

Health Diagnosis by using Machine Learning Algorithm

N. Tejaswini
Dept of CSE
Besant Theosophical College

P. Veeramuthu
Assitant Professor
Dept of CSE
Besant Theosophical College

Abstract:- In medical imaging, Computer Aided Diagnosis (CAD) is a rapidly growing dynamic area of research. In recent years, significant attempts are made for the enhancement of computer aided diagnosis applications because errors in medical diagnostic systems can result in seriously misleading medical treatments. Machine learning is important in Computer Aided Diagnosis. After using an easy equation, objects such as organs may not be indicated accurately. So, pattern recognition fundamentally involves learning from examples. In the field of bio-medical, pattern recognition and machine learning promise the improved accuracy of perception and diagnosis of disease. They also promote the objectivity of decision-making process. For the analysis of high-dimensional and multimodal bio-medical data, machine learning offers a worthy approach for making classy and automatic algorithms. This survey paper provides the comparative analysis of different machine learning algorithms for diagnosis of different diseases such as heart disease, diabetes disease, liver disease, dengue disease and hepatitis disease. It brings attention towards the suite of machine learning algorithms and tools that are used for the analysis of diseases and decision-making process accordingly.

Keywords: Machine Learning, Artificial Intelligence, Machine Learning Techniques

INTRODUCTION:

Artificial Intelligence can enable the computer to think. Computer is made much more intelligent by AI. Machine learning is the subfield of AI study. Various researchers think that without learning, intelligence cannot be developed. There are many types of Machine Learning Techniques that are shown in Figure 1. Supervised, Unsupervised, Semi Supervised, Reinforcement, Evolutionary Learning and Deep Learning are the types of machine learning techniques. These techniques are used to classify the data set.

1) SUPERVISED LEARNING: Offered a training set of examples with suitable targets and on the basis of this training set, algorithms respond correctly to all feasible inputs. Learning from exemplars is another name of Supervised Learning. Classification and regression are the types of Supervised Learning. Classification: It gives the prediction of Yes or No, for example, "Is this tumor cancerous?", "Does this cookie meet our quality standards?" Regression: It gives the answer of "How much" and "How many".

2) UNSUPERVISED LEARNING: Correct responses or targets are not provided. Unsupervised learning technique

tries to find out the similarities between the input data and based on these similarities, un-supervised learning technique classify the data. This is also known as density estimation. Unsupervised learning contains clustering [1]. Clustering: it makes clusters on the basis of similarity.

3) SEMI SUPERVISED LEARNING: Semi supervised learning technique is a class of supervised learning techniques. This learning also used unlabeled data for training purpose (generally a minimum amount of labeled-data with a huge amount of unlabeled-data). Semi-supervised learning lies between unsupervised-learning (unlabeled-data) and supervised learning (labeled-data).

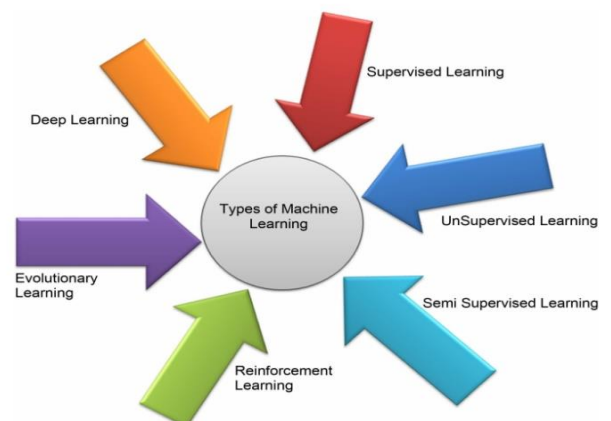


Figure 1. Types of machine learning techniques.

4) REINFORCEMENT LEARNING: This learning is encouraged by behaviorist psychology. Algorithm is informed when the answer is wrong, but does not inform that how to correct it. It has to explore and test various possibilities until it finds the right answer. It is also known as learning with a critic. It does not recommend improvements. Reinforcement learning is different from supervised learning in the sense that accurate input and output sets are not offered, nor suboptimal actions clearly precised. Moreover, it focuses on on-line performance.

5) EVOLUTIONARY LEARNING: This biological evolution learning can be considered as a learning process: biological organisms are adapted to make progress in their survival rates and chance of having off springs. By using the idea of fitness, to check how accurate the solution is, we can use this model in a computer

6) DEEP LEARNING: This branch of machine learning is based on set of algorithms. In data, these learning

algorithms model high-level abstraction. It uses deep graph with various processing layer, made up of many linear and nonlinear transformation.

