International Journal of Electrical and Computer Engineering (IJECE)
Vol. 11, No. 2, April 2021, pp. 1561~1569
ISSN: 2088-8708, DOI: 10.11591/ijece.v11i2.pp1561-1569

An ICA-ensemble learning approaches for prediction of RNA-seq malaria vector gene expression data classification

Micheal Olaolu Arowolo[1], Marion O. Adebiyi[2], Ayodele A. Adebiyi[3], Charity Aremu[4]
[1,2,3]Department of Computer Science, Landmark University, Omu-Aran, Kwara State, Nigeria
[4]Department of Agriculture, Landmark University, Omu-Aran, Kwara State, Nigeria

Article Info

Article history:

Received Feb 4, 2020
Revised Jul 17, 2020
Accepted Sep 23, 2020

Keywords:

Ensemble classifier
ICA
Malaria vector
RNA-seq

ABSTRACT

Malaria parasites introduce outstanding life-phase variations as they grow across multiple atmospheres of the mosquito vector. There are transcriptomes of several thousand different parasites. Ribonucleic acid sequencing (RNA-seq) is a prevalent gene expression tool leading to better understanding of genetic interrogations. RNA-seq measures transcriptions of expressions of genes. Data from RNA-seq necessitate procedural enhancements in machine learning techniques. Researchers have suggested various approached learning for the study of biological data. This study works on ICA feature extraction algorithm to realize dormant components from a huge dimensional RNA-seq vector dataset, and estimates its classification performance, Ensemble classification algorithm is used in carrying out the experiment. This study is tested on RNA-seq mosquito anopheles gambiae dataset. The results of the experiment obtained an output metrics with a 93.3% classification accuracy.

Corresponding Author:

Micheal Olaolu Arowolo
Department of Computer Science
Landmark University
Omu-Aran, Kwara State, Nigeria
Email: arowolo.micheal@lmu.edu.ng

1. INTRODUCTION

Next-generation sequencing technology has created several wide datasets, that allows biologists to examine and determine difficult gene transcripts such as RNA relationships and ailments such as cancer, contagions (malaria), tumors, heredities, biological, among others [1]. In Africa, mosquito anopheles gambiae are blood-sucking parasites with large pathways to Plasmodium Falciparum. Anopheles mosquitoes is a deadly malaria parasite, accountable for thousands of deaths. As battle with antimalaria suppositories banquets upsurges, perceptives for state-of-the-art drugs necessitates improved biological knowledge of these kind. Mosquito anopheles organism approved precise gene expression controls has been a major concern needing an improved quantitative predictive malaria vector transcripts model [2, 3].

RNA-seq learning produces sensitive biological perceptive investigations by recognizing a preliminary biological enhanced sequencing purposeful plan analysis. RNA-seq data includes the removal of the high-dimensionality curses in a data, such as: disturbances, repetitions, inconsistencies, redundancy, irrelevant, incorrect, invalid, among others [4]. Recent innovations have enhanced approaches for designing state-of-the-art healthcare models such as adapted therapies, intelligent health surveillance systems, among other disease diagnoses [5].

Numerous machine learning methods with practical advances have peen developed through the years to analyze the enormous volume of RNA-seq and data expression of next generation gene sequencing by studying the related biologically outlines [6]. Researchers have used machine learning techniques with

variable performance levels for RNA-seq gene expression data [7, 8]. Computational approaches have remained appliecable to large genetic ailments databases of persons, genes can be found responsible for the presence of ailments. Numerous approaches are used in detecting differentially expressed genes (DEG). Procedures of datamining are significant in identifying the differences between genes derived from the human genome. Numerous machine learning methods are emulated and used in examining and identifying expression of various gene profiling diseases. Gene expression profiling and its approaches by means of numerous datamining are indispensable. Research works have been proposed by numerous authors in this area, existing researches are known in studying gene expressions [5]. Blood-based signature gene expression and datamining for diseases in identifying transcripts that can be used in classification is proposed [9]. Using Gene expression omnibus database from RNA data and using machine learning algorithms language tools, works on RNA-seq data have been proposed by dimensionality reduction, clustering and classification by performing an integrated review, that have recently arose as predominant shifts, using indirect and direct methods with reducing sc-RNA-seq data dimension approaches, reporting scRNA-seq data [10].

This study proposes a dimensionality reduction model, by using ICA feature extraction technique, to realize the relevant correlated latent components in a high dimensional dataset in the gene expression data analyzsis, a Sub-space group Ensemble classification system is used in learning discrete biological outlines that helps achieve developed classification accuracy and suggested as an effective procedure for the finding of innovative genes for malaria.

