International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

NGBFA Feature Selection Algorithm-based Hybrid Ensemble Classifier to Predict Cervical Cancer

CH Bhavani[1], Dr. Ch. Sarada[2], Dr. A. Jyothi Babu[3], Dr. Gurrampally Kumar[4], Moligi Sangeetha[5], Patlegar Vijay Kumar [6]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Early diagnosis may cure cervical cancer. Researchers have struggled to prediction the disease's course because there are no early indications. Several machine learning methods have predicted CC in the past decade. Ensemble techniques generate and integrate several models for more accurate results. This contrasts with single-classifier prediction. During this research, we established "Robust Model Stacking: A Hybrid Ensemble." This classifier runs a homogeneous classifier-based classifications at the base level, then a heterogeneous ensemble that predicts additional data using majority voting (soft). This study included 858 patients, 32 risk indicator characteristics, and four CC diagnosis test targets. SMOTE oversampling solved the data imbalance problem. For each of the dataset's four goal variables, accuracy, recall, f1-score, precision, and AUC-ROC were used to assess the model. The proposed biopsy approach is 98% accurate, Hinselmann 97%, Schiller 96.09%, and Citology 93%. Ensemble learning improves prediction accuracy and reduces bias and variation in this study.

Keywords: *feature selection, Machine learning, cervical cancer, Hybrid Ensemble, classification*

I. INTRODUCTION

Cervical cancer, the second-most dangerous malignancy for women and the fourth-most deadly worldwide, is the major cause of death for women due to its lack of early symptoms. Early detection may aid treatment [1]. Women get this malignancy in the uterine cervix. Cancer cells divide abnormally and spread to other organs [2, 3]. Malignant cells divide here. Sexual contact spreads this disease. Women who participate in sexual activity are practically guaranteed to contract human papillomavirus (HPV), which will resolve on its own. In some cases, abnormal cell activity can lead to pre-cancerous and malignant cells [4–6]. To lower cervical cancer risk, women aged 30–49 should get screened annually. Machine learning (ML) classifies data using probabilistic, statistical, and optimization methods. Historical data are employed for these techniques. Medical data benefit from this capability. Machine learning algorithms increase accuracy, make emergency decisions, and provide support, making them handy when medical assistants are scarce. Machine-learning-based systems are effective in diagnosing various cancers [9, 10]. Since context is important, no classifier can always perform at its best [11]. Thus, using a single machine learning model to determine a patient's malignancy does not improve accuracy. Each machine-learning model has pros and cons for classification. Some scientists are enhancing classification effectiveness by using multiple classifiers instead of one. Ensemble classification uses specified combination rules to combine classifiers [12]. Over the last decade, cancer researchers have focused on ensemble approaches and methodologies because of their importance. The ensemble technique aims to develop a global model for faster diagnoses, prognoses, and therapies, providing accurate problem solutions, estimators, and predictions [13]. Based on the association criteria, many classifiers are combined. This desire inspired the creation of the ensemble technique. Combining practical and theoretical verification yields the best prediction performance compared to separate models. Several ensemble learning algorithms have been developed for classification and regression to solve real-world situations [14]. Overall, two categories can be used to classify ensemble techniques. Heterogeneous and homogeneous ensemble approaches exist. Homogeneous ensembles use several classifier variants. These options may work. Bagging and boosting work best for homogenous ensembles. A bootstrap aggregator that bases learners on decision trees is "bagging." An acronym describes bagging. It applies to dataset properties and samples. A majority vote will determine the final projection of previously unseen data [15, 16]. Each basic learner will participate simultaneously in training. Heterogeneous ensemble learning uses multiple classification models as basic learners and aggregates their outputs to improve prediction accuracy. Increasing

---

[1] *Senior Assistant Professor, Dept. of CSE, CVR College of Engineering, Hyderabad, Telangana, India.*

[2] *Assoc. Professor, Dept. of CSE, CVR College of Engineering, Hyderabad, Telangana, India.*

[3] *Professor, Dept. of Computer Applications, School of Computing, Mohan Babu University, Tirupati, Andra Pradesh, India.*

[4] *Assistant Professor, IT Dept., JB Institute of Engineering & Technology, Hyderabad, Telangana, India.*

[5] *Senior Assistant Professor, Dept. of CSE, CVR College of Engineering, Hyderabad, Telangana, India.*

[6] *Assistant Professor, English Dept. of Humanities and Sciences, Annamacharya University, Rajampet, Andhra Pradesh, India.*

E-mail: *ch.bhavani@cvr.ac.in , chintala.sarada@cvr.ac.in, jyothibabuaddanki@gmail.com, grk.040@gmail.com , moligisangeetha@cvr.ac.in, sendtopatlegar@gmail.com*

prediction accuracy is the goal. Stacking is a heterogeneous ensemble. Integrating multiple base learners into a single dataset creates a new dataset with meta-level predictions. Next, a meta-level classifier accurately predictions the newly obtained data [17]. Heterogeneous model building involves merging a specific number of models and ensemble models with different parameterizations for predicting, both yielding the same outcome.

