International Journal of Engineering & Technology, 7 (4.19) (2018) 485-495

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Machine Learning Techniques on Liver Disease - A Survey

V.V. Ramalingam[1], A.Pandian[2], R. Ragavendran[3]

[1,2,3]Department of Computer Science and Engineering, SRMIST, Kattankulathur

Abstract

A Liver infection causing Chronic Hepatitis B virus affects around 257 million people around the globe. About one million people who are infected from chronic infections like HBV die from chronic liver disease. Along these lines, there is a solid requirement for effective, exact and practical framework to foresee the result of such infection. It will be helpful in taking precaution steps and proper treatment. Machine learning assumes an essential job in restorative industry. Experts in machine learning today can guarantee an accurate and definite diagnosis and analysis of disease. Machine learning methodologies have been approached on various liver disease related datasets to predict outcome result. Machine learning calculations are exceptionally useful in giving essential measurements, continuous information, and progressed examination regarding the patients' illness, "lab test results, circulatory strain, family history, clinical preliminary information, and more to" specialists. Motivation behind this paper is to give an overview on relative survey on machine learning techniques that has been used on various liver disease datasets.

Keywords: *Liver Disease; Linear Regression; Support Vector Machines; Decision Tree; Random Forest; Ensemble Models*.

1. Introduction

The liver is an immense, significant organ in the human body weights around 3 pounds. The liver contains two huge portions, called the privilege and the left projections. The gallbladder sits under the liver, nearby parts of the pancreas and stomach related organs. The liver and these organs cooperate to process, ingest, and process sustenance. The liver fundamental job is to channel the hurtful substances in the blood starting from the stomach related framework, before passing it to whatever is left of the body. Liver" disease can be acquired either by outside factors like infections and liquor use or through genetics. Corpulence has additionally been related with this illness. Overtime harm to liver can cause liver failure or at some circumstances even life-threatening condition.

Liver disease is one of the most "death-dealing disease on the planet. The fundamental driver of liver harm are Fatty liver, Liver Fibrosis, Cirrhosis, hepatitis and diseases [1]. In the beginning times of liver" illness, it is exceptionally hard to identify even though liver tissue has already been harmed. It requires numerous specialists to analyze the damage. [2] This can contort pharmaceutical medications, so early finding is essential to spare the patient.

Common Liver Disorders include:

- Fatty liver is an agonizing liver condition portrayed by liver irritation and arrangement of scar tissue, which has numerous conceivable causes, including corpulence, poor nourishment and certain meds, among numerous others. It can happen in individuals with an abnormal state of liquor utilization as well as in individuals who never had liquor.
- Cirrhosis is another important type of liver damage. It is usually the result of long term damage of liver. When liver is damaged for a long time and starts to malfunction this particular type of liver damage occurs.
- Hepatitis is usually caused by an infection that spreads by direct contact with tainted body.
- Liver Cancer risk is higher on those who has cirrhosis. Most often it spreads from liver to other organs.

2. Dimensionality Reduction

2.1. Normalization

Normalization is a technique which is applied for data preparation for machine learning. The purpose of normalization is to manipulate the values of numeric column in a dataset to use a common scale, without having values in different ranges. It is achieved by dividing the dataset into tables and finding relation between the data.

2.2. Feature Selection

Feature selection can be seen as a useful process of selecting attributes that is felt being relevant or meaningful to the problem. It is also used for evaluation measure. It is used for proposing new feature subsets as well as for evaluation measures which scores different feature subset. M. Banu Priya [3] is seen using PSO feature selection method. A subset of liver patient dataset from finish standardized liver patient dataset is created therefore. It comprised of only the meaningful attributes. PSO computation is based on a swarm of processing elements called particles in which each particle represents a solution.

3. Algorithms and Methodology

3.1. Logistic Regression

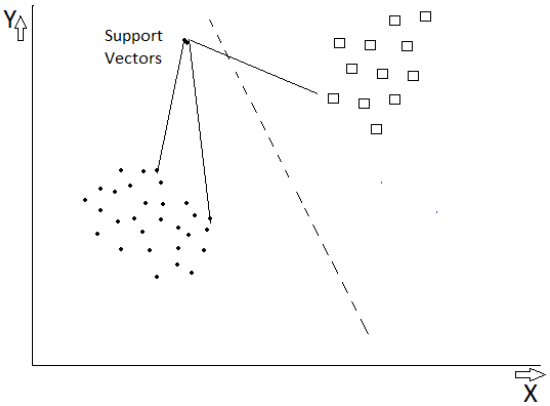
Logistic regression is a type of a supervised machine learning algorithm. It makes a prediction that has binary outcome from the

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

past data. Logistic regression usually returns result in very short time, hence it is preferred being used as an benchmarking model. It has also been evident in [4] that using logistic regression has achieved an accuracy of 72%. The general precision of logistic regression has moved forward from 70% to 72% when the 12 selected indexes from the 24 indexes were used.

3.2. Support Vector Machine

SVM alias support vector machine is a famous algorithm. It is used for both classification and regression. With the given marked prepared information, the calculation yields an ideal hyperplane which arranges new examples.



