International Journal of Research in Advent Technology, Special Issue, NCRCEST 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Machine Learning Based Prediction for Liver Disorder with Optimized Attributes

[1]R.viswanathan, [2]T.Akilan, and [3]K. Kalaivani.
[1,3]Vels Uinversity Chennai, [2]Galgotias College of Engineering and Technology Greater Noida.
rvnathan06@gmail.com, agilanmecse@gmail.com, hodcse@velsuniv.ac.in,

Abstract—Machine learning is a branch of efficient view of uses a range of facts, probabilistic and development techniques that empower PC's to "understand" from past cases and to differentiate hard-to-perceive outline from broad, hysterical or difficult educational records. This boundary is specifically suitable to valuable applications, particularly those that are subject upon tedious proteomic and genomic calculations. In this way, machine learning is usually used as a piece of growth assurance and confession. All over, starting late machine learning has been incorporated with harm outline and approximation. This last approach is particularly captivating as it is a part of a evoking design towards tweaked, perceptive drug. In this paper, a new machine learning algorithm has been chosen from a set of machine learning algorithms which is based on cost effectiveness and is used to minimize the number of attributes in the dataset with minimum error rate and in high accuracy. Moreover, this paper associated with the genuine datasets to predict liver disorders with genuine parameters using Pearson correlation. This paper can be used well on lightweight device like phones or tablets to pick more specifically with fundamental parts.

Keywords—Pearson Correlation, LDD (Liver Issue Dataset), Support Vector Machine (SVM), Artificial neural systems (ANNs)

1.INTRODUCTION

Liver problem alludes to any confusion of the liver [1]. The liver is extensive organs in the upper right abdomen that instruct the processing and expels dissipate items from the blood. Liver ailment incorporates the supplementary conditions as follows:

- Cirrhosis, or scarring of the liver
- Irritation (hepatitis) from irresistible (hepatitis B, hepatitis C) or non-irresistible causes (synthetic or immune system hepatitis)
- Tumors, benevolent and dangerous (liver disease)
- Metabolic scatters

Liquor consumption is one important reason for liver disease [4, 5]. Contaminations, harms, and hereditary conditions can cause maladies of the liver. In most patients with liver illness, with various typical elements of the liver are impeded. Liver illness is a problematic trouble to inspect given the nuance of the signs while in any case time spans. Issues with liver problem are not found until the point that it is as regularly as believable past the last crucial moment as the liver keeps working but not endure when halfway injured [6, 9]. Previous studies can be life-sparing. Not with-standing the way that not identified to even the superior therapeutic specialist. The early warning of these infections can be seen easily. Early finishes of patients can assemble his/her future generously. Along these lines the late outcomes of this examination are fundamental both from the perspective of the PC assessor and the therapeutic ace. Liver illness is any immensity of liver fact of captivity that causes distress. The liver is in-charge of different risky works within the body and it should conclude crippled or hurt, the loss of those points of confinement can make massive trouble to the body. Liver contamination is relatively recommended as hepatic illness. Liver issue is a moderate term that covers all the potential issues that cause the liver to negligence to play out its assigned breaking points. Ordinarily, over 75% of liver tissue should be influenced before a reduction in work happens. This paper deals with the view of liver menace by using a cost propel method for examination for lightweight devices like mobile phones or tablets by dropping the amount of necessary characters required for needy of the ailment without swap over off the bumble rate and accurateness. As the fundamental progress of dealt with the untreated dataset is wiped and prune back by emptying the illogical data events [7]. Then conceiving the data in light of the number (predicting a stimulus in dataset). The dataset for this issue is using LDD (Liver Issue Dataset) which is taken from the UCI Machine Learning Archive. Total number of cases is 345. It is a multivariate informational collection, contain 7 factors, all values are authentic numbers. The qualities of the dataset are as follows:

- Mean Corpuscular Volume (mcv)
- Alkaline Phosphatase (alk)
- Alanine Aminotransferase (Sgpt)
- Aspartate Aminotransferase (sgot)
- Gamma-Glutamyl Transpeptidase (gammagt)
- Drinks number of half-16 ounces counterparts of mixed drinks alcoholic every day (drinks)
- Selector field made by the BUPA scientists to part the information into prepare/test sets (selector)

Discovering the Pearson correlation with the coefficient by making the property as cross section. Then finding the join with high association and store the properties [8]. And the exceptionally corresponded variable will help us in future to lessen the quantity of traits. And afterward we locate the straight relapse (cost streamlining by discovering theta esteems) to get the new qualities by making utilization of very associated properties and assess its mistake rate and exactness. We rehash the means the same number of number of times as required and perform grouping to diminish the

mistake rate and to expand the precision of the recently discovered qualities.

