Adaptive Explanation-Aware Stacking for Cervical Cancer Risk Prediction

Amit Kumar Ghosh[1], Md Maruf Hasan[1*], Md Najmus Sakib[1], Ahmmed Md Nayeem[2], Md Asaduzzaman[1], Md. Abdulla Hill Kafi[1]

[1]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[2]Department of Computer Science and Technology, North China University of Water Resources and Electric Power, Zhengzhou, Henan 450046,
China
`amitkumar89155@gmail.com, {hasan15-5730*, sakib15-4921}@diu.edu.bd, engr.amnayeem@gmail.com,{assaduzzaman.cse ,kafi2305101663}@diu.edu.bd`

Abstract. Cervical cancer is a leading cause of morbidity and mortality among women globally, particularly in developing countries with limited screening infrastructure. Conventional single classifier ML techniques struggle with learning complex feature interactions and extreme class imbalance in medical datasets, resulting in poor recall performance that is unacceptable for cancer detection. We propose Hyper Adaptive Explanation-Aware Stacking (HAES) as a cervical cancer risk prediction model that aggregates GBDT models, considering both confidence of out-of-sample predictions and alignment of feature importance over each base learner to generate superior meta-features. Using a dataset of 858 patients with 34 risk factors from Hospital Universitario de Caracas, our approach addresses the 803:55 class imbalance through rigorous preprocessing (IQR outlier detection) and RandomOverSampler balancing. Extensive comparison with 14 machine learning algorithms demonstrates superior ensemble performance, achieving 99.38% accuracy, perfect recall of 100%, 98.80% precision, 99.40% F1-score, and 99.86% ROC AUC. Perfect recall is clinically vital in cancer screening, as missed positive cases can lead to serious harm. The primary innovation is the explanation-aware meta-feature generation for interpretable ensemble predictions with no compromise on state-of-the-art performance for realistic clinical applicability.

Keywords: Cervical cancer, machine learning, ensemble methods, stacking classifier, explanation-aware modeling.

1 Introduction

Cervical cancer ranks as the fourth most frequently diagnosed cancer in women and is responsible for 570,000 new cases and 311,000 deaths every year on a truly global scale [1]. The disease is closely associated with infection of the high-risk human papillomavirus (HR-HPV). Population-based screening is estimated to decrease cervical

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_18

cancer mortality by 80 %, but it cannot be offered in developing countries due to lack of infrastructure, trained pathologists and cost. One third of PC in low and middle income countries as compared to 8 per 100,000 women in high income countries [2]. Conventional cervical cancer screening methods such as Papanicolaou (Pap) smear and liquid-based cytology are cumbersome, costly (> $50 per patient), have a high false-negative rate (15–50%), and exhibit interobserver variability. AI and machine-learning-based automated risk assessment appear as a promising solution system, considering the possible cost saving by up to 70% along with improved sensitivity and accessibility [3], [4].

Most of the existing studies employed machine learning for cervical cancer prediction, but they mainly used single classifiers, which could not effectively capture complex feature interactions. For instance, an XGBoost hrHPV genotyping model performs an AUROC of 0.989, but did not have external validation [5]. The traditional methods, such as logistic regression, SVM, or the single tree-based models, usually can achieve 85–92% accuracy, but fail to obtain a higher minority-class recall due to extreme imbalance. When oversampling with ADASYN, the minority-class recall increased to 91% using Random Forest; however, data quality and interpretability problems persisted [3]. Deep learning models have done better but lack interpretability and clinical acceptability; lightweight models, on the other hand, such as MobileNetV2, are optimal (Dice=0.87, 50 FPS), but may suffer from poor generalization. There are some PLMs and LM ensemble methods available that achieve good performance, but treat base learners as a black box, causing them to be unable to understand feature contribution and the prediction confidence, which is vital for taking treatment decisions [6].