In [5] has proposed a research work on liver disease prediction using classification techniques. In which SVM classifier outperformed Naïve Bayes classifier. SVM managed to get a higher accuracy rate at 79.66% whereas Naïve Bayes algorithm achieved to get 61.28% of accuracy.
[6] has compared two methodologies out of which SVM managed to get an accuracy of 71%.

3.3. K – Nearest Neighbor

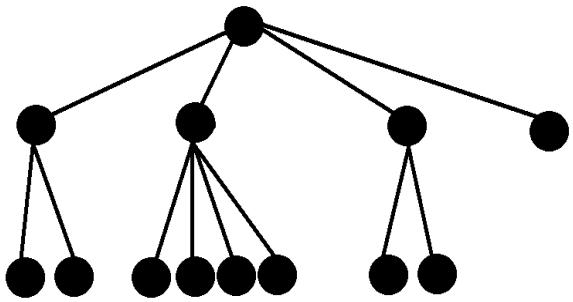
K – Nearest neighbor method is a non-parametric method used for classification. In KNN algorithm the classification happens by determining which of the k instance is at most similar to an input. To achieve this distance measurement is used.
Bendi Venkata Ramana, [7] proposed five arrangement calculations for liver sickness finding. At long last K-Nearest Neighbor acquired a noteworthy precision in their findings. It managed to provide an accuracy of 97.47% in AP Liver ddataset which is the highest accuracy achieved among the other classification techniques. In UCLA dataset it managed to achieve 62.89% accuracy.

3.4. Decision Tree

Decision tree is an algorithm that contains conditional control statements in particular. The outcome might be "true" or "false". Rules can be acquired from the way which begins from the root node and finishes at the leaf node and furthermore uses the nodes in transit as preconditions for the got rule, to foresee the class at the leaf.
The algorithm divides the population into two or more sets. The two steps are done recursively with remaining attributes

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

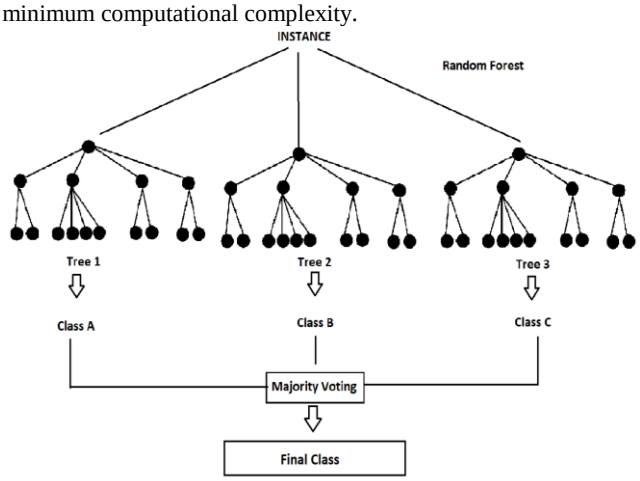
$$Gain(S, A) - Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$



In [8] this research authors have applied several techniques for determination of liver sickness. Among the algorithms decision tree produced a decent accuracy score of 66.14%.. In [9] a decision tree approach was used and it managed to get an accuracy rate of 93.7%. Highest accuracy rate was achieved by using decision tree algorithm.

3.5. Random Forest tree

This method is a form of a supervised machine learning algorithm. It very well may be utilized for both classification and regression. Random forest tree basically creates a forest with multiple trees. Consequently, it is fundamentally a gathering of more than one decision trees. It uses voting system for classification, through this the class is chosen. In regression, it takes the mean of the yields of every decision tree.
In [10] a classification based on random forest tree is used along with feature selection based on random forest tree for Hepatitis B virus reactivation. The accuracy of random forest tree reached 85.15%. In [11] the research was done on a liver disease dataset using random forest and neural networks. The strategy prepares the input features by using the 10-fold cross validation manner. The scheme delivered an accuracy of 80% by using Random Forest Classifier.
A research paper on analysis of decision tree algorithms on hepatitis dataset [12] concludes that random forest tree algorithm produces a accuracy of 87.25% in classification along with minimum computational complexity.



3.6. Neural Network

Neural network is an approach that is motivated by the manner in which natural sensory system works, that incorporate, for example, how the cerebrum procedure data. It is made out of substantial number of interconnected handling components working in solidarity to take care of certain issue. Artificial neural network are finding many uses in medical diagnosis application [15].
In [16], a blended "model of rough set theory and LMBP neural network is presented. It proposes a liver illness conclusion dependent on this blend demonstrate. It states that the combined

model can speed up the training speed of the network and have" a better diagnostic accuracy. It achieved 96.67% accuracy.
[17] In this examination a logical inference with a generalized regression neural network is exhibited for analysis of Hepatitis B liver disorder. In this examination a coherent derivation with a summed up relapse neural system is exhibited for analysis. This research managed to provide an accuracy of 86.3237%