2. RELATED SYSTEM

The several related frameworks exploit algorithms, for example, Support vector machine, Naive-Bayes classification, Random Forest Boost algorithm and Neural net. Witnessing a few frameworks utilizing these methodologies in detail.

A. Support Vector Machine

First, In machine learning, Support vector machines (SVM) are managed to learning the models with corresponding related learning processing that cut-down the information to be utilized for arrangement and deterioration analysis [10, 11]. Given an preparation of arranging cases, each set separately having a position with each one of two classifications, a SVM arranging calculation manufactures a design that distribute out new circumstances to one case or the other case, creating it a non-probabilistic similar direct classifier (despite of the detail that techniques, for example, Platt scaling be existent to access SVM in a probabilistic categorization setting). A SVM display is a interpretation of the cases as emphases in space, correlated with the aim that the circumstances of the various classifications are isolated by a sensible hole that is as widespread as could be likely under the conditions. Recent illustrations are then connected into that similar space and anticipated to have a space with a classification in which side of the hole they fallen.

Nevertheless performing straight procedure, SVMs can productively play out a non-direct representation utilizing, known as the piece trap, verified mapping their assistances to high-dimensional element spaces [14, 15]. At that circumstances the point when data are not named, administered learning isn't acceptable, and an unconfirmed learning approach is needed, which undertake to identify regular gathering of the information to collecting and after that assist new information to these shaped collections. The help vector bunching calculation made by Hava Siegelmann and Vladimir Vapnik, put on to the insights of help vectors, formed in the help vector machines calculation, to make an order for unlabeled information, and is a standout among to the most commonly utilized grouping calculations in mechanical applications.

Michael J Sorich [2] discovered that SVM classifier produces best perceptive execution for the synthetic datasets. SVMs are set of associated managed learning strategies consumed for order and re-order [3]. They have a place with a group of add up direct order. A unique property of SVM will be, SVM all the while limit the experimental order error and enhance the geometric edge. So SVM called Most extreme Edge Classifiers.

The error matrix for support vector machine is given below

TABLE I. SVM

	Predicted		
Actual	1	2	Error
1	8	16	66.7
2	4	23	14.8

Overall error rate is 39.2%

B. Random Forests

Random Forests is a machine learning relapse technique for portrayal that enterprise by forming liver data into an extreme number of choice trees at planning time and resilient the class that is the plan for the classes yield by solitary trees [4]. It is unexcelled in accurateness among present computations. It yields sequence of action profitably on wide liver dataset. It can deal with limitless attributes without variable eradication. It gives evaluations of what factors are necessary in the collection. Random Forests builds up different gathering trees [16]. To make another liver inquiry from a data vector, set the data vector down each individual of the trees in the timberland. Individual tree gives a gathering, and says the tree "votes" for that class. The forest picks the demand that having the salient votes. For the liver disorders dataset the error matrix that is obtained when Random forest is applied to it is as follows

TABLE II. RANDOM FOREST

	Predicted		
Actual	1	2	Error
1	11	13	54.2
2	2	25	7.4

Overall error rate is 29.4%

C. Decision Tree

A decision tree may be a decision help mechanism that uses a tree like graph or model of decisions and their possible outcomes, comprised with happening comes about, resource expenditures, and utility. It's an outline to show related degree algorithmic elect that only contains prohibitive organization expressions. Decision trees or regularly used in examine, especially in call examination, to assist choose a technique possibly to acquire an objective, Moreover, a preferred instrument in machine learning [12, 13]. A decision tree could be a flowchart-like assembly inside which each internal center addresses a "test" on Relate in Nursing attribute, each division addresses the ultimate outcome of the

check, and each leaf center point addresses a grouping stamp. The systems from root to leaf address assemble to runs the show. In call examination, tree whatsoever, thus the solidly related effect outline locality unit used as a noticeable and consistent decision help gadget, wherever the normal estimation of powerful elective region unit figured [17, 18]. Utility limits, decision trees, affect charts and differing call examination mechanical assemblies also, methods district unit taught to class kid understudies in resources of business, success political economy, and overall prosperity, and zone unit trial of research or organization science philosophies. For the liver syndromes dataset the error matrix that is obtained when decision tree is applied to it is as follows

TABLE III. DT

	Predicted		
Actual	1	2	Error
1	14	10	41.7
2	4	23	14.8

Overall error rate is 27.4%.

