Cascading of RBFN, PNN and SVM for Improved Type-2 Diabetes Prediction Accuracy

Krishna Swaroop
Mahindra Ecole Centrale
Bahadurpally, Hyderabad 500043
Email:swaroop14095@mechyd.ac.in

Ramalingaswamy Cheruku
School of Computer Engineering
Mahindra Ecole Centrale
Bahadurpally, Hyderbad 500043
Email: ramalingaswamy.cheruku@mechyd.ac.in

Damoder Reddy Edla
Department of CSE
National Institute of Technology Goa
Email: dr.reddy@nitgoa.ac.in

***Abstract*—Diabetes is a metabolic disorder caused by a defect in insulin secretion or action (or both) leading to hyperglycemia (high glucose levels) . Over time, hyperglycemia damages nerves and blood vessels, leading to complications like heart disease, stroke, kidney disease, blindness, nerve problems, gum infections and amputation. In order to increase the classification accuracy on diabetes data in this paper a dual-stage cascaded ensemble framework is proposed. This frame work has two stages, the first stage consists of simple Radial Basis Neural Network (RBFN) and simple Probabilistic Neural Network (PNN). The results from both the neural networks are combined and serve as inputs to the second stage classifier called support vector machine. The soundness of proposed framework is validated using Pima Indians Diabetes dataset. The Experimental results indicate that the proposed Dual stage network out performs individual as well as state of-the-art models.**

***Keywords*—RBFN, PNN, SVM, Cascading, Ensemble Technique, Diabetes prediction**

I. Introduction

Diabetes is a lifelong condition associated with serious complications.Diabetes is a chronic progressive disease characterized by increased insulin resistance and decreasing pancreatic beta cell function [1], [2]. As per UKPDS study, 40-50% of beta cell is already lost at the time of diagnosis, it continue to decline, progress and worsen with diabetes [3]. Increased insulin resistance, decreased insulin secretion and glucotoxicity leads to decompensated state of glucose metabolism resulting in Diabetes. Diabetes is one of the most challenging health problems in the world in the 21 st century, affecting more than 425 million people. The most prevalent Type 2 diabetes mellitus caused due to insulin resistance is the most common form of diabetes accounting for 90-95 of cases followed by type 1 diabetes mellitus accounting for Diabetes is the 5̂th leading cause of death by disease in the world. As of 2007, some 246 million people worldwide are estimated to have diabetes (International Diabetes Federation). Close to 40 of these patients remain un diagnosed and around remain untreated. Classification and clustering algorithms can play a very important role in prediction of diabetes [4].

Simple Radial Basis Neural Network (RBFN) and simple Probabilistic Neural Network (PNN) [5] can be trained in one single iteration and the performance of these is on par with the multi-layer feed forward neural networks (MLFFNNs) and multi-layer perceptron neural networks (MLPNNs) which are currently the most popular iteratively trained classification techniques among the scientific community [6], [7], [8]. The mediocre performance of the rule based and non-rule based classification techniques and better accuracy of ensemble methods like bagging, boosting and stacking have gained attention.

C. Kaynak et al. [9] proposed the multistage cascading of multiple classifiers.Their focus was not only on limited to improving the accuracy but also to optimize computational and space complexity. The use of single, multi-layer perceptrons and kNN in the proposed cascading model has obtained accuracy of about 77 which is more than comparaed to the accuracy of any individual classifier.

Kandhasamy and Balamurali [10] applied the random forest (RF) classifier on the PID dataset. Bashir et al. [11] proposed the HMV (hierarchical majority voting) ensemble model for disease classification and prediction with a three-layered approach and obtained an accuracy of 77.08 on the PID dataset. Also, Bashir etal. [12] proposed a medical decision support system called HM-BagMoov using a novel weighted multi-layer classifier ensemble framework. The proposed HM-BagMoov obtained an accuracy of 78.21 on the PID dataset.

Minor accuracy improvements translate into saving valuable lives of people. In an effort to increase the classifier predictive accuracy and also optimized the complexity of the model. Our proposed cascaded model comprises of simple Probabilistic Neural Network (PNN) and Simple Radial Basis Neural Network (RBFN) in the first stage followed by a Linear support vector machine (SVM) in the second stage.