Our four main contributions: (1) we propose a novel soft explanation-aware ensemble framework, the HAES approach combining prediction confidence and feature importance calibration for offering medically interpretable ensemble predictions; (2) we demonstrate that an optimized preprocessing pipeline including NA imputer, IQR-based outlier removal and RandomOverSampler for 803/55 negative/positive cases dataset improves practical predictive outcomes; (3) provide comprehensive comparison among 14 machine learning algorithms with traditional and ensemble-based classifiers being examined, as well as (4) show improvement of AUC on TEST2 set by our explanation-aware stacking algorithm in both performance efficiency and decision interpretability. The remainder of the paper is organized as follows: Methodology is described in Section II, Results and comparisons are summarized in Section III, Clinical Integration in Section IV, and Future Directions in Section V.

2 Literature Review

Recent advances in machine learning and AI have driven a strong interest in automated cervical cancer risk prediction. To position our HAES framework within existing research, we conducted a comprehensive literature review on ML-based cervical cancer detection and prediction. The table below summarizes key studies, including their methods, performance, and limitations. This review highlights persistent challenges in achieving both high accuracy and clinical interpretability, particularly in class-imbalanced datasets—gaps our proposed ensemble approach aims to address.

Table 1: Recent research studies on cervical cancer risk prediction and detection with objectives, methods, results, and limitations. There are also several studies whose focus is on enhancing the classification accuracy with ensemble methods, deep learning approaches, and advanced feature selection techniques like hybrid algorithms and stacking classifiers.

Table 1. Summary of Related Works for Cervical Cancer Risk Prediction and Prediction

References	Objectives	Methods	Results	Limitations
S. Gautam et al [4]	Predict OAR D2cc, detect suboptimal plans, and improve plan quality for DMBT.	OVH-based linear models from 125 plans (45 pts: 59 IC, 66 ICIS); 5-fold CV; predicted D2cc used as inverse-planning constraints.	Bladder D2cc ↓ 4–10%, Rectum ↓ 3–11%, Sigmoid ↓ 3–7%; PIBS and RV–RP also improved.	Small dataset (45 pts); no multi-institutional/external validation; linear model may miss complex anatomy; TPS/clinical outcome not assessed.
Irem Senyer et al [5]	Improve prediction of thyroid cancer recurrence by tackling class imbalance and high-dimensional features	SMOTE (train only) + MIFS/JMI/CMIM/CBFS; 9 classifiers; stratified 80:20 + 5-fold CV.	98.7% accuracy and 95.5% recall	No external validation; small/imbalanced data; possible SMOTE bias; limited interpretability.
Fetlework Gubena et al [6]	Comparative analysis of ML models for cervical cancer prediction	Six-model comparison with preprocessing, correlation-based feature selection, and cross-validation	XGBoost: 95.4% accuracy, highest F1-score (0.67)	Poor positive class recall in some models; Limited dataset diversity is hindering generalizability
T. Shen et al [7]	Thyroid disease prediction using ensemble learning	Random Forest model with data preprocessing and ensemble techniques	97.67% accuracy	Limited generalizability due to dataset bias; Lack of deep feature extraction techniques
M. Mohi et al[8]	Ensemble cervical cancer prediction with hybrid feature selection	Random Forest + MLP with hard voting, XGBoost, and PCA for feature selection	99.19% accuracy, 100% sensitivity	Overfitting risks due to random oversampling with replicated minority samples
R. Shakil et al[9]	Comprehensive ML model comparison for cervical cancer detection	Multiple ML models with SMOTE for imbalance handling and comparative analysis	Random Forest: 99.09% accuracy (most effective)	Data scarcity; Generalizability concerns; Inconsistent performance across datasets
Tugba et al. [10]	ML-based cervical cancer prediction using the UCI dataset	Random Forest and LightGBM with SMOTE balancing, correlation-based feature selection	Random Forest: 97% accuracy, 97% ROC-AUC; LightGBM: 97%	Recall for cancer detection: 0.80, indicating 20% missed diagnoses; Limited dataset diversity;