D. Artificial Neural Network

Artificial neural systems (ANNs) are factual models exactly motivated by, and mostly displayed on natural neural systems [19]. They are more appropriate for displaying and controlling nonlinear connections between data sources and yields in parallel. The associated calculations are a piece of the more widespread field of machine learning, and can be consumed as a part of frequent applications as examined.

Artificial neural systems are defined by containing versatile weights along ways between neurons that can be tuned by a taking in calculation that increases from watched information keeping in mind the final goal to improve the model. Notwithstanding the learning calculation only, one must pick an appropriate cost work. The cost function is what is accessed to understand the perfect answer for the issue being illuminated [20, 21, 22]. This comprises deciding the finest esteems for the greater part of the tunable model parameters, with neuron way versatile weights being the essential focus, alongside calculation tuning parameters, for example, the learning rate. It's generally done through optimization methods, for example, inclination drop or stochastic slope plunge. These development methods basically attempt to make the ANN arrangement be as close as conceivable to the ideal arrangement, which when effective implies that the ANN can pay attention to the proposed issue with superior.

TABLE IV. ANN

	Predicted		
Actual	1	2	Error
1	8	10	55.6
2	8	25	24.2

Overall error rate is 35.3%

3. PROPOSED SYSTEM

By inspecting the results of the machine learning algorithms, decision tree algorithm gives us the minimal error rate and most accuracy. So, it has taken on for further process. Numbers of attributes are reduced in the dataset. For attribute reduction, Pearson correlation is found initially among the attributes. The highly correlated variables, this will help to know the essential attributes in the taken dataset. The Pearson correlation for the liver syndromes dataset is given below,

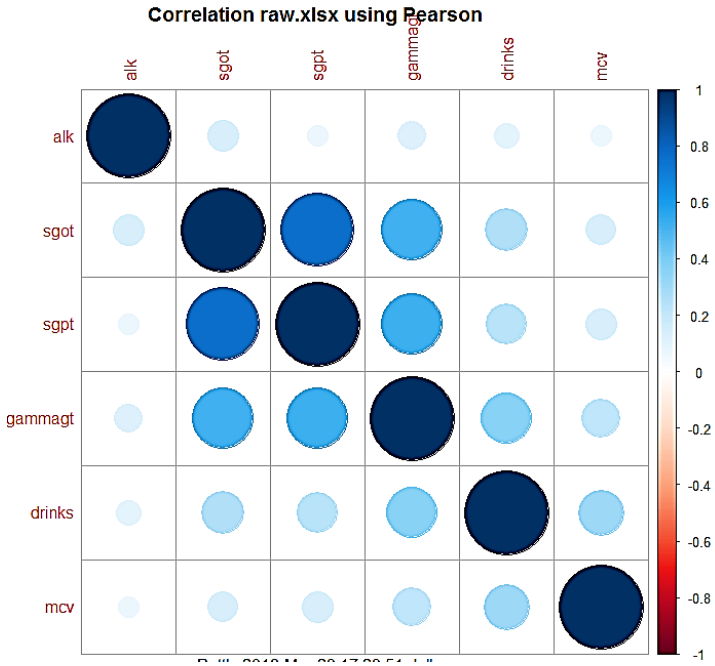

Fig. 1. Correlation Matrix

The Linear Regression is found among the highly correlated attributes and it try to get new attributes using the predefined attributes so that less associated attributes can be reduced. The linear regression is given by the formula:

$$Y = m\,x + c \qquad (1)$$

TABLE V. CLUSTERING WITH DT

	Predicted		
Actual	1	2	Error
1	8	8	50.0
2	3	27	8.4

Overall error rate 27.4%

After the minimization of attributes, the error rate and accuracy is obtained. If the error rate is high Clustering using Decision Tree is performed. The final results are obtained after processing the reduction of attributes and clustering using decision tree is given below:

4.RESULTS

The combined result when all the attributes are taken as input is tabulated below, and figure 2 shows that accuracy of Machine Learning Algorithm (all Attributes) and figure 3 shows the accuracy of Machine Learning Algorithm (alk, sgpt, sgot, gammagt attributes)

TABLE VI. ANN

S. No	Machine learning classification model	Class 1		Class 2		Accuracy		
		Accurate Sample	Total sample	Accurate Sample	Total sample	Class 1	Class 2	Overall accuracy
1	Decision tree	14	24	23	27	0.58	0.85	0.72
2	Neural network	13	24	22	27	0.54	0.81	0.68
3	Random forest	11	24	25	27	0.46	0.93	0.70
4	Support vector machine	8	24	23	27	0.33	0.58	0.50

Overall accuracy of Machine learning algorithm (all attributes) is depicted by using this Figure 2