II. RELATED WORKS

AI, a promising tool, is being harnessed to diagnose disorders in medicine. It's important to note that no classification scheme is a one-size-fits-all solution. The classifiers, identified through 10-fold cross-validation, have shown promising results in lithofacies classification. The author of the reference [19] suggests a strategy for selecting several powerful learners to develop a learning model. These powerful learners include deep neural networks, SVMs, AdaBoost, and Gaussian processes. This study's ensemble, created via fusion using the sum rule on many classifiers, is a testament to the potential of AI. The authors of [20] used a genetic strategy for ensemble pruning on homogeneous ensembles to get the best twenty models. According to a recent study [21], separating poor performers from basic learners created a diverse, potentially effective class, ensuring that the ensemble only contains the best-maintained classifiers. This optimistic outlook on AI's potential in medical diagnosis is a beacon of hope for pediatric cancer researchers and medical professionals. ROC-AUC measures classifier effectiveness. The homogeneous module includes this module. They combined four models using the sum rule and majority voting to create new training sets. It was suggested that CRISPR-DM [23] be used. This stacking-based model uses KNN, SVM, and decision tree algorithms. The model's performance was compared to individual classifiers. The data revealed that the model outperformed the classifications. Pediatric cancer researchers will benefit from this approach. Deep learning can help diagnose cancer using risk markers [25].

Moreover, supervised auto-encoding produced a 68% AUC. Cerv Detect is a hybrid model that [26] developed to detect cardiovascular disease. This model uses shallow neural networks and random forest classifiers. It was 93.9% accurate. With little data, [27] used mean value replacement and ensemble learning to predict CC risk. This analysis was done to accomplish its purpose. Researchers in [28] employed meta-classification techniques and feature selection methods to analyze the dataset. This assessed the dataset's characteristics. Meta-classifiers were five classifiers. The attribute they chose gave their proposal the lowest error rate. After comparing this method to other data balancing strategies, they chose a stacking algorithm for classification to improve two-stage classifier performance. After comparing all the methods, they chose this one. Classifiers were used to predict plant diseases [31]. Mixing classifiers helped them get high-quality findings. While the CC data may be uneven, earlier studies under the prediction did not reconcile it. Only some articles have examined relevant quality selection. Additionally, data imputation was done. By the usage of spatial data mining, the areas with high effected cervical cancer cases can also be identified [32, 33].

III. PROPOSED MODEL

The model is modified to interpret the dataset, as CC risk factor classifications lack some values. An operationally effective solution is expected, removing or replacing missing values. Using missing value strategies, rows were reduced from 858 to 737, reducing the row count. Equations (1) and (2) utilize Numerical average and mode for categorical attributes to decrease features, replacing missing data with the mode, thereby reducing features over records.

Description of Dataset: The University Sitario de Caracas hospital provided an 858-record CC dataset with 33 features, including patient medical data-generated risk variables and four cancer diagnostic tests. The model aims to increase coefficient of determination predicting accuracy. The dataset was created from a patient survey, but due to patient reluctance to provide personal data, it will be incomplete. Feature selection is crucial as the dataset contains properties that could potentially predict CC. The dataset consists of demographics, medical history, and behavior. The first dataset shows a '1' response, while the '0' response indicates a '0' response as shown n Figure 2. Patients often fear privacy and decline important questions, causing anxiety. The data analysis provides multiple options.

Data pre-processing: The model is modified to interpret the dataset, as CC risk factor classifications lack some values. To improve operational efficiency, missing values are removed or replaced. Value strategies are included when working with massive datasets. Deleted and missing data reduces row count from 858 to 737. Numerical average and mode are used to decrease features, replacing missing data with the mode.

$$mean\ \overline{a} = \frac{1}{T}\left(\sum_{p=1}^{n} a_p\right) \tag{1}$$

$$Mode\ C = K + \left(\frac{v_s + v_{s-1}}{(v_s - v_{s-1}) + (v_s - v_{s+1})}\right) Z \tag{2}$$

Z indicates the model class's length, vs. 1 is its predecessor, and vs. 1 is its successor.

Equation 3's basic imputer employs the mean to fill in numerical feature missing values.

$$B = \frac{1}{n}\left(\sum\nolimits_{i=1}^{n} a_i\right) = (a_1 + a_2 + a_3 + \cdots + a_n)n \tag{3}$$

To include missing data in category features, the Simple Imputer model with mode was used. The two datasets were joined to resolve any missing values.

Data balancing

The cervical cancer risk factor dataset has a severely skewed positive and negative classification distribution. The dataset's observations are predominantly non-malignant, accounting for 96% of the total, while only 4% are malignant. Unequal datasets have received little attention in cervical cancer risk factor dataset investigations. ISMOTE addresses the enormous CC dataset data imbalance. SMOTE undersampling and oversampling techniques are used in the ISMOTE resampling strategy.

