Performance Evaluation of Machine Learning Models For Cervical Cancer Prediction

Florence David Onyirimba[1], Abraham Deme[1], Gideon Dadik Bibu[1,5], Olanrewaju A. Olalekan[2],

Olugbenga Silas[3,4], Maduchi Dauda Ayuba[3,4], Jonah Musa[3,4], Maryam Jemila Ali[3,4], Anejo-Okopi Joseph[6]

[1]Computer Science Department, University of Jos, Jos Nigeria

[2]Firstplace Technicals, Kaduna

[3]Department of Medical Science, University of Jos, Jos Nigeria

[4]Jos University Teaching Hospital, Jos Nigeria

[5]Higher College of Technology UAE

[6]Department of Microbiology, University of Jos, Jos Nigeria

ARTICLE INFO	ABSTRACT
Published Online: 25 November 2022	Cervical cancer is exclusively an anatomy of the female genitals involving the cervix and is the common cancer type that appears in all age women groups and the most common cause of death associated with cancer in gynecological practice, yet it is almost completely preventable if precancerous lesions are identified and treated promptly. The need to develop a quick, cheap and efficient method to diagnose a precursor lesion in an environment with high burden of the diseases with a view of reducing the burden of the disease motivated the need to apply Machine Learning (ML) technique towards cancer prediction. The primary objective of the study was to develop a ML model that can predict the occurrence of cervical cancer with a higher degree of accuracy. The cervical cancer dataset used in this study was obtained from Jos University Teaching Hospital (JUTH) and Aids Prevention Initiative in Nigeria (APIN). Several ML techniques were considered which includes Ensemble Bagged Tree, Fine Gaussian SVM, Cubic SVM, Fine Tree, Quadratic SVM, Medium Gaussian SVM, Ensemble Boosted Tree, Ensemble Rusboosted Tree, Medium Tree, Linear SVM, Corase Gaussian SVM and Coarse Tree algorithm. The study shows that Ensemble Bagged Tree and Fine Gaussian SVM gives a higher cervical cancer predictive accuracy of 99.7 percent and 99.6 percent respectively as the best performing predictive models, followed by Cubic SVM and Fine Tree with 98 percent and Fine Tree with 96.6 percent cervical cancer predictive accuracy respectively. The performance evaluation shows that Ensemble Bagged Tree and Fine Gaussian SVM perform excellently well in distinguishing and predicting the cervical classes correctly with the best prediction accuracy.
Corresponding Author: **Olugbenga Silas,** **Maduchi Dauda Ayuba,** **Jonah Musa, Maryam** **Jemila Ali**	

1. INTRODUCTION

Machine learning (ML), is a branch of Artificial Intelligence, that relates the problem of learning from data samples to the general concept of inference (Bishop, 2006). It has also been proven as an interesting area in biomedical research with many applications, where an acceptable generalization is obtained by searching through an n-dimensional space for a given set of biological samples, using different techniques and algorithms (Niknejad and Petrovic, 2013). The main objective of ML techniques is to produce a model which can be used to classify, predict or estimate data sets. To this end, ML technique of prediction has become an interesting area in

biomedical research with many applications in prediction of several medical incidences such as cancer (Niknejad and Petrovic, 2013).

According to WHO (2018), cancer is a large family of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the human body such as the prostate, breast, skin, eye, tongue and cervix among others (Cancer Council, 2019). Cancer in female, known as cervical cancer arises from the cervical epithelium, predominantly from the squamous epithelial cells and is the most common cancer in female and main cause of morbidity and mortality

in Africa and ranks first among all cancers seen in Nigeria (Birbrair et al., 2014).

An estimated 95% of women in developing countries have never been screened for cervical cancer and over 80% of women newly diagnosed with cervical cancer live in developing countries such as Nigeria; most are diagnosed when they have advanced case of the disease. Though this cancer type is preventable, with laudable, effective screening program, it continues to take a negative toll on developing countries like Nigeria. Hence, the need to develop a quick, cheap and efficient method to diagnose a precursor lesion in an environment with high burden of the diseases with a view of reducing the burden of the disease through cancer prediction. The use of ML to predict the occurrence of cancer has been studied by several authors such as Cruz and Wishart (2006); Park et al. (2013); Chih-Jen et al. (2014) and Kourou et al. (2015). However, there is scarce study in current extant literature on the prediction of cervical cancer using various ML techniques.