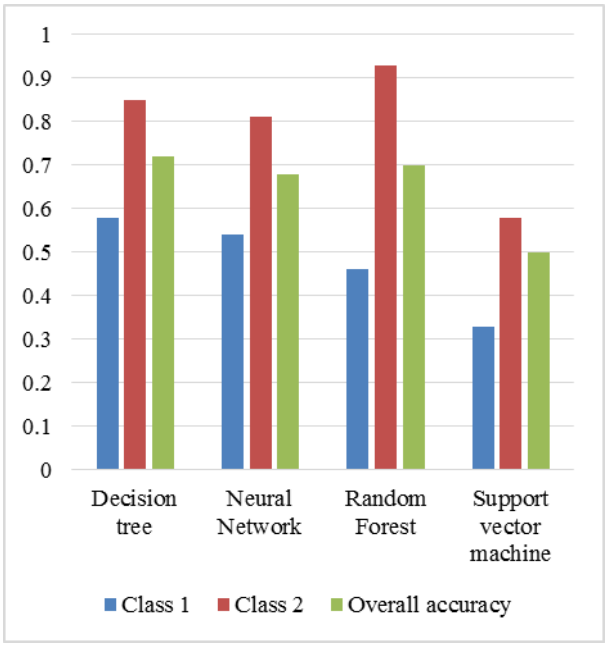

Fig. 2. Accuracy of Machine Learning Algorithm (all Attributes)

The results of the decisional attributes (alk, sgpt, sgot, gammagt) alone are tabulated below:

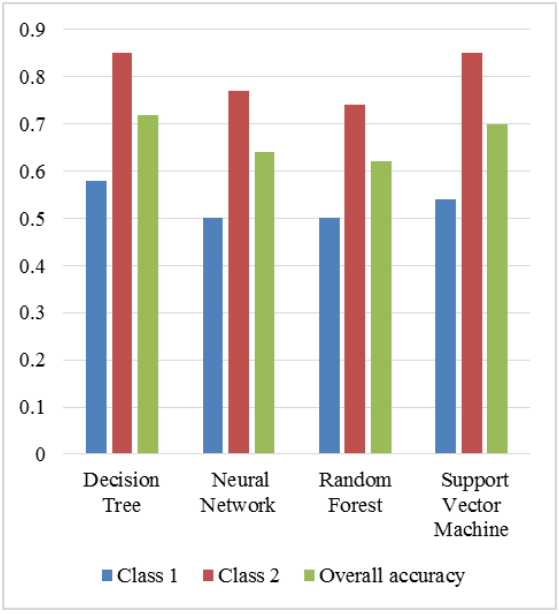

Fig. 3. Accuracy of Machine Learning Algorithm (alk,sgpt,sgot,gammagt attributes)

TABLE VII. ACCURACY OF MACHINE LEARNING ALGORITHM BY CONSIDERING ALK, SGPT, SGOT,GAMAGT, ATTRIBUTES

Serial number	Machine learning classification model	Class 1		Class 2		Accuracy		
		Accurate Sample	Total sample	Accurate Sample	Total sample	Class 1	Class 2	Overall accuracy
1	DECISION TREE	14	24	23	27	0.58	0.85	0.72
2	NEURAL NETWORK	12	24	21	27	0.5	0.77	0.64
3	RANDOM FOREST	12	24	20	27	0.5	0.74	0.62
4	SUPPORT VECTOR MACHINE	13	24	23	27	0.54	0.85	0.70

5.CONCLUSION

From this paper, the decision tree suits the best for the input dataset as it gives the minimal error rate of 27.4%. When compared to others. The other implication of this paper is, it reduce the number of attributes from 7 to 4 without compromising the error rate. As there is no difference between the error rates of all attributes and reduced attributes can use the reduced attributes for lightweight electronic gadgets.

REFERENCES

[1]. BUPA Liver Disorders Dataset. UCI repository of machine learning databases. Available from ftp://ftp.ics.uci.edu/pub/machinelearningdatabases/liver disorders/bupa.data, last accessed: 07 October 2010.

[2]. Michael J. Sorich, John O. Miners, Ross A. McKinnon, David A. Winkler, Frank R. Burden,| and Paul A. Smith, Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-Glucuronosyltransferase Isoforms.

[3]. Sa'sdiyah Noor Novita Alfisahrin, Teddy Mantoro (2013), Data mining Techniques For Optimatization of Liver Disease Classification, *International conference on advanced Computer Science Application and Technologies*,2013.