2. PROPOSED METHOD

In this study, a summarized proposed framework in Figure 1 is adopted, the fundamental idea is to predict machine learning task on high dimensional RNA-seq data, for cells and genes into lower dimensional dataset. The plan is adjusted to fetch out important data in a given dataset by utilizing ICA feature extraction method as a stage. To evaluate the performance of RNA-seq dataset, Ensemble classification algorithms are compared.

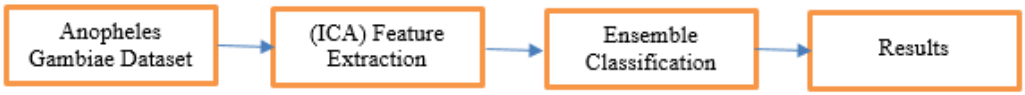

Figure 1. Proposed framework

Numerous approaches on machine learning have been emulated to examine and identify gene expression profiles of several ailments. There is discussion of the necessity for expression of gene profiling and approaches using specific datamining techniques. Numerous investigations carried out by researchers in this area are consulted, recent investigations in analysing gene expressions are reviewed [5]. A supervised machine learning method for variety of RNA-seq segment was proposed by ranking huge sets of segments measured with RNA-seq, using random forest classifier variable rank measurements, specifying the EPS (extreme pseudo-samples) frequency, with variational autoencoder regressors in the RNA-seq extraction ranks of cancer datasets with about 1,210 samples. Results in the RNA-seq training demonstrated a supervised hidden learning-based feature selection method and highlighted the need for gene assortment methods for gene expression analysis [11]. Classification of RNA-seq dataset using supervised model was proposed for a generalized method of highly accurate single cell classifications, by integrating unbiased collection of condensed dimensional space feature selection technique. Sc-Pred was used on RNA-seq pancreatic tissue, colorectal tumour cell removal, mononuclear cells, and mixing dendritic cells datasets. Sc-Pred demonstrated a high classified discrete cells accuracy [12].

RNA-DNA machine learning analysis was proposed on a low expressed genome that could be affected collectively by PAH disease. A state-of-the-art feature selection procedure to classify an irrelevant range of very beneficial genes. Small expression clustered genes were discovered at predicting transformed PAH procedures [13]. Stomach cancer gene expression data classification was developed using deep learning approach, Heatmaps, PCA, and CNN algorithm. RNA-seq gene data expression studied the genes and analysed them, 95.96% and 50.51% were achieved [14]. Transcriptions of RNA-seq malaria data through dissimilarity of techniques to deconvolute disparity transcription for dissimilar malaria parasites were revealed using hidden transcriptional discrete signatures [15]. Supervised datamining approaches such as C4.5, boosted and bagged ensemble classification algorithm for cancer data were proposed on openly available oncogenic microarray data and correlated, the boost and bag ensemble classification outperform the

C4.5 [16]. A diagnostic classification using ensemble algorithm method for genomic cancer data expression was designed using RFE to fetch efficient features for enhanced classification results using AdaBoost [17]. Classification of cancer gene data expression, was carried out using effective ensemble learning method upsurging the performance of the classification of the outcome results, with a reduced amount of dependent on originalities of individual training set [18]. An enhanced ensemble classification learning wrapper-based feature selection and random trees procedure to improve knowledge, makes a subset by using bagged and random trees. Irrelevant features were removed to select the best features for classification, using RF, SVM, and NB with 92% accuracy [19]. Text classification algorithms was proposed using various text dimensionality reduction methods [20].

3. RESEARCH METHOD

Datamining for high dimensional dataset enhancements have been carried out by several authors, Independent component analysis (ICA) and classification using algorithm is proposed for RNA-seq malaria vector data.

3.1. Material

A western Kenya mosquito gene dataset with 7 attributes genes and 2457 instances were used, containing mosquito genes from 2010 to 2012, The profile transcripts contains AGAP003714, AGAP004779, CPLC G3 [AGAP008446], CYP6M2 [AGAP008212], AGAP012984, AGAP002724, AGAP009472 and CYP6P3 [AGAP002865], RNA-seq deltamethrin-resistant transcriptome distinctions and susceptible western Kenyan mosquito Anopheles gambiae genes available dataset from National Institute of Health [21], a summary explanation of the dataset is shown in the Table 1.

Table 1. Dataset description

Dataset	Attributes	Instances
Mosquito Anopheles Gambiae	7	2457

3.2. Methods

The experimental tool used MATLAB to analyze the data obtained [21], using ICA to fetch latent features, and carry out classification using ensemble algorithm approach [22] on the MATLAB tool environment.