2. LITERATURE REVIEW

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data as shown in Figure 1. This study seek to use classification model to predict the occurane of cervical cancer.
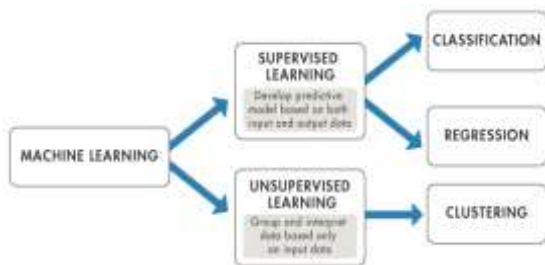

Figure 1: Machine Learning Technique

The application of ML techniques has been a constant over the last couple of decades. It has improved extracted knowledge by learning from result and has the ability to create new valuable information. The growing consciousness of the potential to predict an individual's future risk of cancer has resulted in the development of various algorithms Chih-Jen et al (2014). Such algorithms target to improve the ability of policy makers, doctors and patients to make rational decisions about behavior modification or surveillance, with the expectation that this activity will lead to overall benefit. Such as, diagnosing cancer at its initial or early stages usually provides the best chance for cure.

Screening programs such Pap smear, VIA, and HPV test need to be complemented, using quicker modalities, to ascertain the premalignant lesions spectrum, when it is possible to treat the disease before it becomes untreatable. Figure 2.2 presents some of the risk factors associated with cervical cancer.

Figure 2: Risk factors for cervical cancer (Ruszel *et al.,* 2019).

3. METHODOLOGY

The methodology utilized in this study is depicted in Figure 3. It covers data collection from source data, cleaning/ screening of the collected data to the methodology utilized in the development of the ML model. Figure 3 presents a logical procedure involved in the execution of the model development.
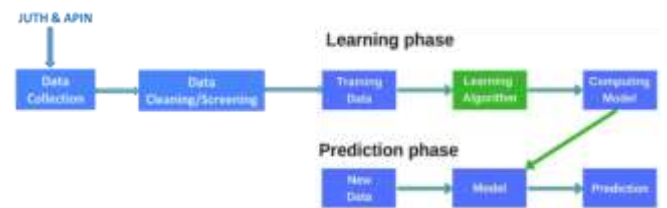

Figure 3: Model Development Procedure

The software used in the course of the study are MATLAB © 2018, MS Excel © 2019 and Filemaker Pro. The cervical cancer datasets used in this study were collected from Jos University Teaching Hospital (JUTH) and Aids Prevention Initiative in Nigeria (APIN), having obtained ethical clearance from the appropriate authority from JUTH. The cervical cancer data collected are pap smear (cytology), containing several features that require cleaning and screening. The dataset collected from the source database were examined for potential errors such as incompleteness, missing information, noise, etc. The collected cervical cancer data was cleaned by removing missing data and duplicates, as well as unwanted information from the raw data. The important features selected for the model development are predictors (Age, Parity, STD History, Number of Sexual Partners, Age at First Sexual Intercourse, Smoking History, Alcohol History and Contraception) and Response (All Diagnosis). After pre-processing and cleaning, the dataset utilized comprises of 15 features (cervical cancer risk factors) used as predictor and 1 target used as response. Detail of the selected data features for model development are presented in Table 1. A 5-fold cross validation was employed to evaluate the predictive model that split the dataset into 5 equally sized partitions called folds, which is then used in training the

model on 4 parts of the dataset, and then evaluate them with the 5th (untampered) part. The predictors and responses were specified and the ML algorithm for all Decision Trees, all Support Vector Machines (SVM), K-Nearest Neighbours, Ensembles, Discriminant Analyses and Logistic Regression models were used to train the data using cross validatiom method to avoid over-fitted or under-fitted model. The accuracy of each of the models were also determined in order to check the accuracy of each and determine the model with the highest accuracy.

Table 1: Selected features for model development

Features	Variables	Data Type
Predictors (Input)	Age	Numerical
	Parity	Numerical
	STD History	Categorical
	Number of Sexual Partners	Numerical
	Age at First Sexual Intercourse	Numerical
	Smoking History	Categorical
	Alcohol History	Categorical
	Virginal Infection	Categorical
	Circumcision	Categorical
	Age at First Menstrual cycle	Numerical
	Household	Numerical
	Age at first child Birth	Numerical
	Last delivery type	Categorical
	Number of virginal deliveries	Numerical
	Contraceptives	Categorical
Response (Target)	All Diagnosis	Categorical

4. RESULTS AND DISCUSSION

The result obtained from the prediction of cervical cancer using Various ML techniques obtained from this study is discussed below.

