AN EXPERT SYSTEM FOR THE DIAGNOSIS OF DIABETIC PATIENTS USING DEEP NEURAL NETWORKS AND RECURSIVE FEATURE ELIMINATION

J. Vijayashree, J. Jayashree
School of Computer Science and Engineering, VIT, Vellore, India

ABSTRACT

The aim of this study is to develop an expert system for the diagnosis of diabetic patients. PIMA Indian Diabetes Dataset is used by the proposed system. The proposed system uses feature reduction techniques like Recursive feature elimination and principal component analysis for enhancing the performance. The expert system is built using Deep Neural Networks and Artificial Neural Networks. This study also compares the performance of proposed deep neural network with artificial neural networks using various performance measures: accuracy, sensitivity, specificity.

Key words: Artificial Neural Networks, Deep Neural Networks, Diabetes mellitus, Feature reduction.


1. INTRODUCTION

Diabetes Mellitus (DM) is a chronic disease which increases the concentration of glucose in blood. This is because of two reasons; one the pancreas inability to produce enough insulin and the other is the inadequacy of body to use the insulin. Type1 or insulin dependent or juvenile onset diabetes is due to lack of insulin and Type2 or insulin independent or adult onset diabetes is due to the inefficiency of body to use insulin. Gestational diabetes is the third type of diabetes caused during pregnancy due to high glucose level in blood. This type of diabetes disappears after delivery but the mother and child are prone to type2 diabetes in future.

According to International Diabetes Federation (IDE) statistics 2015, South East Asia region stands second next to Western Pacific region in diabetes. India, Bangladesh, Bhutan, Sri Lanka, Nepal, Mauritius and Maldives are the seven countries belonging to South East Asia region. About 8.5% that is 78.3 million inhabitants suffer with diabetes of which 52.1% are undiagnosed and 1.2 million deaths occur every year due to diabetes.
According to 2014 estimates of World Health Organization, the population suffering with diabetes has escalated from 108 million in 1980 to 422 million in 2014 and the universal frequency of diabetes amongst adults has escalated from 4.7% in 1980 to 8.5% in 2014. Also in 2030 diabetes will be the 7th leading cause of death.

Family history, unhealthy eating, lack of exercise, increased weight, high blood pressure and history of gestational diabetes are the risk factors which influences diabetes. Most substantial indicators of diabetes are frequent urination, excessive thirst, weight loss and lack of energy. Diabetes may affect other parts of the body like blood vessels, eyes, kidneys and nerves which lead to other health issues like cardiovascular disease, blindness (diabetic retinopathy), kidney failure (diabetic nephropathy) and lower limb amputation (diabetic neuropathy).

Diabetes affects different parts of the body leading to other diseases like heart attack, kidney failure, loss of vision and foot ulcers. Heart attack and kidney failure are the major cause of death worldwide. So it is necessary to diagnose diabetes at an early stage to avoid unnecessary deaths. If diabetes is diagnosed at premature phase, it can be meticulously controlled by means of proper diet, insulin, regular exercise and medication.

Large amount of patient records were available in the hospitals using which medical practitioners find useful patterns. Classification, prediction and diagnosis of the diseases can be done with the help of patterns found in the data set. Evaluations of patient’s record and medical expert’s decision are the most important factors in medical diagnosis. Nowadays expert systems, artificial intelligent techniques and classification systems play a major role in helping medical experts to diagnose the disease like tumors, cancers, hepatitis, lung diseases, etc. in an efficient way.

2. LITERATURE REVIEW


Nongyao Nai-arun and Rungruttikarn Moungmai (2015) conducted a study to perform a comparison between the performances of five classifiers; decision Tree, Artificial Neural Networks, Logistic Regression, Naïve Bayes and Random Forest. Then to improve the robustness of these classifiers bagging and boosting techniques were applied. The study concluded that random forest algorithm is the best.

Performance analysis of four prediction models; J48 Decision Tree, K-Nearest Neighbors, Random Forest and Support Vector Machines has been done by J. Pradeep Kandhasamy*, S. Balamurali (2015) under two situations. One with noisy data and the other without noisy data. Analysis revealed that prediction models KNN and Random Forest produces good results after removing noisy data.


Lukmanto (2015) proposed fuzzy hierarchical model for early detection of diabetes mellitus. In order to overcome the crisp boundary problem in traditional Decision Tree, varma, kamadi VSRP.et.al used Fuzzy decision boundaries and proposed a modified Gini Index-Gaussian fuzzy Decision Tree algorithm where split points are identified using gini index.
Patil (2010) proposed a Hybrid Prediction model to predict type-2 diabetic patients. This model uses K-Means clustering algorithm to remove the incorrectly classified instances. Then classification is performed by constructing the decision tree using C4.5 algorithm.

