

An Optimized Recursive General Regression Neural Network Oracle for the Prediction and Diagnosis of Diabetes

Dana Bani-Hani¹, Pruthak Patel² and Tasneem Alshaikh³

¹ State University of New York at Binghamton

Received: 11 December 2018 Accepted: 2 January 2019 Published: 15 January 2019

Abstract

Diabetes is a serious, chronic disease that has been seeing a rise in the number of cases and prevalence over the past few decades. It can lead to serious complications and can increase the overall risk of dying prematurely. Data-oriented prediction models have become effective tools that help medical decision-making and diagnoses in which the use of machine learning in medicine has increased substantially. This research introduces the Recursive General Regression Neural Network Oracle (R-GRNN Oracle) and is applied on the Pima Indians Diabetes dataset for the prediction and diagnosis of diabetes. The R-GRNN Oracle (Bani-Hani, 2017) is an enhancement to the GRNN Oracle developed by Masters et al. in 1998, in which the recursive model is created of two oracles: one within the other. Several classifiers, along with the R-GRNN Oracle and the GRNN Oracle, are applied to the dataset, they are: Support Vector Machine (SVM), Multilayer Perceptron (MLP), Probabilistic Neural Network (PNN), Gaussian Naïve Bayes (GNB), K-Nearest Neighbor (KNN), and Random Forest (RF). Genetic Algorithm (GA) was used for feature selection as well as the hyperparameter optimization of SVM and MLP, and Grid Search (GS) was used to optimize the hyperparameters of KNN and RF. The performance metrics accuracy, AUC, sensitivity, and specificity were recorded for each classifier. The R-GRNN Oracle was able to achieve the highest accuracy, AUC, and sensitivity (81.14

Index terms— GRNN oracle, data mining, machine learning, genetic algorithm, diabetes, prediction model.

1 Introduction

According to the World Health Organization (WHO), the number of people with diabetes had quadrupled since 1980. Prevalence is increasing worldwide, particularly in low-and middle-income countries. It is estimated that medical costs and lost work and wages for people diagnosed with diabetes is \$327 billion yearly and twice as much as those who do not have diabetes (CDC, 2018). About 422 million people worldwide have the disease. It can lead to serious complications in any part of the body such as kidney disease, blindness, nerve damage, and heart disease (Temurtas et al., 2009), and increases the risk of dying prematurely -diabetes is the seventh leading cause of death worldwide.

There are many factors to analyze to diagnose diabetes in a patient which makes the physician's job difficult. Thus, to save time, cost, and the risk of an inexperienced physician, classification models may be built to help predict and diagnose diabetes based on previous records (Polat et al., 2008). The use of machine learning in medicine has increased substantially. With the exponential growth of big data, manual efforts to analyze such data are impossible, therefore, automated techniques such as machine learning are used. Machine learning is defined as having the ability for a system to learn on its own, by extracting patterns from large raw data (Goodfellow et al., 2016).

The General Regression Neural Network Oracle (GRNN Oracle), developed by Elman et al. in 1998, combines the predictions of individually trained classifiers and outputs one superior prediction by determining the error rate for each classifier from a set of observations in order to assign weights to favor classifiers with lower error rates. The final prediction for an unknown observation is calculated by summing each classifier’s prediction for that unknown observation multiplied by the classifier’s weight; the classifiers with lower error rates have greater influence on the final prediction.

Because of the strong capabilities of the oracle, it has been enhanced to consist of two GRNN Oracles; one within the other. First proposed by Bani-Hani (2017), the first oracle is created through its own combination of algorithms and acts now as a classifier as it has its own predictions and error contribution to a set of unknown observations. It is then combined with other classifiers to create a new, outer oracle that has been named the Recursive General Regression Neural Network Oracle (R-GRNN Oracle). This study is applied on the Pima Indians Diabetes dataset where Genetic Algorithm (GA) is used for feature selection and hyperparameter optimization, and the proposed classifier, the Recursive General Regression Neural Network Oracle (R-GRNN Oracle), is applied along with seven other classifiers, namely Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF), Probabilistic Neural Network (PNN), Gaussian Naïve Bayes (GNB), K-Nearest Neighbor (KNN), and the GRNN Oracle, for the prediction and diagnosis of diabetes. The R-GRNN Oracle was able to achieve the highest accuracy and AUC (area under the Receiver Operating Characteristic (ROC) curve) performance metrics in comparison to the other classifiers used.

The remainder of this paper is organized as follows: Section 2 presents the related work regarding this study. Section 3 explains the methodology adopted in this study. Section 4 shows the experimental analysis and results. Section 5 presents the discussion. And Section 6 presents the conclusion and future work.

2 II.

3 Related Work

Prediction models are vastly implemented in clinical and medical fields to support diagnostic decision-making (Zheng et al., 2015). Many other studies have been carried out on the same dataset, however, due to reporting training accuracies rather than testing and validation accuracies, they have been excluded from the literature review for several reasons including, and most importantly, overfitting, as overfitting generates higher accuracies due to fitting the model too perfectly to the training set making the model not generalized enough. The other studies that have been excluded are those that obtained high accuracies but did not mention whether they obtained it from a training set or a testing or validation set making the results questionable. It is worthy to note that this study applied 4-fold cross validation to train each classifier and were tested on a validation subset that did not take part in neither the training nor testing steps.

