Medical Diagnosis System Using Machine Learning

Dhaval Raval[1], Dvijesh Bhatt[2], Malaram K Kumhar[3], Vishal Parikh[4], Daiwat Vyas[5]

[12345]Nirma University, Computer Science and Engineering Department, Ahmedabad, India

14mcen16@nirmauni.ac.in, dvijesh.bhatt@nirmauni.ac.in,
malaram.kumhar@nirmauni.ac.in,daiwat.vyas@nirmauni.ac.in

Abstract: Disease prediction is one of the critical task while designing medical diagnosis software. Artificial intelligence and neural network are two major techniques which are already used to solve this type of medical diagnosis problem. Recently, Machine Learning techniques have been successfully utilized in a different applications including to assist in medical diagnosis. It is very effortless and on time process for patients to analyze disease based on clinical and laboratory symptoms with appropriate data and give more efficient result for specificdisease. In this paper, first we have observed the current scenario of medical diagnosis system with different data mining techniques and later we have proposed an algorithm to predicate the Swine Flu disease based on several attributes,
Keywords: Machine Learning Algorithm,Data Mining,Medical Data,Diagnosis, Neural network

1. Introduction

In computer science, artificial intelligence is used generate more imaginative machine.Learning system is primary requirement for any imaginative system.There are numbers of researchers presently concede that without learning system, machine can't produce effective outputs.Thus,Machine Learning is dominant branches of artificial intelligence.Machine learning algorithms are used to analysis data again and again to produce most effective results. Presently machine learning provides essential of machine for imaginative data scrutinize.Currently medical clinics are very well furnished with fully automatic machines and those machines are generating huge amount of data, then those data are collected and shared with information systems or with doctor to take required steps.Machine learning techniques can be used for the analysis of medical data and it is helpful in medical diagnosis for sensing different specialized diagnostic problems.Using Machine learning, systems take the patient data like symptoms, laboratory data and some of the important attributes as an input and generates theaccurate diagnosis results. Based on the accuracy of the result, machine will decide which data will be worked as training and trained dataset for the future reference. In current scenario, doctor is collecting all the record of the patient and based on that he will give medicines to patients. With this scenario, huge amount of time is wasted due to several reasons which some time produced disaster in any once life. By using machine learning classification algorithms, for any specific disease, we can improve the accuracy, speed, reliability and performance of the diagnostic on the current system.Machine learning is capable of offering automatic learning techniques to extract common patterns from realistic data and make sophisticate and accurate decisions, based on the different learningbehaviors. But major problem with medical data are,most of medical data have huge number of dimensionality somedical applications find problems of frequently changing of data, human error while entering data manually, rule-based heuristics intractable. In this paper, we tried to resolve the issue of current system. Thus, weproposed the new approach which can predict the Swine Flu dieses.Main purpose of this system is to assist in medical diagnosis of Swine Flu. It is very effortless and on time process for patients to analyze disease based on clinic and laboratory symptoms and data to give the more accurate result of Swine Flu disease. Also it will help to detect the dieses in primary stage.

2. Objectives

As per public documents, Swine flu is the disease of applicability and wide spread of swine flu in the Country

- The H1N1 virus killed many people in last several years since 2009.In 2009, the count was 981.In 2010, it was 1763. The rate decreased up to 75 people in 2011 but again increased to 405,218 and 699 in year 2012, 2013 and 2014, respectively.
- According to health ministry of India by 30 march 2015, there are 31,974 reported cases and 1,895 people had lost their lives to the disease.
- In Gujarat, 6495 Cases were reported and 428 Deaths were occurred due to Swine flu by the date 30 March 2015.
- This disease is highly infectious and can be life threatening to humans. Chances of being epidemic in India are very high.
- The high virulence of Swine flu makes it eligible candidate for development of newer methods of detection for patient and public safety.