II. Preliminaries

A. Simple Probabilistic Neural Network (PNN)

Specht [14] first proposed the PNNs in 1990. The learning speed of the PNN model is very fast, making it suitable in real-time disease diagnosis. A few advantages for the PNN over the conventional MLFFNN and MLPNN are:

- PNNs are computationally faster than the MLFFNN and MLPNNs.
- PNNs provide robust performance on noisy data and easily incorporate additional samples.

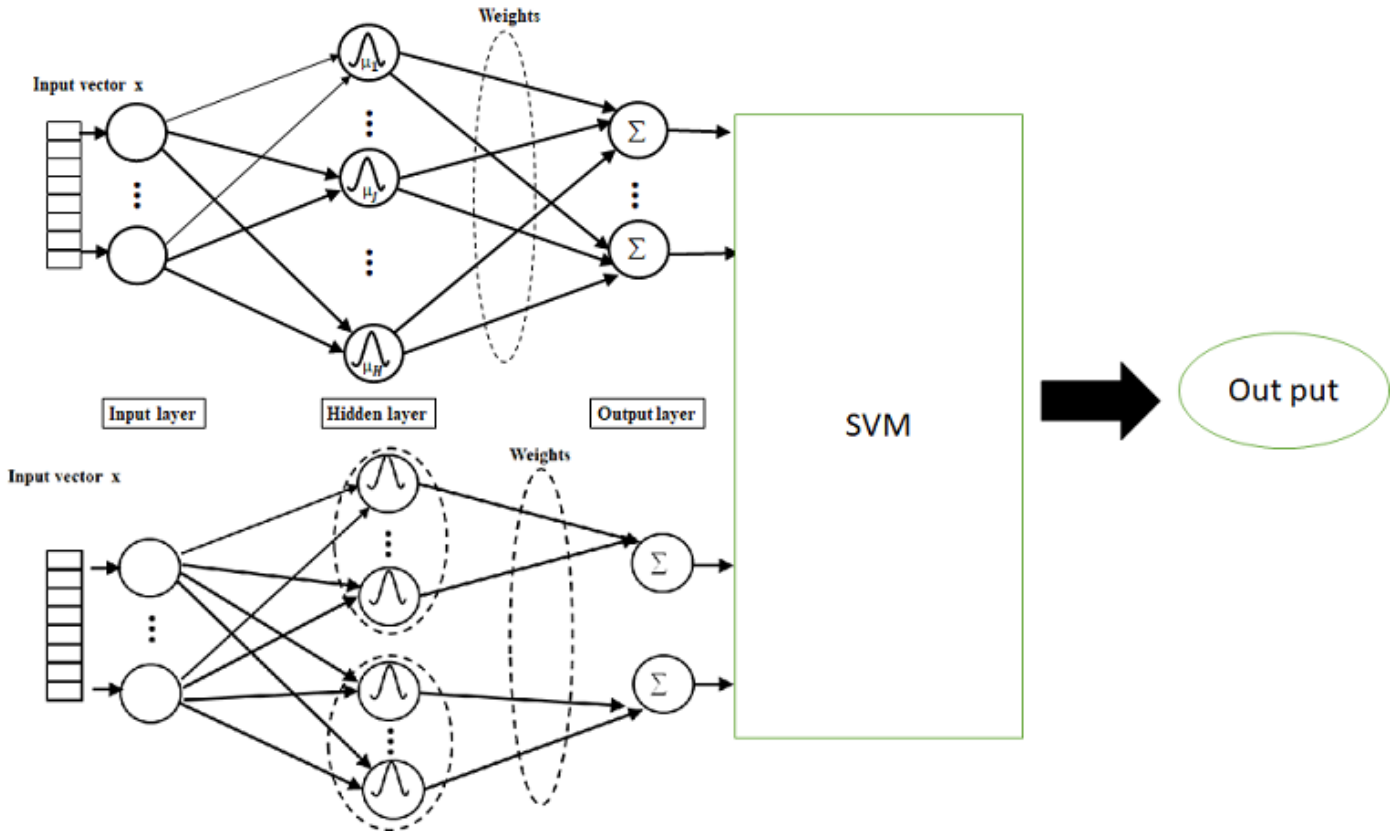

Fig. 1. The proposed cascaded framework.