4 III.

5 Methodology

Six individual classifiers were used in this research: SVM, MLP, RF, PNN, GNB, and KNN, in which some were used to create the GRNN Oracle, and some were combined with the first oracle to create the R-GRNN Oracle. The software and language used for this study was Python 3.6 and the hardware specifications were Intel® Core™ i7-8750H CPU @ 2.20GHz with 32.0 GB RAM. a) Individual Classifiers Support Vector Machine: SVM is a statistical learning method proposed by Vapnik (1995). It is a widely used supervised machine learning algorithm used for both classification and regression. SVM works by finding the hyperplane that maximizes the margin between the classes in the feature space, as seen in Figure 1. Support vectors are observations that help dictate the hyperplane. It classifies new samples based on which side of the boundary they are located on. Multilayer Perceptron: MLP is a feed forward artificial NN that is a modification of the standard linear perceptron. It is an algorithm that does not require a linear relationship between the independent variables and the dependent variable as it is able to solve problems that are not linearly separable through the use of activation functions located in each node. An MLP consists of an input layer, a hidden layer(s), and an output layer. It is a supervised machine learning algorithm that exploits back propagation to train itself to optimize the weights of each edge connecting two nodes. It is the most frequently used NN (Hossain et al., 2017) and is widely used for classification, regression, prediction, and approximation tasks. Figure 2 illustrates an example of an MLP with one hidden layer with five hidden nodes. Gaussian Naïve Bayes: GNB is a supervised learning algorithm that is widely used for classification problems because of its simplicity and accurate results (Farid et al., 2014). It uses Bayes theorem as its framework (Griffis et al., 2016) and has strong independence assumptions between the independent variables. One important advantage of GNB is that it could estimate the parameters necessary for classification by training on a small training set.

K-Nearest Neighbor: KNN is a non-parametric, lazy learning method for classification and regression tasks (Zhang, 2016). k is a user-set parameter that represents the number of known observations closest to the unknown observation mapped out in the feature space. For classification tasks, the class of the new observation is based on the majority class surrounding it; k is typically an odd number. For regression tasks, the new observation is taken as the average of its k neighbors.

diabetes, which indicates the imbalanced property of the dataset. In this dataset, there are eight independent variables (features) and one dependent variable (outcome: diabetes or no diabetes), as presented in Table 1. More detailed attributes distributions and statistical analysis are further shown in Figure 3, where the color orange signifies patients who have diabetes. All patients recorded are females at least 21 years old of Pima Indian heritage.

12 b) Data Preprocessing

The first step taken in the data preprocessing phase was excluding outliers as they can drastically affect the model's predictive ability. Any point that was three standard deviations (3 σ) away from the mean of any given feature was excluded. The original dataset had 768 patients, and after outlier removal, the new dataset contained 709 patients where 243 of them had diabetes. The next step was to correct the imbalanced property of the data. Since only 243 patients from the remaining 709 had diabetes, this is a class imbalance problem where those with diabetes only make up 34% of the data. Thus, an oversampling approach was applied to the minority class. Oversampling was favored over under sampling because the dataset's size concerning the number of observations was already small, and concerning how the R-GRNN Oracle works, it would not be a wise approach to remove observations, as the recursive oracle requires the dataset to be relatively large for the data subsets to be drawn. After this step, a normalization technique was applied to each independent variable in which the variable was scaled to a range between 0 and 1; 0 indicating the lowest value in a particular feature and 1 indicating the highest. The formula of normalization is given in Formula 8 where \min is the minimum value in the set of values in feature x , and \max is the maximum value in feature x . This is performed to ensure each feature has an equal weight so that no one feature would outweigh another before the creation of the prediction model.
$$x_{\text{normalized}} = \frac{x - \min}{\max - \min} \quad (8)$$

13 c) Hyperparameter Optimization

The hyperparameters in any algorithm contributes greatly to the output of the model, therefore, determining the optimal (or near-optimal) combination of hyperparameters would yield the best result. For example, some of a NN's hyperparameters include the number of hidden layers a user sets and the number of hidden nodes in each hidden layer. Hyperparameters are defined as the properties of a model that the user can set the value to. They are different from parameters as parameters are changed internally by the model itself during training rather than set by the user before the training process. An example to a parameter is the weights of a NN, as they are adjusted through back propagation using Gradient Decent (or any other optimizer) rather than by the user.

GA was utilized to optimize the performances of the SVM and MLP, while GS was applied on KNN and RF. The reason that GS was used instead of GA was that both KNN and RF have one parameter of interest: the number of neighbors and the number of DTs, receptively.

Therefore, no combinations of hyperparameters are needed which makes it a straightforward exhaustive search. SVM and MLP however have more than one hyperparameter that need to be optimized simultaneously, which also include continuous values, this is why GA is used.

Formula 9 shows the fitness function (???) used to evaluate each chromosome (each solution). They were evaluated based on their prediction accuracies, where TP is the true positive rate, in which it indicates those who actually have diabetes and were predicted to have diabetes, TN is the true negative rate where those who do not have diabetes were predicted not having diabetes, FP is the false positive rate in which those without diabetes were falsely predicted that they do have diabetes, FN is the false negative rate, where patients have diabetes but were falsely predicted that they don't, and F is the number of folds required for the K-fold cross validation, in which it was set to four for this study.
$$Fitness = \frac{1}{F} \left(\frac{TP}{TP + FP} + \frac{TN}{TN + FN} \right) \quad (9)$$

The hyperparameters that were included in this study relating to SVM were C and gamma (γ), where both take on continuous values, while MLP's hyperparameters included the learning rate (η), momentum, the number of hidden layers, the number of hidden nodes in each hidden layer, and the solver, where η and momentum are continuous, the number of hidden layers and nodes are integers, and the solver is categorical. Figure ?? and Figure ?? show the encoding (genotype) for the SVM and MLP parameters, respectively, where each continuous hyperparameter was encoded with a binary chromosome with a length of 15 alleles. SVMs can handle nonlinear classifications through transforming inputs into feature vectors with the use of kernels. The SVM kernel set for this study is the Radial Basis Function (RBF) in which it is a popular Gaussian kernel function. Some of RBF's greatest advantages are its high accuracy, its fast convergence, and its applicability in almost any dimension. γ is a regularization hyperparameter that determines how correctly the hyperplane between the classes separates the data. It controls the trade-off between model complexity and training error (Joachims, 2002). γ has a serious impact on the classification accuracy as it defines the influence of each training observation (Tuba and Stanimirovic, 2017), with lower values meaning "far" and higher values meaning "close". It can be thought of as the inverse of the radius of influence of observations selected by the model as support vectors. Figure 6-A shows SVM's accuracy achieved by GA in each of the 100 generations run.

