ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

A COMPREHENSIVE ANALYSIS OF CLASSIFICATION MODELS BASED ON MACHINE AND DEEP LEARNING

ALI BADIE, MOHAMMAD AMIN MORAGHEB, ALIREZA DEHGHAN, ALI NOSHAD

Ali Badie [1]
Department of Computer Engineering, Faculty of Engineering, Salman Farsi University of Kazerun, Kazerun, Iran, badie.itsu@yahoo.com
Mohammad Amin Moragheb [2]
Department of Computer Engineering, Noorabad Higher Education Institute, Mamasani, Iran, babakmoragheb@yahoo.com
Alireza Dehghan [3]
Department of Computer Engineering, Faculty of Engineering, Salman Farsi University of Kazerun, Kazerun, Iran, dehghan.itsu@yahoo.com
Ali Noshad [4]
Department of Computer Engineering, Faculty of Engineering, Salman Farsi University of Kazerun, Kazerun, Iran, noshad96.itsu@yahoo.com

Abstract
Having examined many scientific researches, classification can perceive as one of the prominent techniques which utilize widely in a variety of data mining modes. Acquiring insight such as efficiently of performance on large data sets, robustness in handling variety of data sets with corrupted data, and classification ability with aim of producing valid and new knowledge about the functionality of every classification model can exert very beneficial effects on tasks that are addressed with machine learning users. Inspired by these problems, in the current study we introduced a design of conditional Generative Adversarial Network–Deep Neural Network (cGAN-DNN) architecture for classification of records from two hospital that was collected from patients affected with breast cancer and diabetes. Our experimental results show that the model developed here, can reliably and with robust stability classify the subjects with an accuracy of 98.0% (diabetes) and 99.6% (breast cancer). In the terms of sensitivity, precision, and F1-score the obtained results are 98.4%, 98.5%, and 98.2% (diabetes), 99.7%, 99.4%, and 99.6% (breast cancer), respectively. Besides, the proposed model was validated in the comparison with machine learning approaches, which experimental results claim a high inconsistency among machine-learning based approaches. Regarding obtained results from the experiments and evaluation based on metrics, we can demonstrate that the proposed model can be tuned to utilize into various classification tasks.
Keywords: Deep neural networks, Classification, Deep learning, Generative adversarial net

1. Introduction

Recently, studies on text classification have gain more importance due to the availability of a significant number of electronic documents from a variety of sources. These resources include the web, government databases, news articles, biological databases, digital libraries, e-mails, and blog databases. Therefore, proper classification and discovery of knowledge from these sources are considered to be vital for study and research. Researchers are trying to use machine learning techniques in order to automatically classify and discover patterns in digital documents. Enable users to extract information and knowledge from textual sources is one of the prerequisite goals of classification [1-3]. The number of machine learning algorithms is increasing daily, and classification can be performed through different algorithms [4]. Classification algorithms use a variety of techniques to find relations among the features, and these relationships are summarized in a model that can be utilized for datasets whose classes are unknown [5]. Classification or prediction comes with a major functionality and it's the most widely used technique in a variety of data mining modes. Classification algorithms are based on supervised learning, in which they looking for hidden relationships among the target classes and independent variables [6]. Classification algorithms or supervise learning in general assign labels to observations so that unsupervised data can be classified based on the training data. The aim, model structure, performance, search, and data management methods are the main components of any algorithms [7]. Photo and pattern recognition, medical diagnosis, loan verification, error detection, financial trends are some of the most popular classification tasks [8]. Before using a model produced by a classification algorithm, the model should be evaluated based on certain criteria. This model is likely to lead to certain errors, so the data scientist should consider this possibility when selecting the specific model [9]. Having

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

examined many reports, the accuracy or percentage of items that are correctly classified by the model is the most common decision criterion for further evaluation of the models [7]. Admittedly, a variety of criteria can be utilized to make a comprehensive comparison and evaluation of the models. Berson et al. [10] defined these concepts of evaluation as accuracy, explanation, and integration capabilities. Maimon and Rokach [6] introduced these comparison criteria as generalization errors for the model: computational complexity or amount of CPU consumed by the inductor, comprehensibility or ability to run efficiently on larger databases, robustness or the ability to work on corrupted data, the stability or ability to produce reproducible results in different datasets, and finally the interesting nature of the ability to classify in the order of producing valid and new knowledge. Gaining new information about the functionality of every algorithm can exert very beneficial effects on tasks that are addressed with machine learning users [4, 11].

