects for causing liver malady. The GA registers wellness esteem that compute each instance of liver sickness research center elements to anticipate whether the individual experiences liver infection or not. This half breed technique accomplished 94.19% of most noteworthy exactness. [16] utilized SVM and Naïve Bayes Algorithms for the forecast. This work expressed that utilizing Naive classifier limit the required highlights of the patient's therapeutic record. [17] offered another flame fly calculation (FA) and winged animal mating streamlining (BMO) based MLP strategy for the infection visualization. FA includes in looking at determined yield and the normal yield to discover the mistake rate, for example, mean square blunder and total of squared mistake. The UCI respiratory liver infection dataset is assessed utilizing this methodology and MLP – BMO assessed as better. [18] connected versatile Neuro-Fuzzy deduction framework for liver issue characterization. Five layers are utilized for assurance and back spread learning principle is connected to alter the parameters in the shrouded layer. The sound and undesirable liver information were isolated viably. [19] considered the significant elements that predominantly immediate to the Liver issue. This work contemplated the helpfulness of various information mining calculations such like Naïve Bayes, Decision Tree, Multi-Layer Perceptron and k-NN. It gave a reasonable course map and assessing criteria to characterize the Liver ailments. [20] Fuzzy-harsh case choice expel the clamor of information and works quick for little preparing sets. At that point consistent quality based characteristic grouping is connected. It managed the vulnerability information great and helped with loud information. It endures with enormous preparing information. [21] offered Hybrid methodology of irregular woods classifier and highlight determination practice characterizes the ailment information. If Good trait choice methodology yet littler gatherings are favored over bigger gatherings. [22] GA is utilized for removing from accommodating and vital highlights. Discrete individual and various classifier coordination were utilized to make accurate technique for illness recording. [23] Rich arrangement of consecutive principles are utilized for the characterization calculation. Least help check is utilized to coordinate the successive examples.

The significant disadvantages of this methodology is conflicting and inadequate records because of unpredictable visits of patients influences consecutive process. It additionally endures with Imbalance and Post-examination issue.

From the overview, the serious issues that influences the early analyze of liver issue are dissected, and furthermore the aim of the past works are explored with its very own benefits and bad marks. Normally, the current instruments accepted that the exactness of forecast is accomplished.

However, this wasn't the situation at that point, so it must be improved further to build the arrangement precision. Additionally, some other research works tended to these issues by presenting effective blend

Models dependent on highlight determination and arrangement, yet it raised some different issues such like

- More Time multifaceted nature of calculation.
- Some methodologies are appropriate just for little information.
- Some mix of classifier over fit with informational collection and some will be under fit with informational index.
- The reliance of occasions and credited are uncategorized.
- Some methodologies are not adoptable for continuous gathering of database usage.

To determine the issues that were called attention to on this study, this paper attempts to build up another cross breed approach (HANN) by executing Bio-enlivened developmental calculation and counterfeit neural Network for untimely distinguishing proof of Liver Disorder robotized framework to the welfare of human culture.

### III. PROPOSED METHOD

Fig.1 represents the feature selection methodology for the attributes using M-PSO algorithm .Generally, The PSO is used to select the optimal features . The features are chosen dependent on its velocity factor and the difference will be calculated based on the velocity factor. Here we present investigating conditions of properties. In the event that a trait has higher impact in causing liver issue, at that point the reliance factor is broke down which likewise has higher impact. The inertia weight are given to the features with respect to its impact in causing liver disease. Another change should be possible with suitable parameter modification that supports molecule swarm streamlining to upsurge the viability of highlight choice. Each features has numerical weight that speaks to the impact among two features. Loads which are positive point towards fortification and the loads which are negative means restraint. The weight is standardized among speed with loads with for calculating its accurate relating point This gives a far reaching answer for early liver issue determination. M-PSO is the furthestmost detectable data driven methodology among different populace focused stochastic enhancement strategies. It is anything but difficult to apply, gives the great clarification of every single yield

and there is no over fitting. Streamlining of loads of qualities and arrangement of pertinent events for M-PSO is done at the same time in this proposed model. The restorative data is given as contribution to the framework. M-PSO chooses the ideal highlights and store it in the learning database. The