EXISTING SYSTEM: Diagnosis of Diseases by Using Different Machine Learning Algorithms Many researchers have worked on different learning algorithms for disease diagnosis. Researchers have been accepted that machine-learning algorithms work well in diagnosis of different diseases. Figurative approach of diseases diagnosed by Machine Learning Techniques is shown in Figure 2. In this survey paper diseases diagnosed by MLT are heart, diabetes, liver, dengue and hepatitis.

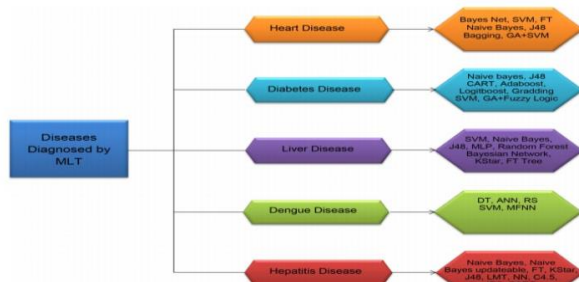


Figure 2. Diseases diagnosed by MLT.

HEART DISEASE:

Otoometal. [5] presented a system for the purpose of analysis and monitoring. Coronary artery disease is detected and monitored by this proposed system. Cleveland heart data set is taken from UCI. This data set consists of 303 cases and 76 attributes/features. 13 features are used out of 76 features. Two tests with three algorithms Bayes Net, Support vector machine, and Functional Trees FT are performed for detection purpose. WEKA tool is used.

DIABETES DISEASE:

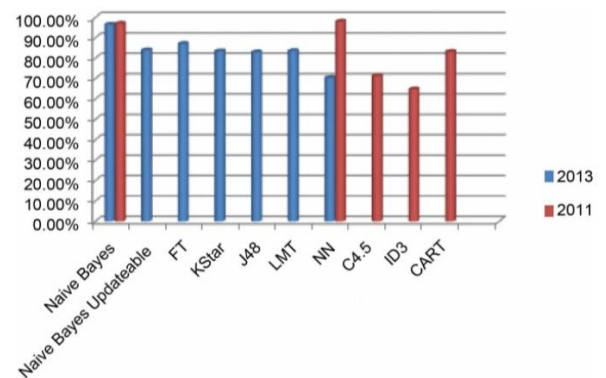
Iyer et al. [11] has performed a work to predict diabetes disease by using decision tree and Naive Bayes. Diseases occur when production of insulin is insufficient or there is improper use of insulin. Data set used in this work is Pima Indian diabetes data set. Various tests were performed using WEKA data mining tool. In this data-set percentage split (70:30) predict better than cross validation. J48 shows 74.8698% and 76.9565% accuracy by using Cross Validation and Percentage Split Respectively. Naive Bayes presents 79.5652% correctness by using PS. Algorithms shows highest accuracy by utilizing percentage split test.

LIVER DISEASE:

Vijayarani and Dhayanand [17] predict the liver disease by using Support vectormachine and Naive bayes Classification algorithms. ILPD data set is obtained from UCI. Data set comprises of 560 instances and 10 attributes. Comparison is made on the basis of accuracy and time execution. Naive bayes shows 61.28% correctness in 1670.00 ms. 79.66% accuracy is attained in 3210.00 ms by SVM. For implementation, MATLAB is used. SVM shows highest accuracy as compared to the Naive bayes for liver

disease prediction. In terms of time execution, Naives bayes takes less time as compared to the SVM.

Proposed system :Discussions and Analysis of Machine Learning Techniques For diagnosis of Heart, Diabetes, Liver, Dengue and Hepatitis diseases, several machine-learning algorithms perform very well. From existing literature, it is observed that Naive Bayes Algorithm and SVM are widely used algorithms for detection of diseases. Both algorithms offer the better accuracy as compare to other algorithms. Artificial Neural network is also very useful for prediction. It also shows the maximum output but it takes more time as compared to other algorithms. Trees algorithm are also used but they did not attain wide acceptance due to its complexity. They also shows enhanced accuracy when it responded correctly to the attributes of data set. RS theory is not widely used but it presents maximum output.



CONCLUSION:

Statistical models for estimation that are not capable to produce good performance results have flooded the assessment area. Statistical models are unsuccessful to hold categorical data, deal with missing values and large data points.

All these reasons arise the importance of MLT. ML plays a vital role in many applications, e.g. image detection, data mining, natural language processing, and disease diagnostics. In all these domains, ML offers possible solutions. This paper provides the survey of different machine learning techniques for diagnosis of different diseases such as heart disease, diabetes disease, liver disease, dengue and hepatitis disease. Many algorithms have shown good results because they identify the attribute accurately. From previous study, it is observed that for the detection of heart disease, SVM provides improved accuracy of 94.60%. Diabetes disease is accurately diagnosed by Naive Bayes. It offers the highest classification accuracy of 95%. FT provides 97.10% of correctness for the liver disease diagnosis. For dengue disease detection, 100% accuracy is achieved by RS theory. The feed forward neural network correctly classifies hepatitis disease as it provides 98% accuracy. Survey highlights the advantages and disadvantages of these algorithms. Improvement graphs of machine learning algorithms for prediction of diseases are presented in detail.