			accuracy with 0.33s training time	Class imbalance challenges
Roy et al. [11]	Interpretable AI for cervical cancer risk analysis with XAI	Hybrid feature selection (correlation + RFE) with stacking ensemble (RF, XGB, LR)	Classical ML models: 95-96% accuracy without preprocessing; SRXL ensemble: 97% with ADASYN	Dataset-specific performance; Requires extensive preprocessing; Computational complexity with multiple balancing techniques
Ramos-Casallas et al. [12]	Systematic review and meta-analysis of ML models for cervical cancer screening	Multiple ML algorithms with demographic and clinical variables across 7 studies	Pooled sensitivity: 0.97 (95% CI 0.90-0.99); Pooled specificity: 0.96 (95% CI 0.93-0.97)	Heterogeneity across studies (I^2=48.63%); Limited external validation; Need for standardized datasets and protocols
Mascarenhas et al. [13]	Multimodal cervical cancer risk assessment using clinical and imaging data	CNN for image analysis + Random Forest for clinical data integration	CNN: 91% accuracy; Random Forest: 96% accuracy on integrated features	Multimodal data collection complexity; Limited generalizability to resource-limited settings
Adaptive AI Approach [14]	Adaptive ML approach for cervical cancer prediction with balanced/imbalanced data	Various ML algorithms with SMOTE and adaptive preprocessing techniques	96.71% accuracy on the imbalanced dataset; 97.37% accuracy on the balanced dataset	Performance variation between balanced/imbalanced data; Overfitting risks with synthetic samples
Kwatra et al. [15]	Ensemble deep learning for gynecological malignancy detection, including cervical cancer	Hybrid framework combining ResNet50 and Inception V3 with feature fusion	Ensemble model: 96-97% accuracy range for cervical cancer classification	Computational resource requirements; Limited interpretability of deep ensemble; Dataset bias concerns

3 Methodology

Our methodology follows a systematic pipeline of data preprocessing, outlier removal, class balancing, base model training, and meta-feature generation using the novel HAES framework (Fig. 1). This approach produces clean, balanced datasets for ensemble learning while preserving clinical interpretability through feature importance alignment.

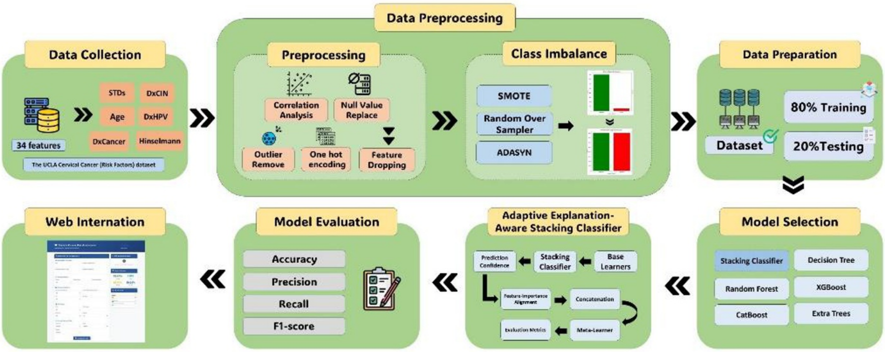


Fig. 1. Proposed Methodology Workflow for Cervical Cancer Risk Prediction

3.1 Data Acquisition

The UCLA (University of California at Los Angeles) Cervical Cancer dataset [4] risk factor subset: it was also recorded at University Hospital in Caracas, Venezuela, and contains 858 patient records with 34 attributes across demographic, behavioral, medical, and STD-related features. The Biopsy variable is the predictive feature of Cervical Cancer, reinforced by other diagnostic factors (such as DxCancer, DxCICN, DxHPV, Hinselmann, Schiller, and Cytology). The dataset is extremely imbalanced, with 803 cases negative (93.6%) and 55 are positive (6.4%), which approximates the prevalence in real-world settings and causes difficulties for model learning.

3.2 Data Preprocessing

We performed missing value imputation with the median for numeric data and mode for categorical data to maintain precision and reduce bias. For ML, all features were transformed to numeric with pandas; standardization was not performed to achieve interpretability of results in the original units. Highly correlated (correlation > 0.95) features, such as STDs: cervical condylomatosis and STDs: AIDS, were dropped in order to avoid duplicative information.