2. Literature review

The predictive power of machine learning classification algorithms has been perceived attractive for many years. Therefore, numerous studies have focused on presenting new classification models or comparing the existing models and the prominent factors influencing model performance. Hacker and Ahn [12] conducted a comparison experiment that focused on user selection. They compared many methods and proposed a new classification model called the relative SVM (Support Vector Machine) that outperformed the others. In another study, the classification algorithms such as AIRS2P, C4.5, CART, CSCA, ExCHAID, IBK, Logistics, Logitboost, MLP, MLVQ, Naïve Bayesian, QUEST, RSES, and SVM were implemented in the ten datasets. Factors affecting the performance and speed of classification algorithms based on experimental results of difference, correlation, and regression tests were emphasized [7]. Another study was conducted by Korde and Mahender, in which they introduced the classifications, text classification process, as well as a general review of classification techniques and comparison of Rocchio's Algorithm, KNN, Naïve Bayes, Decision tree, Decision Rule, SVM, Neural Network, LLSF, Voting, Associative classifier and Centroid based classifier, were performed based on a few criteria including time complexity and performance [13]. In another research conducted by Shah and Jivani was aimed at the diagnosis and prevention of breast cancer in patients. This study was performed on the Decision tree, Bayesian Network, and KNN algorithms based on accuracy, time-consuming, kappa statistics, relative absolute error, and relative square root error. In this comparison, the Naïve Bayes model was superior [4]. In another study conducted by Nabi and Ahmad [14], a comparison was made between the three classification algorithms KNN, Bayesian Network, and Decision tree. In this paper, the strength and accuracy of each algorithm for classification in terms of performance efficiency and time complexity were evaluated. In another study, Kim et al. [15] proposed two new methods: text normalization and feature weighting, which eliminates the flaws of the Bayesian algorithm in text classification tasks. They claim that their methods work well compared to many new methods such as SVM. Brazdil et al. [16] proposed a meta-learning method to assist in the algorithm selection process. They used the KNN algorithm to identify datasets that are very similar to each other, and the performance and speed of the selected algorithms in that dataset were used to collect a ranking for the user. In another research, Maindonald points to the problems and complexities of comparing algorithms and emphasizes the fact that users who have more experience with a particular model tend to produce the best results with that model. Thus, published functionality results are very broad indicators and depend on the dataset [17].

3. Methodology

This study aimed to provide results with reliability, repetition, and a strong evaluation in between machine learning and deep learning on text classification. The details of the steps of organizing this research are as follows: first, the selected datasets for the experiments are introduced, and then according to the conditions of testing and models used, pre-processing is performed on the data, and in the next step, the common algorithms used in the machine learning is considered and then the proposed model is introduced. Results obtained in the tests are displayed based on accuracy, sensitivity, specificity, precision, and F1-score, and the approaches are compared based on the results obtained in the tests. Besides, there is a look at comparing common approaches used in machine learning with each other, and at the end of the discussion, new ideas have been developed in the field of deep learning.

3.1. Data Description

Two datasets were used for this comparison. The first data set includes 5 features for diagnosing breast cancer in women based on the average radius, average tissue, mean environment, average area, and uniformity. This dataset contains 570 patient records of the University of Wisconsin Hospital. Finally, the last dataset is related to patients with diabetes, which examines 8 characteristics of pregnancies, glucose, blood pressure, skin thickness, insulin, BMI

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

(body mass index), diabetes pedigree function, and age to diagnose diabetes in patients. This collection contains 2000 records of patients from Frankfurt Hospital, Germany. Table 1 summarizes the attributes of each of the datasets.

Table 1: Dataset characteristics used for experiments

Dataset name	Number of variables	Target class types	Number of instances
Breast Cancer	5	2	570
Diabetes	8	2	2000

3.2. Data Pre-processing

After the collection of the datasets, pre-processing was performed on data. This process was divided into two stages: cleaning and review. Data cleaning can consider being identification, filing, and deletion of incomplete, inaccurate, or ambiguous data. Having examined the nature of the datasets, it can be seen all of the records are complete and integers and real values. Next, mean normalization was performed on the datasets. The far difference in between features quantity can lead to slow training and convergence, to address this problem, the datasets were normalized into a range of 0-1 using mean normalization. The normalization was done column-wise using the following equation (1):