With MLP, the activation function set in this study was the Rectified Linear Unit (ReLU). Activation functions are operations which map an output to a set of inputs. They are used to impart non-linearity to the network

structure (Acharya et al, 2017). Because ReLU returns a positive number, i.e. $\max(0, x)$, the two major advantages of it are sparsity and the reduced likelihood of the "vanishing gradient" problem, as adding as many hidden layers as one would like would not cause the gradient multiplication to reach a very small number that it will likely "vanish" with more layers to add. Solvers in NNs train and optimize the weights connecting the nodes between two-adjacent layers. The two solvers considered for this study are Stochastic Gradient Decent (SGD) and Adam, a variant of SGD. The other two important hyperparameters are η and momentum. η controls how fast the network learns during training and momentum helps to converge the data (Acharya et al, 2017). They can be thought of the stepping size and direction in the search space. Feature selection plays an important role in classification for several reasons (Luukka, 2011). First, it can simplify the model's complexity which helps reduce computational cost, and when the model is taken for practical use fewer inputs are needed. Second, by removing redundant features from the dataset one can also make the model more transparent and more comprehensible, providing better explanation of suggested diagnosis, which is an important requirement in medical applications. Feature selection process can also reduce noise in which it may enhance classification accuracy.

GA was applied for feature selection through SVM and its optimized hyperparameters from the previous step. The solution representation for feature selection was embodied by a chromosome of eight binary values (i.e. 0's and 1's). An allele of the value 0 indicates that feature x_i was not included while an allele of 1 indicated that it was included; x_i is the i th feature in the dataset. To explain further, Figure 7 illustrates an example where the encoding of the selected features #1, #2, #3, and #6, out of eight features is shown. Chromosomes with a subset of features selected are then evaluated based on their accuracy. The subset of features that attained the highest accuracy was selected for further analysis. Formula 9 was also used as the fitness function for chromosome evaluation. For the first GRNN Oracle (the inner oracle), the classifiers fed into it were SVM, GNB, and RF. The accuracy and AUC for SVM were 79.72% and 85.79%, respectively. GNB had 79.09% and 85.56%, respectively, and RF at 77.50% and 81.15%. The performance of the first oracle had an accuracy of 79.54%, AUC of 85.16%, sensitivity of 59.60%, and specificity of 88.51%. MLP, PNN, and KNN were not chosen because of their inferior performances when compared to the other models. All models were run 15 times and the average of the performance metrics were taken.

For the R-GRNN Oracle, the first GRNN Oracle, which now acts as a classifier with its own predictions, was combined with SVM. Since SVM had a better performance than others, it was chosen as a match with the first oracle to create the second oracle. Figure 8 illustrates the classifiers being fed into each one of the two oracles. The first oracle achieved an accuracy of 79.54%, however, it was surpassed by SVM (79.72%), but the recursive model had the highest accuracy at 81.14% and highest AUC at 86.03%. Although it was able to reach the highest sensitivity too (63.80%) in comparison to the rest, MLP had the highest sensitivity (89.71%), where the recursive model came in third with 89.14% after MLP and SVM. However, since detecting TPs is of great importance (those who have diabetes), the sensitivity metric, where the R-GRNN was the highest, has a higher significance than specificity. Table 2 shows the accuracy, AUC, sensitivity, and specificity of all the classifiers: six individual classifiers (performing on their own), the GRNN Oracle, and the R-GRNN Oracle. The performances can also be seen in Figure 9 and Figure 10. Figure 9 shows the recursive model's 15 runs where the best, average, and worst performance were recorded, 86.47%, 81.14%, and 76.15%, respectively. It is worthy to mention that the dataset was shuffled each time the classifiers were run to ensure the robustness of the model, as no matter how it the data was shuffled, it always yielded better performance than the rest of the classifiers. Shuffling the data is the reason behind the high variation seen in Figure 9. Also, as a reminder to what was mentioned earlier, 4-fold cross validation was applied to train and test the models, but the actual Year 2019 () validation of each model was applied on a validation subset that was not involved in neither the training nor testing steps of each model.

14 Discussion

While the accuracy of the proposed model was not the highest in the literature, it still came in third when compared to all the publications studied (Table 3). It also bested the traditional oracle, SVM, MLP, RF, PNN, GNB, and KNN. However, as a slight remark, the studies did not confirm whether their accuracies were from conducting several runs and taking the average or not. As in this study, the highest accuracy achieved by the recursive model was 86.47%; one could simply report it as the highest achieved, therefore, it is wise if several runs are conducted and the average was taken.

15 Conclusion and Future Work

This study presented the R-GRNN Oracle and was applied on the Pima Indians Diabetes dataset. It was applied along with seven other classifiers in which their final performances were compared. The other classifiers included are the traditional GRNN Oracle, SVM, MLP, RN, PNN, GNB, and KNN. GA was used to optimize the hyperparameters of SVM and MLP, and GS was used on RF and KNN. The models were run 15 times and the dataset was shuffled each run to ensure robustness. 4-fold cross validation was adopted as the validation method. Compared to the other models, the recursive oracle achieved the highest accuracy, AUC, and sensitivity at 81.14%, 86.03%, and 63.80%, respectively. It, however, came in third for specificity at 89.14% where optimized MLP had the highest at 89.71%.

Future research may include applying feature selection and hyperparameter optimization simultaneously rather than applying feature selection based on the optimized hyperparameters from all the features. It can also include using other metaheuristics, such as Particle Swarm Optimization (PSO) for hyperparameter optimization.