From analysis, it can be clearly observed that these algorithms provide enhanced accuracy on different diseases. This survey paper also provides a suite of tools that are developed in community of AI. These tools are very useful for the analysis of such problems and also provide opportunity for the improved decision making process.

REFERENCES:

- [1] Marshland, S. (2009) Machine Learning an Algorithmic Perspective. CRC Press, New Zealand, 6-7.
- [2] Sharma, P. and Kaur, M. (2013) Classification in Pattern Recognition: A Review. International Journal of Advanced Research in Computer Science and Software Engineering, 3, 298-313.
- [3] Rambhajan, M., Deepanker, W. and Pathak, N. (2015) A Survey on Implementation of Machine Learning Techniques for Dermatology Diseases Classification. International Journal of Advances in Engineering & Technology, 8, 194-195.
- [4] Kononenko, I. (2001) Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. Journal of Artificial Intelligence in Medicine, 1, 89-109.
- [5] Ootom, A.F., Abdallah, E.E., Kilani, Y., Kefaye, A. and Ashour, M. (2015) Effective Diagnosis and Monitoring of Heart Disease. International Journal of Software Engineering and Its Applications, 9, 143-156.
- [6] Vembandasamy, K., Sasipriya, R. and Deepa, E. (2015) Heart Diseases Detection Using Naive Bayes Algorithm. IJSET-International Journal of Innovative Science, Engineering & Technology, 2, 441-444.
- [7] Chaurasia, V. and Pal, S. (2013) Data Mining Approach to Detect Heart Disease. International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2, 56-66.
- [8] Parthiban, G. and Srivatsa, S.K. (2012) Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients. International Journal of Applied Information Systems (IJASIS), 3, 25-30.
- [9] Tan, K.C., Teoh, E.J., Yu, Q. and Goh, K.C. (2009) A Hybrid Evolutionary Algorithm for Attribute Selection in Data Mining. Journal of Expert System with Applications, 36, 8616-8630. <https://doi.org/10.1016/j.eswa.2008.10.013>
- [10] Karamizadeh, S., Abdullah, S.M., Halimi, M., Shayan, J. and Rajabi, M.J. (2014) Advantage and Drawback of Support Vector Machine Functionality. 2014 IEEE International Conference on Computer, Communication and Control Technology (I4CT), Langkawi, 2-4 September 2014, 64-65. <https://doi.org/10.1109/i4ct.2014.6914146>
- [11] Iyer, A., Jeyalatha, S. and Sumbaly, R. (2015) Diagnosis of Diabetes Using Classification Mining Techniques. International Journal of Data Mining & Knowledge Management Process (IJDKP), 5, 1-14. <https://doi.org/10.5121/ijdkp.2015.5101>
- [12] Sen, S.K. and Dash, S. (2014) Application of Meta Learning Algorithms for the Prediction of Diabetes Disease. International Journal of Advance Research in Computer Science and Management Studies, 2, 396-401.
- [13] Kumari, V.A. and Chitra, R. (2013) Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications (IJERA), 3, 1797-1801.
- [14] Sarwar, A. and Sharma, V. (2012) Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2. Special Issue of International Journal of Computer Applications (0975-8887) on Issues and Challenges in Networking, Intelligence and Computing Technologies-ICNICT 2012, 3, 14-16.
- [15] Ephzibah, E.P. (2011) Cost Effective Approach on Feature Selection using Genetic Algorithms and Fuzzy Logic for Diabetes Diagnosis. International Journal on Soft Computing (JSC), 2, 1-10. <https://doi.org/10.5121/ijsc.2011.2101>
- [16] Archana, S. and DR Elangovan, K. (2014) Survey of Classification Techniques in Data Mining. International Journal of Computer Science and Mobile Applications, 2, 65-71
- [17] Vijayarani, S. and Dhayanand, S. (2015) Liver Disease Prediction using SVM and Naïve Bayes Algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4, 816-820.
- [18] Gulia, A., Vohra, R. and Rani, P. (2014) Liver Patient Classification Using Intelligent Techniques. (IJCSIT) International Journal of Computer Science and Information Technologies, 5, 5110-5115.
- [19] Rajeswari, P. and Reena, G.S. (2010) Analysis of Liver Disorder Using Data Mining Algorithm. Global Journal of Computer Science and Technology, 10, 48-52.
- [20] Tarmizi, N.D.A., Jamaluddin, F., Abu Bakar, A., Othman, Z.A., Zainudin, S. and Hamdan, A.R. (2013) Malaysia Dengue Outbreak Detection Using Data Mining Models. Journal of Next Generation Information Technology (JNIT), 4, 96-107.