M. Durairaj, G. Kalaiselvi (2015) developed diabetes prediction model using three algorithms; Naïve Bayes, Artificial Neural Networks & K_Nearest Neighbors and their efficiency was calculated. The result indicate that ANN was the best followed by Naïve Bayes and then KNN.

M. Durairaj, G. Kalaiselvi (2015) reviewed various data mining techniques like Artificial Neural Network, Support Vector Machine, K-Nearest Neighbors & C4.5 and their applications in different medical data sets. Review results showed that Artificial neural networks has got the highest predicted accuracy among the four data mining techniques compared.

3. MATERIALS AND METHODS

3.1. Dataset

The benchmark dataset used in this study is the Pima Indian Diabetes Dataset from University of California, UCI repository of machine learning databases. The dataset contains 768 instances where each instance contains 8 attributes and 1 class variable. The values of all the 8 attributes are of numeric type and the class variable has two values 1 and 0. Class value 1 is inferred as tested positive for diabetes and class value 0 is inferred as tested negative for diabetes.

3.2. Proposed Work

![Figure 1 Process flow of proposed model](image)
Figure 1 shows the process flow of the proposed method. The proposed method uses PIMA Indian diabetes dataset. The dataset contains some missing values. In order to remove the missing values pre-processing is done by filling the missing values using null value. The dataset consists of eight features and all the eight features may not have utmost importance in diagnosing the disease. To find the best features that could diagnose the disease feature reduction techniques like recursive feature elimination and principal component analysis is used. After feature reduction only four features were selected and they form the input to the classification networks. In this paper, the proposed deep neural network is compared with the artificial neural networks for accuracy, sensitivity and specificity.

3.3. Feature Reduction Strategies

Data and learning algorithm are the two main characteristics that distress the performance of a classifier. Problems with regard to data are either too much data or to little data or noisy data. Problems with regard to the choice of learning algorithm may affect the learned results, predictive accuracy, speed of learning and comprehensibility. In feature reduction both data and the classifier are considered to overcome the hindrances caused by data as well as to learn efficiently. Feature selection deals with the problems regard to data.

Feature selection is a process that chooses an optimal subset of features according to a certain criterion (Liu and Motoda, 1998). It removes redundant or irrelevant features from the dataset which can reduce the classification accuracy and clustering quality. The main advantage of feature selection is to reduce the cost and complexity, and to improve the accuracy and visualization of the classification model (Sebban and Nock (2002).

Feature selection method is of three types based on how they combine selection algorithm and model construction; Filter, Wrapper and Embedded method. Filter method selects a subset of variables independent of the later applied machine learning algorithm that shall subsequently use them. Wrapper method is a feedback method that selects a subset of variables by incorporating the machine learning algorithm in feature selection process. Whereas in embedded method the feature selection process is built into the machine learning algorithm itself.

3.3.1. Recursive Feature Elimination

RFE algorithm aims to find the best performing feature subset. First the model is trained on the initial set of features and weights are assigned to each one of the features. Then the feature whose weight is smallest among all is detached from the initial of features. Again re-compute the weight on the existing set of features. This procedure is recursively repeated until the desired number of features is reached which is the stop criterion. Recursive feature elimination works as follows.

Step 1 - Train the model using initial set of features
Step 2 - Compute weight (w) on all the features in initial set
Step 3 - Detach the feature with the smallest weight (wi) from the initial set
Step 4 - Train the model using existing set of features
Step 5 - Re-compute weight on the features in the existing set
Step 6 - Detach the feature with the smallest weight from the existing set
Step 7 - Until stop criterion is met go to step 4

3.3.2. Principal Component Analysis

Principal component analysis method is for identifying the important “directions” in the data. This can rotate data into (reduced) coordinate system that is given by those directions. Find
direction (axis) of greatest variance. Find direction of greatest variance that is perpendicular to previous direction and repeat. Implementation: find eigenvectors of covariance matrix by diagonalization. Eigenvectors (sorted by eigenvalues) are the directions.

3.4. Classification Techniques
3.4.1. Artificial Neural Networks
Artificial neural Network refers to the modelling of the brain in two aspects which are functions and structure. The models are composed of many computing elements, usually denoted by neurons; each neuron has a number of input and output as in figure 2. The input to the neuron is obtained as the sum of the synaptic weight of the input. The input value is compared to the threshold value associated with the neuron to determine the activation signal of the neuron. This activation signal is then passed through an activation function to produce the output of the neuron. Sigmoid activation function was chosen for this research work due to its soft switching operation and also it helps artificial neural network to represent a more complex problem.

3.4.2. Deep Neural Networks

![Figure 2 Artificial Neural Network Structure](image)

![Figure 3 Deep Neural Networks Structure](image)
J. Vijayashree, J. Jayashree

A deep neural network (DNN) in figure.3 is an ANN with multiple hidden layers of units between the input and output layers. Similar to shallow ANNs, DNNs can model complex non-linear relationships. The extra layers enable composition of features from lower layers, giving the potential of modelling complex data with fewer units than a similarly performing shallow network. DNNs are prone to over fitting because of the added layers of abstraction, which allow them to model rare dependencies in the training data.