4.1 Machine Learning Algorithm for the Prediction of Cervical Cancer

Several ML classifier techniques (Decision Trees, Support Vector Machines (SVM) and Ensembles) were considered in the study and the accuracy of each of the model was also determined, in order to establish the model with the highest accuracy. **Table 2.** presents the accuracy of the various ML classifier techniques considered in this study.

Table 2 shows the various ML algorithms that were trained in this study with the corresponding model accuracy. Fine Gaussian SVM and Ensemble Bagged Tree gives the highest accuracy of 99.6% and 99.7% respectively, followed by Cubic SVM and fine Fine Tree with 98% and 96.6% respectively. The highest accuracy of 99.7% observed with

Bagged Tree Ensemble ML technique could be due to the combination of predictions from several base models (particularly Random Forest Bag and Decision Tree learners) to produce one optimal predictive model with high accuracy (Dawngliani *et al.*, 2020). This value is higher than 93.33% accuracy reported for SVM algorithm for cervical cancer prediction by Asadi *et al.* (2020), using 145 data samples and 23 features per sample and also, higher than 93.33% accuracy reported for SVM algorithm for pancreatic cancer prediction by Bodkhe (2017), using real genomic data having 22,763 samples and 154 features per sample.

Table 2: ML classifiers techniques accuracy

S/No.	ML Techniques	Accuracy (%)
1	Ensemble Bagged Tree	99.7
2	Fine Gaussian SVM	99.6
3	Cubic SVM	98
4	Fine Tree	96.6
5	Quadratic SVM	93.2
6	Medium Gaussian SVM	90.7
7	Ensemble Boosted Tree	87.1
8	Ensemble Rusboosted Tree	77.7
9	Medium Tree	77.2
10	Linear SVM	69.2
11	Corase Gaussian SVM	65.4
12	Coarse Tree	61.8

The second highest accuracy of 99.6% observed with Fine Gaussian SVM could be attributed to the fact that Fine Gaussian SVM makes finely detailed distinctions between classes by setting kernel scale to sqrt(P)/4 in order to maximize the distance between the decision hyperplane and the nearest data point, called the marginal distance which is in turn used to finding the best-suited hyperplane (Kourou *et al.*, 2015; Ude, 2019), and also because SVMs are generally more suitable for non-linear classification problems (Bazazeh and Shubair, 2016). The obtained accuracy of 99.6% for Fine Gaussian SVM in this study is slightly higher than 98% for SVM and 98.6% for CNN (Zhang *et al.*, 2018) and higher than 85% for Genetic Algorithms (Aličković and Subasi, 2017), 90 – 95% for Convolutional Neural Network (CNN) (Bora *et al.*, 2016), and 93.671% for SVM (Al-Wesabi *et al.*, 2018) reported for cervical cancer prediction in previous studies. Also, 98% accuracy obtained for Cubic SVM in this study is higher than several accuracies reported in previous studies.

Also, Table 2 shows that Fine Tree, Quadratic SVM, Medium Gaussian SVM and Ensemble Boosted Tree give relatively high accuracies of 96.6%, 93.2%, 90.7% and 87.1% respectively. This could be due to the fact that, the Fine Tree model has high flexibility, with the ability to make many good distinctions between classes by breaking complex decision-making process into a simpler set of decisions that can easily be interpreted (Mirzajani1 and Salimi, 2018). The

accuracy obtained for the Fine Tree (96.6%) model is higher than 95.55%, 95.55%, 95.45% and 93.33% accuracy reported for QUEST Tree, C&R Tree, Radial Basis Function-ANNs and SVM model by Asadi *et al.* (2020) for cervical cancer prediction using ML algorithms with the best performing model to be QUEST Tree.

The accuracy obtained for the Fine Tree (96.6%) model is comparable to 95.69 – 95.93% for Voting with and without PCA and 95.93 96.62 – 96.73% for Voting with and without PCA model before SMOTE and after SMOTE for cervical cancer prediction (Alsmariy *et al.*, 2020). However, the model accuracy obtained in this study for Fine Tree (96.6%), Quadratic SVM (93.2%), Medium Gaussian SVM (90.7%) and Ensemble Boosted Tree (87.1%) were all higher than 73.33% for Naïve Bayes, 75.87% for Functions-Based Logistic SMO, 74.59% for Lazy-Based LWL, 77.97% for Meta-Based Iterative Classifier Optimizer, 69.11% for Rules-Based Decision Table and 77.97% for Trees-Based Decisions reported b,y Singh and Sharma (2019) for cervical cancer prediction using UCI data repository, and also higher than 85% for Random Forests (RF) and 87.50% for SVM reported by Abdullah *et al.* (2018) using cervical cancer dataset obtained from the Gynecologic Oncology Group Tissue Bank (PA, USA), based on gene expression profiling data.