$$Normalized\ Data = \frac{X - mean(x)}{stdev(x)} \qquad (1)$$

3.3. Machine Learning Algorithms

In this step, a brief overview of the causes and benefits of using machine learning algorithms is given: support vector machines are commonly used for classification, regression, or ranking performance. This algorithm is based on statistical learning theory and the law of minimizing structural risk is used to determine the optimal boundary setting for optimal class separation. This algorithm is one of the most widely used classification algorithms in bioinformatics due to its performance, efficiency, and capability in large data space and a large number of characteristics [18-21]. The logistics regression algorithm is widely used in machine learning for classification problems. In this algorithm, L1 regularization is used to prevent over-fitting, when the number of learning parameters is high. regularized logistic regression has shown good performance in classifying cases with many features [22,23]. Random forest classification algorithms are widely used in machine learning problems due to their accurate and reliable performance in diagnosis and classification [24]. Simplicity in comprehension and interpretation, even for non-expert users, is one of the reasons why this algorithm has been widely used in decision-making programs [25,26]. It is a case-based classifier that acts based on the distance or similarity of performance to the Euclidean distance or the measurement of the cosine similarity. This algorithm has been used in many applications due to its effectiveness, non-parametricity, and easy execution [27,28]. Today, Bayes classifiers are used in many applications due to their simplicity and high accuracy. Using this algorithm in a large data set gives us results with high speed and accuracy so that the presented results are competitive with the results obtained through more accurate methods [28-31].

3.4. The Proposed Model Architecture

The development of deep learning algorithms for complex tasks has relied on the availability of large labeled training datasets, usually containing thousands of examples. However, it is challenging to construct large labeled datasets specially in medical sciences [37]. In the current research, a hybrid approach was proposed for the purpose of constructing a robust classification model. It is common knowledge that the more data a machine learning algorithm has access to, the more effective it can be. Even when the data is of lower quality, algorithms can actually perform better, as long as useful data can be extracted by the model from the original dataset [38]. In that order, we construct an extension of the generative adversarial net to a conditional setting [40, 41] for tackling the shortage of the examples in the datasets.

The GAN framework establishes two distinct players, a generator and discriminator, and poses the two in an adversarial game. The discriminator is tasked with distinguishing between samples from the model and samples

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

from the training data; at the same time, the generator is tasked with maximally confusing the discriminator. We can state the objective here using a minimax value function [39]:

$$\min G \max D \left(\mathbb{E}|_{X \sim p_{data}(X)} [\log D(X)] + \mathbb{E}|_{Z \sim p_Z(z)} \left[\log \left(1 - D(G(z)) \right) \right] \right) \quad (2)$$

The terms of the equation in prose:

1. Train the discriminator to maximize the probability of the training data.
2. Train the discriminator to minimize the probability of the data sampled from the generator. At the same time, train the generator on the opposite objective.

In the current research, in order to add a conditioning ability to this framework, and generate data points for specific classes of data, an arbitrary condition **y** for generation can be established. This procedure restricts both the generator in its output and the discriminator in its expected input. For this purpose, the generator model can be defined as $G : (Z \times Y) \rightarrow X$, which takes noise data $z \in Z$ with an embedding $y \in Y$ as an input and it generates an example $x \in X$. Regarding the discriminator, the model can be described as $D : (X \times Y) \rightarrow [0,1]$, which takes an example x and a specific y condition (label) and outputs the probability based on the condition y which x came from the empirical data distribution rather than from the generator model. The objective function of two-player minimax game can be described as [41]:

$$\min G \max D \; V(D,G) = \left(\mathbb{E}|_{X \sim p_{data}(X)} [\log D(X|Y)] + \mathbb{E}|_{Z \sim p_Z(z)} \left[\log \left(1 - D(G(z|y)) \right) \right] \right) \quad (3)$$

We constructed a conditional adversarial net for the selected datasets conditioned on their class labels. In the proposed generator model, a noise prior z with dimensionality 100 was drawn from a uniform distribution. The class labels y were represented by dense vectors of size 10 in which each of the classes for the selected datasets (0 or 1) will map to a different 10-element vector representation. Then, both z and y are mapped to hidden layers with Rectified Linear Unit (ReLU) activation function. Layers with size of 30, 15 and 30 neurons, respectively. At the end, a sigmoid unit layer was considered as our output for generating the feature size dimensional examples.

Similarly, in the proposed discriminator model, the class labels y were represented by dense vectors of size of 10 in which each of classes will map to a different 10-element vector representation. Then, it will be mapped to a hidden layer with the size of the features. Both of the x and y will then concatenate as one single vector, and mapped to hidden layers with size of 20, 10 and 20, respectively. A sigmoid unit was utilized for obtaining the final probability of the model.