Feature Selection Algorithm: NBGFA

The study used correlation analysis to identify redundant features in a dataset, removing duplicate attributes to improve model correctness and avoid bias. The SMOTE synthetic minority over-sampling strategy was employed to prevent overfitting. The dataset was used for training at 80% and testing at 20%, with 1282 records added to ensure equal representation of cancer patients and cancer-free individuals. The NGBFA, combining genetic operation and firefly binary version technology, was used to overcome overfitting [34].

Hybrid strategy stacking robust model

There is no consensus on a categorizing algorithm that always produces the best results. Due to its higher accuracy than classifiers, ensemble learning has become prominent. This method uses resilient model stacking, a hybrid ensemble strategy. This method has the following steps:

Homogenous Ensembling

Step 1: First, divide the N-occurrence dataset into two data sets. These are the training and testing data sets.

Step 2: Use an *'n'* classifier for training data. Perform the first classifier. Repeat this method m times with different classifier settings to get a group of size m with a classifier accuracy over 75%. The goal can only be reached this way.

Step 3: Maximum voting can be used to get predictions from the homogenous ensemble from Step 2.

Step 4: After stages 2 and 3, the remaining classifier algorithms from 2 to n are applied.

Heterogenous Ensembling

Step 5: Apply step 3 predictions to all 'n' classifiers with an accuracy greater than 75% using maximum voting.

Step 6: Step 5 predictions should be used to assess the classifier's accuracy. Figure 4 depicts the intended model architecture.

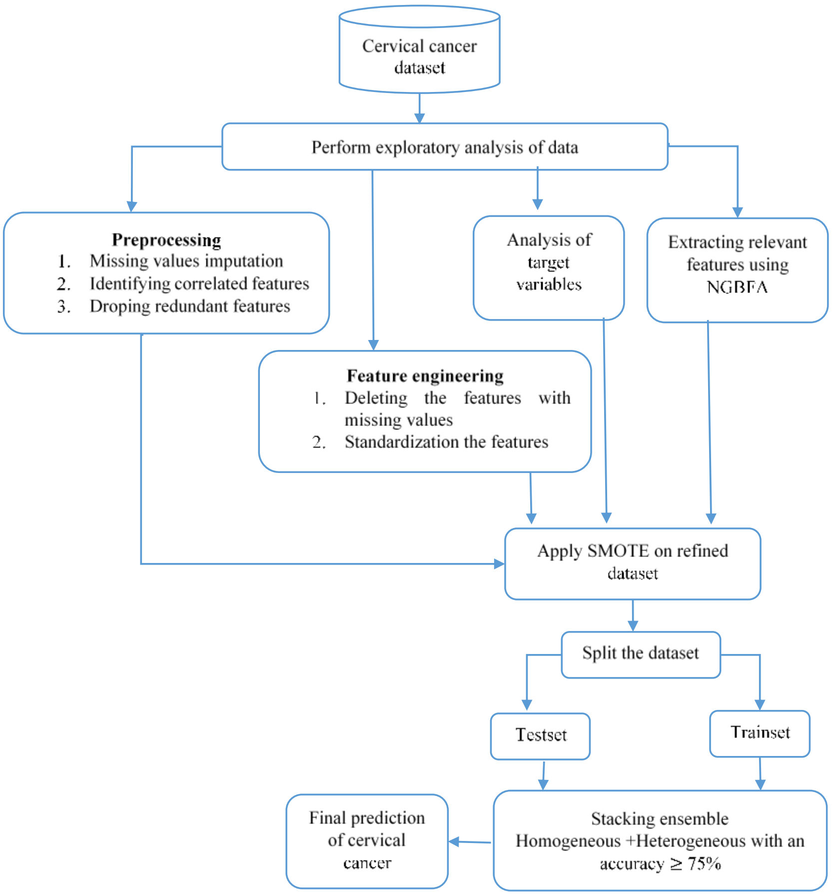

Fig. 1: Architecture of the proposed work

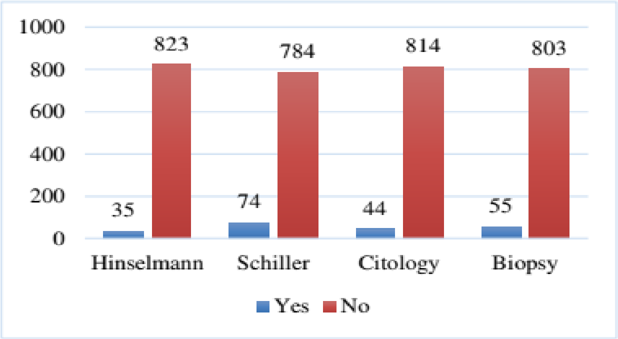

Fig. 2: The original dataset's target variable results in a distribution

Pseudo Code

The proposed model's pseudo code is as follows:

An Algorithm Pseudo Code Hybrid: Consider a Reliable Model Stacking Mechanism

```
Input:
$D$: Files
$N$: instances
$n$: classifiers
Output: In addition to data classification, judgments and forecasts are supplied.
Begin
Step 1: Homogenous Ensemble
    Execute the function for 'n' classifiers if i is between 1 and n.
    Do this for j between 1 and m. For various parameter sets
    Classifier$ij$($D$) should be used.
    Classifier I must be part of an ensemble to exceed 75% accuracy.
Step 2: Heterogenous Ensemble
    The output of I's step 1 grouped classifiers must be paired with a soft voting mechanism with classifier accuracies above 75% to make final predictions. Accurate projections require this.
Step 3: Final model accuracy
End
```

IV. RESULTS AND DISCUSSIONS

Python was used for machine learning model building, with code executed on Google Colab using a 2.65-GHz Intel Core i5 processor and 8GB of system RAM. Eight algorithms were implemented, including KNN, SVM, LR, RF, DT, XGBoost, and AdaBoost. The implementation process was repeated five times with different parameter values. A homogeneous assembly strategy with soft voting achieved level 1 ensemble precision. A heterogeneous ensemble at level 2 used each classifier ensemble's level 1 test data predictions. The hybrid ensemble-based resilient model stacking beats advanced classification methods.

$$Accuarcy = \frac{\mathrm{TP}}{\mathrm{TP+TN+FP+FN}} \quad (4)$$

$$\frac{\text{Sensitivity}}{\text{recall}} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (5)$$

$$\frac{\text{Specificity}}{\text{precision}} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (6)$$

$$\text{F1}-\text{score} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} \quad (7)$$

$$ROC-AUC = 1 - \frac{1}{p^{+}p^{-}} \sum_{a^{+}\epsilon P^{+}} \sum_{a^{-}\epsilon P^{-}} \left(\left(f(a^{+}) < f(a^{-})\right) + \frac{1}{2}\left(f(a^{+}) = f(a^{-})\right) \right) \quad (8)$$

Machine learning metrics alone cannot adequately assess a model's science and impartiality. We use three statistical criteria to evaluate the model's statistical efficacy in this study.

Cohen's kappa: This common agreement measure measures the agreement between actual discoveries and expected outcomes in the categorization problem. According to [35], it measures the parties' agreement.

$$k = \frac{(q_i - q_j)}{(1 - q_j)} \quad (9)$$

Matthews' correlation coefficient (MCC): The MCC evaluates binary class models in machine learning. In essence, it is a -1 to +1 correlation coefficient. [36]: [36] Our MCC calculation method:

$$MCC = \frac{\mathrm{TP \times TN - FP \times FN}}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (10)$$

Comparing NGBFA-HMS with Traditional Ensemble Classifiers

TABLE I. Comparison of NGBFA-MSE biopsies with homogenous and heterogeneous classifiers.

Classifier	Accuracy	Precision	Recall	F1-Score
E – KNN	87.1	89.2	87.6	86.4
E – SVM	93.3	93.3	93.2	93.2
E - LR	59.8	60.01	60.1	59.8
E – RF	96.1	96.2	96.3	96.2
E – DT	88.2	88.3	88.4	88.2
E – XGBOOST	96.4	96.1	96.1	96.2

