ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-14

(Approved and Registered with Govt. of India)

INVESTIGATION OF DISEASE USING THE MACHINE LEARNING APPROACH

Vipin Rai [1], Dr. P.K. Bharti[2] Dr. Rakesh K. Yadav[3]

Ph.D. Research Scholar [1], Professor & Vice Chancellor[2] , Professor & Director[3]
School of Engineering & Technology ı
Shri Venkateshwara University, Gajraula, UP [1, 2]
KCC Institute of Technology & Management, Greater Noida, U.P, India[3]

Abstract:The prediction of disease using Machine learning, Deep learning and neural network techniques now a day most energizing research within the scientific community. The conclusive statement obtained from the doctor or any type of suggestive things like Integrated Soft-computing tools, data mining and neural network analysis always is predictive. Always there is chance to improvises the expected result through better disease treatment process. All of these processes need very huge study of patient's data. It is quite tough to arrange huge and real data sets for present's research. So this paper introduces some Deep learning and machine learning technique for predicting the critical disease using the data sets provided by UCI and Kaggle.This will provide helping hand in self-generating suggestion of patients' record values. Neural network itself is capable and sufficient generator of all kind of data type. It just needs the type of attributes and respective range and further it feed to training sets. In this paper it is assuring to the patient have better interface system that have the power of judgement critical disease. This includes the implementation of deep learning algorithms that will enhance the make of the console more decisive.Other Algorithms like A-Priori, K-Mean and PSO is also be used for making the earlier tools in console. Integrated Machine learning tools will provide the key information and attributes to the doctors and patients

Keywords: Neural Network, Epoch, Machine learning classifiers, Multi Level Perceptron, Integrated Machine learning tools Approach, PSO

1. INTRODUCTION

Machine learning
Machine learning is a collection of methods designed to take advantage of tolerances for inaccuracies and uncertainties to achieve manageability, robustness, and low solution cost. Its main components are fuzzy logic, neural computing and probabilistic reasoning. Machine learning may play an increasingly important role in many applications, including software engineering. The example of machine learning is the human mind. "

Machine learning (ML) is an emerging field consisting of complementary elements of fuzzy logic, neural computing, evolutionary computing, machine learning and probabilistic reasoning. Machine learning technology has been widely used due to its strong learning and cognitive ability and good tolerance for uncertainty and inaccuracy.

Machine learning is not a homomorphism of concepts and technologies. Rather, it is a series of methodologies that reflect the guiding principles of machine learning in one form or another: using susceptibility to inaccuracy, uncertainty, and partial truth to achieve manageability, robustness, and low Solution cost. From a slightly different perspective, Machine learning is a methodological alliance that provides an effective tool for the development of intelligent systems, or either combination.

Dr. Zadeh who is a pioneer of fuzzy logic. He pointed out that "the guiding principle of Machine learning is to use the tolerance of inaccuracy, uncertainty and partial truth to achieve manageability, robustness, low solution cost, and better with reality. Harmony." Due to its intelligent control, nonlinear programming, optimization and decision

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-14

(Approved and Registered with Govt. of India)

support, Machine learning has become popular and has attracted research interest from people of different backgrounds. It is becoming more and more difficult to control increasingly complex modern machinery using traditional control system technology.

Machine learning tools (neural networks, genetic algorithms, k-NN, LI-KNN, GI-KNN, GA, PSO, and ACO)

Artificial Neural Network

ANN is a parallel distributed very large information processing structure consisting of many nonlinear processing units called neurons. Neurons run as mathematical processors that perform specific mathematical operations on their inputs to generate outputs. It trains recognition patterns and identifies incomplete patterns, similar to human brain processes that identify information, literally masking noise and correctly retrieving information. In terms of modeling, significant progress has been made over the past few decades to improve the ANN. ANN is a strong interconnected system of neurons, with simple behavior, but they can solve complex problems when connected. Can be further changed to improve its performance

Genetic algorithms

An evolutionary algorithm (EA) was invented to mimic some of the processes observed in natural evolution. Evolution takes place on chromosomes - organic devices used to encode biological structures. The natural selection process then drives the chromosomes that encode the successful structure to regenerate more frequently than the chromosomes that encode the failed structure.