3.7. Ensemble Method

Ensemble classification techniques is being broadly utilized in therapeutic field, for obtaining better accuracy than using one particular classification. It is basically a machine learning method that uses multiple machine learning algorithm to obtain a better predictive performance. Ensemble classifiers have been used in Neuroscience, proteomics and medical diagnosis.
[13] proposed a modified random forest algorithm to calculate the accuracy of the classification algorithms on UCI liver dataset. It uses multi-layer perception classification algorithm and random subset feature selection method. The results showed that the algorithm produces 74.7826% accuracy. [14] In this particular paper a framework is created and is specifically evaluated on different disease datasets. On which it includes two liver dataset, one dataset on parkinson's disease, liver related disease dataset that is hepatitis dataset, two diabetes dataset, two breast cancer dataset and two heart disease dataset. Among the liver disease dataset, this particular framework achieved an accuracy of 71.53% on Indian liver disease patient dataset and 67.54% accuracy on Bupa liver disease dataset.

4. Conclusion

The main focus of this paper is to throw light on the importance of different classification techniques for disease prediction, particularly liver related diseases. The dataset that has been reviewed is considered in so many existing techniques. The datasets that we have discussed are liver disease related like hepatitis and hepatocellular carcinoma.

References

[1] D.A. Saleh F. Shebl M. Abdel-Hamid etal. "Incidence and risk factors for hepatitis C infection in a cohort of women in rural Egypt" Trans. R. Soc. Trop. Med.Hyg, vol. 102 pp. 921928, 2008.
[2] S. A. Gonzalez dan E. B. Keeffe, "Acute liver failure," dalam Handbook of Liver Disease Third Edition, Philadelphia, Elsevier, 2012, pp. 20-33
[3] M. Priya et al, "Performance Analysis of Liver disease prediction using Machine Learning Algorithms", IRJET e-ISSN: 2395-0056 Vol. 05, Issue: 1, Jan 2018.
[4] N. Li, Y. Jiang, G. Gong, G. Han and J. Ma, "Non-Invasive Assessment Model of Liver Disease Severity by Serum Markers Using Cloud Computing and Internet of Things," in IEEE Access, vol. 6, pp. 33969-33976, 2018.
[5] Dr.S.Vijayarani, Mr.S.Dhayanand, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms",International Journal of Science, Engineering and Technology Research(IJSETR), 2015.
[6] Sumedh Sontakke, J. L. (20l7). "Diagnosis of Liver Diseases using Machine Learning." ICEI.
[7] "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis " Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu, Prof. N. B. Venkateswarlu -(IJDMS), Vol.3, No.2,May 2012
[8] "Data Mining Techniques for optimization of liver disease classification." Sadiyah Noor Novita Alfisahrin, Teddy Mantoro Electronic ISBN: 978-1-4799-2758-6 DOI:
[9] 10.1109/ACSAT.2013.81-IEEE
[10] "Prediction of Liver Fibrosis stages by Machine Learning model: A Decision Tree Approach*" Heba Ayeldeen,, Olfat Shaker , Ghada Ayeldeen , Khaled M. Anwar, Electronic ISBN: 978-1-4673-9669-1IEEE-2016 https://doi.org/10.1109/ICoCS.2015.7483212.
[11] H. Wang, Y. Liu and W. Huang, "Random forest and Bayesian prediction for Hepatitis B virus reactivation," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017, pp. 2060-2064.
[12] M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza and M. K. Hasan, "Performance Evaluation of Random Forests and Artificial Neural Networks for the Classification of Liver Disorder," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, 2018, pp. 1-5.doi: 10.1109/IC4ME2.2018.8465658
[13] M. Ramasamy, S. Selvaraj and M. Mayilvaganan, "An empirical analysis of decision tree algorithms: Modeling hepatitis data," 2015 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, 2015, pp. 1-4. doi:
[14] 10.1109/ICETECH.2015.7275013
[15] Bendi Venkata Ramanaland Prof.M.Surendra Prasad Babu, "Liver Classification Using ModifiedRotation Forest", Internationa Journal of Engineering Research and Development ISSN: 2278-067X,Volume 1, Issue6, (June 2012), PP.17-24.
[16] Bashir S., Qamar U., Khan F. H., Naseem L. HMV: a medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science. 2016;13:10–25. doi: 10.1016/j.jocs.2016.01.001. [CrossRef]
[17] Al-Shayea, Qeethara. (2011). Artificial Neural Networks in Medical Diagnosis. Int J Comput Sci Issues. 8. 150-154.
[18] D. Xu, H. Fu and W. Jiang, "Research on Liver Disease Diagnosis Based on RS_LMBP Neural Network," 2016 12th International Conference on Computational Intelligence and Security (CIS), Wuxi, 2016, pp. 646-649.doi: 10.1109/CIS.2016.0156
[19] G. S. Uttreshwar and A. A. Ghatol, "Hepatitis B Diagnosis Using Logical Inference And Generalized Regression Neural Networks," 2009 IEEE International Advance Computing Conference, Patiala, 2009, pp. 1587-1595.doi: 10.1109/IADCC.2009.4809255