Medical Data Classification using Machine Learning Techniques

Koby Bond
Computer Science Department,
Southern Connecticut State University
501 Crescent Street,
New Haven, CT 06515

Alaa Sheta
Computer Science Department,
Southern Connecticut State University
501 Crescent Street,
New Haven, CT 06515

ABSTRACT

Medical data classification is a challenging problem in the data mining field. It can be defined as the process of splitting (i.e., categorizing) data into appropriate groups (i.e., classes) based on their common characteristics. The classification of medical data is a significant data mining problem explored in various real-world applications by numerous researchers. In this research, we provide a detailed comparison between several machine learning classification approaches and explored their predictive accuracy on several datasets. They include Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Decision Trees (DT). The quality of the developed classifiers was evaluated using several criteria such as Precision, Recall, and F-Measure. Several data set from the UCI Machine Learning Repository (i.e., Pima Indians Diabetes and the Breast Cancer Coimbra datasets) was used for this study. The experimental results reveal that the ANN-based classifier was the most accurate classification in all cases, with its ROC area being the highest.

Keywords

Medical Data Classification, Machine Learning, Neural Networks, Support Vector Machines, Decision Trees

1. INTRODUCTION

Recently, Medical Data Classification (MDC) has been a major area of research with the growth of machine learning techniques. These techniques develop learning models based upon gathered medical datasets; thus, improving the quality of medical diagnosis. Medical Data includes clinical measures coupled with environmental, social, and behavioral information that is important to the patient's wellness. Classification of medical data for error-free judgment is a challenging field of research because the data is abundantly large and complex in size. The classification of this medical data is used in the process of clinical coding which transforms medical procedures into a standardized coding system. This system is universally utilized by healthcare providers, government health programs, software developers, and other organizations. These organizations use these systems for a variety of applications in the medical field which may include statistical analysis of diseases, knowledge-based and decision support systems [1]. All this data is collected, interpreted, and utilized when patients interact with healthcare systems.

Diabetes refers to a group of diseases that affect how your body uses blood sugar (glucose). Early detection and treatment of diabetes are important so that patients can begin to manage the disease early and potentially prevent or delay serious disease complications that can decrease the quality of life of a patient. This is important because According to the U.S. Department of Health and Human Services approximately 5 million of the 18 million people with diabetes are unaware they have diabetes [2].

According to the American Cancer Society (ACS), Breast Cancer is one of the most common cancers in women. 1 in 8 women in the U.S. have the chance of developing breast cancer sometime in their life and 1 in 36 women in the U.S. will die from breast cancer. Also, according to Catherine Tuite, the Section Chief of Breast Radiology at Fox Chase Cancer Center, "A heightened awareness of the disease has led to a greater number of women being screened for breast cancer which can catch the disease when it is most treatable." This is important because patients can experience better outcomes as a result of early diagnosis, treatment options, and less extensive surgery [3].

The ability of healthcare providers to efficiently supply not only general services and information regarding a patient but also statistical analysis on healthcare issues and therapeutic measures is crucial. Since the data is normally collected through the measures of physician notes, laboratories, and other procedures the data becomes abundantly complex and contributes to poor clinical decision making. These poor decisions cause errors that affect the quality of services provided to patients. Medical data classification has the potential to generate an efficient environment that can significantly improve the quality of clinical decisions and enable providers the ability to locate data efficiently for proper implementation.

The goal behind this work is to emphasize the need for Machine Learning in the Health Care Field. Demonstrate the classification capabilities of the Support Vector Machine, Artificial Neural Network, and Decision Tree classification techniques on medical data. As well as provide a comparison between the classification techniques when classifying medical data.

This paper is organized as follows. In Section 2, we describe related work in the area of medical data classification. In Section 6, we describe the classification software employed during the experiment. Section 3 describes the classification techniques that were under study. Section 4 describes the evaluation metrics used to measure the quality of the models. In Section 5, we describe each dataset utilized during testing. Section 7 provides descriptions and visuals for the results of the experiment. Finally, Section 8 describes the conclusion summarizing all the findings within the research.

2. RELATED WORK

There have been numerous studies on the various performances of the SVM, ANN, and Decision Tree-based approaches on medical data for classification purposes. Decision Trees are a popular algorithm because of their intelligibility and simplicity. This is a diagram that uses nodes to represent a test on the dataset. There are numerous decision tree algorithms but amongst the most well-known include the C4.5 algorithm. An Artificial Neural Network is typically used for extracting knowledge from a dataset in the form of rules such as clustering, regression, and classification [4]. Support Vector Machine is a technique that uses kernel-based methods. This approach finds a hyperplane within the data to locate support vectors for linear classification [5].