Disease Diagnosis Using Machine learning

The viability and sustainability of a country's economic and social development depends to a large extent on the country's vibrant health care sector. In the absence of an appropriate health care system, computerized medical diagnostic systems can maintain stable economic growth. Medical diagnosis depends to a large extent on the doctor, especially the novel or junior doctor, and it may take years to accumulate enough experience. The actual experience of doctors cannot be inherited in the next generation. It need some diagnostic systems that will help you store years of experience. Several disease diagnosis models have been proposed to help doctors diagnose problems urgently.

Importance of machine learning

For many years, machine learning has played an important role in computer-aided disease diagnosis in physician decision-making. In fact, the role model of machine learning is the human mind. The basic principles of Machine learning are: inaccuracy, ambiguity, partial authenticity and guesswork for tractability, strength and low solution cost. In the past few decades, several Machine learning methods have been proposed for medical related fields. A Machine learning based diagnostic system uses symptoms to identify the disease. Symptoms may be clinical parameters such as blood pressure, blood sugar, scan reports, etc., or language expressions such as nausea, weakness, rejection of social gatherings, etc. In the proposed model, the attributes of the selection are rarely considered, and these attributes are shown as symptoms by suspected people.

A-priori: A-priori algorithm was introduced by **Agrawal et al. 1993. It is most important algorithm** for mining frequent item-sets. It is used for findings of all frequent or maximum occurrence items sets. This algorithm will extract the maximum occurred item sets as well desired frequency of the specific items within the group of items.

It could be understand with simple seven (taken for example) transactional situations, T1 to & T7. Within this transection five different items A, B, C, D and E (taken for example) could come with different pattern. Suppose T1 come up with (A, B, D) , T2(A, E) , T3 (E), T4(B,C) , T5(C,D,E) , T6 (A,B,E) and T7 (B,C,D,E). Further if there is need to check the transactional status of "A" then it could say that the occurrence of A= (T1, T2, T6). So it is clear understand as that the A has occurred in three transection out of total seven transections. Similarly when the combination of two items need for this findings then we go to again similar work. Suppose the situation demand the occurrence of A, B together in item sets. For the execution of real data set of disease it need to run through software for find the desired result easily and efficiently. This whole process has to need the proper algorithm which is as below.

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-14

(Approved and Registered with Govt. of India)

Algorithm

S_i: Disease data item-set of size i
F_i: frequent item-set of size i
F_1 = {frequent items};
For (i = 1; F_i! = Φ; i++) do begin
C_{i+1} = Disease data set generated from F_i;
Each transaction occurs "t" in database do
Increment the count of all Disease data set in C_{i+1} that are
Contained in t
L_{i+1} = medical data in C_{i+1} with min_support
End
Return U_i L_i;

PSO

Particle Swarm Optimization (PSO) has been successfully applied in data mining field to extract rule based classification systems. The PSO is very usefull in the extract a set of rule for diagnosis of various kind of disease. As per N. G. Hedeshi in 2011, he develop the optimization algorithm for extract a set of rules for diagnosis of coronary artery disease. As this study find that the PSO optimization for individual velocity (weight) andposition (Finding through the mean Distance) of each attributes comes under the disease diagnosis of large sets of data, …….

Initialize ();
For each value of disease attribute
for N run as N fold approach
// generate train and test data sets
generate_train set ();
generate_test set ();
// start pso set_PSO_Parameters ();
generate_random_particles ();
// generate_random_velocity ();
// main pso loop Evaluate the particles' fitness
Compute p-best and g-best in the population modify the velocity and position of each particle
// end main pso loop best_solution, best_Fitness end end cal_accuracy_test set ();

Softcomputing in Healthcare Systems

The relationship between the patient and the doctor is the key to the entire medical system and the key to providing the highest quality of effective care while keeping costs within limits. According to CNN International, the shortage of doctors and increased demand may lead to a crash healthcare of the medical system in the next few years. With the severe shortage of excellent practitioners, patients are waiting in the emergency room of the hospital to wait for their turn to consult the primary care physician. Therefore, an appropriate health care system, if not supplemented, will become a requirement of time. Researchers are looking for a cost-effective solution that delivers the best results without side effects. Regular monitoring of important parameters and correct treatment based on this data is another major aspect involved in healthcare. This is where Machine learning technology finds its relevance.