E-ADABOOST	94.3	94.2	94.2	94.2
Proposed	**98.13**	**97.71**	**98.12**	**97.6**

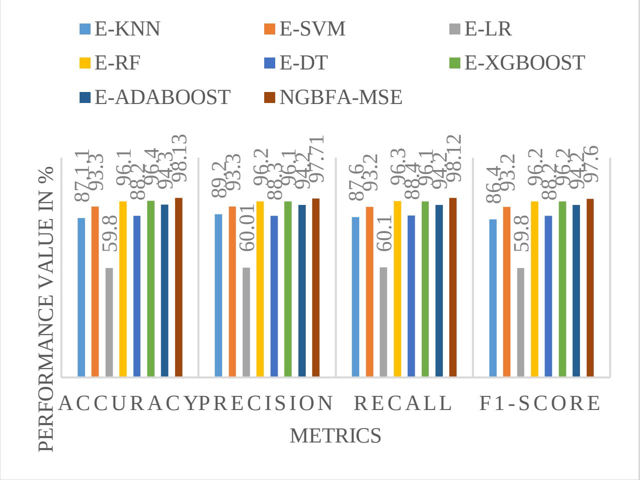

Fig. 3: Comparison of NGBFA-MSE with Homo & Heterogenous Classifiers for Biopsy

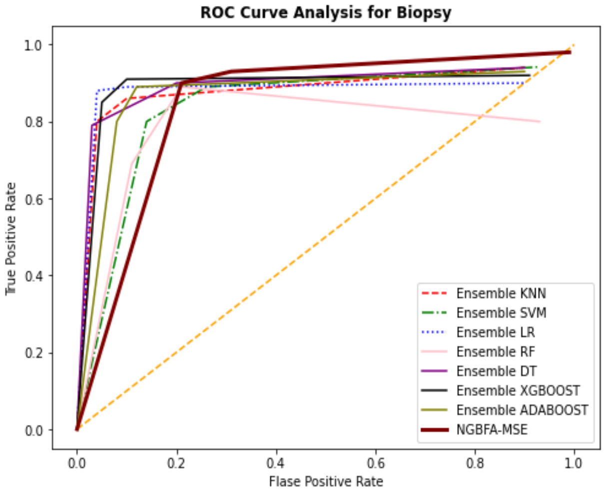

Fig.4: Analysis of ROC Curve with NGBFA-MSE for Biopsy

Comparing NGBFA-HMS with Traditional Ensemble Classifiers

Table II. Comparison of NGBFA-MSE with Hinselmann's homogeneous and heterogeneous classifiers.

Classifier	Accuracy	Precision	Recall	F1-Score
E - KNN	91.02	91.6	91.08	91.4
E - SVM	94.8	95.02	95.01	95
E - LR	73.02	74.7	72.87	73.1
E - RF	94.2	94.9	95.03	95.01
E - DT	92.8	93.6	93.8	94.01
E – XGBOOST	95.8	96.01	96	96
E - ADABOOST	94.9	95.7	95	95.01
Proposed	**97.32**	**96.012**	**97**	**97.01**

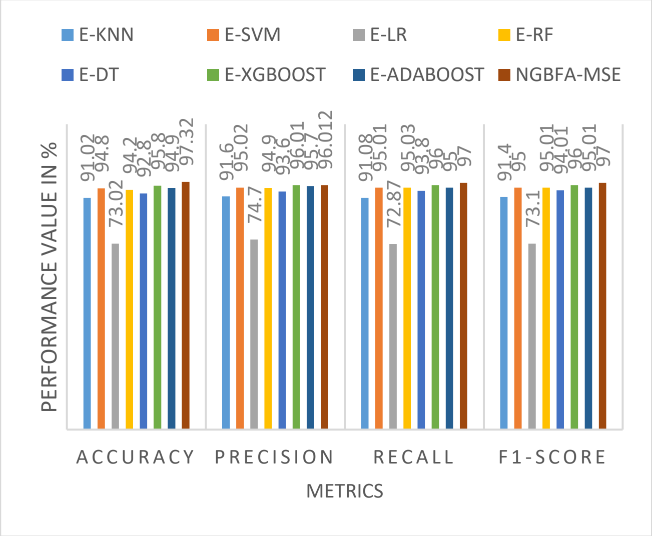

Fig.5: Comparison of NGBFA-MSE with Homo & Heterogenous Classifiers for Hinselmann

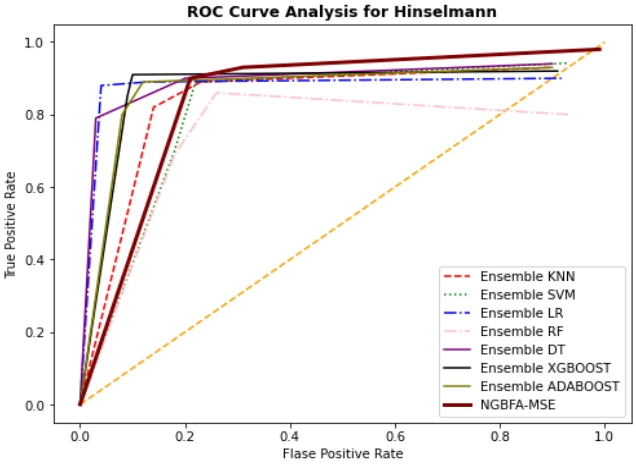

Fig. 6: Analysis of ROC Curve with NGBFA-MSE for Hinselmann

Result Analysis of target feature: Schiller

Analysis of Individual Classifier Performances

Age, count of pregnancies, count of contraceptive years, count of IUD years, number of years smoked, number of packs smoked per year, number of STDs, HIV, vulvo-perineal condylomatosis, and Hepatitis B. Eleven ideal attributes were chosen for the Schiller target feature. We tallied Schiller target variable trial data using the prescribed model and achieved 96% accuracy. Tables 3 and Figure 7 exhibit homogeneous and heterogeneous classifier results. The proposed NGBFA-hybrid model stacking is compared to these results. Online ROC-AUC curve values are shown in Figure 8.

Table III. Schiller's evaluated homogeneous and heterogeneous classifiers using the NGBFA-MSE.