The authors in [4] proposed medical data classification using the J48 Decision Tree and Multilayer Perceptron (MLP) algorithms. The paper compares the performance of both algorithms against the following datasets that were retrieved from the UCI Machine Learning Repository: Balance-Scale, Diabetes, Glass, Lymphography, and Vehicle. Using various accuracy measures they found Multilayer Perceptron (MLP) algorithm performed the best in most cases for data classification. R. Chitra et al. [6] proposes medical data classification using Neural Networks. The paper emphasizes the importance of early detection of Heart Disease. Using various accuracy measures they found that Neural Network algorithms are good for disease prediction in the early stages of the disease. B. Dennis et al. [7] proposed medical data classification based on an Adaptive Genetic Fuzzy System (AGFS). The paper describes and compares the difference between the AGFS system and current existing systems. Following the results, they found that the AGFS system obtained better accuracy results when compared to the existing systems. A. S. Galathiya et al. [8] proposed medical data classification using Decision Trees. The paper compares the C4.5, C5.0, and the ID3 algorithms on several medical datasets. Using various accuracy measures they found that the C5.0 algorithm gives sufficient classification in less computation time compared to the other classifiers. S. Singaravelan et al. [5] proposed medical data classification using the J48 and Sequential Minimal Optimization (SMO) algorithms. The paper compares the performance of both algorithms against the following datasets that were retrieved from the UCI Machine Learning Repository: Diabetes, Iris, Tic-Tac-Toe, and Yuta-Selection. Using various accuracy measures they found Sequential Minimal Optimization (SMO) algorithm performed the best in most cases for data classification.

M. L. Samb et al. [9] proposed a modified RFE-SVM based features selection method for medical data classification. For improved results, they combined the algorithm method with local search operators. Using various accuracy measures, they found that the reuse of features that were removed during the algorithm process improved the classification results. S. Khanmohammadi et al. [10] proposed a machine learning algorithm that could be used for developing clinical decision support systems (CDSS). Using various accuracy measures, they concluded that SVM models were the most desirable classification algorithms for developing clinical decision support systems. Patricio et al. [11] proposed medical data classification utilizing the SVM, logistic regression, and random forest to develop a model that could be a measure for breast cancer. Utilizing the Breast Cancer Coimbra Dataset retrieved from the UCI Machine Learning Repository, they found SVM-based models allowed for the best predictor for breast cancer within patients.

D. Chicco and G. Jurman [12] proposed medical data classification using the following machine learning algorithms: random forests, decision trees, gradient boosting, linear regression, one rule, artificial neural networks, naïve Bayes, SVM radial, SVM linear, and k-nearest neighbors. After applying those algorithms against the datasets they found the random forest was the best classifier. E. Weitschek et al. [13] proposes that the main issue with medical data classification is due to missing values, various measuring scales, and data collection procedures. The paper describes the main challenge is retrieving relevant information from the abundant medical data.

3. CLASSIFICATION TECHNIQUES

Classification models have supervised learning methods implemented to generate a set of rules to predict the values of attributes for training and testing the dataset [14]. These models provide numerous approaches and these approaches can learn from experience.

3.1 Artificial Neural Network (ANN)

Originating in the 1950s, the feed-forward neural network (FF-NN) is a biologically intuitive classification algorithm inspired by the structure of brain cells [15]. Neural Networks are comprised of a vast number of neurons that work as a processing unit. The neurons are in layers that have numerous neurons based on the complexity of the classification model under study. Each layer is connected by a link that has a selected weight. Weights determine the learning experience a network has [16]. The weights are computed based on a learning algorithm, with one of the most common being the Backpropagation (BP) learning algorithm [17].

Each neuron in the network has an activation function f_i. The Sigmoid Activation and the Hyperbolic Tangent ($\tanh$) are the most common and are shown in Figure 1. Equation 1 shows the Sigmoid activation function [18]:

$$f_c(x) = \frac{1}{1 + exp(-x)} \tag{1}$$

Equation 2 shows the Hyperbolic tangent sigmoid function:

$$f_b(x) = \tanh(x) \tag{2}$$

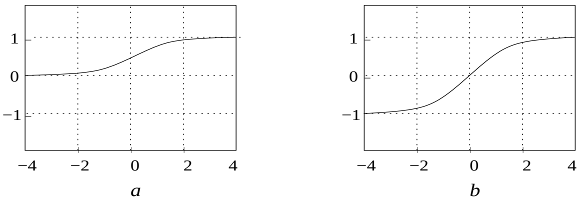


Fig. 1. a) Sigmoid ; b) Hyperbolic Tangent

Figure 2 shows a fully Connected two-layers Feedforward Network model with input, hidden, and output nodes.