16 References Références Referencias

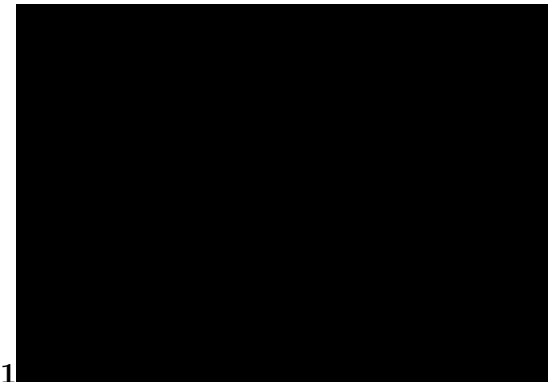


Figure 1: Figure 1 :

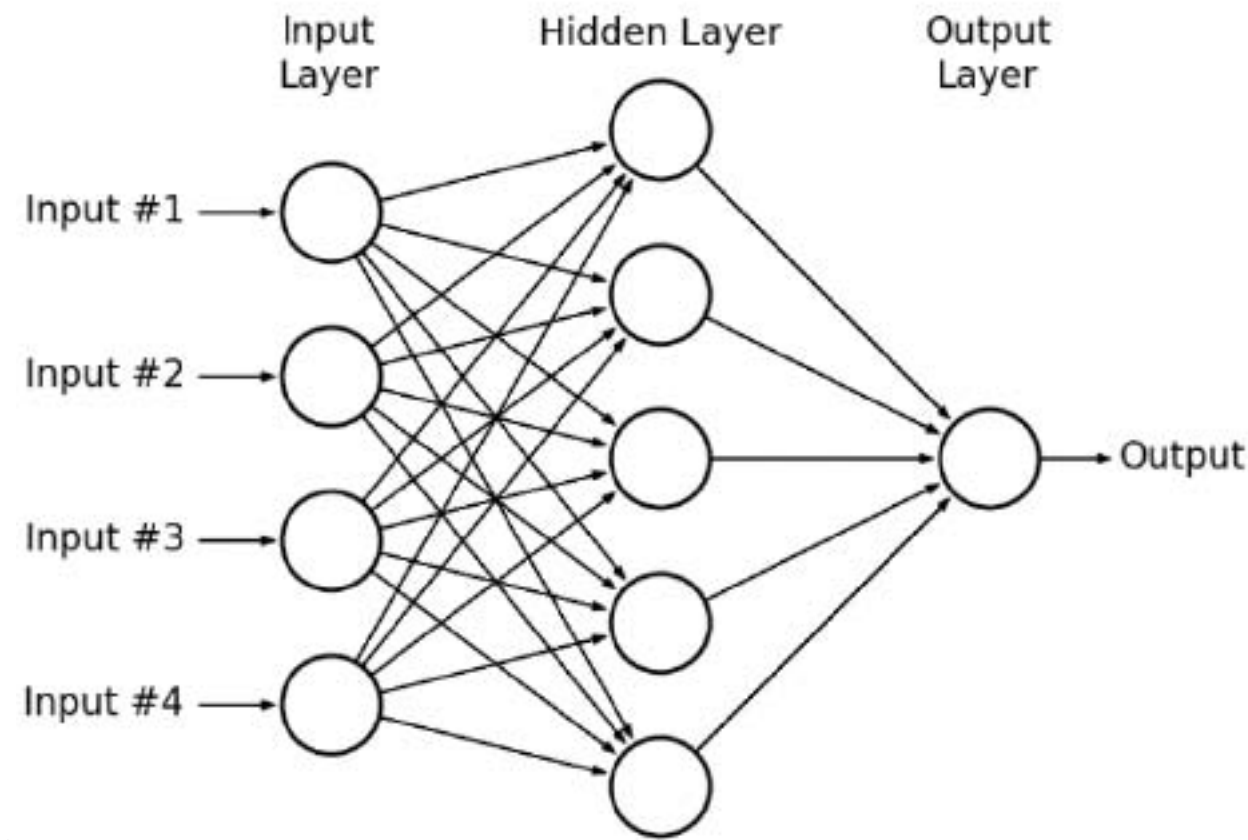


Figure 2: Figure 2 :

1 2

¹() D © 2019 Global Journals
²© 2019 Global Journals

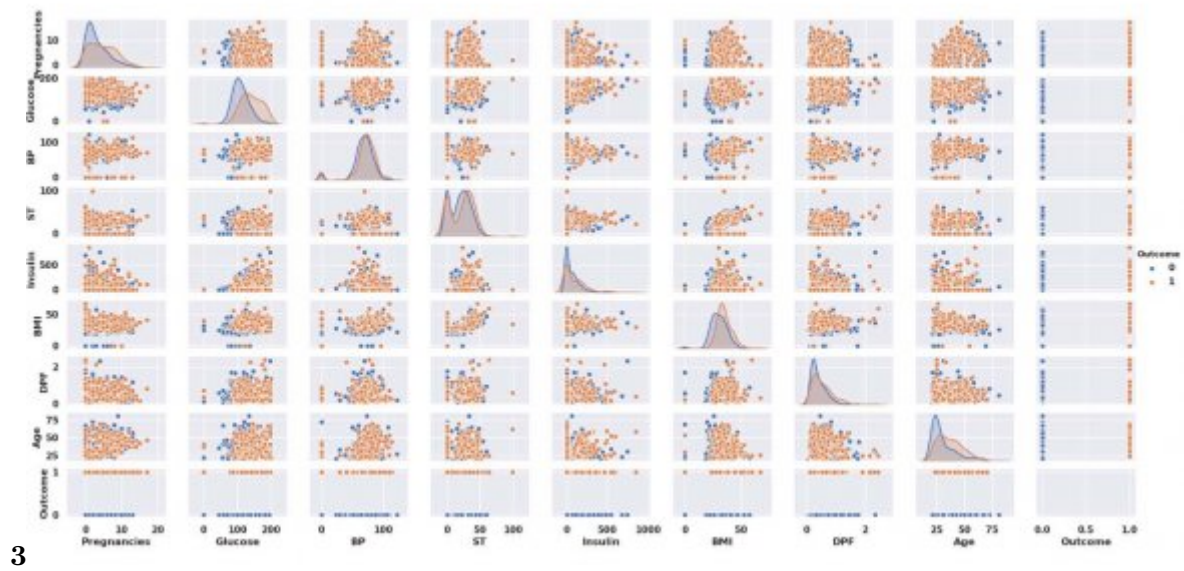


Figure 3: Figure 3 :