A. **Data Requirements**
- **Dataset:** We have required data set of Swine Flu for the simulations. Its data collected from the Civil Hospital Ahmadabad.
- **Data Collections:** The following details are collected from the suspected Swine flu patients:
  - Name: It is defined the patient's name.
  - Age: It is defined the age of the patient's in data set.
  - City: It is defined the city/village of the patient.
  - State: It is defined the state of the patient.
  - Clinical Symptoms: It is defined the Clinical symptoms like Fever,Fever Duration,Stuffy Nose,Sore Throat,Headache,Respiratory Distress.
  - Laboratory Indicators: It is defined the Laboratory symptoms like N/L Ratio and WBC Count.
  - Diagnostic result with Swine Flu Positive cases and Swine Flu Negative cases.

B. **Swine Flu Detection Time**

There are two categories, i.e Major and Minor, where most of the time are spent to detect the Swine flu.
- **Minor Time**
  - Sample collections from patients
  - Sample sending to respective laboratory
- **Major Time**
  - Sample Differentiation
  - Analytical Media Preparation
  - Analysis Time
  - Data Interpretation
  - Report Generation

3. Literature Survey

Disease prediction is most important for medical system to make the best possible medical care decisions.Incorrect decisions are likely to purpose suspensions in medical treatment or even loss of life.A number of disease prediction models are used in medical diagnosis system which are using data mining and machine leaning techniques like Bayesian classification,Decision Tree,Regression model,Neural Network,Single best model,Ensemble model etc. In normal medical diagnosis system, it predicates the disease based on the patients symptoms and laboratory databefore analyzing the disease. [1] This prediction techniques give the good performance and with less accuracy using medical dataset. Neural Network is improving the observation capacity of information systems over the training of a limited number of neural network's nodes and collecting their results. Proposed system is enhancing the performance and training of neural networks for the classification with using cross-validation tool for the optimizing the network parameters and architecture. Using the artificial neural network (ANN) technique for disease prediction are compared below. Limitations of this scenario are as mentioned below: [1]

    i. System is only using one data set for validation which does not predictable enough to generate outcomes.

    ii. System is only exploring the common predictable performance of their models without considering the F-score and precision as measures.

    iii. Most studies do not provide statistical test results to demonstrate the level of significance of their experimental results.

    iv. Most studies related to ensemble classifier do not compare the performance difference between individual classifiers and an ensemble classifier consisted of individual classifiers.[1]

Already some researched in automated imaginative systems for medical applications is an essential and impressive. The classification of automated decision support system is a feasible to collaborating physician rapidly and accurate diagnose patients. [2]Automated systems are useful in giving fast and accurate results. It is also helpful in reducing cost and time. It uses patient database to give enhanced results. [2] The Fuzzy Min-Max network for medical diagnosis system shows how the Classification, Regression Tree (CART) and Random Forest (RF) models are integrated to make a hybrid intelligent system. Rule Extracting is the pros of the CART and it is in a tree based structure. It is not more flexible to perform same accuracy on medical data samples. Absences of capability of predictions is the pros of the FMM. In medical system, accuracy of the DSS is more decisive. The prospective method is not gain the high precision, sensitivity, and specificity rates, but still to contribute description for its prognosis in the structure of a decision tree. [3] Using Fuzzy Hierarchical Approach to Medical Diagnosis we can improve the results by following ways: [4]
- Complexity in diagnosis process.
- Simple fuzzy logic which does not provide hierarchical structure.
- Uncertainty occurred by different diagnosis system.

- Problem occurred in other diagnosis system where grammatical labels comprehend to actual code in a period a numbers of values sensible process can be solved by this approach.