3.2.1. Independent component analysis (ICA)

ICA is a valued PCA extension that has remained conservative since the visor parting of independent bases from their direct grouping [20]. The original fact of ICA is the possessions of uncorrelation of the general PCA. Built n x p on data medium X, whose rows ri $(j = 1 \dots, n)$ reckon toward variables observed also whose cj $(j = 1 \dots, p)$ columns are the entities of matching variables, the ICA X model, written as follows:

$$X = AS \tag{1}$$

With complete overview, A is a nxn fusion matrix, where S is a nxp is a basis matrix under the need of being statistically independent as conceivable. Independent components are the innovative variables kept in the rows of S, to wit, the variables detected are linearly composed independent components. The independent components achieved by learning the precise linear groupings of the variables observed, subsequently mixing can be inverted as:

$$U = S = A - 1X = WX \tag{2}$$

3.2.2. Ensemble classifier

Ensemble classifiers can be proficient using on unrelated subsets of the data training, diverse classification constraints, or with diverse subset features in random subspace model [23]. Ensemble classifier comprises of integrating fallouts of assorted classifiers to produce a concluding decision, it is frequently used for gaining highly accurate results. Ensemble classifiers are relatively common in machine learning complications, and can be employed in bioinformatics field. Classification decision is achieved by merging the decision of each classifier [24]. Ensemble approaches is machine learning techniques combines decisions to advance the performance of the general classification. Several terms have been discovered in the literature to signify comparable connotations such as; multi-strategy learning, aggregation, integration multiple

classifiers, classifier fusion, combination, committee, and so on. Ensemble classifier may have complete and improved performance than discrete base classifiers. The efficiency of ensemble approaches is extremely dependent on the unconventionality of error devoted by discrete learner. Ensemble approaches performance hinge on the accuracy and variety of the base learners, ensemble classification has common techniques;

Boostrap aggregating (Bagging) employs training the data by arbitrarily changing the unique training T by items N data. The training auxiliary sets are called bootstrap duplicates with some occurrences not appearing while others give the impression more than once. The C*(x) final classifier is built by combining Ci(x). All Ci(x) takes an equivalent division.

Adaptive boosting (AdaBoost) technique effects the training data. Originally, the procedure allocates all xi instance by means of an equivalent mass. In individual iteration i, knowledge algorithm attempts to diminish the training set weighted error and a classifier $Ci(x)$ is yielded. The $Ci(x)$ weighted error is calculated and useful in informing the training instances xi weights. xi weight rises giving to its effects on the performance of classifier's that allots a weight higher for a misclassified xi and a small weight aimed at an acceptably classified xi. The concluding classifier C*(x) is created by a discrete Ci(x) weighted vote rendering to its built accuracy on the training weighted set [19]. Adopting Kamran, *et al*. [20], they showed how a boosting algorithm works for datasets, then trained by multi-model designs (ensemble learning). These advances resulted in the adaptive boosting (AdaBoost). Presume constructing Dt such that $D1(i) = \frac{1}{m}$ given Dt and ht:

$$D_{i+1}\{i\} = \frac{D_t(i)}{Z_t} X \begin{cases} e^{-\alpha_t} \; if y_i = \; h_t(x_i) \\ e^{\alpha_t} \; if y_i \neq \; h_t(x_i) \end{cases} \quad (3)$$

$$\frac{D_t(i)}{Z_t} \exp(-\alpha y_i h_t(x_i)) \quad (4)$$

where Z_t states to the normalization factor and α_t is as follows;

$$\alpha_t = \frac{1}{2} in(\frac{1-\epsilon_t}{\epsilon_t}) \quad (5)$$

Basic ensemble classification techniques namely: The max voting (MV), weighted averaging (WA) and Averaging. Max voting (MV) exists [25-27] Ensemble learning have three combinational methods: stacking (STK), blending (BLD), bagging (BAG) and boosting (BOT) [28-31].

3.3. Evaluation performance

Datamining model performance evaluation requires metrics of validations, classification algorithms uses the confusion matrix in analyzing four features known as the; true positive (TP), false positive (FP), true negative (TN) and the false negative (FN). These features recognize the correctly and incorrectly classified instances from the given sample of dataset used in testing the model [5, 32].

