

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka



A Justification Report

Submitted by
Rahul M (1DT15CS091)
Surya Gangaraj K (1DT15CS115)
Mohit Surana(1DT15CS069)
Sanketh Bennur(1DT15CS102)



DAYANANDA SAGAR ACADEMY OF TECHNOLOGY & MANAGEMENT

22 Mile , Udayapura Opp. To Art of Living Kanakapura Road, Bengaluru
- 500082

Department of Computer Science and Engineering

Remarks:

**Signature of the Project
Coordinator**

Signature of the HOD

Medical Diagnosis Using Machine Learning

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 specific disease.

This model takes the symptoms form the user and using **machine learning** model the app will analysis the diseases. After analyzing it will suggest best remedies or prevention methods to the user.

Advantages

- Machine learning and new smartphones makes it faster and easy to use for the user.
- This user will be able to use this given informations to take quick actions to prevent the disease occuring.

Disadvantages

- It was very difficult to fetch the data and analyze it. Without proper trained model it is impossible to get an model of this level. Now we can archive it using cloud computing and these servers are used for training the model.

Applications

- This method can be further expanded and modified so that end users will get benefited more.
- This model can be implemented in android and watchOs as an app so that the user can carry it with him.
- This app will keep on reminding the user about remedies and prevention methods for that diseases.

Medical Diagnosis System Using Machine Learning

Summary:

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 specific disease. 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 predict the Swine Flu disease based on several attributes.

Main contributions and strengths:

- Reduces complexity in diagnosis process.
- Logic is made clearer and provides a hierarchical structure.
- Uncertainty occurred by different diagnosis system is eliminated.
- 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 using this approach.

Main Weaknesses:

- System is only using one data set for validation which does not predictable enough to generate outcomes.
- System is only exploring the common predictable performance of their models without considering the F-score and precision as measures.
- Most studies do not provide statistical test results to demonstrate the level of significance of their experimental results.
- Most studies related to ensemble classifier do not compare the performance difference between individual classifiers and an ensemble classifier consisted of individual classifiers.

Other Interesting thoughts raised by the paper: like,

- In normal medical diagnosis system, it predicates the disease based on the patients symptoms and laboratory data before analyzing the disease.
- In medical system, accuracy of the DSS is more decisive.

A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles.

Summary

This paper discusses a technique of Computer-based clinical decision support systems (CDSSs) in which proposed to address deficiencies and have significantly improved clinical practice over the past decade. However, they remain limited to clinics and hospitals, and do not take advantage of patient data that are obtained on a daily basis using wearable medical sensors (WMSs) that have the ability to bridge this information gap. WMSs can collect physiological signals from anyone anywhere anytime.

Main contributions and strengths:

- Need for reliable, accurate, and intelligent out-of-clinic decision support complement for current CDSSs.
- Avoiding non-uniformity of diagnosis offered by doctors.
- Supervised MLAs make predictions using mathematical rules learned from a labeled training dataset .

Main Weaknesses:

- Collecting WMS data from individuals to tackle more challenging tasks.

Other Interesting thoughts raised by the paper: like,

- This method consists of automatic DDM generation procedure that can monitor various diseases in parallel.

Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction

Summary

Despite of abundant data available within the healthcare systems, there is a lack of effective analysis tools to discover hidden relationships and trends in data. This research paper uses data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction.

Three different supervised machine learning algorithms i.e. Naive Bayes, KNN, Decision List algorithm have been used for analysing the dataset. Tanagra tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared.

After number of experiments, outcome reveals that Decision Tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The second conclusion is that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute enough for heart disease prediction.

Main strengths and contributions:

- Decision support systems aids in achieving clinical tests at a reduced cost.
- Genetic algorithm used reduces the actual data size to get the optimal subset of attributes enough for heart disease prediction.
- Tanagra tool used is a powerful system that contains clustering, supervised learning, meta supervised learning, feature selection, data visualization supervised learning assessment, statistics, feature selection and construction algorithms.

Main weaknesses:

- Attribute values need to be normalized before clustering to avoid domination between high and low value attributes.
- Real data is not used to validate the algorithms.

Other interesting thoughts raised by the paper: like,

- Incorporation of genetic algorithm in decision tree outperforms the other data mining techniques and gives the best accuracy.