The architecture of the PNN is displayed as four layers and is shown in Figure 1. The figure displays a PNN that recognizes two classes and extended to multi-class problems [6].

- **Input layer** : The input neurons supply the same input values to the hidden layer neurons. The size of this layer is determined by the dataset dimensionality (D).
- **Hidden layer**: There is one neuron per training pattern. The response of each hidden layer neuron is computed using the equation below.
- **Output layer**: This layer has one neuron for each class. Each output neuron receives the output from the hidden layer neurons associated with a given class, and the summation is carried out as follows:
- **Decision layer**: The size of this layer is one. This layer determines the class label of the given input vector (X) present at the input layer using Eq. (3).

B. Simple Radial Basis Neural Networks (RBFN)

Proposed by Broomhead and Lowe in 1988 [3] the RBFN has its foundation in the conventional approximation techniques. The RBFN is much more faster and efficient than MLP models but has equivalent capabilities. RBFN performs classification by measuring the inputs similarity to examples from the training set. Each RBFN neuron stores a prototype, which is just one of the examples from the training set. When we want to classify a new input, each neuron computes the Euclidean distance between the input and its prototype. Roughly speaking, if the input more closely resembles the class A prototypes than the class B prototypes, it is classified as class A.
The RBFN consists of three layers:

- **Input layer**: It functions similar to the PNN.
- **Hidden layer**: It also functions similar to the PNN. The output value of each hidden layer neuron is computed using Eq.
- **Output layer**: The output layer is made up of two neurons, where two is the number of distinct classes. The response of the output layer neuron is a weighted sum of the hidden layer outputs, which is computed using Eq. (4)

III. Proposed Methodology

In this paper we are proposing two stage cascaded ensemble framework. In cascaded network outputs of first stage in

TABLE I. PIMA INDIANS DIABETES DATA SET ATTRIBUTES DESCRIPTION.

Feature	Description	Feature	Description
1	Number of times pregnant	5	Serum insulin
2	Plasma glucose concentration	6	Body mass index
3	Diastolic blood pressure	7	Diabetes pedigree function
4	Triceps skin fold thickness	8	Age
Class 0 or 1 : 0 - Diabetes negative; 1-Diabetes positive			

TABLE II. COMPARATIVE ANALYSIS

Model	Accuracy(%)	Specificity(%)	Sensitivity(%)
RBFNN	65.22	0	100
ORBFNN	74.78	70.00	77.33
PNN	68.26	57.50	74.00
OPNN	63.04	70.00	59.33
Our Model	85.65	90.00	68.89

TABLE III. COMPARISON OF PROPOSED METHOD WITH OTHER ENSEMBLE APPROACHES.

Classifiers	PID dataset			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Reference
Casc	76.92 $\pm$0.6	–	–	C. Kaynak et al. (2000) [9]
RF	71.74	53.81	80.40	J.P. Kandhasamy et al. (2015) [10]
AdaBoost	76.43	52.99	89.00	S. Bashir et al. (2016) [11]
Bagging	77.99	75.96	85.00	S. Bashir et al. (2016) [11]
Majority voting	76.30	50.00	**90.40**	S. Bashir et al. (2016) [11]
Accuracy-Weighting	77.00	65.54	85.55	S. Bashir et al. (2016) [11]
HMV	77.08	78.93	88.40	S. Bashir et al. (2016) [12]
HM-BagMoov	78.21	78.65	92.60	S. Bashir et al. (2016) [12]
Proposed model	**85.65**	**90.00**	68.89	**This study**

feed-ed as inputs for next stage. In the proposed approach we have used Simple Probabilistic Neural Network (PNN) and Simple Radial Basis Neural Network (RBFN) in the first stage as they require single shot training. Next, outputs of these two networks as supplied to another classifier called support vector machine which is in the second stage of framework. The proposed framework is shown in below figure.