Classifier	Accuracy	Precision	Recall	F1-Score
E - KNN	90.01	90.8	90.6	90.26
E - SVM	92.8	93.01	93	93.01
E - LR	69.32	69.9	70	69.6
E - RF	94.67	95.1	95.06	95
E - DT	89.01	89.1	89.07	89.08
E – XGBOOST	96.1	95.9	96	96.02
E-ADABOOST	94.87	95	95.01	95
Proposed	**96.13**	**95.6**	**96.04**	**96.08**

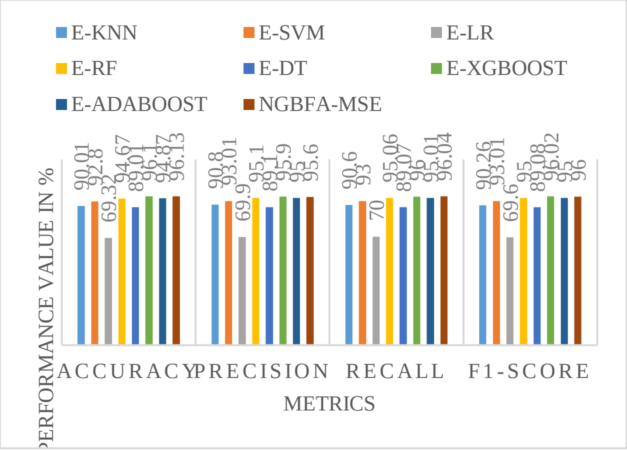

Fig.7: Comparison of NGBFA-MSE with Homo & Heterogenous Classifiers for Schiller

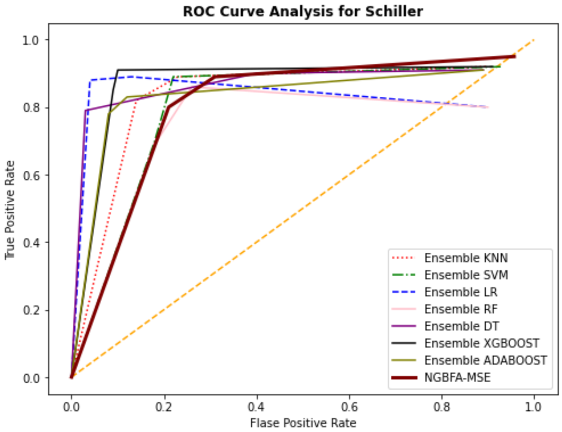

Fig. 8: Analysis of ROC Curve with NGBFA-MSE for Schiller

Result Analysis of target feature: Citology

Analysis of Individual Classifier Performances

We chose five ideal attributes to determine citation aim feature traits. Table 4 shows that the proposed model has 94.04% accuracy when tabulating experimental results for the citation target variable. Unlike the hybrid model stacking. The ROC-AUC curve values are shown in Figure 10.

Table IV. Comparing different Homo & Heterogenous Classifiers with NGBFA-MSE for Citology

Classifier/Metric	Accuracy	Precision	Recall	F1-Score
E - KNN	88.2	89.7	88.3	87.4
E - SVM	85.9	86.02	86.02	86.1
E - LR	57.9	61.01	60.9	58.7
E - RF	91.2	92.6	93.01	92.6
E - DT	91.2	91.9	92.02	91.7
E – XGBOOST	90.4	90.1	90.01	90.4
E – ADABOOST	88.9	90.01	89.7	89.3
Proposed	**94.08**	**91.8**	**92**	**92.1**

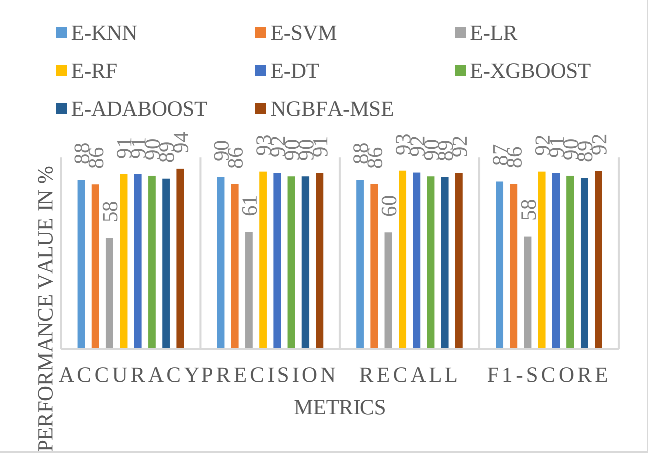

Fig. 9: Comparison of NGBFA-MSE with Homo & Heterogenous Classifiers for Citology