From Table 2 the accuracy obtained in this study for Ensemble Rusboosted Tree, Medium Tree, Linear SVM, Corase Gaussian SVM and Coarse Tree models are 77.7%, 77.2%, 69.2%, 65.4% and 61.8% respectively for cervical cancer prediction. These values are lower compared to values reported in previous studies (Bora *et al.*, 2016; Bodkhe, 2017; Al-Wesabi *et al.*, 2018; Mirzajani1 and Salimi, 2018; Zhang *et al.*, 2018; Abdullah *et al.*, 2018; Singh and Sharma, 2019; Asadi *et al.*, 2020; Alsmariy *et al.*, 2020; Dawngliani *et al.*, 2020). On the basis of model predictive accuracy, Ensemble Bagged Tree and Fine Gaussian SVM with 99.7% and 99.6% accuracies respectively gives the best performing predictive model for cervical cancer prediction, followed by Cubic SVM and Fine Tree with 98% and Fine Tree with 96.6% cervical cancer predictive accuracy respectively.

4.2 Performance Evaluation of Models Prediction

The performance of each model was evaluated based on True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), using confusion matrix plot to understand the ML classifier model performance in each class and areas where the classifier performed poorly or better. The confusion matrix was calculated using the predictions on the held-out observations (cross validation technique). Figures 4 to 6 show the confusion matrix plot for the established models with the highest accuracy (Ensemble Bagged Tree, Fine Gaussian SVM and Cubic SVM).

Figure 4 shows the confusion matrix plot for Ensemble Bagged Tree cervical cancer predictive model. The rows show all the true classes while the column shows the predicted classes. From the top row, over 99% of Negative

for IL Malignancy class are correctly classified while less than 1% are falsely classified as Positive and Repeat After 6 Month, representing that the true positive rate for correctly classified Negative for IL Malignancy is over 99%. Likewise, the second row shows that less than 1% of Positive class are wrongly classified as Negative for IL Malignancy while over 99% was correctly classified as Positive class and none was falsely classified as Repeat After 6 Month, signifying that the true positive rate for correctly classified Positive class is over 99%. Equally, the third row shows that 1% of Repeats after 6 Months are wrongly classified as Negative for IL Malignancy while none was falsely classified as Positive class and 99% are correctly classified as Repeat After 6 Month class, demonstrating that the true positive rate for correctly classified Repeat After 6 Month class is 99%. Generally, the true positive rate for Negative for IL Malignancy and Positive classes are over 99% while that of Repeat After 6 Month class is 99%, which further corroborate the higher cervical cancer predictive accuracy obtained with Ensemble Bagged Tree model.
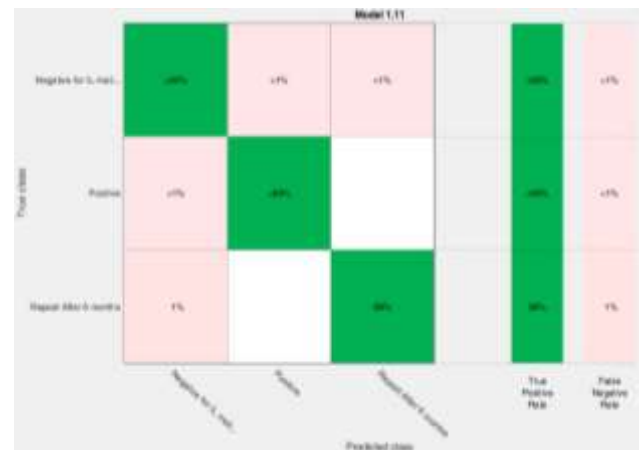

Figure 4: Confusion Plot for Ensemble Bagged Tree Model

Figure 5 shows the confusion matrix plot for Fine Gaussian SVM cervical cancer predictive model. The rows show all the true classes while the column shows the predicted classes. From the top row, 100% of Negative for IL Malignancy class are correctly classified while none was falsely classified as Positive and Repeat After 6 Months, indicating that the true positive rate for correctly classified Negative for IL Malignancy is 100%. Also, the second row shows that 1% of Positive class are wrongly classified as Negative for IL Malignancy while 99% was correctly classified as Positive class and none was falsely classified as Repeat After 6 Month, indicating that the true positive rate for correctly classified Positive class is 99%. Similarly, the third row shows that 1% of Repeat After 6 Month class are wrongly classified as Negative for IL Malignancy while none was falsely classified as Positive class and 99% was correctly classified as Repeat After 6 Month class, indicating that the true positive rate for correctly classified Repeat After 6 Month class is 99%. Overall, the true positive rate for Negative for IL Malignancy, Positive and Repeat After 6 Month class are 100%, 99% and