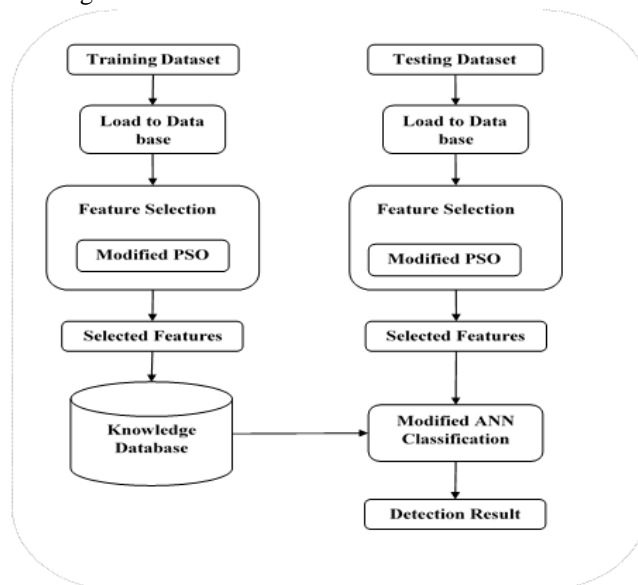


Figure 1. Workflow of the Proposed System

help of this calculation is decreased unpredictability and execution time. H-ANN calculation includes in the analysis of liver issue with the medicinal dataset through the M-PSO chose highlights. The selected dataset training and testing are provided as the contribution to the changed arrangement calculation. It discover the people who are influenced by the ailment. The rest of the piece of this segment depicts the feature selection and classification in the next work.

#### A. Altered Particle Swarm Optimization

Highlight space of an order issue is a key factor inciting the presentation of a grouping calculation. Without earlier learning, the previously mentioned assignment is trying to figure out which highlights are valuable. In this way, a colossal measure of highlights are normally familiarize with the dataset, including suitable, improper and laid off highlights. Notwithstanding, immaterial and laid off traits are not valuable in the interest of grouping. The table.1 speak to the significant traits that are typically utilized in the (LFT). M-PSO is prescribed to decrease the amount of highlights that could be utilized for the improvement of the order calculation. The wellness esteem calculation is the essential advance of appraisal practice. The situation of a molecule symbolizes a chose trait subset. The unsullied highlights that does not fulfill the wellness of the benchmarks is disposed of from the preparation set and the preparation set is changed to another preparation set. The characterization execution of the named highlights is evaluated on the changed preparing set.

## Feature Selection for Liver Disease using Particle Swarm Optimization Algorithm

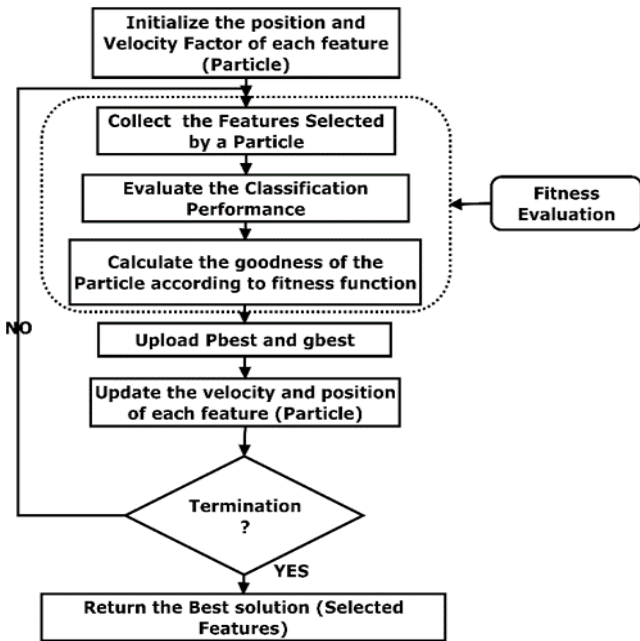


Figure 2. working model of M-PSO

Through the predefined wellness work the ideal goodness worth is determined for each every quality (Particles) with respect to the order execution. After wellness assessment of whole particles, the framework fill in the p best and g best, and after that raise to date the speed and position of each property. The calculation stops when a predefined consummation condition, for example the careful going measure of emphases or an ideal wellness esteem, has been experienced.

Fig- 2 displays the working methodology of our M-PSO calculation. In our m-pso include determination, the portrayal of an extricated molecule set is characterized by F. N be the no of attributes for each record. The billet an incentive in the measurement j is meant has P\_fit of display probability is characterized by nearness designated. The limit  $[(max)]_{BF}$  is used to

### Modified PSO algorithm

**Input:** Extracted Features

**Output:** Best Selected Features

**Procedure:**

F = Features Extracted

N be the number of Features for each record

M be the Number of Iterations

$V_{Fr}$  = Velocity Factor

// Compute the minimum and maximum range of the particles available.

For i= 1: M were M is the number of Feature Attributes

For j= 1: N were N is the number of particles

$$Par_{List} \leftarrow F(M)_N$$

End for j

$$RF_{mn} \leftarrow \text{Min}(Par_{List}) \quad // \text{Computing}$$

minimum value of particles in each attributes

$$RF_{mx} \leftarrow \text{Max}(Par_{List}) \quad // \text{Computing maximum}$$

value of particles in each attributes

End for i

// Generate Fitness Parameters

For k=1: size (RF)

$$V(F)_{mx} \leftarrow RF(k)_{mx} * V_{Fr}$$

$$V(F)_{mn} \leftarrow (-RF(k)_{mx} * V_{Fr})$$

End k

Initialize min and max weights as  $W_{g_{mn}}$  and  $W_{g_{mx}}$  and also C1, C2 and other random numbers