2. Research Background

M. Evanchalin Sweety, G. Wiselin Jiji (2014) proposed feature reduction and decision tree classifier technique for Alzheimer's disease (AD) which leads to death. This paper classifies feature reduction and decision tree classifier based on the proposed particle swarm optimization algorithm (PSO). Early AD detection was performed in three phases. In the first stage shows thatbasic features such as feature vector, feature brain, mean, variance, scene, kurtosis, standard deviation, area, perimeter, and eccentricity are extracted from the MRI image. Later on the second

stage shows the feature reduction is performed by Particle Swarm Optimization (PSO), and in the third stage, the decision tree classifier is used to detect whether the brain image is affected by Alzheimer's disease.

Shereen A. Taie, Wafaa Ghonaim (2017) proposed automated framework brain tumor detection, which detects and classifies brain tumors in MR imaging. The proposed framework brain tumor detection is an important tool for detecting tumors, and because it is capable of measuring regional variation characteristics in the brain reflecting disease progression, it distinguishes patients diagnosed as certain brain tumors and probable brain tumors. The framework consists of four steps: segmentation, feature extraction and feature reduction, classification, and finally using the optimization algorithm flock optimization algorithm (CSO) to dynamically optimize the parameter values of the classifier. The flock optimization algorithm is a bio-stimulus optimization algorithm, particle swarm optimization algorithm. The experimental results show that the proposed system does not use 7 times the population, the average accuracy is 71.63%, the cost time is 4.70 seconds, and 10 times, the average accuracy is 84.55%, and the cost time is 4.63 seconds. Using swarm optimization will increase the accuracy of the proposed system to 94.39%. In addition, increasing the training data set maximizes the accuracy of the proposed system to 99.9%. Finally, a comparative analysis between PSO and CSO shows that CSO outperforms PSO because it has the same accuracy and reduces cost time by 82.39%.

Miss. Sneha Joshi, Prof. Megha Borse (2016) wrote an article about diabetes chronic diseases. Recently it is estimated that in 2015, 415 million people worldwide suffer from diabetes, and turn in death condition are 1.5 million to 5 million. There are some ways to generate accurate predictions and artificial neural networks using back-propagation neural networks is one of them. The neural network has an input layer with 8 parameters, a hidden layer with 10 neurons and an output layer to produce good results. The GUI was developed to make the tool user-friendly so that patients can get accurate test results from the assistant even without a doctor. The program will help doctors get patient records in seconds, saving time for further treatment. This article summarizes the implementation and development of software tools in MATLAB that will predict whether someone has diabetes. The performance of BPNN used to predict diabetes was 81%, indicating an improvement in previous work. This is a better way than a finger, which is very painful if done in more time.This paper develops the architecture of neural networks on the basis of input disease data. The back propagation neural network structure has a hidden layer in which there are 10 neurons, 8 input nodes and 1 output node in the hidden layer as the structure for binary classification. The output can be 0 or 1, which shows 0 as a normal patient and 1 as a diabetic. The results of diabetic and non-diabetics can be displayed on the screen. The GUI is developed so that the physician can load input parameter readings and can apply training as needed and can obtain results for single or multiple patients. Additional options are available for single and multiple patients. There is a text box on the GUI that the doctor can get the results of any patient by entering the patient serial number. Randomly use small positive or negative values from 0 to 1 to train the network from the input to the hidden output layer. The performance plot shows that the mean square error (MSE) in the third epoch is 0.107, very close to zero, indicating that the target value is close to the actual value. The regression plot shows the best fit of 0.5. The accuracy obtained is 81% and the number of iterations is small. As the number of iterations decreases, time consumption is reduced compared to previous papers.