99% respectively, which further confirms the higher accuracy of the Fine Gaussian SVM cervical cancer predictive model.
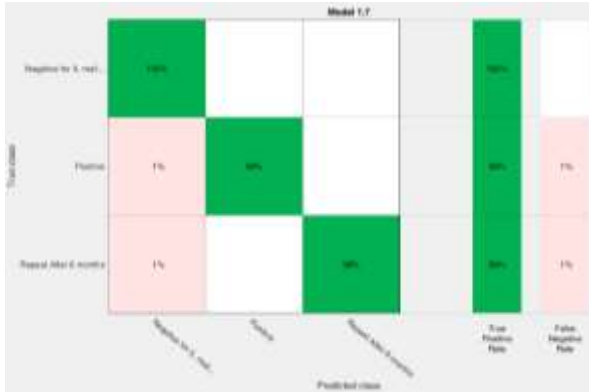

Figure 5: Confusion Matrix Plot for Fine Gaussian SVM Model

Figure 6 shows the confusion matrix plot for Cubic SVM cervical cancer predictive model. From the top row, 99% of Negative for IL Malignancy class are correctly classified while 1% are falsely classified as Positive and Repeat After 6 Month class, indicating a true positive rate of 99% correctly classified as Negative for IL Malignancy. In same manner, the second row shows that 1% of Positive class are wrongly classified as Negative for IL Malignancy and Repeat After 6 Month while 99% was correctly classified as Positive class, suggesting that the true positive rate for correctly classified Positive class is 99%. Equally, the third row shows that 3% of Repeat After 6 Month class are wrongly classified as Negative for IL Malignancy while none was falsely classified as Positive class and 97% are correctly classified as Repeat After 6 Month class, signifying a true positive rate of 97% for correctly classified Repeat After 6 Month class. Hence, the true positive rate for Negative for IL Malignancy, Positive and Repeat After 6 Month class are 99%, 99% and 97% respectively which also conform with the high predictive accuracy obtained for Cubic SVM cervical cancer predictive model.

Furthermore, Receiver Operating Characteristic (ROC) and Area Under Curve (AUC) were used to visualize how well a ML classifier model performs (Majnik and Bosnic, 2013; Yang and Berdine, 2017).
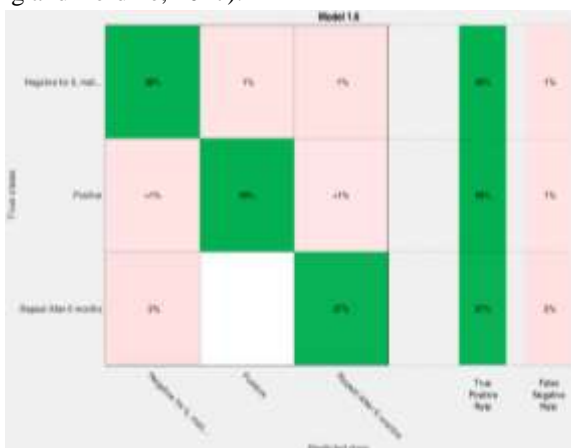

Figure 6: Confusion Plot for Cubic SVM Model

The ROC curve was used as an evaluation metric to examine how well the models were able to separate the 'signal' from the 'noise' at various threshold values of true positive rate (TPR) against false positive rate (FPR) while the AUC was used to measure the ability of the various ML classifier models to distinguish between classes. Figure 7 – 9 shows the ROC and AUC for Ensemble Bagged Tree, Fine Gaussian SVM and Cubic SVM models.
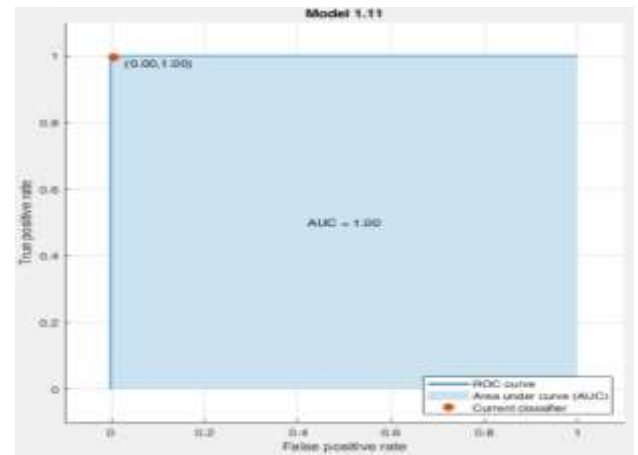