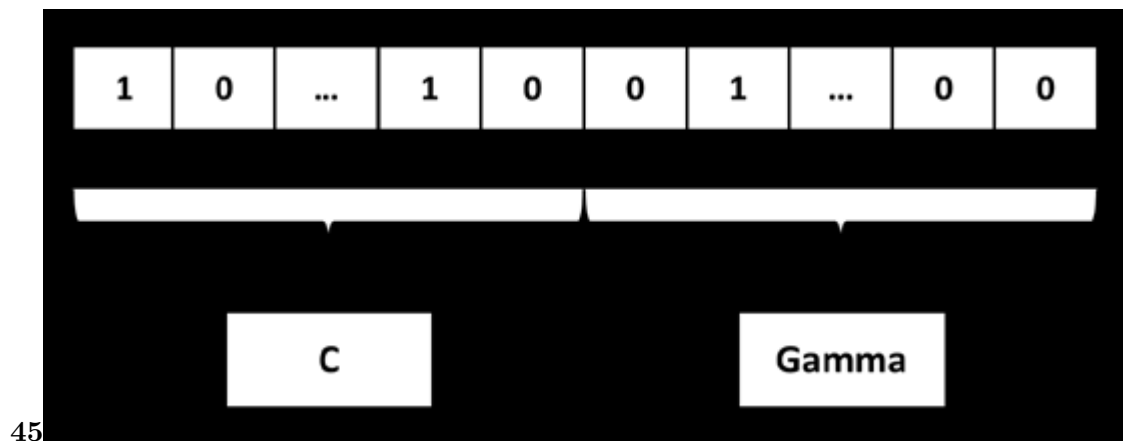


Figure 4: Figure 4 :Figure 5 :

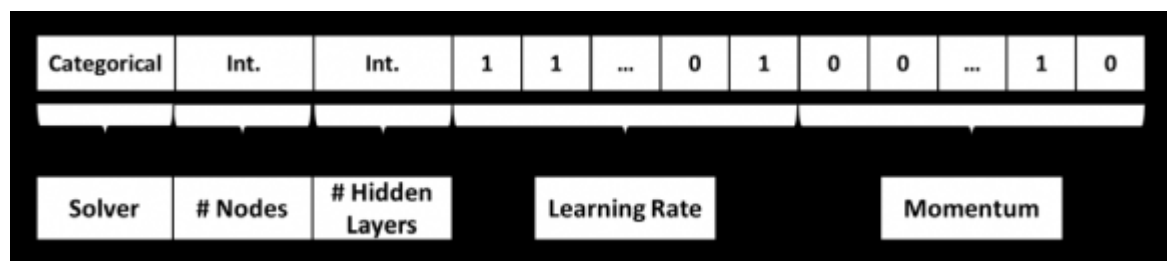


Figure 5:

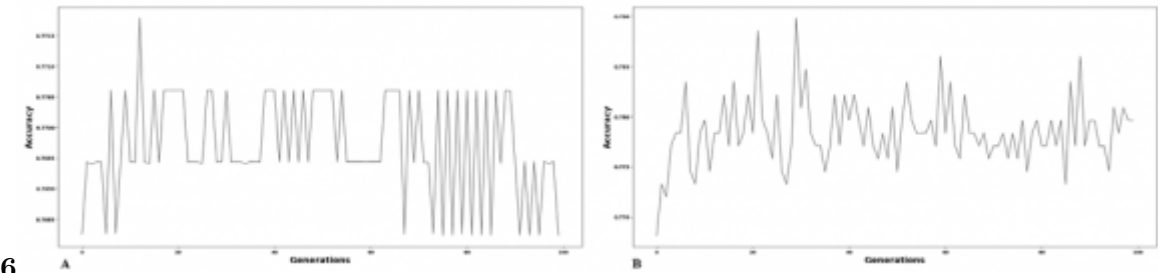


Figure 6: Figure 6 :

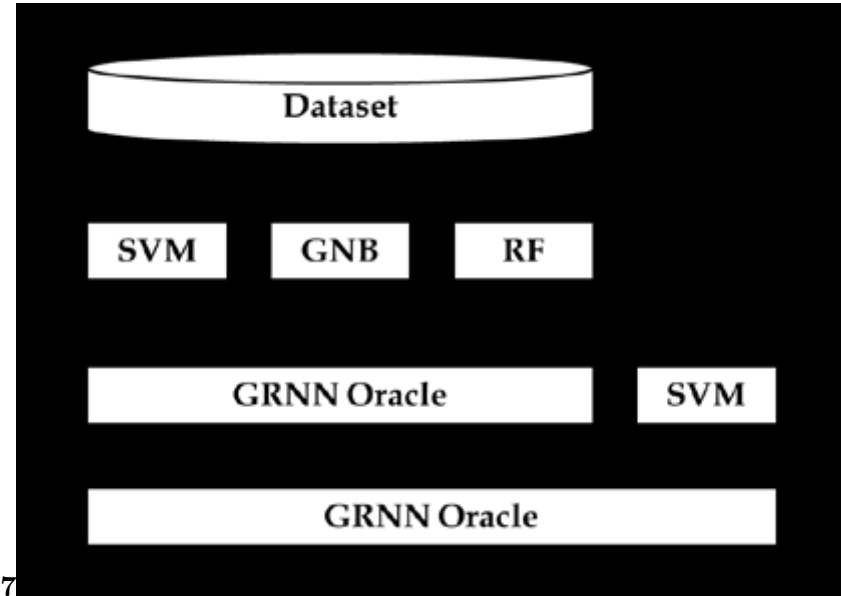


Figure 7: Figure 7 :