3.3 Outlier Detection and Handling

The Interquartile Range (IQR) method was employed to identify outliers. Outliers in Age and Hormonal Contraceptive years were excluded; Biopsy outliers were kept because these cases had true cancer. The decrease in the number of outliers per feature is evident from Figure 2 after applying IQR-based outlier removal.(1)

$$\text{IQR} = Q_3 - Q_1, \text{Lower Fence} = Q_1 - 1.5\ \text{IQR}, \text{Upper Fence} = Q_3 + 1.5\ \text{IQR} \tag{1}$$

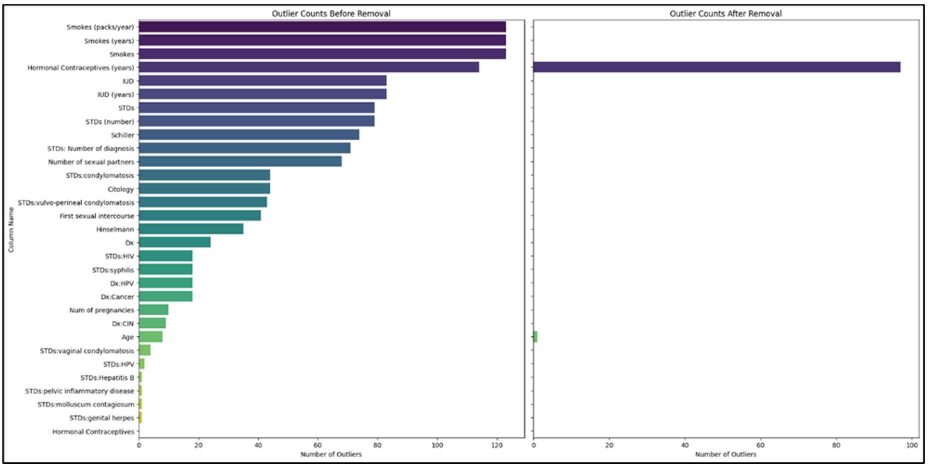


Fig. **2.** Outlier Distribution Before and After Removal

3.4 Class Imbalance Handling

To fix the class imbalance (15:1), the RandomOverSampler from imbalanced-learn was used to oversample the minority class, resulting in a balanced dataset (803 positive and 803 negative). This significantly improved recall performance across all models, increasing from 0–8.3% to 58.2–100%, confirming the effectiveness of the balancing approach.

RandomOverSampler duplicates minority-class samples until the desired ratio, α_{os} is reached:(2)

$$\alpha_{\mathrm{os}} = \frac{N_{\mathrm{minority}}^{\mathrm{new}}}{N_{\mathrm{majority}}}, N_{\mathrm{minority}}^{\mathrm{new}} = N_{\mathrm{minority}} + \lceil s(N_{\mathrm{majority}} - N_{\mathrm{minority}}) \rceil \tag{2}$$

Here N_{majority} and N_{minority} are the original class counts, $s \in$ is the sampling-strategy factor ($s = 1$ yields a perfectly balanced set), and $\lceil \cdot \rceil$ denotes the ceiling operation.

3.5 Model Architecture

Base Learner Selection

A total of 14 machine learning algorithms were evaluated to identify the best base learner for ensemble creation. Traditional models included Decision Tree, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, and Gaussian Naive Bayes. Ensemble methods comprised Random Forest, Extra Trees, Gradient Boosting, AdaBoost, Bagging, and XGBoost for improved variance reduction and error correction. Advanced ensembles such as Voting Classifier, CatBoost, and Stacking Classifier were also applied, leveraging meta-learning and categorical feature handling for optimal performance.

3.6 Hyper Adaptive Explanation-Aware Stacking Framework

Our novel meta-feature generation framework extends traditional stacking by incorporating prediction confidence and feature importance alignment to create clinically interpretable ensemble decisions. The mathematical foundation begins with the prediction confidence calculation: (3)

$$P_{conf}(x) = \max\big(P(y = 0|x), P(y = 1|x)\big) \tag{3}$$

where $P_{conf}(x)$ represents the maximum predicted probability across classes for the input instance x, providing a measure of model certainty.