Figure 7: ROC and AUC Curve for Ensemble Bagged Model

Figure 7. shows the ROC and AUC for Ensemble Bagged model in predicting cervical cancer. The classifier marker (0.00, 1.00) shows a 0% FP rate which implies that no observations were incorrectly classified and a 100% TP rate which implies that all observations are correctly classified, indicating a perfect prediction ability. Equally, the AUC for Ensemble Bagged model was found to be 1, signifying an outstanding ability of the model to classify the classes accurately as an AUC value closer to 1 portends excellent classification ability (Yang and Berdine, 2017; Teodorescu, 2017). Hence the ROC and AUC further confirms that Ensemble Baggged model performs excellently in predicting cervical cancer.
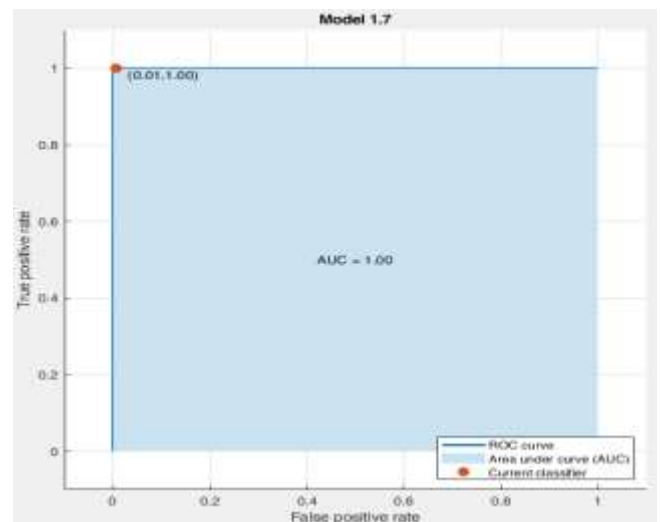

Figure 8: ROC and AUC Curve for Fine Gaussian SVM Model

Figure 8 shows the performance of the Fine Gaussian SVM Model in predicting cervical cancer and the marker on the plot shows the values of the FP rate and the TP rate for the Fine Gaussian SVM classifier model. From Figure 8, the classifier marker (0.01, 1.00) indicates a FP rate of less than 1% observations incorrectly predicted and a TP rate of almost 100% observations are correctly predicted. Conversely, the AUC for Fine Gaussian SVM Model was found to be 1, indicating an outstanding ability of the model to discriminate the classes accurately because the closer the AUC is to 1, the more perfect the performance of the model in distinguishing the classes accurately (Yang and Berdine, 2017; Teodorescu, 2017). Hence the ROC and AUC further confirms that Fine Gaussian SVM model also performs excellently well in predicting cervical cancer.
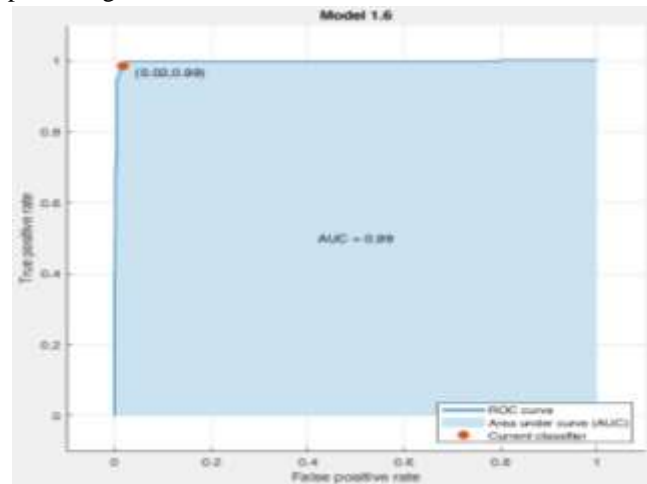

Figure 9: ROC and AUC Curve for Cubic SVM Model

Figure 9 shows the ROC and AUC for Cubic SVM model in predicting cervical cancer. The classifier marker (0.02, 0.99) implies a 2% FP rate which suggests that almost 2% of all observations were incorrectly classified and about 99% TP rate which implies that almost 99% of observations are correctly classified, indicating a better prediction ability. The AUC for Cubic SVM model was determined to be 0.99, signifying also a better ability of the model to distinguish the classes accurately as an AUC value closer to 1 signifies better classification ability (Yang and Berdine, 2017; Teodorescu, 2017). Hence the ROC and AUC further confirms that Cubic SVM model give a good cervical cancer prediction ability.