3.4. Applications

An enhanced path of gene expression analysis in dentifying RNA-seq data discoveries for related genes can be helpful in the development of various applications such as modified treatment, diseases detection, genes and drug discovery, tumor recognition, ailments, among others. Datamining technique is used in identifying the designs and possesses fantastic applicable algorithms tools. In this study, MATLAB tool is used to carry out the program due to its user-friendly environment [16], to predict RNA-seq technology for the prognosis and dialnosis of malaria ailments using an 8GB RAM size, 64-bit System, iCore2 processor and MATLAB 2015A tool.

4. RESULTS AND DISCUSSIONS

This study determines RNA-seq innovation of 2457 instances mosquitoes' data. ICA algorithm was applied to fetch out latent components from the anopheles' data, the ICA feature extraction distinguishes and removes uncorrelated variables, to choose the determinant variance with a reduced number of independent components to give important useful gene evidence valuable for supplementary examination. Ensemble AdaBoost classification algorithm is applied on the extracted ICA 45 latent significant features of genes realized in 7.8486 Seconds. 10-folds cross validation is used to evaluate the classification execution performance, using 0.05 parameter holdout to training the data and 5% for testing the classification accuracy.

Assessment learning procedure classification is used to train, test and evaluate the experiment using a 10-fold cross validation in eliminating the sampling partialities. Result evaluation is carried out on the computational time and performance metrics [32]. Classification of the models, using AdaBoost ensemble classifier is carried out with 93.3% performance accuracy. The results and procedures are shown in the figures below. ICA feature extraction algorithm is used to extract the hidden features from mosquito anopheles data shown in Figure 2. The extracted features are classified using ensemble algorithm, the scattered plot and results are shown in the figures below using the confusion matrix to give a result to the performance metrics.

In Figure 3 a scattered plot is shown for the classification, the correctly classified and misclassified using dots and cross signs to represent values for the variables, indicating values for individual data points, this plot is used in observing the relationships between the classified variables. Figure 4 and Figure 5 shows the confusion matrix for the classifications of the experiment, using bagged and boosted ensemble classifiers. The confusion matrix table is then used in describing the performance of the classification model of the sets of the tested data with the known true values with the confusion matrix represented with true positive, false positive, true negative and false negative values.