We have trained the model using Adam optimizer with mini-batches of size 128 on a total of 1200 epochs. Binary cross entropy loss was considered as loss function. The learning rate and momentum were used with initial value of 0.0002 and 0.5, respectively. Fig. 1 demonstrates the architecture of the proposed conditional generative adversarial net.

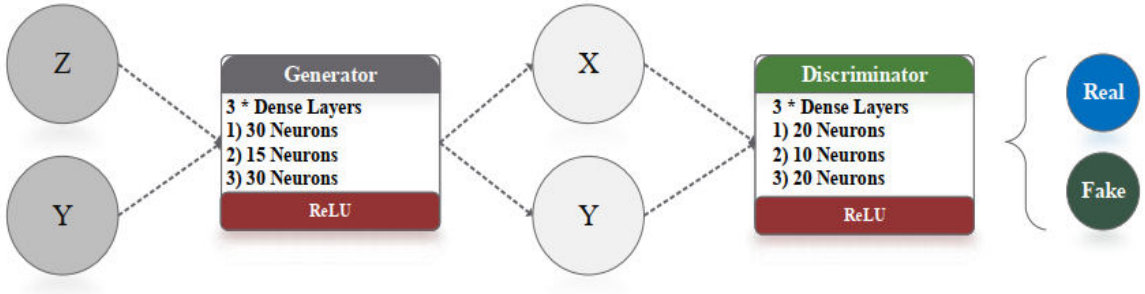

Fig. 1: The architecture of the proposed conditional generative adversarial net.

Recently, artificial neural networks or deep neural networks (DNNs) have been showing high performance [32, 33]. DNN classifiers have recently shown their superiority over other classical classifier approaches based on feature vector classification [34, 35]. In that order, in the current study for the classification task, a 7-layer DNN was developed. During the experiments, a variety of architectures were tested on the datasets, and improvements were taken place on each test in the order of producing a sensitive and robust architecture that can be utilized in a wide range of classification problems. The proposed classification model consists of 7 hidden layers in which each layer

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

encompasses a different quantity of neurons which were tested in a variety of ranges as the number of hidden layers and then bests were selected for the experiments, in this model, the neurons in these 7 layers numbered 400, 350, 300, 250, 300, 350, and 400 respectively. The number of inputs corresponds to the number of features which is varied in different datasets. The number of neurons in the output layer on the other hand was the same, as the result of 2 classes in each dataset. An illustration of the proposed model used in this research is shown in Fig. 2.