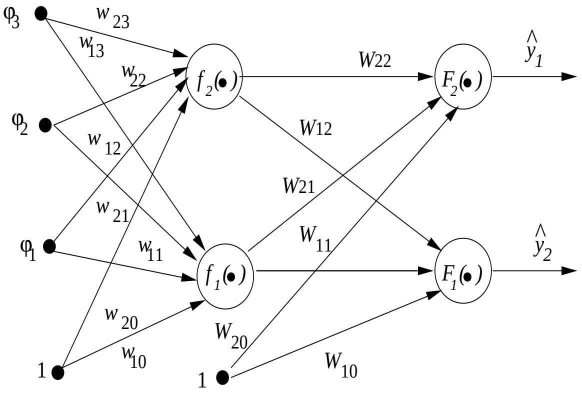


Fig. 2. A Fully Connected Two Layer Feedforward Network

3.2 Multilayer Perceptron (MLP)

In this research, we use the Multilayer Perceptron (MLP) as the learning algorithm for the Artificial Neural Network, which is provided by Weka [19]. The algorithm maps input data onto the appropriate desired outputs. This network consists of an output layer, input layer, and multiple trainable hidden layers consisting of perceptrons and sigmoid nodes that are connected and interact to produce an output [4].

We can present the learning algorithm of the fully connected MLP mathematically by Equation 3 [20]:

$$\begin{aligned}\hat{y}_i(t) &= g_i[\varphi, \theta] \\ &= F_i\left[\sum_{j=1}^{n_h} W_{i,j} f_j\left(\sum_{l=1}^{n_\varphi} w_{j,l}\varphi_l + w_{j,0}\right) + W_{i,0}\right]\end{aligned} \tag{3}$$

where:
$\hat{y}_i$: Output Signal.
g_i: Function within the network.
θ: Parameters such as weights $w_{j,l}$, and biases $W_{i,j}$.

Neural Networks can be trained for performance improvement. Weights are adjusted until the output $\hat{y}$ matches the input φ. Typically, this is achieved by training input/output pairs [21].

3.3 Support Vector Machine (SVM)

Originating from Russia, the SVM is used to solve binary classification problems [22, 23]. The SVM shows a reasonable generalization performance in malware detection [24], stock market prediction [25], Web page classification [26], manufacturing process modeling [27], and diagnosis of sleep apnea [28]. The SVM commonly uses numerical quadratic programming (QP) for classification and training the SVM is known to be slow especially when dealing with large-sized problems [29].

The SVM learning process works as follows; η is the training observations where:

$$\eta = \{(x_i, y_i) | x_i \in R^l, y \in \{1, -1\}\}_{i=1}^{n} \tag{4}$$

where y_i indicates the class where x_i is fitted. The SVM locates the optimal hyperplane that separates the points having $y_i = 1$ from those having $y_i = -1$. Equation 5 shows the hyperplane.

$$f(x) = w^T \psi(x) + b \tag{5}$$

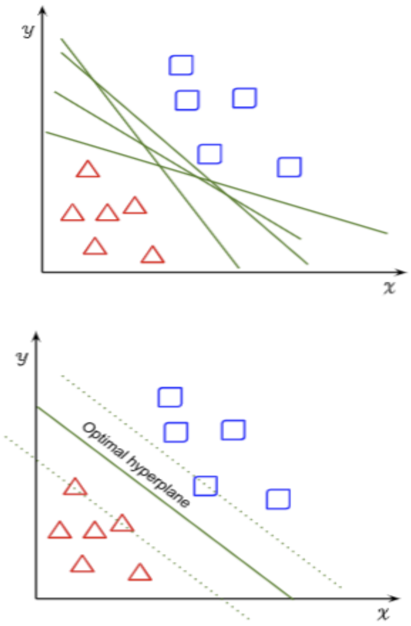


Fig. 3. SVM Optimal Hyperplane [25]

Figure 3 describes the design of the optimal hyperplane within the SVM. In the top section of the figure, the lines of separation between the data have small margins while in the lower section the line of separation between the data has the maximum margins.

For the hyperplane to be drawn between the training examples, the training data has to be linearly separable in the feature space of $\psi(x)$. The SVM maps the training vector x_i using the function ψ to find as many hyperplanes that maximize the margin. Kernel functions $K(x, y)$ are functions that take input data and transform it for processing [30]. When implemented the system only considers the data instances that are close to the hyperplane which is referred to as support vectors. In this research, we explore the use of the Polynomial Kernel Function. Equation 6 illustrates the equation for the Polynomial Kernel Function.

$$K(x, y) = (\gamma x^T y + r)^d > 0 \tag{6}$$

where γ, r, and d are kernel parameters.