Diagnosis of Ischaemic Heart Disease

Summary

This paper discussed about diagnosis of a heart disease called ischaemic using machine learning. They have conducted many iterations (about 327) on the same patient to get better accuracy in the classification. And apart from that each and every data is evaluated by physicians. Algorithms used are following naive Bayesian classifier, backpropagation learning of neural nets with weight elimination and algorithms for induction of decision.

The accuracy of the model is calculated using the following formula :

accuracy = (true positive+true negative test results) / all patients

sensitivity = true positive test results / all patients with disease

specificity = true negative test results / all patients without disease

Main strengths and contributions:

- Machine Learning techniques
- Have taken around 327 samples from a single patient and each sample is verified with physician.
- Implemented better formula for getting accurate predictions.

Main weaknesses:

- Real data is not used to validate the algorithms.

Other interesting thoughts raised by the paper: like,

- The accuracy of the model plays an important role in the prediction, it's better to recheck each and every data which we are entering in to model. Here in this team verified each sample with the help of a good physician.

Medical Diagnosis System Using Machine Learning

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 specific disease. 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,

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 the accurate 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 learning behaviors. But major problem with medical data are, most of medical data have huge number of dimensionality some medical 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, we proposed the new approach which can predict the Swine Flu diseases. 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 diseases in primary stage.

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.

Data Requirements

- **Dataset:** We have required data set of Swine Flu for the simulations. Its data collected from the Civil Hospital Ahmedabad.
- **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

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**

- O Sample collections from patients
- O Sample sending to respective laboratory

- **Major Time**

- O Sample Differentiation.
- O Analytical Media Preparation o Analysis Time.
- O Data Interpretation.
- O 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 learning 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 data before 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]

Good Performance

Dealing with missing data Dealing with noisy data 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].

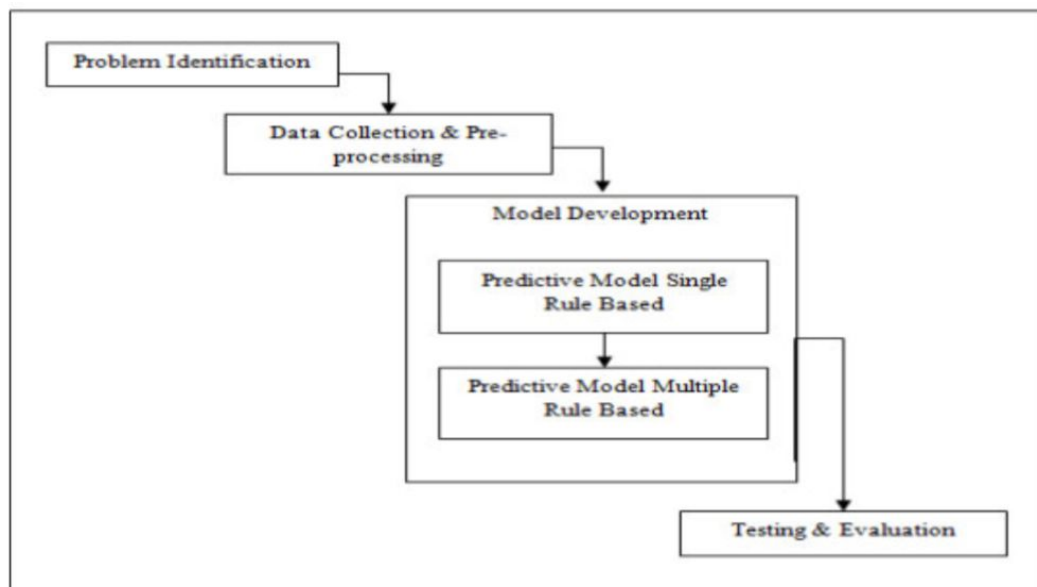
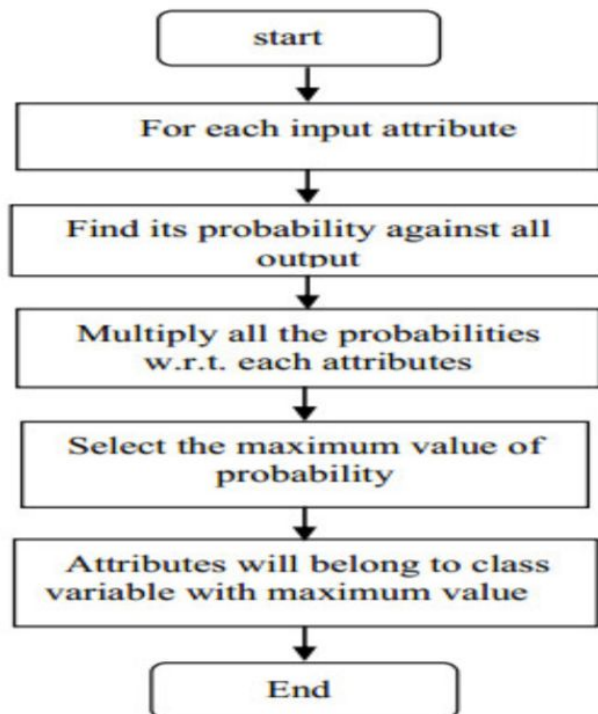