For Z= 1: No of Itr

For x=1: No of Features

For y=1: No of Particles

Compute Fitness of each particle by the below

equation

$$P_{Fit} = [V(F)_{mx}(x, y) - V(F)_{mn}(x, y)] * P_y$$

// Compute particle velocity by the below

equation

$$\text{Velocity}(V_t) = (W_{g_{mn}} * V_t) + (C1 * r1 * (P_{Best} - P_{Fit})) + (C2 * r2 * (G_{Best} - P_{Fit}))$$

$$\text{If } P_{Fit} > P_{Best}$$

$$P_{Best} = P_{Fit}$$

End if

$$\text{If } P_{Fit} > G_{Best}$$

$$G_{Best} = P_{Fit}$$

End if

End for y

// Compute clustering & choose best Fit Particles

Initial Fit threshold ( $F_t$ ),

Set best Fitness=  $G_{Best}$

$$\text{Compute } \min_{BF} = b_{Fit} - \left(b_{Fit} * \frac{F_t}{100}\right)$$

$$\text{Compute } \max_{BF} = b_{Fit} + \left(b_{Fit} * \frac{F_t}{100}\right)$$

$$P_{Fit} \text{ Is between } \min_{BF} \text{ and } \max_{BF}$$

$$\text{Increment N}(P_{Fit})$$

$$\text{If } N(P_{Fit}) > (P_{Size} * (1 - F_t))$$

$$b_{index} \leftarrow \text{Add}(P_{index})$$

End if

End for x

End for Z.

decide if an element's determined best fit worth achieves the most extreme worth or not.

On the off chance that  $N(P\_Fit) > (P\_Size * [(1-F)]_t)$ , at that point the  $j^{th}$  Feature is named. Generally the element isn't nominated. The genuine PSO is altered with two optimized contemplations.

1. The interrelation of the highlights are resolved dependent on reliance examination in the progression of figuring least and greatest estimation of particles in every quality.

2. The places of the previously mentioned particles that depends in the plausible space just needs to ascertain pBest and gBest values. For the most part the result of PSO relies upon the parameter settings such like inactivity weight and speeding up variables  $c_1$  and  $c_2$ . The outcomes could be effectively enhanced by making appropriate characteristic qualities. In the event that the parameter adjustment is excessively little, the molecule development is in addition little. This set-up besides make the outcome as valuable information, however appeared to be additional tedious. In the event that the change is extraordinary, molecule development will likewise be outrageous. This will make the methodology break down right on time, with the goal that a valuable list of capabilities can't be procured. Henceforth, suitable parameter alteration underpins molecule swarm improvement to upsurge the adequacy of highlight choice.

We center the factor latency weight  $W$  that looks as significant for the PSO's combination conduct. The latency weight is connected to oversee the impact of the previous vestige of speeds continuing the present speed. As a result, the parameter  $w$  alters the parity among the worldwide ideal and nearby examination capacities of the swarm. So we require to set huge dormancy weight at first to inspect the new looking through regions. This worth ought to be diminished step by step to accomplish refined arrangement. The loads of the traits could be acquired from the LFT and their extents. This trait esteems demonstrates their effect in liver issue. The time unpredictability of the tuned PSO is  $O(I \times J \times K)$ , where  $I, J, K$  are the measure of emphases, the measure of Features, the measure of particles separately. In the most pessimistic scenario event, if the quantity of properties stays unaltered and the amount of cycle achieves the decided emphasis number, the time multifaceted nature is  $O(I \times J \times K)$ . The quantity of highlights will be diminished to a critical level after specific cycles, subsequently the time multifaceted nature is  $O(\sum_{i=1}^J J \times K)$ , where  $1 \leq J \leq K$

#### IV. CONCLUSION AND FUTURE WORK

The programmed early conclusion of liver issue is a critical certifiable therapeutic issue in which order of liver ailment assumes a significant job and can be utilized in the restorative field for right characterization. This work presents optimized procedures for early expectation of Liver issue by feature extraction It actualizes the near examination of proposed and existing strategy for enhancing expectation of liver sickness patients in which preprocessing method is utilized to expel the clamor from the given Liver infection dataset. M-PSO is utilized for credit choice to decide the careful highlights that had lead to make a conclusion of illness precisely. The proposed strategy is assessed by persuaded execution metric, for example, exactness, affectability and particularity to quantify the arrangement execution. The exploratory outcomes demonstrate that proposed strategy demonstrates effective arrangement of liver sicknesses and they show the early forecast of liver ailment when contrasted with other existing strategies, the

goal of this paper is to extract features and in next work to classify the those optimized features.

#### REFERENCES

1. T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, *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 Liver Diseases and the European Association for the Study of the Liver," *Hepatology*, vol. 60, pp. 715-735, 2014.
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.

### AUTHOR PROFILE



**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 and infectious disease



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 between VIT University and industries such as CDAC and DLink. 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