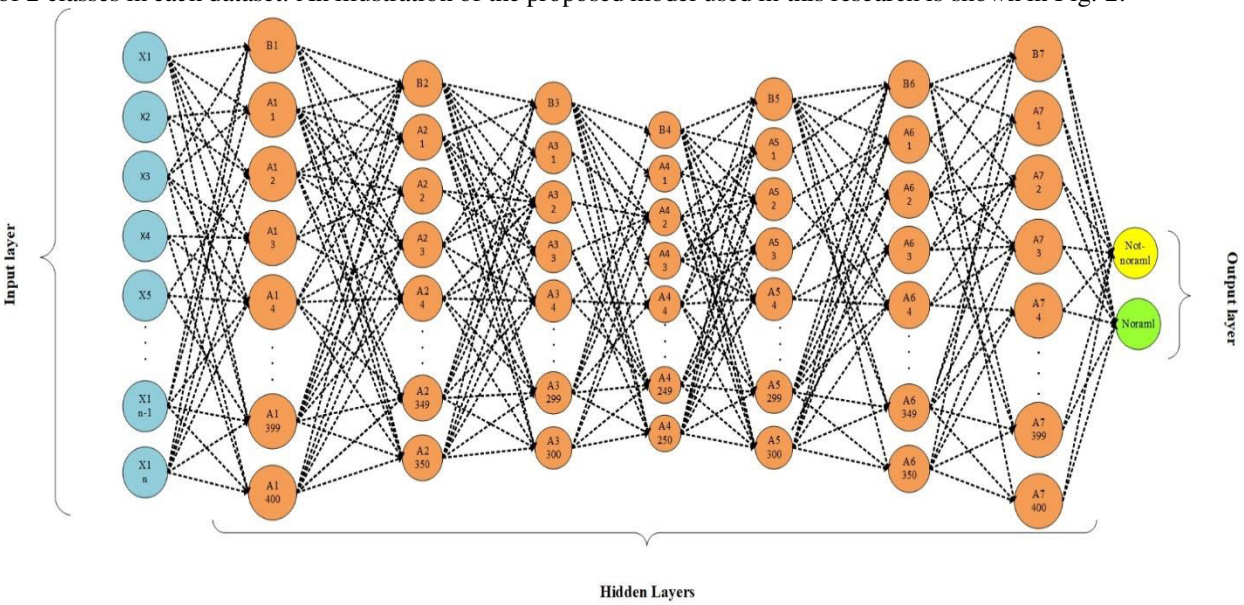

Fig. 2: The architecture of the proposed model

ReLU (Rectified Linear Unit) was used as an activation function for hidden layers and for the output, the sigmoid function was utilized. The datasets are relatively small and so is the test set, and as result, it is expected for the proposed model to stick to one of the local minimums, for tackling this problem $Adam$ optimizer was considered as a suitable approach. The selected optimizer method is deliberately searching for a global minimum to explore the most optimized approaches. For faster training and convergence, the weights of the networks and the bias values were randomly initialized and the learning rate was set as 0.001.

4. Performance measures

In this study, the performance classification of models in terms of precision, sensitivity (Recall), F1-score (F-measure), Loss, Accuracy, and AUC (Area Under Curve) were examined. These criteria are based on the number of true classified positive samples (TPs), the number of true classified negative samples (TNs), the number of false classified positive samples (FPs), and the number of false classified negative samples (FNs) are defined. The accuracy of the ratio of cases to the correct classification is estimated from the total cases (Formula 3).

$$(3)\quad Precision = \frac{TP}{TP + FP}$$

Sensitivity: This shows the correct percentage of results obtained by the classification algorithm. The following ratio shows the true classified positive samples (Formula 4):

$$(4)\quad Sensitivity = \frac{TP}{TP + FN}$$

F1-score: F measurement is a criterion for measuring the accuracy of a test that in this study, has been used for advanced classification evaluation and is defined based on the following formula (Formula 5):

$$(5)\quad F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Classification accuracy will be the number of correct predictions as a ratio of all predictions made (Formula 6). This is the most common evaluation criterion for classification problems that are often used [36].

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

$$(6)\ Accuracy = \frac{TP\ + TN}{TP + FN + FP + TN}$$

Area Under Curve is one of the most widely used metrics for evaluation. It is used for binary classification problems. *AUC* of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example. As evident, *AUC* has a range of [0, 1]. The greater the value, the better is the performance of our model.

5. Results and Discussion

5.1. Experimental Results

All of the records in both datasets were considered for the training of the proposed cGAN. The proposed cGAN was trained at a total of 1200 and 2000 epochs on diabetes and breast cancer records, respectively. After the training, the proposed cGAN was generated 20000 new records for each dataset, and the total number of records reach to 22000 and 20569 for diabetes and breast cancer datasets, respectively.

The experiments were performed for classifying records in each dataset. In order to show the reliability of the model, the KFold cross-validation technique was utilized. Having examined many researches in this field, we can easily conclude that a variety of classification models are suffering from overfitting problem in which the model demonstrates a good performance on the training set and it cannot perform well on the test set, for the avoidance of this situation all of the samples were divided to k sets and in every iteration, $k-1$ sets were going to use as training sets and the rest is used for test and validation. In the current study after testing a wide range of quantities, $k=5$ split is considered to be the most effective value for this task. In the order of a comprehensive evaluation, a statistical assessment was conducted for the eight representative classification models on the descriptors mention in the performance measures section. Sensitivity, precision, F1-score, and accuracy results for the binary classification of records in the diabetes dataset, are given in Table 2.

Table 2: Sensitivity, precision, F1-score, and accuracy values obtained for machine learning base models and proposed model on diabetes dataset.

Models	**Performance Metrics (%)**			
	Precision	Sensitivity	F1-score	Accuracy
SVM	70.5%	55.2%	61.8%	76.8%
Naïve Bayes	65.1%	58.6%	61.5%	75.1%
Random Forest	92.6%	92.8%	92.7%	94.9%
Decision Tree	97.3%	96.2%	96.8%	97.4%
K-NN	79.5%	57.9%	66.7%	80.3%
Logistics Regression	71.0%	52.5%	60.3%	76.4%
Proposed DNN	**98.5%**	**98.4%**	**98.2%**	**98.0%**

The shown Table 2 illustrates the results of experiments and evaluation on the diabetes dataset. As can be observed, the proposed model, Random Forest and Decision Tree showed superior performance in comparison with other models. The proposed model was highly sensitive in detecting patients with this disease with a sensitivity of 98.4% and the accuracy of 98.0%, Decision Tree and Random Forest were performed with modest differences with a sensitivity of 96.2% and 92.8%, and the accuracy of 97.4%, and 94.9%, respectively. In terms of Precision and F1-score, these models were showed similar performance in these ranges. Regarding other models, results indicate that the performance was weaker. By the way of illustration, in Naïve Bayes, the performance was at the lowest level in comparison to others, in which the accuracy was approximately 75%, SVM and Logistic Regression have followed a similar pattern with a modest difference of around 1% in terms of accuracy, and they showed insignificant sensitivity in the detection of patients with diabetes.