Uma Rani K and Mallikarjun S. Holi (2013) study Acoustic speech analysis and measurement methods may provide useful biomarkers for diagnosing the sound of neurological disorders. This paper describes a method for automatically detecting the neurological disturbances such as Parkinson's disease, cerebellar demyelination and stroke using the characteristics of the Mel frequency red coefficient (MFCC). The extracted features are provided to a multi-layer neural network and trained to classify whether the sound is a neurological disorder or a normal subject. There is no risk in capturing and analyzing speech signals because it is inherently non-intrusive and, in the case of careful control, it can provide a large amount of meaningful data. The data collected in this work included 137 continuous vowels of which 73 were from patients with different neurological diseases and 64 vocalizations from controlled subjects, including male and female subjects. . The 13 MFCC features were used as input to the best designed artificial neural network (ANN) for classification. 112 voices were used to train the network and 25 voices for testing. The best classification accuracy achieved was 92%.

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-14

(Approved and Registered with Govt. of India)

The paper concludes that MFCC has proven to be a good parameter for detecting speech disorders with a classification accuracy of 92%. This may be due to the fact that the MFCC calculation is a robust process that does not require previous pitch estimation and also maintains information about the spectral envelope of the speech signal. Therefore, the automatic detection of neurological disorders can lead to the development of expert systems to provide important feedback to the clinician or the patient's own speech therapy. This can also improve the individual treatment and contribute to the inconvenience and cost of the patient's actual visit to the clinic. In addition, voice measurements are non-intrusive, inexpensive and easy to manage. In the future, in order to improve the accuracy of the classification, a better classifier can be considered for the same feature.

S. Nandha Kumar, D. Dinesh, T. Siddharth, S. Ram Kumar, S. Nikhill and R. Lavanya (2017) It is proposed that diabetic retinopathy (DR) is a major factor leading to decades of blindness. The main cause of vision loss is damage to the vasculature in the retina. An early sign of DR is a micro-aneurysm, which appears as a tiny red dot on the retina. Early detection of this indicator helps ophthalmologists detect DR, which helps prevent blindness. In this work, a series of image processing algorithms, including preprocessing and coarse segmentation using mathematical morphology, were used to detect initial candidates for micro-aneurysms. This is followed by fine segmentation, where Particle Swarm Optimization (PSO) is used to estimate a set of best features. The classification performance of Naive Bayes and Support Vector Machine (SVM) is compared. Compared with PSO-naïve Bayes, the accuracy of the 19 features selected using PSO-SVM is 99.92%, with 22 features, and the accuracy rate is 93.31%. The proposed system can be used to accurately and quickly detect micro-aneurysms, thereby greatly reducing the workload and time of the ophthalmologist.

The proposed system concludes that the PSO algorithm for feature selection and the SVM for classification. This method resulted in the selection of 19 features. The classification accuracy is close to 100%. Therefore, the CAD system that has been developed will prove to be reliable in helping ophthalmologists diagnose DR. In addition, the reduction in feature set is used to achieve the primary purpose of the CAD system, namely to reduce the time it takes for the ophthalmologist to diagnose. Detecting symptoms during the screening process will be faster and more accurate. In the future, pixel-based segmentation can be performed to improve segmentation accuracy, which in turn affects classification accuracy.

Chun-Ling Lin, Sheng-Ta Hsieh and You-Jhong Hu (2013) suggested that because certain diseases are characterized by subjective perception of the symptoms described, if the symptoms are not obvious, the doctor can easily mistake them for other diseases. In order to help doctors diagnose the results quickly and accurately, this study proposes a medical diagnostic assistant system. The system uses fuzzy systems, back propagation neural networks (BPNN) and fuzzy neural networks (FNN) as the core engines of the influenza diagnostic expert system. The three systems were compared, and the inferred output of the expert system was used as a prognostic data for the disease, providing doctors with a diagnostic reference and reducing the diagnostic error rate to ensure early detection and treatment by doctors to prevent more serious diseases.