3.3.1 Sequential Minimal Optimization. In this research, we use the Sequential Minimal Optimization (or SMO) as a learning algorithm for the SVM, which is provided by Weka [19]. The SMO algorithm trains a support vector classifier and is used within the SVM to solve the numeric quadratic programming (QP) problem. The SVM [31, 32] requires the solution of the QP problem:

$$\begin{aligned}&Minimize && \textstyle\sum_{i=1}^{N} \frac{1}{2}||w||^2 \\ &Subject\ to && y_i - w^T\psi(x_i) - b \leq \epsilon\end{aligned} \tag{7}$$

Where x_i is the training sample and y_i is the target value. $\hat{y}_i$is the prediction values which is computed by $w^T\psi(x_i) + b$. ϵ is the threshold parameter. To solve the QP problem the SMO algorithm splits the problem up to avoid a time-consuming numerical QP optimization technique [5].

3.4 Decision Trees

A Decision Tree is a classification technique that builds models that are in a structure of a tree when applied to a dataset. This technique breaks down a dataset into subsets until the full tree is developed. The resulting tree structure consists of branches, a root node, decision nodes, and leaf nodes. Branches depict one of the possible alternatives or courses of action available at each node. A root node is the highest node in the tree structure that depicts the best predictor. Decision nodes consist of multiple branches and leaf nodes represent the classification of a decision [33]. Figure 4 illustrates how a decision tree is constructed.

3.5 J48 Decision Tree

In this research, the J48 Decision Tree was utilized which is provided by Weka [19]. This Decision Tree is the java implementation of the C4.5 algorithm. This algorithm is an Iterative Dichotomiser 3 algorithm extension that builds decision trees from the training data. The algorithm utilizes entropy and information gain to build the decision tree. By assessing the information gain and selecting the variable that maximizes the information gain. This in turn minimizes the entropy and splits the dataset into groups for classification. This splitting process ends when all instances in a subset belong to the same class [33]. The pseudocode for the algorithm is presented in Algorithm 1.

Algorithm 1 C4.5 Algorithm Pseudo Code

```
procedure C4.5 ALGORITHM
    Check for base case
    For each feature $f$
    Find Information Gain 'G' by splitting based on $f$
    Assume $f_{best}$ is the attribute with the best gain 'G'
    if $f_{best} = found$ then
        Create Decision Node
        Re-cure on the sub-lists and add children nodes
        Repeat until all features are used
    else
        Stop (Best Tree Found)
```

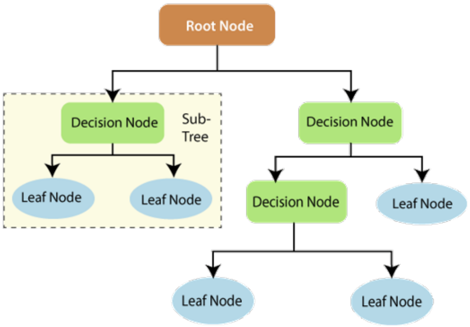


Fig. 4. Decision Tree Model [34]

4. EVALUATION METRICS

Evaluation metrics are utilized to measure the quality of a model. The metrics that were explored in this research to evaluate the performance are as follows:

4.1 Precision

Equation 8 represents the equation for the Precision metric. This metric informs the user of the proportion of prediction positives that is truly positive. In medical terms, this measures the percentage of people with a positive diagnostic test who have the disease. Lastly, how often is the model correct, when the result of the test is yes?

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

4.2 Recall

Equation 9 represents the equation for the Recall metric. This metric informs the user of the proportion of actual positives that are correctly classified. In medical terms, the proportion of patients that have both the disease and a positive result in testing. Lastly, when the answer is yes, how often does the model predict yes.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

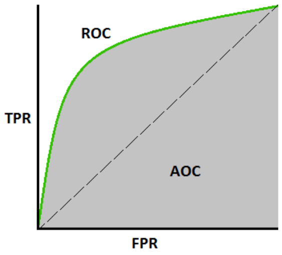
4.3 F-Measure (F1 Score)

Equation 10 represents the equation for the F-Measure metric. This metric maintains the balance between the recall and precision for the classifier because it's difficult to compare two models if there wasn't any balance. In other words, this is a weighted average of the recall and precision values.

$$F1Score = \frac{2(Precision)(Recall)}{Precision + Recall} \tag{10}$$

4.4 ROC Curve

This curve is a graph that illustrates the performance ability of the model at each classification threshold. To achieve this the curve plots the True Positive Rate and the False Positive Rate at different thresholds. And the area under the curve measures the performance of each possible threshold. So, this curve can illustrate the performance ability of the model during the entire experiment [35]. Figure 5 illustrates a ROC curve.



Curve.png

Fig. 5. ROC Curve

5. MEDICAL DATASETS

In this research, we utilized two UCI Machine Learning Repository datasets [36]. Table 1 represents the datasets that were deployed.