Table 3: Sensitivity, precision, F1-score, and accuracy values obtained for machine learning base models and proposed model on Breast Cancer dataset.

Models	**Performance Metrics (%)**			
	Precision	Sensitivity	F1-score	Accuracy
SVM	70.5%	55.2%	61.8%	76.8%

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

Naïve Bayes	65.1%	58.6%	61.5%	75.1%
Random Forest	92.6%	92.8%	92.7%	94.9%
Decision Tree	97.3%	96.2%	96.8%	97.4%
K-NN	79.5%	57.9%	66.7%	80.3%
Logistics Regression	71.0%	52.5%	60.3%	76.4%
Proposed DNN	**98.5%**	**98.4%**	**98.2%**	**98.0%**

Contrary, in the experiments on the Breast Cancer dataset, results were showed significant differences in the comparison with the diabetes dataset, as can be viewed in Table 3. In these experiments, all of the models showed similar performance and the proposed model showed robust stability in the contrast to other models. By the way of explanation, the highest performance in terms of accuracy was achieved by the proposed model, Naïve Bayes and Decision Tree and attained the results of 99.6%, 90.2%, and 90.0%, respectively. However, in terms of sensitivity which is considered to be the prime importance for detecting the positive case of patients diagnosed with Breast Cancer, the proposed model, Naïve Bayes, and SVM outperformed the others with negligible differences. Other models, followed similar patterns with insignificant variation in these terms.

Table 4: Obtained results from the experiments on two datasets based on AUC and Logloss.

Models	**Diabetes**		**Breast Cancer**	
	AUC	Logloss	AUC	Logloss
SVM	0.827	0.492	0.937	0.286
Naïve Bayes	0.822	0.590	0.958	0.368
Random Forest	0.982	0.320	0.922	1.877
Decision Tree	0.972	0.829	0.898	3.277
K-NN	0.905	0.846	0.914	1.680
Logistics Regression	0.825	0.494	0.937	0.288
Proposed DNN	**0.999**	**0.042**	**0.999**	**0.050**

By paying attention to Table 4, AUC produced results of the proposed model showed a noticeable difference in the experiments on the record of patients in two datasets. For example, the proposed model, Random Forest, and Decision Tree by far were the highest in patients diagnosed with diabetes with AUC of 0.999, 0.982, and 0.972, respectively. However, in other models, there were 8% to more than 10% variation. Naïve Bayes on experiments on diabetes patients obtained the lowest AUC results compared to other classification models, although, this trend proved negligible and on the second dataset (Breast Cancer) outperformed with AUC of 0.922. These results indicated the inconsistency among the machine learning classification methods among the different datasets. Regarding extracted values for Logloss, again the proposed model showed a strong and superior performance in two experiments with modest differences (0.042 for diabetes and 0.050 for breast cancer). In the experiments that were conducted on the diabetes dataset besides from the proposed model, the lowest Logloss was obtained from the Random Forest classifier. In contrast, in the second dataset, this model came with one of the highest values in this term, and the values were at the lowest belonged to SVM and Logistics Regression with 0.286 and 0.288, respectively.

5.2. Discussion

In this study, two well-known type of neural networks, cGAN-DNN, were combined to provide a comprehensive view of the role of diagnosing and text classification in comparison with conventional machine learning methods. The results indicate that there was significant variation in terms of accuracy, sensitivity, precision, and F1-score in-between machine learning models, whereas the proposed model showed considerable stability in two datasets, by the way of explanation, the superiority of machine learning algorithms were changed significantly by the second experiment which was on the fewer records, however, the proposed model remained its position among models. When all trials are taken into account, the performance of algorithms will differ significantly. The best classifiers out of all trials were the proposed model, Decision Tree, and Random Forest that predicted well. However, SVM, KNN, and Naïve Bayes had the lowest predictive power and they were showed an inconsistency among the experiments.