[4]. Onwodi Gregory (2015), Prediction of Liver Disease (Biliary Cirrhosis) Using Data Mining Technique, *International Journal of Emerging Technology & Research*, ISSN (E):2347-5900, ISSN (P):2347-60791.

[5]. Al Nuaimi, Z.N.A.M., & Abdullah, R. (2017) Neural network training using hybrid particle move artificial bee colony algorithm for pattern classification. Journal of Information and Communication Technology, 16 (2), pp. 314-334.

[6]. A. K. R. Mujahid & C. Thirumalai. (2017). Pearson Correlation Coefficient Analysis (PCCA) on Adenoma carcinoma cancer. International Conference on Trends in Electronics and Informatics (ICEI), pp. 492-495.

[7]. C. Thirumalai & M. Senthilkumar. (2017). An assessment framework of intuitionistic fuzzy network for C2B decision making. 4th International Conference on Electronics and Communication Systems (ICECS), pp. 164-167.

[8]. C. Thirumalai, G. V. SaiSharan, K. V. Krishna & K. J. Senapathi. (2017) Prediction of diabetes disease using control chart and cost optimization-based decision. International Conference on Trends in Electronics and Informatics (ICEI), pp. 996-999.

[9]. Gourav Dwivedi, Rajiv K Srivastava, & Samir K Srivastava. (2018). A generalized fuzzy TOPSIS with improved closeness coefficient. Expert Systems with Applications, Vol. 96, pp. 185-195.

[10]. Kalaiarassan, G., Krishan, Somanadh M., Chandrasegar, Thirumalai., & Senthilkumar, M. (2017). One-Dimension Force Balance System for Hypersonic Vehicle an experimental and Fuzzy Prediction Approach. Elsevier, ICMMM.

[11]. Kasim, M.M., Kashim, R., & Khan, S.A.M.N. (2017). A linear programming based model to measure efficiency and effectiveness of undergraduate programs. Journal of Information and Communication Technology, 16 (2), pp. 394-407.

[12]. K. Sharma, B. Muktha, A. Rani and C. Thirumalai. (2017) Prediction of benign and malignant tumor. International Conference on Trends in Electronics and Informatics (ICEI), pp. 1057-1060.

[13]. Mokhtar, S.A., Wan Ishak, W.H., & Norwawi, N.M. (2016). Modeling reservoir water release decision using Adaptive Neuro Fuzzy Inference System. Journal of Information and Communication Technology, 15 (2), pp. 141-152.

[14]. Othman, M., & Azahari, S.N.F. (2016). Deseasonalised forecasting model of rainfall distribution using fuzzy time series. Journal of Information and Communication Technology, 15 (2), pp. 153-169.

[15]. Ponnurangam, Dhavachelavan, and G. V. Uma. (2005) Fuzzy complexity assessment model for resource negotiation and allocation in agent-based software testing framework. Expert Systems with Applications, pp. 105-119.

[16]. Ramli, R., Jamaluddin, F., Bakar, E.M.N.E.A., Alias, M.Y., Mahat, N.I., & Karim, M.Z.A. (2013). Assignment of spectrum demands by merits via analytic hierarchy process and integer programming. Journal of Information and Communication Technology, 12 (1), pp. 39-53.

[17]. R. Poovarasan, S. Keerthi, & K. Yuvashree and C. Thirumalai. (2017). Analysis on diabetes patients using Pearson, cost optimization, control chart. International Conference on Trends in Electronics and Informatics (ICEI), pp. 1139-1142.

[18]. Srisawat, C., & Payakpate, J. (2016). Comparison of MCDM methods for intercrop selection in rubber plantations. Journal of Information and Communication Technology, 15 (1), pp. 165-182.

[19]. Vaishnavi B., K. Yarrakula, Karthikeyan J. and C. Thirumalai. (2017). An assessment framework for Precipitation decision making using AHP. 11th International Conference on Intelligent Systems and Control (ISCO), pp. 418-421.

[20]. S. Ansari, I. Shafi, A. Ansari, J. Ahmad, S. I. Shah, "Diagnosis of liver disease induced by hepatitis virus using Artificial Neural Networks", *IEEE 14th International Multitopic Conference*, 2011.

[21]. M. Neshat, M. Yaghobi, M.B. Naghibi, A. Esmaelzadeh, "Fuzzy Expert System Design for Diagnosis of liver disorders", *International Conference Symposium on Knowledge Acquisition and Modeling IEEE*, 2008.

[22]. L. Ozyilmaz, T. Yildirim, "Artificial Neural Network for Diagnosis of Hepatitis Disease", *Proceedings of the International Joint Conference on Neural Networks*, pp. 586-589, 2003.