Based on the ROC and AUC, Ensemble Bagged Tree and Fine Gaussian SVM model performs excellently well in distinguishing the classes while a, the Cubic SVM model also shows a good performance in predicting cervical cancer. Comparatively, on the basis of accuracy, confusion matrix and ROC and AUC, Ensemble Bagged Tree and Fine Gaussian SVM model perform excellently well in distinguishing and predicting the classes correctly and therefore gives the best predictive model for cervical cancer prediction while also Cubic SVM and Fine Tree performs better in predicting cervical cancer compared to some previously reported ML model. Therefore, Ensemble Bagged

Tree and Fine Gaussian SVM model performs excellently in predicting cervical cancer with higher accuracy.

5. CONCLUSION

The study examines the development and implementation of ML classification model that can predict the occurrence of cervical cancer with a higher degree of accuracy using cervical cancer pap-smear (cytology), dataset obtained from the database sponsored by Operation Stopped Cervical Cancer Gynecologist Clinic in the Gynecology unit in collaboration with APIN in JUTH. The reviewed literature shows that several authors have thoroughly examined cervical cancer stage prediction using different ML techniques with cancer data from different sources. However, there is scarce work on cervical cancer prediction using ML technique that utilizes data obtained from Nigeria while a lot of the study focuses on cervical cancer developmental stages and not primarily on its detection.

Several techniques were examined which include Ensemble Bagged Tree, Fine Gaussian SVM, Cubic SVM, Fine Tree, Quadratic SVM, Medium Gaussian SVM, Ensemble Boosted Tree, Ensemble Rusboosted Tree, Medium Tree, Linear SVM, Corase Gaussian SVM and Coarse Tree. All ML technique examined are capable of predicting cervical cancer using pap-smear dataset involving 3 response classes (positive, negative and repeat) with at least 14 features and a combination of numerical and categorical data except discriminant analysis, logistic regression which are suitable for only two possible classes and KNN.

The study shows that Ensemble Bagged Tree and Fine Gaussian SVM gives a higher cervical cancer predictive accuracy of 99.7% and 99.6% respectively as the best performing predictive model, followed by Cubic SVM and Fine Tree with 98% and Fine Tree with 96.6% cervical cancer predictive accuracy respectively. The performance evaluation further based on TP, TN, FP and FN using confusion matrix plot, further confirms the higher predictive accuracy of Ensemble Bagged Tree and Fine Gaussian SVM model in predicting cervical cancer, followed by Cubic SVM cervical cancer predictive model. The ROC and AUC also showed that Ensemble Bagged Tree and Fine Gaussian SVM model performs excellently well in distinguishing the classes followed by Cubic SVM model which also shows a good performance in predicting cervical cancer. Comparatively, on the basis of accuracy, confusion matrix and ROC and AUC, Ensemble Bagged Tree and Fine Gaussian SVM model perform excellently well in distinguishing and predicting the classes correctly, and therefore gives the best predictive model for cervical cancer prediction while also Cubic SVM and Fine Tree performs better in predicting cervical cancer compared to some previously reported ML models. Therefore, Ensemble Bagged Tree and Fine Gaussian SVM model performs excellently in predicting cervical cancer with higher accuracy.

ACKNOWLEDGEMENT

REFERENCES

1. Abdullah, A. A., Abu-Sabri, N. K., Khairunizam, W., Zunaidi, I., Razlan, Z. M., and Shahriman, A. B. (2019). Development of predictive models for cervical cancer based on gene expression profiling data. IOP Conf. Series: Materials Science and Engineering 557: 1 – 8. doi:10.1088/1757-899X/557/1/012003.

2. Alam, T. M., Khan, M. M. A., Iqbal, M. A. Wahab, A. and Mushtaq, M. (2019). Cervical Cancer Prediction through Different Screening Methods using Data Mining. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 10, No. 2: 388 – 396.

3. Aličković, F and Subasi, A. (2017). "Breast cancer diagnosis using GA feature selection and Rotation Forest." *Neural Computing and Applications*, Vol. 28: 753 – 763.

4. Alsmariy, R., Healy, G. and Abdelhafez, H. (2020). Predicting Cervical Cancer using Machine Learning Methods, *International Journal of Advanced Computer Science and Applications, (IJACSA)*, Vol. 11, No. 7, p. 173 – 183.