Figure 2. Naïve Bayes Algorithm [9]

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 auscultatory blood pressure signals are used to examine two real world classification problems. The outcome of the computation of performance metrics 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 learning techniques which are used in medical field. Table 1: Techniques are used in medical data analysis

Table 1: Techniques are used in medical data analysis

Techniques	Description
Support Vector Machine	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Complexity increases. Time consuming process.
Random Forest	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Simultaneously change cluster based on symptoms. It's not give the accurate result for the number of diseases.
Logistic Regression	<ul style="list-style-type: none"> Recursive Procedure. Time consuming process.
Backpropagation	<ul style="list-style-type: none"> Predefined hidden units. Time complexity increase.

Proposed Solution

Here we are proposing the algorithm using the neural network with feed-forward network. Backpropagation algorithm is used for learning procedure and for 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 mentioned below.

Error Function: $\mathbf{E(n)} = 1/2 \sum_{k \in P} e_k^2(n)$

Where e_k = 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.

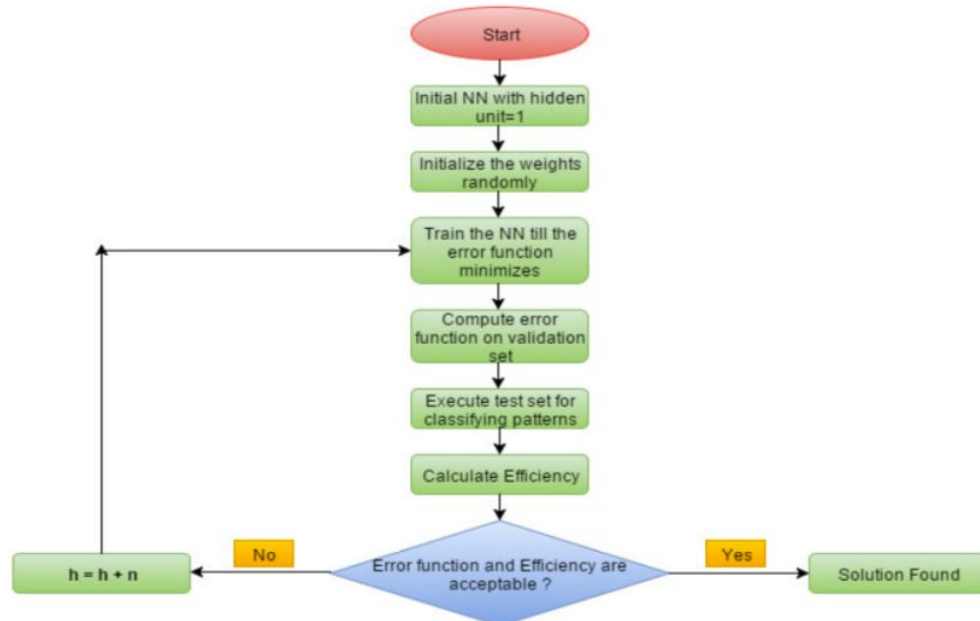


Figure 3: Flow chart of proposed system

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 priorities using information gain and using hidden layer of neural network we tried to improve the accuracy and result of current working system. 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 and Informatics*, 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. Manjeevan Seera, "A hybrid intelligent system for medical data classification," *Expert Systems with applications* s41 (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 of the 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 in Biomedicine, 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 *Electrical Engineering 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," in *Advances in Recent Technologies in*

Communication and Computing(ARTCom), 2010 International Conference on, pp. 101–105, IEEE, 2010.