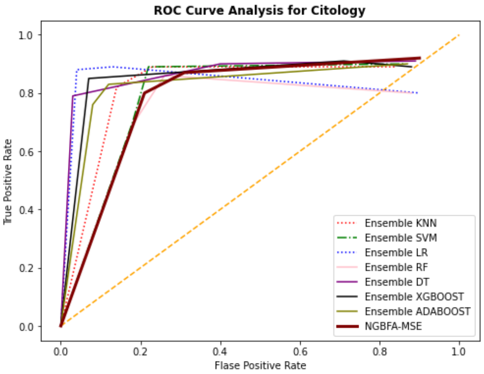

Fig.10: Analysis of ROC Curve with NGBFA-MSE for Citology

V. CONCLUSION

This paper used a mathematical machine learning-based algorithm to evaluate many cervical cancer scenarios. The analysis is done using statistical techniques, machine learning, and cervical cancer detection methods. The proposed hybrid ensemble classifier from eight classification algorithms to predict CC accurately. We thoroughly researched and evaluated all modeling methods to determine the most effective design outcomes. The proposed hybrid approach classifier with two layers was utilized to test the methodology's accuracy and other metrics for each of the four target variables. The method uses resilient model stacking and the analysis matched the previously gathered dataset. The first level uses homogeneous ensembles of varied classifiers to predict cervical cancer. In contrast, the second uses heterogeneous classifiers with soft voting to predict cervical cancer using UCI repository risk data. Both tiers have classifiers. The imbalance problem in the dataset was solved using SMOTE. Despite using all four target variables, the NGBFA-HMS outperformed other classification algorithms. Sociodemographic features may include the sample data's location and education level. Consider many more sociodemographic parameters. Educational institutions and schools can inform students' families about improved healthcare.

REFERENCES

[1] Organization CDC, Centers for Disease Control and Prevention retrieved from https://www.cdc.gov/cancer/cervical/basic_info/index.htm#:~:text=When%20cancer%20starts%20in%20the,at%20risk%20for%20cervical%20cancer. Last accessed on 06/04/2022.

[2] W. H. O. R. Health, W. H. O. C. Diseases, and H. Promotion, Comprehensive Cervical Cancer Control: A Guide to Essential Practice, World Health Organization, Geneva, Switzerland, 2006.

[3] Organization W H 2022 retrieved from https://www.who.int/health-topics/cervicalcancer#tab=tab_1 last accessed 06-04-2022.

[4] Organization W H 2021 retrieved from https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer

[5] Bosch FX, de Sanjosé S. Chapter 1: Human papillomavirus and cervical cancer--burden and assessment of causality. J Natl Cancer Inst Monogr. 2003;(31):3-13. doi: 10.1093/oxfordjournals.jncimonographs.a003479. PMID: 12807939.

[6] [6] Ch. Bhavani, Dr. A. Govardhan. Supervised Algorithms of Machine Learning in the Prediction of Cervical Cancer: A Comparative Analysis. Annals of the Romanian Society for Cell Biology, 1380–1393. 2021, Retrieved from https:// www.annalsofrscb.ro/index.php/journal/article/view/4580

[7] Organization ASCO retrieved from https://www.cancer.net/cancer-types/cervical-cancer/symptoms-and-signs last accessed 06/04/2022.

[8] Jaswinder Singh & Sandeep Shrama. (2019). Prediction of Cervical Cancer Using Machine Learning Techniques. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11 (2019) pp. 2570-2577.

[9] Akter, L., Ferdib-Al-Islam, Islam, M.M. et al. Prediction of Cervical Cancer from Behavior Risk Using Machine Learning Techniques. SN COMPUT. SCI. 2, 177 (2021). https://doi.org/10.1007/s42979-021-00551-6.

[10] Mohamed Hosni, Ibtissam Abnane, Ali Idri, Juan M. Carrillo de Gea, José Luis Fernández Alemán,Reviewing ensemble classification methods in breast cancer, Computer Methods and Programs in Biomedicine, Volume 177, 2019, Pages 89-112, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2019.05.019. (https://www.sciencedirect.com/science/article/pii/S0169260719301907)

[11] A. Idri, M. Hosni, A. Abran, Systematic Mapping Study of Ensemble Effort Estimation, in: Proc. 11th Int. Conf. Eval. Nov. Softw. Approaches to Softw. Eng., 2016: pp. 132–139. doi:10.5220/0005822701320139.

[12] R.E. Schapire, Measures of Diversity in Classifier Ensembles, Mach. Learn. 51 (2003) 181–207. doi:10.1049/ic:20010105.

[13] Pintelas P, Livieris IE. Special Issue on Ensemble Learning and Applications. Algorithms. 2020; 13(6):140. https://doi.org/10.3390/a13060140.

[14] Karlos, S.; Kostopoulos, G.; Kotsiantis, S. A Soft-Voting Ensemble Based Co-Training Scheme Using Static Selection for Binary Classification Problems. Algorithms 2020, 13, 26.

[15] Sabzevari, M., Martínez-Muñoz, G. & Suárez, A. Building heterogeneous ensembles by pooling homogeneous ensembles. Int. J. Mach. Learn. & Cyber. 13, 551–558 (2022). https://doi.org/10.1007/s13042-021-01442-1.

[16] J. L. Fernández-Alemán, J. M. Carrillo-de-Gea, M. Hosni, A. Idri and G. García-Mateos, "Homogeneous and heterogeneous ensemble classification methods in diabetes disease: a review," 2019 41st Annual International Conference of the IEEE Engineering in

Medicine and Biology Society (EMBC), 2019, pp. 3956-3959, doi: 10.1109/EMBC.2019.8856341.