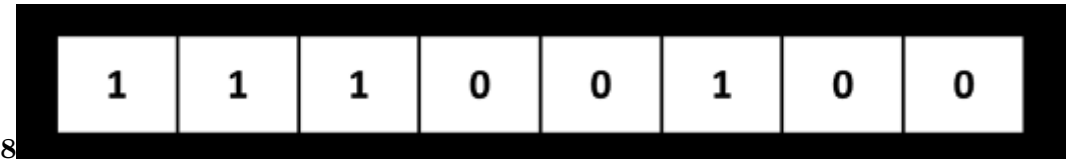


Figure 8: Figure 8 :

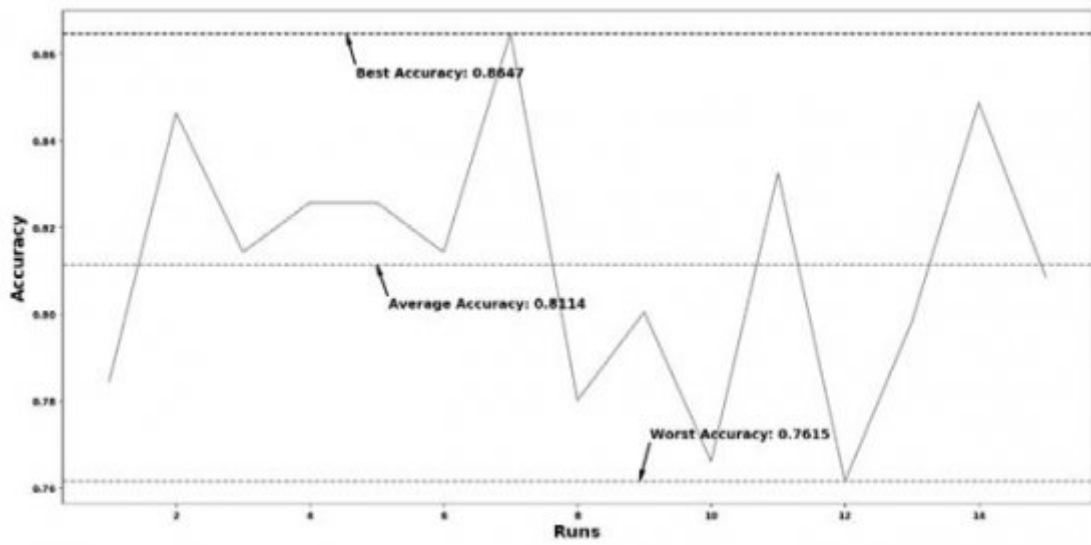


Figure 9: Figure 9 :

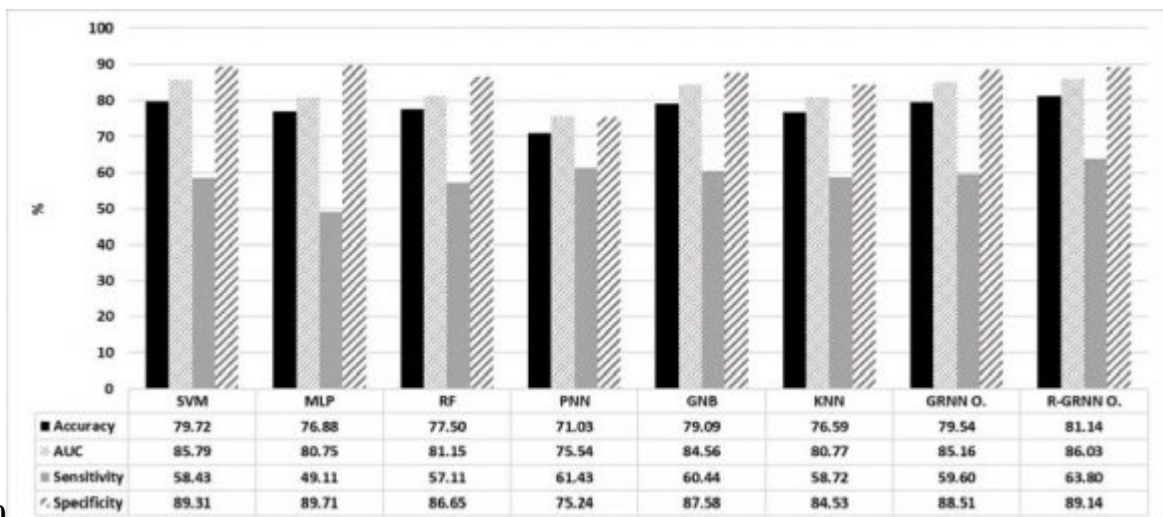


Figure 10: Figure 10 :

achieve the highest accuracy of 78.4% using a two-layer MLP. A hybrid of Artificial Neural Network (ANN) and Fuzzy Neural Network (FNN) was proposed by Kahramanli and Allahverdi in 2008. Their approach resulted in an accuracy of 84.2%. Lekkas and Mikhailove (2010) applied Evolving Fuzzy Classification (EFC) to two datasets including Pima Indians Diabetes dataset. They were able to reach an accuracy of 79.37%. Miche et al. (2010) presented the Optimally Pruned Extreme Learning Machine (OP-ELM) and compared its performance to a MLP, SVM, and Gaussian Process (GP) on several regression and classification datasets. Regarding the dataset concerning this study, the GP had the highest accuracy among the classifiers tested with an accuracy of 76.3%. Huang et al. (2004) was able to achieve an accuracy of 77.31% using SVM, although their paper proposed an algorithm called Extreme Machine Learning (EML). Kumari and Chitra (2013) used SVM and obtained an accuracy of 78.2%. Al Jarullah (2011) also found the accuracy to be 78.2% using Decision Trees (DTs). Bradley and Mangasarian (1998) applied Feature Selection via Concave (FSC), SVM, and Robust Linear Program (RLP) in which the RLP had the highest accuracy on the Pima Indian Diabetes dataset at 76.16%. Using a novel Adaptive Synthetic (ADASYN) sampling approach, He et al. (2008) achieved an accuracy of 68.37%. Zahan et al. (2005) proposed a new artificial immune system named Attribute Weighted Artificial Immune System (AWAIS) in which they attained a classification accuracy of 75.87%. Luukka (2011) used Similarity-Based (S-Based) classifier with fuzzy entropy measures

Square Support Vector Machine (LS-SVM) for the prediction of diabetes through Generalized Discriminant Analysis (GDA).