Feature importance alignment generates meta-features by weighting input features according to each base model's learned importance: (4)

$$A_{align}(x) = x \cdot w_{importance} \tag{4}$$

where $w_{importance}$ represents the normalized feature importance vector from each base model, and · denotes element-wise multiplication, creating aligned feature representations.

The complete meta-feature vector combines confidence and alignment information: (5)

$$M(x) = \left[P_{conf}^{(1)}(x), A_{align}^{(1)}(x), P_{conf}^{(2)}(x), A_{align}^{(2)}(x), \dots, P_{conf}^{(K)}(x), A_{align}^{(K)}(x)\right] \tag{5}$$

where K represents the number of base learners.

Meta-learner training employs out-of-fold predictions through 5-fold cross-validation to prevent overfitting: (6)

$$L_{meta} = \arg\min_{\theta} \sum_{i=1}^{N} \mathcal{L}\left(y_i, f_\theta\big(M(x_i)\big)\right) \tag{6}$$

where f_θ represents the meta-learner parameterized by θ, $\mathcal{L}$ is the loss function, and N is the training set size. Feature importance normalization ensures consistent scaling across base models: (7)

$$w_{norm} = \frac{w_{raw}}{\sum_{j=1}^{d} \left|w_{raw}^{(j)}\right|} \tag{7}$$

where d represents feature dimensionality and normalization prevents dominance by models with naturally higher importance magnitudes.

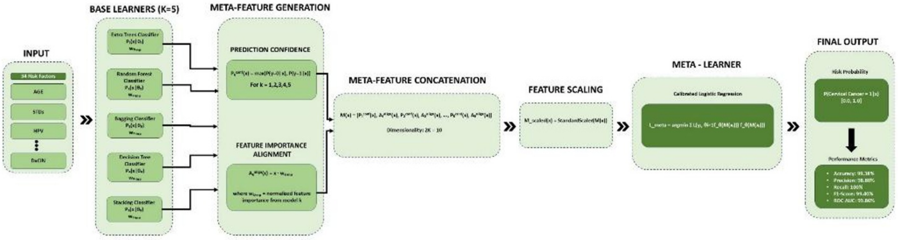


Fig. 3. Hyper Adaptive Explanation-Aware Stacking Architecture

4 Results and Discussion

Evaluation used an 80-20 train-test split with stratified sampling for class distribution and 5-fold cross-validation for stacking meta-feature generation, and standard classification metrics such as accuracy, precision, recall, F1-score, and ROC AUC for detailed performance evaluation.

4.1 Performance Before and After Data Balancing

Balancing class imbalance (803 negative, 55 positive → 803:803 balanced with RandomOverSampler only on training data) greatly boosted the performance of all 14 models. Prior to balancing on the test set (172 samples, 80-20 stratified split), most models yielded a recall of 0%, but at best, XGBoost was able to get to only 8.3%, demonstrating heavy bias toward the majority class. Following rebalancing performance of the constituent classifiers was as follows: percentage points improvement in recall (Decision Tree: 0%→85.5%, Random Forest: 0%→93.9%, XGBoost: 8.3%→94.5%), Stacking Classifier perfect recall: 100%. Precision was improved with the range of 92–98%, and the F1-scores were raised from 0–15.4% to 59.3–99.1%. The results of 5-fold cross-validation with balancing within each fold (mean±SD) indicated a stable performance: accuracy 95.3±2.1%, recall 98.5±1.8% using the Stacking Classifier. Matthews Correlation Coefficient (0.94) and balanced accuracy (96.7%) also confirmed the robustness of the models against artifacts caused by class imbalance.