It is concluded that an expert system based on fuzzy neural networks has been proposed for influenza diagnosis. 50 virtual subjects with different symptoms were established for the experiment. According to the experimental results, the fuzzy neural network (FNN) expert system not only has higher patient diagnostic accuracy than the back propagation neural network (BPNN) and the fuzzy system. FNN is the right way to build a disease diagnosis system.

Pranjali B. Padol, S. D. Sawant (2016) it is suggested that grapes are one of the most widely grown fruit crops in India. Manual observation experts are used in practice to detect leaf disease, which requires more time for further control actions. If there is no accurate diagnosis of the disease, appropriate control measures cannot be taken at the appropriate time. This is where modern agricultural technology needs to detect and prevent leaves from coming from different diseases. This paper aims to introduce a new method for image detection of grape leaf disease using image processing technology, which can minimize the loss caused by automation and increase profits. In this system, classification is performed using support vector machine (SVM) and artificial neural network (ANN) classification. A new classifier is proposed by using fusion classification technology, which recombines SVM and ANN classifier for

basic classifier regeneration of grape leaf disease detection. Based on disease detection, the grape farmers will be provided with a suitable mixture of fungicides.

P. Ravivarma, B. Ramasubramanian, G. Arunmani B. Babumohan (2014) suggested that diabetic retinopathy is the leading cause of blindness in many diabetic patients. Automated detection of exudate in retinal images can help early screening for diabetic retinopathy. Several techniques can achieve good performance on high quality retinal images. But when image quality is low, they need a new approach. In this paper, they propose a new method for detecting exudates in low-quality retinal images. The color retinal image is preprocessed by a hyperbolic median filter and then segmented using a fuzzy mean clustering algorithm. After segmenting the image, a set of features based on color, size, and texture is extracted. These features are then optimized using Particle Swarm Optimization (PSO) techniques. Finally, features are classified using a Recursive Support Vector Machine (SVM) classifier. The proposed method achieves 98% accuracy and 98.5% productivity for the identification of secretions.

The conclusion of this paper is that exudate is one of the early signs of diabetic retinopathy. Low-quality color fundus images are enhanced using Contrast Finite Adaptive Histogram Equalization (CLAHE). Noise is removed from the image using a hyperbolic median filter. The enhanced color fundus image is segmented using a fuzzy C-means clustering algorithm. Compared to K-means, fuzzy C-means clustering takes less computation time. It provides more color information from which to improve classification results. Finally, in order to classify these segmented images into normal and exudates, a set of features based on texture and color is extracted using a gray level co-occurrence matrix (GLCM). Selected features using PSO are classified as secretions and non-secretion using a recursive support vector machine (rSVM) classifier.

Poonam Undre, Harjeet Kaur, Prakash Patil (2015) proposed that diabetes is now a major health problem for people of all ages. The main problem with this type of death is its prediction. It was found that if diabetes is detected at an early stage, it can be cured. So early detection of diabetes is very important. There are different techniques that can help detect diabetes early. In this paper, a combination of three different methods for early detection of diabetes is presented. These three methods are fuzzy systems, neural networks, and case-based reasoning. By using a combination of all these methods, it was found that diabetes can be detected at an early stage. The benefit of using these systems is that the prediction rate is more accurate than other technologies.

This article describes the method of diabetes detection that combines fuzzy, neural network and CBR methods. Due to the combination of all these techniques, the prediction rate was found to be largely improved. Therefore, early detection of diabetes can be achieved using this method.

Vidita Tilva, Jignesh Patel, Chetan Bhatt (2013) Integrated pest management (IPM) is an integrated approach that combines practices to minimize the loss of agricultural products due to the optimal use of pesticides by pests and pathogens. Early detection and control of pests is an aspect of IPM. Weather-based prediction is a recognized method. Various meteorological data such as temperature, humidity, and leaf wet duration (LWD) play a crucial role in the growth of microorganisms that cause disease. Effectively predicting such diseases based on climate data can help farmers take timely action to suppress disease. This can also rationalize the use of pesticides, which is one of the reasons behind land pollution. The weather-based forecasting system can be considered as part of the Agricultural Decision Support System (ADSS), which is a knowledge-based system (KBS). This paper proposes the structure of plant disease prediction system based on fuzzy logic. It has been shown that the proposed method can be implemented with minimal weather data such as temperature and humidity.