. It is a widely used optimization technique inspired by nature, more specifically, evolution and survival of the fittest. It finds solutions throughout the search space using two main operators: crossover and mutation. Every solution is represented as a chromosome with several alleles encoded with genetic material that measure the fitness value of the

[Note: objective function. Crossover produces two somewhat different chromosomes, called offspring, from two parents. The mutation operator is applied on the offspring at a given probability to create diversity in the Grid Search:]

Figure 12:

1

	Description	Type
X1 No. of Pregnancies		Discrete
X2 Plasma Glucose Concentration		Continuous
X3 Diastolic Blood Pressure		Continuous
X4 Skin Thickness		Continuous
X5 2-hr Serum Insulin		Continuous
X6 BMI		Continuous
X7 Diabetes Pedigree Function		Continuous
X8 Age		Continuous
Y	Outcome: Dia- betes/No Diabetes	Discrete

Figure 13: Table 1 :

2

	Accuracy	AUC	Sensitivity	Specificity
SVM	79.72	85.79	58.43	89.31
MLP	76.88	80.75	49.11	89.71
RF	77.50	81.15	57.11	86.65
PNN	71.03	75.54	61.43	75.24
GNB	79.09	84.56	60.44	87.58
KNN	76.59	80.77	58.72	84.53
GRNN O.	79.54	85.16	59.60	88.51
R. GRNN O.	81.14	86.03	63.80	89.14

Figure 14: Table 2 :

3

Method	Accuracy
--------	----------

Figure 15: Table 3 :

voice
Technology (Disease Neuroimaging
Classification of Alzheimer’s disease and prediction
of

data. In Advanced Communication
Initiative (2017).

mild cognitive impairment-
to-
Alzheimer’s

conversion from structural magnetic resource
imaging using feature ranking and a genetic
algorithm. Computers in biology and medicine, 83,
109-119.

7. Belgiu, M., & Dr?gu?, L. (2016). Random forest in
remote sensing: A review of applications and future
directions. ISPRS Journal of Photogrammetry and
Remote Sensing, 114, 24-31.

8. Bhardwaj, A., & Tiwari, A. (2015). Breast cancer
diagnosis using genetically optimized neural
network

Applications, 42(10), 4611-4620.

9.

model. Expert Systems with

Figure 16:

283 [Centers for Disease Control and Prevention. Diabetes quick facts] , <[https://www.cdc.gov/diabetes/](https://www.cdc.gov/diabetes/basics/quick-facts.html)
284 [basics/quick-facts.html](https://www.cdc.gov/diabetes/basics/quick-facts.html)>.Accessedon2018-09-08 *Centers for Disease Control and Prevention.*
285 *Diabetes quick facts*

286 [American Medical Informatics Association] , *American Medical Informatics Association*

287 [Park and Edington ()] , J Park , D W Edington . 2001.

288 [9th International Conference on] *9th International Conference on*, IEEE. p. .

289 [Polat et al. ()] *A cascade learning system for classification of diabetes disease: Generalized discriminant analysis*
290 *and least square support vector machine. Expert systems with applications*, K Polat , S Güne? , A Arslan .
291 2008. 34 p. .

292 [Hossain et al. ()] ‘A comparative study of vibrational responsebased impact force localization and quantification
293 using radial basis function network and multilayer perceptron’. M S Hossain , Z C Ong , Z Ismail , S Y Khoo
294 . *Expert Systems with Applications* 2017. 85 p. .

295 [Temurtas et al. ()] ‘A comparative study on diabetes disease diagnosis using neural networks’. H Temurtas , N
296 Yumusak , F Temurtas . *Expert Systems with applications* 2009. 36 (4) p. .

297 [Hsu et al. ()] *A practical guide to support vector classification*, C W Hsu , C C Chang , C J Lin . 2003.

298 [Holland and Henry ()] *Adaptation in natural and artificial systems: An introductory analysis with applications*
299 *to biology, control, and artificial intelligence*, John Holland , Henry . 1992. MIT press.

300 [He et al. (2008)] ‘ADASYN: Adaptive synthetic sampling approach for imbalanced learning’. H He , Y Bai , E
301 A Garcia , S Li . *IEEE International Joint Conference on*, 2008. June. 2008. IJCNN 2008. IEEE. p. . (Neural
302 Networks)

303 [Zhu et al. ()] ‘An improved early detection method of type-2 diabetes mellitus using multiple classifier system’.
304 J Zhu , Q Xie , K Zheng . *Information Sciences* 2015. 292 p. .

305 [Masters et al. (1998)] ‘An oracle based on the general regression neural network. In Systems, Man, and
306 Cybernetics’. T Masters , W H Land , S Maniccam . *IEEE International Conference on*, 1998. October.
307 1998. 1998. IEEE. 2 p. .

308 [Carpenter and Markuzon ()] ‘ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases’. G
309 A Carpenter , N Markuzon . *Neural Networks* 1998. 11 (2) p. .

310 [Mohamed et al. ()] ‘Assessment of artificial neural network for bathymetry estimation using High Resolution
311 Satellite imagery in Shallow Lakes: case study El Burullus Lake’. H Mohamed , A Negm , M Zahran , O C
312 Saavedra . *International water technology conference*, 2015. p. .

313 [Zheng et al. ()] ‘Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support
314 vector machine algorithms’. B Zheng , S W Yoon , S S Lam . *Expert Systems with Applications* 2014. 41 (4)
315 p. .

316 [Sun et al. ()] ‘Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer
317 in CT based on a multi-dimensional data set’. T Sun , J Wang , X Li , P Lv , F Liu , Y Luo , Y Gao , Q Zhu
318 , X Guo . *Computer methods and programs in biomedicine* 2013. 111 (2) p. .