Table 1. Sample Datasets [36]

Dataset	Instances	Attributes	Task	Type
Pima Indians	768	9	Classification	Multivariate
Breast Cancer	116	10	Classification	Multivariate

5.1 Pima Indians Diabetes

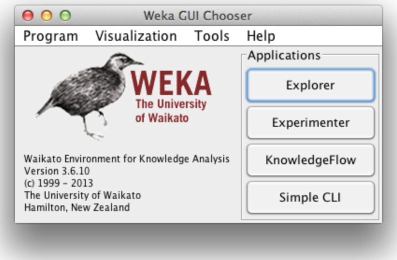
This Dataset originates from the National Institute of Diabetes and Digestive and Kidney Diseases [36]. The dataset contains 768 records of female patients that are at least 21 years old of Pima Indian heritage. The purpose of this data is to predict whether a patient has diabetes based on the diagnostic measurements presented [37].

5.2 Breast Cancer Coimbra

The Breast Cancer Coimbra Dataset [36] contains 116 clinical instances in which 64 being patients diagnosed with breast cancer and 52 being healthy patients. The purpose of this data is to indicate the presence or absence of breast cancer based on predictors and dependent variables. These predictors are data and parameters that were gathered in routine blood analysis [11].

6. CLASSIFICATION SOFTWARE

In this research, we explored the use of WEKA, which is a Machine Learning Workbench. This is a toolkit that contains a wide range of learning algorithms to identify information. Weka also has visualization tools and graphical user interfaces for ease of access and functionality [14]. Figure 6 illustrates Weka and its four main interfaces.



Waikato Environment for Knowledge Analysis
Version 3.6.10
(c) 1999 - 2013
The University of Waikato
Hamilton, New Zealand

Fig. 6. Weka Workbench [38]

- **Explorer**: This interface provides the user a method to operate Weka interactively. Many tasks are completed within this interface such as data visualization and preprocessing, simulations, execution of classifiers, training, and testing the datasets. The results can be represented as simple statistics, textual and graphical models, evaluation measures and predictions for each instance [14].
- **Experimenter**: This interface provides the user a method to construct machine learning experiments that can be saved and repeated for future instances and execute simulations. However, there is no data visualization or pre-processing, and the ability to train and test the datasets is limited. The results are typically represented through tables in this interface [14].
- **Knowledge Flow**: This interface provides the user a method to set up machine learning experiments. The results are obtained through graphical representation of data [14].
- **Simple CLI**: This interface provides the user a method to operate Weka within the command line for direct access to all of Weka's features and commands. The result representation within this interface is similar to the explorer interface [14].

7. EXPERIMENTAL RESULTS

In this experiment, We examined the performance of the SVM, ANN, and Decision Tree-based classifiers against multiple datasets to generate an overall accuracy comparison. The following tables represent the individual accuracy measurements of a classification model upon each dataset and the following figures are graphical representations of their corresponding tables. The individual accuracy measurements include TP Rate, FP Rate, Precision, Recall, F-Measure, and ROC Area. Once more, TP Rate measures how often the model correctly predicts classifications. While FP Rate measures how often the model incorrectly predicts classifications. Precision measures how often the model is correct when predicting positive patterns. Recall measures how often the model correctly classifies positive patterns. F-Measure is the weighted average between the recall and precision metrics. Lastly, ROC Area is the area represented by ROC curves, and it represents the performance ability of the model during the entire experiment. Each table and figure gives an insight on the overall accuracy of each model as well as illustrates how each classification model performed on each dataset respectfully.

7.1 Classification Models for Pima Indians Dataset

Testing on the Pima Diabetes Dataset shows that the ANN-based algorithm was the best predictor of how accurately it classified patients with or without diabetes. Table 2 shows the evaluation statistics for each test on the Pima Diabetes Dataset. It also emphasizes that the ANN-based algorithm contained the greatest ROC Area during the entire test. This means that the ANN-based algorithm shown the best predictive ability throughout the experiment.Figure 7 represents the Decision Tree and Figure 8 represents the Neural Network developed during the experiment.

Table 2. Accuracy Measures on
Pima Indians Diabetes Dataset

Parameter	DT	SVM	ANN
TP Rate	.712	.775	.751
FP Rate	.338	.341	.310
Precision	.715	.772	.749
Recall	.712	.775	.751
F-Measure	.713	.763	.750
ROC Area	.729	.717	**.809**

7.2 Classification Models for Breast Cancer Dataset

Testing on the Breast Cancer Dataset shows that the ANN-based algorithm was the best predictor of how accurately it classified patients with or without breast cancer. Table 3 shows the evaluation statistics for each test on the Breast Cancer Dataset. It also emphasizes that the ANN-based algorithm contained the greatest ROC Area during the entire test. Meaning the ANN-based algorithm showed the best predictive ability throughout the experiment.