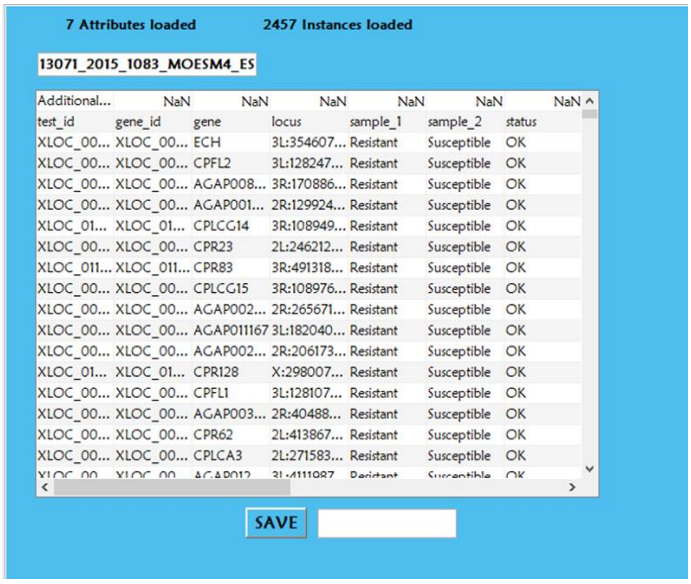

Figure 2. The mosquito anopheles gambiae dataset

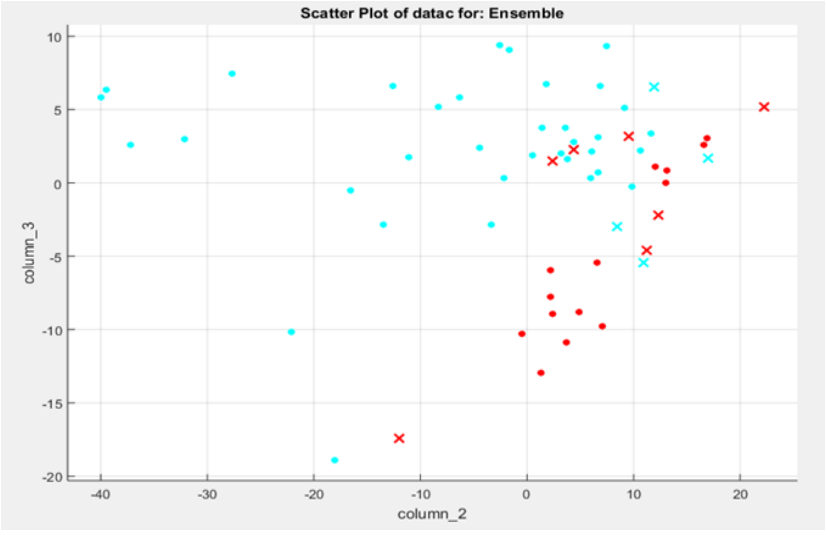

Figure 3. Classification scattered plot

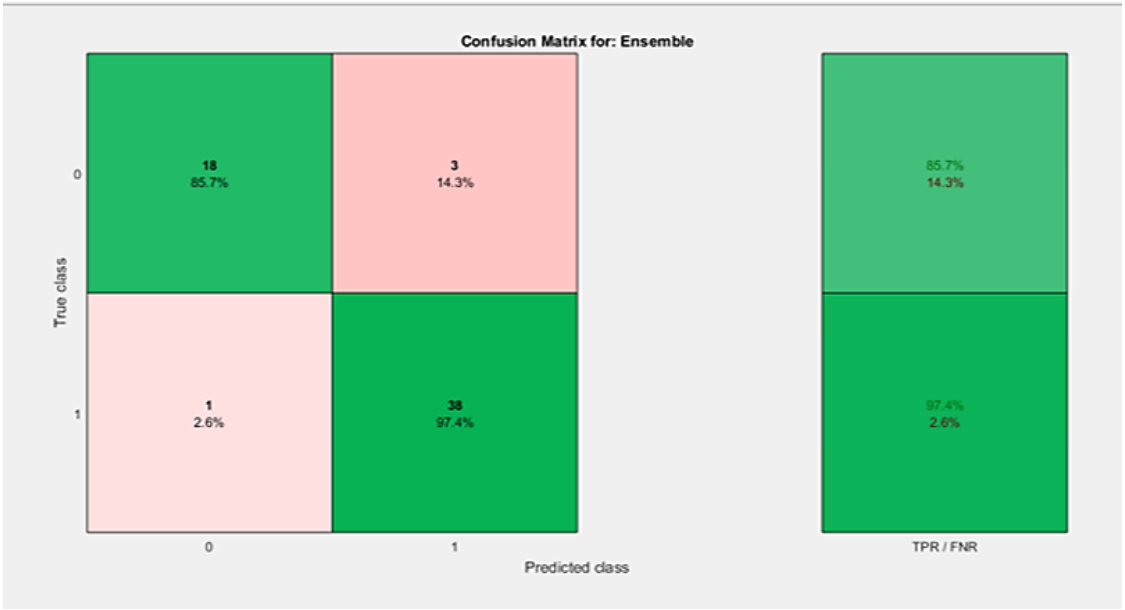

Figure 4. Confusion matrix for ensemble subspace discriminant classification TP=38; TN=18; FP=3 FN=1

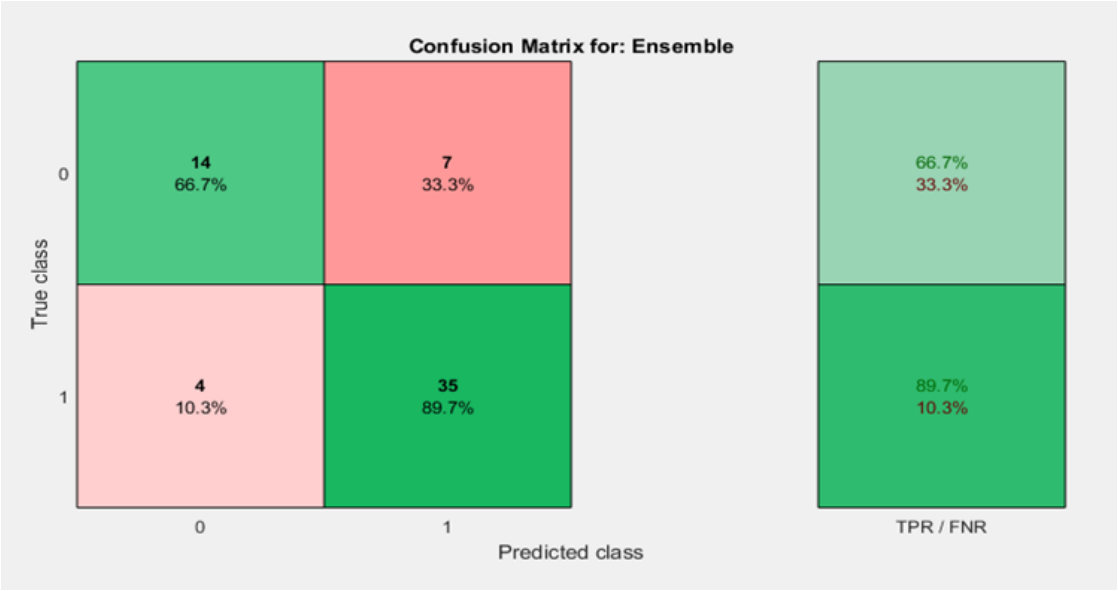

Figure 5. Confusion matrix for ensemble bagged tree classification TP=35; TN=14; FP=7; FN=4