319 [Jayasurya et al. ()] ‘Comparison of Bayesian network and support vector machine models for two-year survival
320 prediction in lung cancer patients treated with radiotherapy’. K Jayasurya , G Fung , S Yu , C Dehing-Oberije
321 , D De Ruysscher , A Hope , W De Neve , Y Lievens , P Lambin , A L A J Dekker . *Medical physics* 2010.
322 37 (4) p. .

323 [Zhang et al. ()] ‘Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis
324 detection: decision tree, k-nearest neighbors, and support vector machine’. Y Zhang , S Lu , X Zhou , M
325 Yang , L Wu , B Liu , P Phillips , S Wang . *Simulation* 2016. 92 (9) p. .

326 [Ramírez et al. ()] ‘Computer-aided diagnosis of Alzheimer’s type dementia combining support vector machines
327 and discriminant set of features’. J Ramírez , J M Górriz , D Salas-Gonzalez , A Romero , M López , I Álvarez
328 , M Gómez-Río . *Information Sciences* 2013. 237 p. .

329 [Christina and Santiago ()] *Decision Support System for a Chronic Disease-Diabetes*, S S Christina , N Santiago
330 . 2018.

331 [Gil and Manuel ()] ‘Diagnosing Parkinson by using artificial neural networks and support vector machines’. D
332 Gil , D J Manuel . *Global Journal of Computer Science and Technology* 2009. 9 (4) .

333 [Sakumura et al. ()] ‘Diagnosis by volatile organic compounds in exhaled breath from lung cancer patients using
334 support vector machine algorithm’. Y Sakumura , Y Koyama , H Tokutake , T Hida , K Sato , T Itoh , T
335 Akamatsu , W Shin . *Sensors* 2017. 17 (2) p. 287.

336 [Tuba and Stanimirovic (2017)] ‘Elephant herding optimization algorithm for support vector machine parameters
337 tuning’. E Tuba , Z Stanimirovic . *Electronics*, 2017. June. ECAI. p. 2017.

- [Huang et al. (2004)] ‘Extreme learning machine: a new learning scheme of feed forward neural networks’. G B Huang , Q Y Zhu , C K Siew . *IEEE International Joint Conference on*, 2004. July. 2004. IEEE. 2 p. . (Neural Networks)
- [Vaishali et al. (2017)] ‘Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset’. R Vaishali , R Sasikala , S Ramasubbareddy , S Remya , S Nalluri . *Computing Networking and Informatics (ICCNI*, 2017. October. IEEE. p. . (2017 International Conference on)
- [Goodfellow et al. ()] I Goodfellow , Y Bengio , A Courville , Y Bengio . *Deep learning*, (Cambridge) 2016. MIT press. 1.
- [Farid et al. ()] ‘Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks’. D M Farid , L Zhang , C M Rahman , M A Hossain , R Strachan . *Expert Systems with Applications* 2014. 41 (4) p. .
- [Thirugnanam et al. ()] *Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. Procedia engineering*, M Thirugnanam , P Kumar , S V Srivatsan , C R &nerlesh . 2012. 38 p. .
- [Haller et al. ()] ‘Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results’. S Haller , S Badoud , D Nguyen , V Garibotto , K O Lovblad , P R Burkhard . *American Journal of Neuroradiology* 2012.
- [Joachims et al. ()] ‘Learning to classify text using support vector machines: Methods, theory and 36’. T Joachims , T Manninen , H Huttunen , P Ruusuvuori , M Nykter . *PloS one* 2002. 2013. 8 (8) p. e72932. (Leukemia prediction using sparse logistic regression)
- [Fang and Grzymala-Busse (2006)] ‘Leukemia prediction from gene expression data-a rough set approach’. J Fang , J W Grzymala-Busse . *International conference on artificial intelligence and soft computing*, (Berlin, Heidelberg) 2006. June. Springer. p. .
- [Miche et al. ()] ‘OP-ELM: optimally pruned extreme learning machine’. Y Miche , A Sorjamaa , P Bas , O Simula , C Jutten , A Lendasse . *IEEE transactions on neural networks* 2010. 21 (1) p. .
- [Ramesh et al. ()] ‘Optimal predictive analytics of pima diabetics using deep learning’. S Ramesh , H Balaji , N C S Iyengar , R D &caytiles . *diabetes* 2017. (9) p. 10.
- [Zheng et al. ()] ‘Predictive modeling of hospital readmissions using metaheuristics and data mining’. B Zheng , J Zhang , S W Yoon , S S Lam , M Khasawneh , S Poranki . *Expert Systems with Applications* 2015. 42 (20) p. .
- [Breiman ()] ‘Random forests’. L Breiman . *Machine learning* 2001. 45 (1) p. .
- [?ahan et al. (2005)] ‘The medical applications of attribute weighted artificial immune system (AWAIS): diagnosis of heart and diabetes diseases’. S ?ahan , K Polat , H Kodaz , S Güne? . *International Conference on Artificial Immune Systems*, (Berlin, Heidelberg) 2005. August. Springer. p. .
- [Vapnik ()] *The nature of statistical learning theory*, V Vapnik . 2013. Springer science & business media.
- [Smith et al. (1988)] ‘Using the ADAP learning algorithm to forecast the onset of diabetes mellitus’. J W Smith , J E Everhart , W C Dickson , W C Knowler , R S Johannes . *Proceedings of the Annual Symposium on Computer Application in Medical Care*, (the Annual Symposium on Computer Application in Medical Care) 1988. November. p. 261.
- [Griffis et al. ()] ‘Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans’. J C Griffis , J B Allendorfer , J P Szaflarski . *Journal of neuroscience methods* 2016. 257 p. .
- [World Health Organization. 10 facts on diabetes] *World Health Organization. 10 facts on diabetes*, <<http://www.who.int/features/factfiles/diabetes/en/>>. Accessed on 2018-09-08