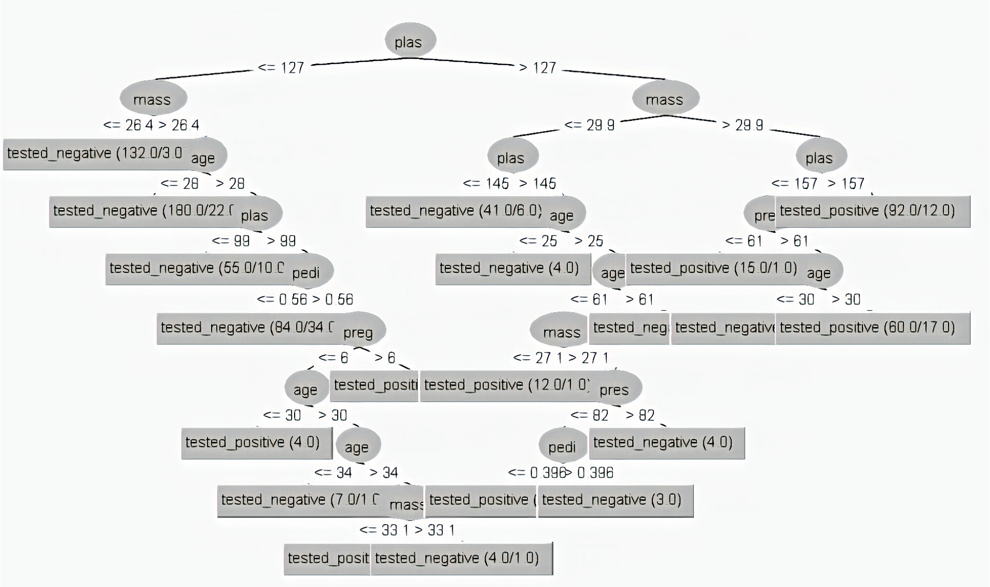


Fig. 7. Developed Tree for Pima Indians Diabetes Dataset

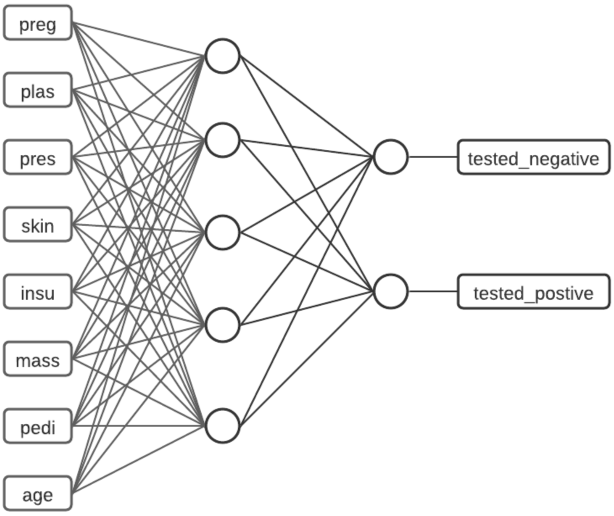


Fig. 8. Developed ANN for Pima Indians Diabetes Dataset

Figure 9 represents the Decision Tree and Figure 10 represents the Neural Network developed during the experiment.

Table 3. Accuracy Measures on Pima Indians Diabetes Dataset

Parameter	DT	SVM	ANN
TP Rate	.698	.741	.707
FP Rate	.317	.271	.296
Precision	.697	.741	.708
Recall	.698	.741	.707
F-Measure	.696	.740	.707
ROC Area	.693	.735	**.759**

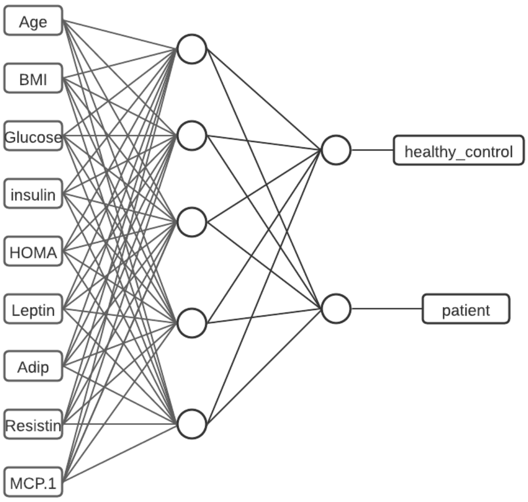


Fig. 10. Developed ANN for Breast Cancer Dataset

8. CONCLUSIONS

In this study, we presented our initial idea in exploring the use of several machine learning classifiers to classify medical data. ANN, SVM, and DT were adopted to classify a medical data set of Pima Indians Diabetes and the Breast Cancer Coimbra datasets. Several experiments were implemented to compare these algorithms to get the best results. ANN shows better classification accuracy than SVM and DT. We plan to extend this research to explore a possible