Testing the datamining learning performance methods, the RNA-seq data was copied from the https://figshare.com/articles/Additional_file_4_of_RNAseq_analyses_of_changes_in_the_Anopheles_gambiae_transcriptome_associated_with_resistance_to_pyrethroids_in_Kenya_identification_of_candidateresistance_genes_and_candidateresistance_ SNPs/4346279/1 repository. ICA feature extraction technique was used on the 2457 genes features, and extracted 1572 features with 45 latent components. Ensemble classification is used to predict the performance. Result demonstrated the efficiency of datamining approached in genes. The performance results for the proposed approach are revealed and related in Table 2. The outcome shows that Subspace Discriminant ensemble classification outperforms bagged tree ensemble in terms of accuracy.

In this study, an improved investigation of the classification of malaria vector data is carried out, numerous works have been proposed by investigators, the figure and tables above have shown and demonstrated that, dimensionality reduction model with ICA feature extraction methods can progress ensemble classification results, Figure 6 shows the performance chart for comparing the output results. This study

proposed a prediction and detection model for malaria disease in human. The proposed method used an ICA dimensionality reduction and ensemble classification datamining procedures, the investigation and performance assessment of the results gotten were shown in the tables and figures below.

Table 2. Performance metrics table for the confusion matrix

Performance Metrics	Ensemble Subspace Discriminant Classification	Ensemble Bagged Tree Classification
Accuracy (%)	93.3	81.7
Sensitivity (%)	97.4	89.7
Specificity (%)	85.7	66.7
Precision (%)	92.7	83.3
Recall (%)	97.4	89.7
F-Score (%)	95.0	86.4

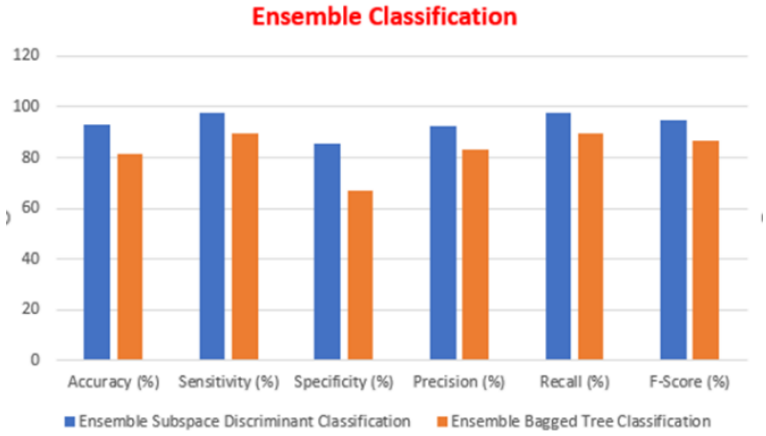

Figure 6. Performance metrics graph

5. CONCLUSION

An enhanced classification approach for malaria prognosis and diagnosis using dimensionality reduction and classification algorithm was proposed, numerous works by researchers in this area has been reviwed, results of the experiment have demonstrated ICA feature extraction dimensionality reduction can support the advancement of ensemble classification. Recent and future works can be enhanced using other ensemble classifiers with other feature extraction algorithms.