Using Limitation of the profession and aspect for medical diagnosis system using machine learning we can improve the result as it: [5]

- Provides an analysis of the automated data scrutiny.[6]
- Significance the naive Bayesian, neural network, decision trees.
- Specific requirement for machine learning systems[7]
   i. Good Performance
   ii. Dealing with missing data
   iii. Dealing with noisy data
   iv. Reduction of the number of tests

Predictive models can be used for dengue outbreak detection. It is working as shown in the figure 1. Predictive model uses different rule based classifiers for detection. Classifiers such as Rough set classifier, Naïve Bayes classifier, Associative classifier, Decision Tree classifier are used for dengue detection. [8].
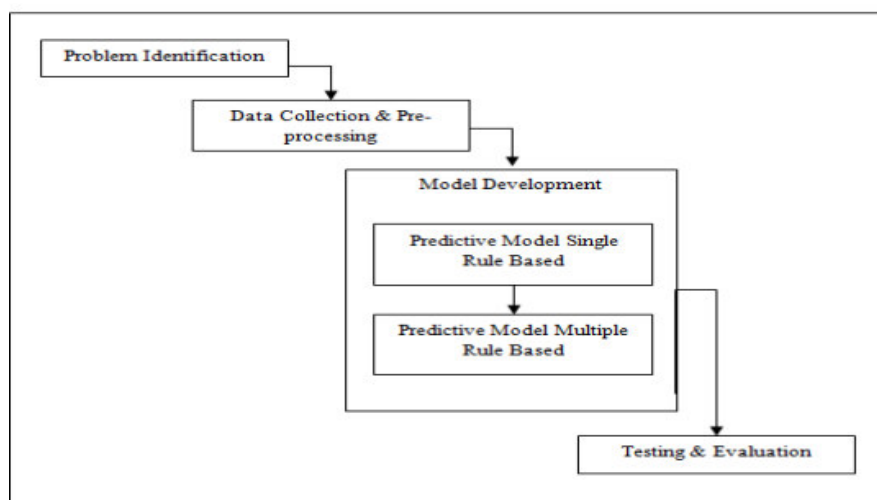

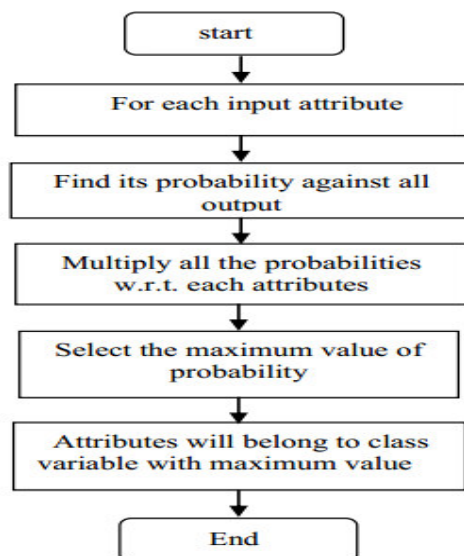Figure 1.   Existing Framework [8]

Figure 2. Naïve Bayes Algorithm [9]

Naive Bayes Classifier, as per figure 2, is used previously for prediction of Swine Flu Disease [9]. Swine flu transmission generally occurs due to sneeze or cough droplets of people. The droplets which

contains virus are inhaled by other people and thus the transmission occurs. [9][10] In second method which uses neonatal screening of dried blood spots and protein microarray to monitor the trends of the 2009 influenza A (H1N1) virus[11]. Electrocardiogram (ECG) and auscultator blood pressure signals are used to examine two real world classification problems. The outcome of the computation of performance matrices such as accuracy, sensitivity,specificity and the area under receiver operating characteristic curve suggests that with original dataset logistic regression models are good but for noisy dataset, ensemble machine learning models are more appropriate.[12]. In one more approach in which researchers used the data mining along with non-wearable sensor hardware is also proposed to give the difference between on and off medication states. [13]

Exciting data mining and Machine learningtechniques which are used in medical field.