5. Al-Wesabi, Y. M. S., Choudhury, A. and Won, D. (2018). Classification of Cervical Cancer Dataset. *Proceedings of the 2018 IISE Annual Conference*, Binghamton University, USA, p. 1 – 6.

6. Asadi, F., Salehnasab, C. and Ajori, L. (2020) Supervised Algorithms of Machine Learning for the Prediction of Cervical Cancer. *Journal of Biomedical and Physical Engineering*. 10(4):513-522. doi: 10.31661/jbpe.v0i0.1912-1027.

7. Bazazeh, D., and Shubair, R. (2016). Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis, p. 2–5.

8. Birbrair, A., Zhang, T., Wang, Z. M., Messi, M. L., Olson, J. D., Mintz, A. and Delbono, O. (2014). Type-2 pericytes participate in normal and tumoral angiogenesis. *American Journal of Physiology & Cell Physiology*, 307 (1): 25-38.

9. Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

10. Bodkhe, A. (2017), "Predicting Pancreatic Cancer Using Support Vector Machine" (2017). Master's Projects. 535. DOI: https://doi.org/10.31979/etd.9w5j-j4ax. Available from https://scholarworks.sjsu.edu/etd_projects/535, accessed: 20/02/2021.

11. Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K. and Das, A. K. (2016). "Pap Smear Image Classification Using Convolutional Neural Network", *Tenth Indian Conference on Computer Vision, Graphics and Image Processing*.

12. Cancer Council (2019). Understanding Cervical Cancer: A Guide for People with Cancer, Their Families and Friends, Cancer Information: 4 – 68.

13. Cruz, J. A. and Wishart, D. S. (2006). Applications of Machine Learning in Cancer Prediction and Prognosis, *Cancer Informatics*, 2: 59– 78.

14. Dawngliani, M. S., Chandrasekaran, N., Lalmuanawma, S. and Tangkhanhau, H. (2020). Prediction of Breast Cancer Recurrence Using Ensemble Machine Learning Classifiers. Springer Nature Switzerland AG, p. 232 – 243. http://doi.org/10.1007/978-3-030-46828-620.

15. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction, *Computational and Structural Biotechnology Journal*, 13: 8–17.

16. Majnik, M. and Bosnic, Z. (2013). ROC Analysis of Classifiers in Machine Learning: A Survey. Intelligent Data Analysis. DOI: 10.3233/IDA-130592.

17. Mirzajani1, S. S. and Salimi, S. (2018). Prediction and Diagnosis of Diabetes by Using Data Mining Techniques. Avicenna Journal of Medical Biochemistry. Vol. 6 (1), p. 3 – 7. doi:10.15171/ajmb.2018.02/.

18. Niknejad, A and Petrovic, D. (2013). Introduction to computational intelligence techniques and areas of their applications in medicine. *Med Appl Artif Intell*.: 51.

19. Park, K., Ali, A., Kim, D., An, Y., Kim, M. and Shin, H. (2013). Robust predictive model for evaluating breast cancer survivability. *Engl. Applied Artificial Intelligence*, 26: 2194–205.

20. Siegel, R. L., Miller, K. D. and Jemal, A. (2017). "Cancer statistics, 2017," CA: *A cancer Journal for Clinicians*, vol. 67: 7 – 30.

21. Singh, J. and Sharma, S. (2019). Prediction of Cervical Cancer Using Machine Learning Techniques, *International Journal of Applied Engineering Research*, Vol. 14, No. 11, p. 2570 – 2577. ISSN 0973-4562. http://www.ripublication.com

22. Teodorescu, M. H. (2017). Machine Learning Methods for Strategy Research. Harvard Business School, HBS Working Paper 18-011: 1 – 17.

23. Ude, A. A. (2019). Classification of Breast Cancer using Logistic Regression A Thesis Presented to the Department of Computer Science. M.Sc. Thesis, Department of Computer Science, African University of Science and Technology, p. 10 – 30.

24. World Health Organization (2006), Comprehensive Cervical Cancer Control A guide to essential practice.

25. World health organization (2018) Cancer Fact Sheet, February 2018. Retrieved 21 March 2018. https://en.m.wikipedia.org/wiki/Cancer#cite_ref-WHO2018_2-7. Accessed: 11/11/2018.

26. Yang, S. and Berdine, G. (2017). The Receiver Operating Characteristic (ROC) Curve. *The Southwest Respiratory and Critical Care Chronicles*, 5(19), p. 34 – 36.

27. Zhang, L., Lu, L., Nogues, I., Summers, R. M. and Liu, S. (2018*). DeepPap: Deep Convolutional Networks for Cervical Cell Classification*, arXiv:1801.08616v.