REFERENCES

[1] Shanwen S., Chunyu W., Hui D., Quan Z., "Machine Learning and its Applications in Plant Molecular Studies," *Briefings in Functional Genomics*, vol. 19, no. 1, pp. 40-48, 2019.
[2] David F. R., Kate C., Yank Y. L., Karine G., Roch L., "Predicting Gene Expression in the Human Malaria Parasite Plasmodium Falciparum Using Histone Modification, Nucleosome Positioning, and 3D Localization Features," *PLOS Computational Biology*, vol. 15, no. 9, pp. 1-23, 2019.
[3] Alistair Miles, et al., "Genetic diversity of the African malaria vector Anopheles gambiae," *Nature*, vol. 552, no. 7683, pp. 96-100, 2017.
[4] Arowolo M. O., Adebiyi M., Adebiyi A., "A Dimensional Reduced Model for the Classification of RNA-seq Anopheles Gambiae Data," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 23, pp. 3487-3496, 2019.
[5] Karthik S. and Sudha M., "A Survey on Machine Learning Approaches in Gene Expression Classification in Modelling Computational Diagnostic System for Complex Diseases," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 2, pp. 182-191, 2018.
[6] Johnson N. T., Dhroso A., Hughes K. J., Korkin D., "Biological classification with RNA-seq data: Can alternatively spliced transcript expression enhance machine learning classifiers?," *RNA*, vol. 24, no. 9, pp. 1119-1132, 2018.
[7] Libbrecht M. W. and Noble W. S., "Machine learning applications in genetics and genomics," *Nature Reviews Genetics*, vol. 16, pp. 321-332, 2015.
[8] Jagga Z. and Gupta D., "Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms," *BMC Proceedings*, vol. 8, 2014, pp. 1-7.
[9] Oh D. H., Kim I. B., Kim S. H., Ahn D. H., "Predicting Autism Spectrum Disorder Using Blood-based Gene Expression Signatures and Machine Learning," *Clin Psychopharmacology Neuroscience*, vol. 15, no. 1, pp. 47-52, 2017.

[10] Ren Q., Anjun M., Qin M., Quan Z., "Clustering and Classification Methods for Single-cell RNA-seq Data," *Briefings in Bioinformatics*, vol. 21, no. 4, pp. 1-13, 2019.
[11] Stephen W. and Ruhollah S., "Using Supervised Learning Methods for Gene Selection in RNA-seq Case-Control Studies," *Frontiers in*, vol. 9, no. 297, pp. 1-6, 2018.
[12] Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nquyen Q., Powell J. E., "scPred: Accurate Supervised Method for Cell-type Classification from Single-cell RNA-seq Data," *Genome Biology*, vol. 20, no. 264, pp. 1-17, 2019.
[13] Cui S., Wu Q., West J., Bai J., "Machine Learning-based Microarray Analyses Indicate Low-Expression Genes Might Collectively Influence PAH Disease," *PLOS Computational Biology*, vol. 15, no. 8, pp. 1-25, 2019.
[14] Shon H. S., Yi Y. G., Kim K. O., Cha E. J., Kim K. A., "Classification of Stomach Canacer Gene Expression Data Using CNN Algorithm of Deep Learning," *Journal of Biomedical Translation Research*, vol. 20, no. 1, pp. 15-20, 2019.
[15] Adam J. R., et al., "Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites," *eLIFE, Tools and Resources*, vol. 7, pp. 1-29, 2018.
[16] Tan A. C. and Gilbert D., "Ensemble Machine Learning on Gene Expression Data for Cancer Classification," *Applied Bioinformatics*, vol. 2, no. 3, pp. 75-83, 2003.
[17] Song N., Wang K., Xu M., Xie X., Chen G., Wang Y., "Design and Analysis of Ensemble Classifier for Gene Expression Data of Cancer," *Advancement in Genetic Engineering*, vol. 5, no. 1, pp. 1-7, 2016.
[18] Tarek S., Elwahab R. A., Shoman M., "Gene Expression Based Cancer Classification," *Egyptian Informatics Journal*, vol. 18, no. 3, pp. 151-159, 2017.
[19] Li K., Zhou, G., Zhai, J., Li F., Shao M., "Improved PSO_AdaBoost Ensemble Algorithm for Imbalanced Data," *Sensors*, vol. 19, no. 6, pp. 1-18, 2019.
[20] Kamran K., Kiana J. M., Mojtaba H., Sanjana M., Laura B., Donald B., "Text Classification Algorithms: A Survey," *Information MDPI*, vol. 10, no. 150, pp. 2-68, 2019
[21] Mariangela B., Eric O., William A. D., Monica B., Yaw A., Guofa Z., Joshua H., Ming L., Jiabao X., Andrew G., Joseph F., Guiyun Y., "RNA-seq analyses of changes in the Anopheles gambiae transcriptome associated with resistance to pyrethroids in Kenya: identification of candidate-resistance genes and candidate-resistance SNPs," *Parasites and Vector*, vol. 8, no. 474, pp. 1-13, 2015.
[22] James G., Witten D., Hastie T., Tibshirani R., "An introduction to statistical learning with application in R," *New York (NY): Springer*, 2013.
[23] Nagi S. and Bhattacharyya D. K., "Classification of Microarray Cancer Data Using Ensemble Approach," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 2, pp. 159-173, 2013.
[24] Sarah M., Ahmed I. S., Labib M. L., "Classification Techniques in Gene Expression Microarray Data," *International journal of Computer Science Mobile Computing*, vol. 7, no. 11, pp. 52-56, 2018
[25] Guzman E., El-halaby M., Bruegge B., "Ensemble Methods for App Review Classification: An Approach for Software Evolution," *2015 30th IEEE/ACM Int. Conference of Automative Software Engineering*, Lincoln, NE, pp. 771-776, 2015.
[26] Ren Y., Suganthan P. N., Srikanth N., "Ensemble methods for wind and solar power forecasting: A state-ofthe-art review," *Reneweable Sustainable Energy Revolution*, vol. 50, pp. 82-91, 2015.
[27] Flennerhag S., "Machine Learning Ensemble," 2017. [Online]. Available: http://flennerhag.com/2017-04-18-introduction-to-ensembles/,
[28] Tsai C. F., Hsu Y. F., Yen D. C., "A comparative study of classifier ensembles for bankruptcy prediction," *Application Soft Computing Journal*, no. 24, pp. 977-984, 2014.
[29] Mayr A., Binder A., Gefeller O., Schmid M., "The Evolution of Boosting Algorithms from Machine Learning to Statistical Modelling," *Methods Informatics and Medicine*, vol. 53, no. 6, pp. 419-427, 2014.
[30] Nisioti A., Mylonas A., Yoo P. D., Member S., Katos V., "From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3369-3388, 2018.
[31] Hafizah S., Ariffin S., Muazzah N., Latiff A., Khairi M. H. H., Ariffin S. H. S., et al., "A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN)," *Engineering, Technology and Applied Science Research*, vol. 8, no. 2, pp. 2724-2730, 2018.
[32] Arowolo M. O., Abdulsalam S. O., Isiaka R. M., Gbolagasde K. A.," A Comparative Analysis of Feature Selection and Feature Extraction Models for Classifying Microarray Dataset," *Computing and Information System*, vol. 22, no. 2, pp. 29-38, 2018.