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

The nature of each experiment stage helped us explain the differences between algorithm performances. For example, in the basic implementations step, the noticeable performance was achieved by the Decision Tree, however, in the second stage which the experiments were conducted on the smaller size dataset, the performance of this algorithm is heavily reduced, this indicates that many models will suffer from the lack of training data. On the other hand, SVM and Naïve Bayes were showed much better performance on the fewer records, and also the training process was accelerated by these models. These results indicate that the classifier's performance highly depends on the size of the dataset, but in the proposed model these variations were noticeably insignificant.
Regarding all of these points that were made in the current study, lead us to consider the proposed model classifier as a strong classifier with a high level of stability.
There were numerous limitations that the authors of the current study were encountered, one of which was the inconsistency of the research on selected datasets. By the way of illustration, various researches that were focused on this area of study used a variety of datasets which as result made it quite challenging to perform a comprehensive evaluation. Another hinder was the limitation of scientific literature for evaluation specific models in some metrics, however, the current study was aiming at producing a valid comparison and experiments despite these limitations.

6. Conclusion and Future Scope

The results obtained from this study indicate both machine learning and deep learning approaches can show considerable performance in classification tasks. Future more, as shown in this study, the use of the deep learning approach in the classification, has more robust and stable performance than the conventional machine learning methods.
In the research that was conducted by Maimon and Rokach, they concluded that “no algorithm can be best in all possible domains”, and as result, they developed the “No Free Lunch Theorem”. This theory implies that if one algorithm outperforms the others in the specific domains, then there are necessarily other domains which convey the reversed implications. Having examined the previous researches, determining which classification model to use can be considered to be one of the most significant challenges that researchers now encountering, and this choice even became more challenging when other criteria are involved, such as mentioned metrics like sensitivity or even complexity. Although selecting a classification model in a variety of options for a specific problem is not going to be an easy task, this study demonstrates the importance of model selection and introduces a design of a deep neural network which showed robust stability in the experiments on different datasets with huge differences on the size. As a result, these results obtained from this study stand in contrast side of previous researches that proposed contrary theories. Admittedly, needless to mention, the nature of the data determines which classification algorithm will provide the best solution to a given problem. However, having examined many pieces of research many proposed models and methods can be utilized to minimize these variations and gain suitable performance on different classification tasks.
The algorithm can differ concerning certain criteria such as sensitivity or specificity, or other metrics, however, based on research that was conducted by Dogan and Tanrkulu [7], in practice, the suitable approach is to develop or tune several models and select the best model for implementation. The obtained results in the current study lie in agreement with the study of Dogan and Tanrkulu.
Given that the use of deep learning approaches in the classification of images is a recurring trend, in future work, we will try to, by adjusting the number of neurons and changing the structure of the model, including adding layers and strengthening the data pre-processing step, obtain the accuracy of the more the 90% in the image classification tasks.

Reference

[1] Khan A, Baharudin B, Lee LH, Khan K. A review of machine learning algorithms for text-documents classification. Journal of advances in information technology. 2010 Feb;1(1):4-20.
[2] Ikonomakis M, Kotsiantis S, Tampakas V. Text classification using machine learning techniques. WSEAS transactions on computers. 2005 Aug 8;4(8):966-74.
[3] Sebastiani F. Text categorization. InEncyclopedia of Database Technologies and Applications 2005 (pp. 683-687). IGI Global.
[4] Shah C, Jivani AG. Comparison of data mining classification algorithms for breast cancer prediction. In2013 Fourth international conference on computing, communications and networking technologies (ICCCNT) 2013 Jul 4 (pp. 1-4). IEEE.
[5] Kesavaraj G, Sukumaran S. A study on classification techniques in data mining. In2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) 2013 Jul 4 (pp. 1-7). IEEE.