Table 1: Techniques are used in medical data analysis

Techniques	Description
Support Vector Machine	• SVM method works based on particular disease and gives the accurate result also. When it was used for another disease, it was not produced the accurate result because for modification in current algorithm is required for particular disease.
Naïve Bayes	• Probability based checked and found the disease. • Using the class conditional probability. • Which disease of probability high than it bias on that disease and give the result of the highly probable disease.
Decision Tree	• Complexity increases. • Time consuming process.
Random Forest	• Forest is made using more than one decision tree • Splitting criteria of decision tree is random attribute selection. • Random vectors are sampled independently and with equal distribution among all the trees. • The class selection process includes voting of every tree and majority voting class is returned.
Clustering Method	• Simultaneously change cluster based on symptoms. • It's not give the accurate result for the number of diseases.
Logistic Regression	• Recursive Procedure. • Time consuming process.
Backpropagation	• Predefined hidden units. • Time complexity increase.

4. Proposed Solution

Here we are proposing the algorithm using the neural network with feed-forward network. Backpropagation algorithm is used for learning procedure andfor training the multilayer feed-forward network. It can be utilized for purpose like medical diagnosis, pattern classification, image processing, character recognition etc. But Traditional approach of this algorithm is need to be determined the number of units in the hidden layer before training is started in the neural network. To overcome this difficulty many algorithms, that construct a network dynamically, had been proposed. Out of them, the well-known constructive algorithms are dynamic node creation (DNC), feed-forward neural network construction (FNNC) algorithm and the cascade correlation (CC) algorithm. We have used FNNC. In figure 3, we are displaying our algorithm flow. In first step, we are creating and initializing neural network with one hidden layer and attributes of dataset and give random weight to all the feature and hidden layer nodes links. Using the backpropagation on first hidden layer, we try to minimize the error function. Error function is used as mention below.

Error Function: $£(n) = 1/2 \sum_{k \in P} ek^2(n)$ ….. (1)

Where ek= desired output from k^{th} neuron - actual output from k^{th} neuron

Once after minimizing the error function, we train the data and then calculate the error function on validate set and try to execute the test to generate the classify patterns with efficiency. If calculated efficiency is acceptable means we get accurate result then we stop the execution and find the predicted value else we will add n hidden layers and calculate weights again and initialize it until we get accurate result.
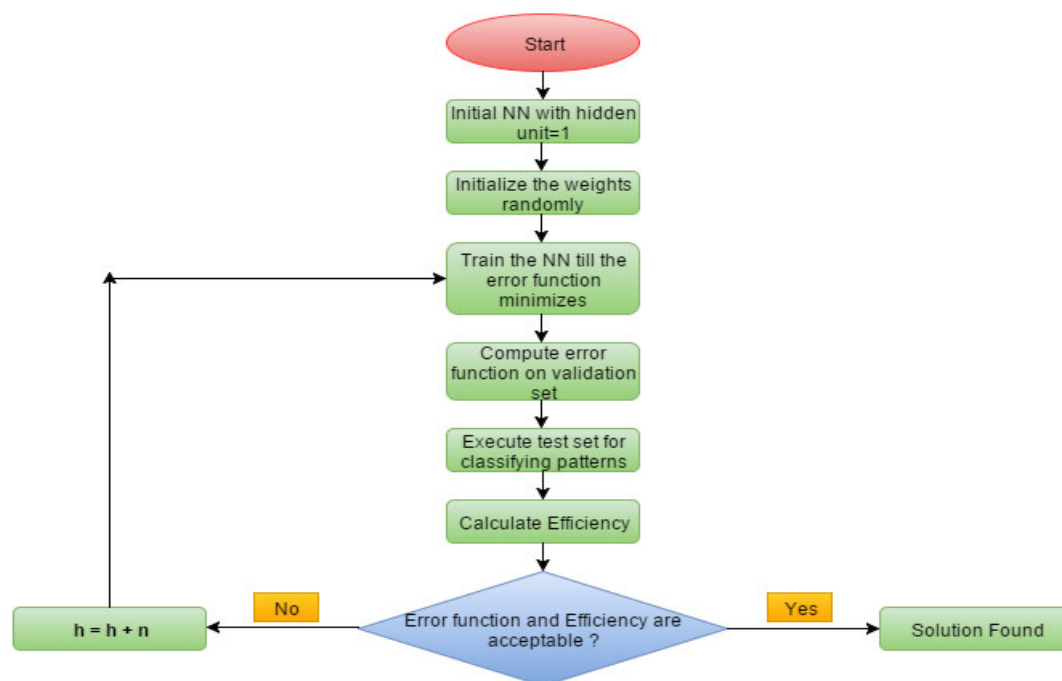
Figure 3: Flow chart of proposed system