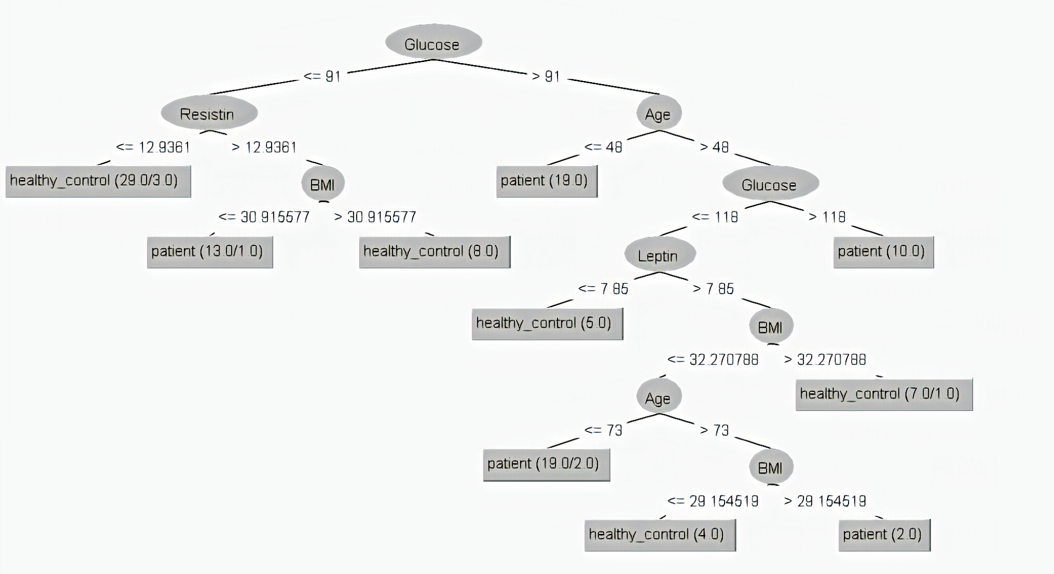


Fig. 9. Developed Tree for Breast Cancer Dataset

hybrid ANN and metaheuristics search algorithm to enhance the ANN classification accuracy.

Acknowledgement

Authors would like to acknowledge the financial support of the Computer Science Department, Southern Connecticut State University, CT, USA

9. REFERENCES

[1] B. Tarle, "Medical data classification using different optimization techniques: A survey," 09 2016.

[2] "The importance of early diabetes detection," Feb 2017. [Online]. Available: https://aspe.hhs.gov/report/diabetes-national-plan-action/importance-early-diabetes-detection#:~:text=Earlydetectionandtreatmentof,limbamputations,andkidneyfailure.

[3] Feb 2021. [Online]. Available: https://www.foxchase.org/blog/why-breast-cancer-awareness-so-important

[4] R. Arora and S. Suman, "Comparative analysis of classification algorithms on different datasets using weka," *International Journal of Computer Applications*, vol. 54, pp. 21–25, 09 2012.

[5] S. S., D. Murugan, and S. Mayakrishnan, "A study of data classification algorithms j48 and smo on different datasets," *Asian Journal of Research in Social Sciences and Humanities*, vol. 6, p. 1276, 01 2016.

[6] R. Chitra and V. Seenivasagam, "Review of heart disease prediction system using data mining and hybrid intelligent techniques," in *SOCO 2013*, 2013.

[7] B. Dennis and S. Muthukrishnan, "Agfs: Adaptive genetic fuzzy system for medical data classification," *Appl. Soft Comput.*, vol. 25, pp. 242–252, 2014.

[8] A. P. G. A.S. Galathiya and C. K. Bhensdadia, "Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning," 2012.

[9] M. L. Samb, F. Camara, S. N'Diaye, Y. Slimani, M. Esseghir, and C. Anta, "A novel rfe-svm-based feature selection approach for classification," 2012.

[10] S. Khanmohammadi and M. Rezaeiahari, "Ahp based classification algorithm selection for clinical decision support system development," in *Complex Adaptive Systems*, 2014.

[11] M. Patrício, J. Pereira, J. Crisóstomo Silva, P. Matafome, M. Gomes, R. Seiça, and F. Caramelo, "Using resistin, glucose, age and bmi to predict the presence of breast cancer," *BMC Cancer*, vol. 18, 12 2018.