[6] Maimon O, Rokach L. Introduction to knowledge discovery and data mining. InData mining and knowledge discovery handbook 2009 (pp. 1-15). Springer, Boston, MA.
[7] Dogan N, Tanrikulu Z. A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness. Information Technology and Management. 2013 Jun 1;14(2):105-24.
[8] Dunham MH. Data mining: Introductory and advanced topics. Pearson Education India; 2006.
[9] Cios KJ, Swiniarski RW, Pedrycz W, Kurgan LA. The knowledge discovery process. InData Mining 2007 (pp. 9-24). Springer, Boston, MA.
[10] Berson A, Smith S, Thearling K. Building data mining applications for CRM. McGraw-Hill Professional; 1999 Dec 1.
[11] Li C, Wang J, Wang L, Hu L, Gong P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. Remote sensing. 2014 Feb;6(2):964-83.
[12] Hacker S, Von Ahn L. Matchin: eliciting user preferences with an online game. InProceedings of the SIGCHI Conference on Human Factors in Computing Systems 2009 Apr 4 (pp. 1207-1216).
[13] Korde V, Mahender CN. Text classification and classifiers: A survey. International Journal of Artificial Intelligence & Applications. 2012 Mar 1;3(2):85.
[14] Abd AL-Nabi DL, Ahmed SS. Survey on classification algorithms for data mining:(comparison and evaluation). International Journal of Computer Engineering and Intelligent Systems. 2013;4(8):18-27.
[15] Kim SB, Han KS, Rim HC, Myaeng SH. Some effective techniques for naive bayes text classification. IEEE transactions on knowledge and data engineering. 2006 Sep 25;18(11):1457-66.
[16] Brazdil PB, Soares C, Da Costa JP. Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. Machine Learning. 2003 Mar 1;50(3):251-77.
[17] Maindonald J. Data mining methodological weaknesses and suggested fixes. InConferences in Research and Practice in Information Technology Series 2006 Nov 30 (Vol. 245, pp. 9-16).
[18] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
[19] Vapnik VN. The nature of statistical learning. Theory. 1995.
[20] Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 1;2(2):121-67, 1998.
[21] Ben-Hur A, Weston J. A user's guide to support vector machines. InData mining techniques for the life sciences 223-239, 2010.
[22] Goodman J. Exponential priors for maximum entropy models. InProceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004, 305-312, 2004.
[23] Ng AY. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. InProceedings of the twenty-first international conference on Machine learning, 4 (78), 2004.
[24] Yan W. Application of random forest to aircraft engine fault diagnosis. InThe Proceedings of the Multiconference on" Computational Engineering in Systems Applications", (Vol. 1, pp. 468-475). IEEE, 2006.
[25] Chen H, Zhan Y, Li Y. The application of decision tree in Chinese email classification. In2010 International Conference on Machine Learning and Cybernetics, Vol. 1, pp. 305-308). IEEE, 2010.
[26] Lewis DD, Ringuette M. A comparison of two learning algorithms for text categorization. InThird annual symposium on document analysis and information retrieval, Vol. 33, pp. 81-93, 1994.
[27] Wang L, Zhao X. Improved KNN classification algorithms research in text categorization. In2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet); 21 (pp. 1848-1852). IEEE, 2012.
[28] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. KNN model-based approach in classification. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Berlin, Heidelberg, (pp. 986-996). 2003.
[29] Aghila G. A Survey of Na\" ive Bayes Machine Learning approach in Text Document Classification. arXiv preprint arXiv:1003.1795. 2010.
[30] McCallum A, Nigam K. A comparison of event models for naïve bayes text classification. InAAAI-98 workshop on learning for text categorization, (Vol. 752, No. 1, pp. 41-48), 1998.
[31] Shah C, Jivani AG. Comparison of data mining classification algorithms for breast cancer prediction. In2013 Fourth international conference on computing, communications and networking technologies (ICCCNT); (pp. 1-4). IEEE, 2013.
[32] Taormina, R., Chau, K.W., 2015. Neural network river forecasting with multi-objective fully informed particle swarm optimization. J. Hydroinform. 17, 99–113.

ISSN- 2394-5125 VOL 8, ISSUE 02, 2021

[33] A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control. Trans. Neural Netw. Learn. Syst. 27, 416–425.
[34] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", Nature, vol. 521, pp. 436-444, 2015
[35] Caliskan A, Badem H, Basturk A, Yuksel ME. Diagnosis of the parkinson disease by using deep neural network classifier. Istanbul University-Journal of Electrical & Electronics Engineering. 2017 Jan 1;17(2):3311-8.
[36] Brownlee, J. (2016). Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End. Machine Learning Mastery.
[37] Rajpurkar P, Park A, Irvin J, Chute C, Bereket M, Mastrodicasa D, Langlotz CP, Lungren MP, Ng AY, Patel BN. AppendiXNet: Deep learning for diagnosis of appendicitis from A small dataset of CT exams using video pretraining. Scientific reports. 2020 Mar 3;10(1):1-7.
[38] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621. 2017 Dec 13.
[39] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. arXiv preprint arXiv:1406.2661. 2014 Jun 10.
[40] Gauthier J. Conditional generative adversarial nets for convolutional face generation. Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester. 2014;2014(5):2.
[41] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. 2014 Nov 6.