5.   Conclusion and Future Work

In this paper, the issue of current medical diagnosis system and various data mining algorithms are used for the medical prediction is explained. The focus is on using different algorithms and consolidation of certain target attributes to predict swine flu effectively using data mining and neural network. For predicting swine flu, significantly 12 attributes are listed and give priories using information gain and using hidden layer of neural network we tried to improve the accuracy and result of current working sysemt. We have already apply the data mining techniques like SVM,Naive Bayes,KNN on swine flu actual data to get the optimal outputs. The proposed work will be further increased developed for the automation of the swine flu disease prediction more accurately.

References

[1].   C.-H. Weng, T. C.-K. Huang, and R.-P. Han, "Disease prediction with different types of neural network classifiers," Telematics andInformatics, vol. 33, no. 2, pp. 277–292, 2016.

[2].   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, no. 8, pp. 43–48, 2011.

[3].   C. L. ManjeevanSeera, "A hybrid intelligent system for medical data classification," Expert Systems with applications41 (5), pp. 2239–2249, 2014.

[4].   S. Zahan, C. Michael, and S. Nikolakeas, "A fuzzy hierarchical approach to medical diagnosis," in Fuzzy Systems, 1997., Proceedings ofthe Sixth IEEE International Conference on, vol. 1, pp. 319–324, IEEE, 1997.

[5].   I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," Artificial Intelligence in medicine, vol. 23, no. 1, pp. 89–109, 2001.

[6].   R. Isola, R. Carvalho, M. Iyer, and A. K. Tripathy, "Automated differential diagnosis in medical systems using neural networks, knn and som," in Developments in E-systems Engineering (DeSE), 2011,pp. 62–67, IEEE, 2011.

[7].   V. S. H. Rao and M. N. Kumar, "A new intelligence-based approach for computer-aided diagnosis of dengue fever," Information Technology inBiomedicine, IEEE Transactions on, vol. 16, no. 1, pp. 112–118, 2012.

[8].   A. A. Bakar, Z. Kefli, S. Abdullah, and M. Sahani, "Predictive models for dengue outbreak using multiple rule base classifiers," in ElectricalEngineering and Informatics (ICEEI), 2011 International Conference on, pp. 1–6, IEEE, 2011.

[9].   A. Borkar and P. Deshmukh, "Naive bayes classifier for prediction of swine flu disease," vol. 5, pp. 120–123, 2015.

[10].   B. Thakkar, M. Hasan, M. Desai, et al., "Health care decision support system for swine flu prediction using naïve bayes classifier," inAdvances in Recent Technologies in Communication and Computing(ARTCom), 2010 International Conference on, pp. 101–105, IEEE, 2010.

[11].   E. de Bruin, J. Loeber, A. Meijer, G. M. Castillo, M. G. Cepeda, M. R. Torres-Sepulveda, G. Borrajo, M. Caggana, Y. Giguere, M. Meyer, ´et al., "Evolution of an influenza pandemic in 13 countries from 5continents monitored by

protein microarray from neonatal screening bloodspots," Journal of Clinical Virology, vol. 61, no. 1, pp. 74–80, 2014.

[12].   M. Seera, C. P. Lim, W. S. Liew, E. Lim, and C. K. Loo, "Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models," Expert Systems with Applications, vol. 42, no. 7, pp. 3643–3652, 2015.

[13].   C. S. Tucker, I. Behoora, H. B. Nembhard, M. Lewis, N. W. Sterling, and X. Huang, "Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors,"Computers in Biology and Medicine, vol. 66, pp. 120–134, 2015.