[17] Esra'a Alshdaifat, Malak Al-hassan, Ahmad Aloqaily, Effective heterogeneous ensemble classification: An alternative approach for selecting base classifiers, ICT Express, Volume 7, Issue 3, 2021, Pages 342-349, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2020.11.005. (https://www.sciencedirect.com/science/article/pii/S2405959520304872).

[18] de Oliveira JM, dos Santos EM, Carvalho JRH, de Vasconcelos Marques LA (2013) Ensemble of heterogeneous classifiers applied to lithofacies classification using logs from different wells. International joint conference on neural networks. IJCNN, Dallas, TX, USA, pp 1–6.

[19] Nanni L, Brahnam S, Ghidoni S, Lumini A (2015) Toward a general-purpose heterogeneous ensemble for pattern classifcation. Comput Intell Neurosci 2015:85.

[20] Haque MN, Noman N, Berretta R, Moscato P (2016b) Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classifcation. PloS One 11(1):e0146116.

[21] Alshdaifat E, Al-hassan M, Aloqaily A (2020) Effective heterogeneous ensemble classification: an alternative approach for selecting base classifiers. ICT Express, South Korea.

[22] Luong, Anh Vu, Vu, Trung Hieu, Nguyen, Phuong Minh, Van Pham, Nang, McCall, John, Liew, Alan Wee-Chung, Nguyen, Tien Thanh, "A homogeneous-heterogeneous ensemble of classifiers.", , 2020.

[23] Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. Z. (2022). An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. Biomedical Signal Processing and Control, 72, 103279.

[24] Feng Y, Wang X, Zhang J. A heterogeneous ensemble learning method for neuroblastoma survival prediction. IEEE J Biomed Health Inform. 2021 Apr 13;PP. doi: 10.1109/JBHI.2021.3073056. Epub ahead of print. PMID: 33848254.

[25] Fernandes, K., Chicco, D. and Cardoso, J.S. (2018) Supervised deep learning embeddings for the prediction of cervical cancer diagnosis. PeerJ, 4, 1–20.

[26] Mehmood M, Rizwan M, Gregus ml M and Abbas S (2021) Machine Learning Assisted Cervical Cancer Detection. Front. Public Health 9:788376. doi: 10.3389/fpubh.2021.788376.

[27] Pinar Yildirim. 2018. Classification with incomplete data and ensemble learners for the prediction of cervical cancer risk. In Proceedings of the International Conference on Intelligent Science and Technology (ICIST '18). Association for Computing Machinery, New York, NY, USA, 1–5. DOI:https://doi.org/10.1145/3233740.3233741.

[28] Barua, Lipika, M. Sharif, and Tania Akter. "Analyzing Cervical Cancer by using an Ensemble Learning Approach based on Meta Classifier." International Journal of Computer Applications 182 (2019): 29-33.

[29] Gupta, S., & Gupta, M. K. (2021). Computational prediction of cervical cancer diagnosis using ensemble-based classification algorithm. The Computer Journal.

[30] CH. Bhavani, A. Govardhan, Cervical cancer prediction using stacked ensemble algorithm with SMOTE and RFERF, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3451-3457, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.269. (https://www.sciencedirect.com/science/article/pii/S2214785321051282)

[31] PALETI, L., NAGASRI, A., SUNITHA, P., SANDYA, V., SUMALLIKA, T., KANDUKURI, P., & KUMAR, K. K. (2023). SUGAR CANE LEAF DISEASE CLASSIFICATION AND IDENTIFICATION USING DEEP MACHINE LEARNING ALGORITHMS. Journal of Theoretical and Applied Information Technology, 101(20).

[32] Madhavi, P., Supreethi, K.P. (2023). Multi-dimensional STAQR Indexing Algorithm for Drone Applications. In: Morusupalli, R., Dandibhotla, T.S., Atluri, V.V., Windridge, D., Lingras, P., Komati, V.R. (eds) Multi-disciplinary Trends in Artificial Intelligence. MIWAI 2023. Lecture Notes in Computer Science(), vol 14078. Springer, Cham. https://doi.org/10.1007/978-3-031-36402-0_57, Pages 611-619.

[33] Madhavi, P., and S.K. P. (2023) A Metaphorical Analysis of Different Encoding Techniques for Spatial-Temporal Data. International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 11s, pp. 302-308, https://ijisae.org/index.php/IJISAE/article/view/3473.

[34] CH Bhavani, A. Govardhan, A Novel Genetic Inspired Binary Firefly Algorithm For Feature Selection in the Prediction of Cervical Cancer, Advances in Data Science and Adaptive Analysis, doi: 10.1142/S2424922X23500031.

[35] Peng L, Zhang H, Zhang H, Yang B (2017) A fast feature weighting algorithm of data gravitation classification. Inf Sci 375:54–78. https://doi.org/10.1016/j.ins.2016.09.044

[36] Hsu W-Y (2018) A decision-making mechanism for assessing risk factor significance in cardiovascular diseases. Decis Support Syst 115:64–77. https://doi.org/10.1016/j.dss.2018.09.004.