BIOGRAPHIES OF AUTHORS

Arowolo Micheal Olaolu, is a faculty of the Department of Computer Science at Landmark University, Omu-Aran Nigeria. He holds a Bachelor Degree from Al-Hikmah University, Ilorin, Nigeria and a Masters Degree from Kwara State University, Malete Nigeria, he is presently a PhD Student of Landmark University, Omu-Aran Nigeria. His area of research interest includes Machine Learning, Bioinformatics, Datamining, Cyber Security and Computer Arithmetic. He has published widely in local and international reputable journals, he is a member of IAENG, APISE, SDIWC, and an Oracle Certified Expert.

Marion Olubunmi Adebiyi, is a faculty of the Department of Computer Science at Landmark University, Omu-Aran, Nigeria. She holds a B.Sc Degree from University of Ilorin, Ilorin Nigeria. She had her M.Sc and Ph.D Degree in Computer Science from Covenant University, Nigeria respectively. Her research interests include, Bioinformatics of Infectious (African) Diseases/ Population, Organism's Inter-pathway analysis, High throughput data analytics, Homology modellin and Artificial Intelligence. She has published widely in local and international reputable journals She is a member of Nigerian Computer Society (NCS), the Computer Registration Council of Nigeria (CPN) and IEEE member.

Adebiyi, Ayodele Ariyo, is a faculty and former Head of Department of Computer and Information Sciences, Covenant University, Ota Nigeria. He is currently the Head of Department of Computer Science at Landmark University, Omu-Aran, Nigeria, a sister University to Covenant University. He holds a BSc degree in Computer Science and MBA degree from University of Ilorin, Ilorin Nigeria. He had his MSc and PhD degree in Management Information System (MIS) from Covenant University, Nigeria respectively. His research interests include, application of soft computing techniques in solving real life problems, software engineering and information system research. He has successfully mentored and supervised several postgraduate students at Masters and PhD level. He has published widely in local and international reputable journals. He is a member of Nigerian Computer Society (NCS), the Computer Registration Council of Nigeria (CPN) and IEEE member.

Charity Aremu, is a faculty at the Department of Agriculture and the Dean, School of Postgraduate Studies, Landmark University, Omu-Aran, Nigeria. Charity does research in Crop Environment and Improvement, Plant Breeding, Ecology, Evolutionary Biology and Genetics. She is an International scholar with scholarly publications.