IV. EXPERIMENTS RESULTS AND DISCUSSION

A. Experimental Setup

Pima Inidans Diabetes Dataset obtained from the UCI repository consists of 768 records of which 500 records correspond to class 0(non-Diabetic Patient) and the remaining 268 records correspond to Class 1 (Diabetic Patient) [13]. The details of dataset is furnised in Table I.

For experimental purpose we have divided the dataset into training and testing datasets. Our training dataset consists of 500 records among which 350 records are Class 0 and remaining 150 records are Class1 Our testing dataset consists of 268 records(150 class 0 and remainder class 1). The performance of proposed model is evaluated on test dataset. And the proposed model is compared with the individual classifiers these resuts are furnished in Table II. It is clear from table results that the proposed network outperformed the individual classifiers in terms of accuracy, sensitivity, and specificity. Finally, the proposed model is compared with stat-of-the-art models in the literature. These results are shown in Table III. It is clear from the results that the proposed network outperformed the other state-of-the-art methods in terms of accuracy and sensitivity.

Overall, the proposed cascaded framework achieved best accuracy among all on diabetes dataset.

V. CONCLUSION

In this paper, to improve the Diabetes prediction accuracy a dual stage cascaded framework is designed. The proposed framework is constituted by combining simple RBFNN and PNN in first stage and keeping SVM in second stage. Hence, it possess advantages of both RNFNN and PNN. in the proposed cascaded framework outputs of first stage classifiers are supplied as inputs to second stage classifier.

Next, the proposed cascaded framework is experimented on Pima Indians Diabetes dataset. The experimental results proved that the proposed framework achieved more accuracy than individual accuracies of RBFNN, PNN and SVM, and state-of-the-art models.

ACKNOWLEDGMENT

The authors would like to thank Director of Mahindra École Centrale, Hyderabad, Telangana, India and Dean Research of Mahindra Ecole Centrale for supporting this work.

REFERENCES

[1] Alberti, Kurt George Matthew Mayer, and PZ ft Zimmet. "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation." Diabetic medicine 15.7 (1998): 539-553.

[2] World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. No. WHO/NCD/NCS/99.2. Geneva: World health organization, 1999.

[3] Clarke, P., et al. "The impact of diabetesrelated complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65)." Diabetic Medicine 20.6 (2003): 442-450.

[4] Wild, Sarah, et al. "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030." Diabetes care 27.5 (2004): 1047-1053.

[5] Chandra, Bala, and KV Naresh Babu. "An improved architecture for probabilistic neural networks." Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE, 2011.

[6] Cheruku, Ramalingaswamy, et al. "PSO-RBFNN: a PSO-based clustering approach for RBFNN design to classify disease data." International Conference on Artificial Neural Networks. Springer, Cham, 2017.

[7] Cheruku, Ramalingaswamy, Damodar Reddy Edla, and Venkatanareshbabu Kuppili. "Diabetes classification using radial basis function network by combining cluster validity index and bat optimization with novel fitness function." Int. J. Comput. Intell. Syst 10.1 (2017): 247-265

[8] Yegnanarayana, B. Artificial neural networks. PHI Learning Pvt. Ltd., 2009.

[9] Kaynak, Cenk, and Ethem Alpaydin. "Multistage cascading of multiple classifiers: One man's noise is another man's data." ICML. 2000.

[10] Kandhasamy, J. Pradeep, and S. Balamurali. "Performance analysis of classifier models to predict diabetes mellitus." Procedia Computer Science 47 (2015): 45-51.

[11] Bashir, Saba, Usman Qamar, and Farhan Hassan Khan. "IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework." Journal of biomedical informatics 59 (2016): 185-200.

[12] Bashir, Saba, et al. "HMV: a medical decision support framework using multi-layer classifiers for disease prediction." Journal of Computational Science 13 (2016): 10-25.

[13] Asuncion, Arthur, and David Newman. "UCI machine learning repository." (2007).

[14] PNNs provide robust performance on noisy data and easily incorporate additional samples.