[12] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, 12 2020.

[13] E. Weitschek, G. Felici, and P. Bertolazzi, "Clinical data mining: Problems, pitfalls and solutions," *2013 24th International Workshop on Database and Expert Systems Applications*, pp. 90–94, 2013.

[14] T. Smith and E. Frank, "Introducing machine learning concepts with weka," *Methods in molecular biology (Clifton, N.J.)*, vol. 1418, pp. 353–378, 03 2016.

[15] G. D. McCann, J. L. Barnes, F. Steele, L. Ridenour, and A. W. Vance, "An evaluation of analog and digital computers," in *Proceedings of the February 4-6, 1953, Western Computer Conference*, ser. AIEE-IRE '53 (Western). New York, NY, USA: Association for Computing Machinery, 1951, p. 19–48.

[16] R. D. D. Veaux, R. D. De, V. Lyle, and H. Ungar, "A brief introduction to neural networks."

[17] J. Li, J.-h. Cheng, J.-y. Shi, and F. Huang, "Brief introduction of back propagation (bp) neural network algorithm and its improvement," in *Advances in Computer Science and Informa-*

tion Engineering, D. Jin and S. Lin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 553–558.

[18] N.Chiras, C.Evans, and D.Rees, "Non-linear gas turbine modeling using feedforward neural networks," *Proceedings of ASME TURBO EXPO June 3-6, Amsterdam, The Netherlands GT-30035, University of Glamorgan, publisher of Electronics, Pontypridd, CF37 1DL, Wales, UK*, 2002.

[19] E. Frank, M. A. Hall, G. Holmes, R. Kirkby, B. Pfahringer, and I. H. Witten, *Weka: A machine learning workbench for data mining.* Berlin: Springer, 2005, pp. 1305–1314.

[20] M.Norgaard, O.Ravn, Poulsen, and L.K.Hansen, *Neural Networks for Modelling and Control of Dynamic Systems.* Springer, London, 2000.

[21] C.Wu.Rebecca, "Neural network models: Foundations and applications to an audit decision problem," vol. 75, pp. 291–301, 1997.

[22] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[23] V. N. Vapnik and A. Y. Chervonenkis, "A class of algorithms for pattern recognition learning," *Avtomat. i Telemekh.*, vol. 25, no. 6, p. 937–945, 1964.

[24] G. Dai, J. Ge, M. Cai, D. Xu, and W. Li, "Svm-based malware detection for android applications," in *Proceedings of the 8th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec '15. New York, NY, USA: Association for Computing Machinery, 2015.

[25] A. Sheta, S. Ahmed, and H. Faris, "A comparison between regression, artificial neural networks and support vector machines for predicting stock market index," *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, pp. 55–63, 07 2015.

[26] A. Zubiaga, V. Fresno, and R. Martínez, "Is unlabeled data suitable for multiclass svm-based web page classification?" in *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*, ser. SemiSupLearn '09. USA: Association for Computational Linguistics, 2009, p. 28–36.

[27] A. Rodan, A. F. Sheta, and H. Faris, "Bidirectional reservoir networks trained using SVM+ privileged information for manufacturing process modeling," *Soft Comput.*, vol. 21, no. 22, p. 6811–6824, Nov. 2017.

[28] C. Haberfeld, A. F. Sheta, M. S. Hossain, H. Turabieh, and S. Surani, "SAS mobile application for diagnosis of obstructive sleep apnea utilizing machine learning models," in *11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, UEMCON 2020, New York City, NY, USA, October 28-31, 2020.* IEEE, 2020, pp. 522–529.

[29] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Tech. Rep. MSR-TR-98-14, April 1998. [Online]. Available: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/

[30] D. Boswell, "Introduction to support vector machines," 2002.

[31] B. E. Boser and et al., "A training algorithm for optimal margin classifiers," in *In Proceedings of the 5 th Annual ACM Workshop on Computational Learning Theory.* ACM Press, 1992, pp. 144–152.

[32] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[33] S. Kiranmai and L. Ahuja, "Data mining for classification of power quality problems using weka and the effect of attributes on classification accuracy," *Protection and Control of Modern Power Systems*, vol. 3, 12 2018.

[34] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, p. 81–106, Mar. 1986.

[35] "Classification: Roc curve and auc — machine learning crash course." [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

[36] [Online]. Available: https://archive.ics.uci.edu/ml/datasets.php

[37] "Pima indians diabetes database - dataset by data-society," Dec 2016. [Online]. Available: https://data.world/data-society/pima-indians-diabetes-database

[38] J. Brownlee, "What is the weka machine learning workbench," Aug 2020. [Online]. Available: https://machinelearningmastery.com/what-is-the-weka-machine-learning-workbench/