The conclusion of this paper is that Integrated Pest Management (IPM) is a method of controlling disease that allows optimal use of pesticides. Weather-based disease prediction is one of the IPM methods. Various meteorological data such as temperature, relative humidity, and leaf humidity duration and wind speed are used to detect disease probabilities in crops early. The characterization of various meteorological data in linguistic variables is logical for describing favorable climatic conditions for the disease. Fuzzy logic provides the possibility to define linguistic variables. The proposed expert system estimates the likelihood of disease in plants. This is a reflection of the fuzzy

logical structure of weather-based plant disease prediction. Implementing and implementing an expert system with most plant disease estimates is a task for the future.

3. APPROACH AND ALGORITHM

This paper approaches three different stages as follows.

- Classification of Data Sets
- Categorized the Data Sets as per the impact fullness
- Put the data Sets for defining critical variables.

4. CONCLUSIONS

The proposed structure greatly improves the way of investigation of disease from the integrated console.It provides the user or patient to monitor disease without consulting doctors. It is also capable to provide the suggested medicine if find problem and also suggest doctors of respective disease.An Integrated Machine learning tools technique provide the specific keyword that will be useful for finding the desired attributes. It will be collected through various social sites or any other means by internet source. So large dataset attributes can find out with number of occurrence is also been used for making the console.In future the availability of doctor to large population is a very challenging task for human kind as well a change in life style is also very deeply influenced the life of urban population. It increases the chance of occurring the chronic disease.The proposed work can be very useful to handle this situation or can minimize the future problems.

REFERENCES

1. Sweety, M.E., & Jiji, G.W. (2014). Detection of Alzheimer disease in brain images using PSO and Decision Tree Approach. *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 1305-1309.
2. Taie, Shereen & Ghonaim, Wafaa. (2017). Title CSO-based algorithm with support vector machine for brain tumor's disease diagnosis. 183-187. 10.1109/PERCOMW.2017.7917554.
3. Dey, Rajeeb & Bajpai, Vaibhav & Gandhi, Gagan & Dey, Barnali. (2009). Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus. 1 - 4. 10.1109/ICIINFS.2008.4798367.
4. Uma Rani, K & Holi, Mallikarjun. (2013). Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. 76-79. 10.1109/PHT.2013.6461288.
5. S. N. Kumar, D. Dinesh, T. Siddharth, S. Ramkumar, S. Nikhill and R. Lavanya, "Selection of features using Particle Swarm Optimization for microaneurysm detection in fundus images," *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 2017, pp. 140-144. doi: 10.1109/WiSPNET.2017.8299735.
6. Ballini, Rosangela & Gomide, F. (2002). A recurrent fuzzy neural network: Learning and application. 153. 10.1109/SBRN.2002.1181460.
7. Padol, P.B., & Sawant, S.D. (2016). Fusion classification technique used to detect downy and Powdery Mildew grape leaf diseases. *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 298-301.
8. P. Ravivarma, B. Ramasubramanian, G. Arunmani and B. Babumohan, "An efficient system for the detection of exudates in colour fundus images using image processing technique," *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, 2014, pp. 1551-1553. doi: 10.1109/ICACCCT.2014.7019366.
9. Undre, Poonam & Kaur, Harjeet & Patil, Prakash. (2015). Improvement in prediction rate and accuracy of diabetic diagnosis system using fuzzy logic hybrid combination. 1-4. 10.1109/PERVASIVE.2015.7087029.
10. Tilva Vidita, Patel Jignesh, & Bhatt Chetan (2013). Weather based Plant Diseases Forecasting System using Fuzzy Logic. NUiCONE-13. Ahmedabad, India.
11. N. G. Hedeshi and M. S. Abadeh, "An ensemble PSO-based approach for diagnosis of coronary artery disease," *2011 International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Tehran, 2011, pp. 77-82.doi: 10.1109/AISP.2011.5960975