4. PERFORMANCE EVALUATION OF THE PROPOSED SYSTEM

The diagnostic performance is usually evaluated in terms of confusion matrix, classification accuracy, sensitivity, specificity, precision, recall, F-score, ROC. Confusion matrix is given in table 1. The results of various performance measures is displayed in table 2 and the comparison of Precision, Recall and F-score performance measures is given in table 3.

Table 1 Confusion Matrices

<table>
<thead>
<tr>
<th></th>
<th>DNN with RFE</th>
<th>DNN with PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted</strong></td>
<td>Actual</td>
<td>Predicted</td>
</tr>
<tr>
<td>P</td>
<td>36</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>12</td>
<td>P</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ANN with RFE</th>
<th>ANN with PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted</strong></td>
<td>Actual</td>
<td>Predicted</td>
</tr>
<tr>
<td>P</td>
<td>35</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>P</td>
</tr>
</tbody>
</table>

Table 2 Results of Various Performance Measures

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>DNN with RFE</th>
<th>DNN with PCA</th>
<th>ANN with RFE</th>
<th>ANN with PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>82.67%</td>
<td>76.77%</td>
<td>78.62%</td>
<td>70%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>72%</td>
<td>66.67%</td>
<td>68.63%</td>
<td>38.78%</td>
</tr>
<tr>
<td>Specificity</td>
<td>88%</td>
<td>82.65%</td>
<td>83.33%</td>
<td>86.81%</td>
</tr>
<tr>
<td>Positive Likelihood Ratio</td>
<td>6.00</td>
<td>3.84</td>
<td>4.12</td>
<td>2.94</td>
</tr>
<tr>
<td>Negative Likelihood Ratio</td>
<td>0.32</td>
<td>0.40</td>
<td>0.38</td>
<td>0.71</td>
</tr>
<tr>
<td>Positively Predicted Value</td>
<td>75%</td>
<td>69.09%</td>
<td>66.04%</td>
<td>61.29%</td>
</tr>
<tr>
<td>Negatively Predicted Value</td>
<td>86.27%</td>
<td>81%</td>
<td>84.91%</td>
<td>72.48%</td>
</tr>
</tbody>
</table>

Table 3 Comparison of Precision, Recall and F-score performance measures

<table>
<thead>
<tr>
<th></th>
<th>DNN with RFE</th>
<th>DNN with PCA</th>
<th>ANN with RFE</th>
<th>ANN with PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tested negative for diabetes</strong></td>
<td>Precision</td>
<td>Recall</td>
<td>F-Score</td>
<td>Precision</td>
</tr>
<tr>
<td>Tested negative for diabetes</td>
<td>0.86</td>
<td>0.88</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Tested positive for diabetes</td>
<td>0.75</td>
<td>0.72</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td><strong>Tested positive for diabetes</strong></td>
<td>Precision</td>
<td>Recall</td>
<td>F-Score</td>
<td>Precision</td>
</tr>
<tr>
<td>Tested negative for diabetes</td>
<td>0.85</td>
<td>0.83</td>
<td>0.84</td>
<td>0.66</td>
</tr>
<tr>
<td>Tested positive for diabetes</td>
<td>0.72</td>
<td>0.79</td>
<td>0.79</td>
<td>0.61</td>
</tr>
</tbody>
</table>
An Expert System for the Diagnosis of Diabetic Patients using Deep Neural Networks and Recursive Feature Elimination

5. RESULTS AND DISCUSSION

Accuracy, Sensitivity, Specificity and ROC curve were the performance parameters used to compare the two classifiers: Artificial neural networks and Deep neural networks. Accuracy percentage of DNN and ANN for two different feature selection methods: RFE and PCA are shown in figure 4. Sensitivity and Specificity comparison for deep neural networks and artificial neural networks are shown in figure 5. Receiver Operating Characteristic (ROC) Curve is shown in figure 6.

![Figure 4](image-url)  
**Figure 4** Comparison of Accuracy percentage

![Figure 5](image-url)  
**Figure 5** Comparison results of sensitivity and specificity

![ROC curve for DNN with RFE](image-url)  
ROC curve for DNN with RFE

![ROC curve for DNN with PCA](image-url)  
ROC curve for DNN with PCA

http://www.iaeme.com/IJCIET/index.asp  
639  
editor@iaeme.com
6. CONCLUSIONS
Among the two features reduction techniques recursive feature elimination selects the best features for classification. The classification accuracy of proposed deep neural networks is better than artificial neural networks.

REFERENCES
An Expert System for the Diagnosis of Diabetic Patients using Deep Neural Networks and Recursive Feature Elimination