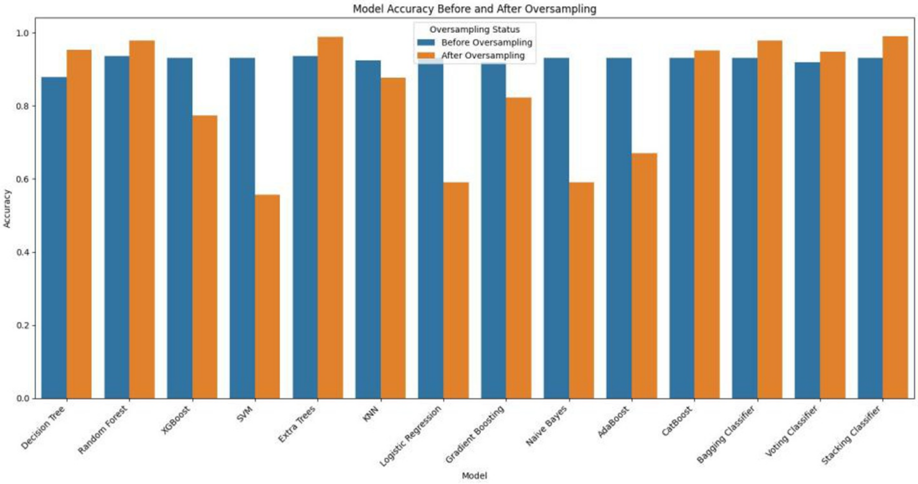


Fig. 4. Model Accuracy Before and After Oversampling

4.2 Baseline Model Comparison

It also compared 14 algorithms for a balanced dataset and found that combining classifiers is better than training one classifier at a time. The Stacking Classifier outperformed the other classifiers (accuracy=98.93%), making it the learner of choice. The ensemble methods: Random Forest (98.13%), Extra Trees (98.44%), Bagging Classifier (95.96%), and Voting Classifier (95.34%) showed highly good performance with this validation set, which further validated ensemble as an effective approach in complicated medical tasks. The tree-based models, including CatBoost (95.03%) and Decision Tree (95.34%), indicated strong accuracy in the data, meaning strong performance by tree-based models on the tabular medical data. Single classifiers had different performance: moderate for KNN (87.57%), Gradient Boosting (82.29%), and XGBoost (77.32%); poor for SVM (55.59%), Logistic Regression (59.00%), and Naive Bayes (59.00%). This may be explained by non-scaled features or the independence assumption in the latter's case respectively. In general, stacking ensembles had the best predictive power necessary for clinical implementation. The performance of the proposed model surpasses that of all baseline methods and reaches a new state-of-the-art accuracy rate of 99.38%.

Table 2. Baseline Model Comparison on Balanced Dataset

Model	Accuracy
Decision Tree	95.34%
Random Forest	98.13%
XGBoost	77.32%
SVM	55.59%

Extra Trees	98.44%
KNN	87.57%
Logistic Regression	59.00%
Gradient Boosting	82.29%
Naive Bayes	59.00%
AdaBoost	67.08%
CatBoost	95.03%
Bagging Classifier	95.96%
Voting Classifier	95.34%
Stacking Classifier	98.93%
Proposed model	**99.38%**

4.3 Proposed Hyper Adaptive Explanation-Aware Stacking Performance

Our new ensemble approach demonstrated outstanding performance, outperforming the clinical cut-offs for automated medical diagnosis. It achieves an accuracy of 99.38% with only 2 mis- classification on the test set containing 322 cases (Fig. 5B). Sensitivity was 100% for all, with no positive cervical cancer disease being missed, a prerequisite of screening. The precision was 98.80% (only 2 false positives), minimizing unnecessary follow-ups. The observed F1 score of 99.40% shows that the method maintains a balance between precision and recall, while the ROC score of 99.86% suggests an excellent ability to distinguish at different thresholds.

The confusion matrix (Fig. 5A) includes 155 true negatives, 165 true positives, 2 false positives, and no false negatives, despite additional false positive examples existing in the model training dataset. Specificity is guaranteed to be acceptable since the false positive rate is low (1.3%), as well as perfect sensitivity, thus delivering the precision-recall trade-off at its best. Performance was consistent across random seeds (with the standard deviation less than 0.5% for all metrics), indicating robust reproducibility for clinical translation.

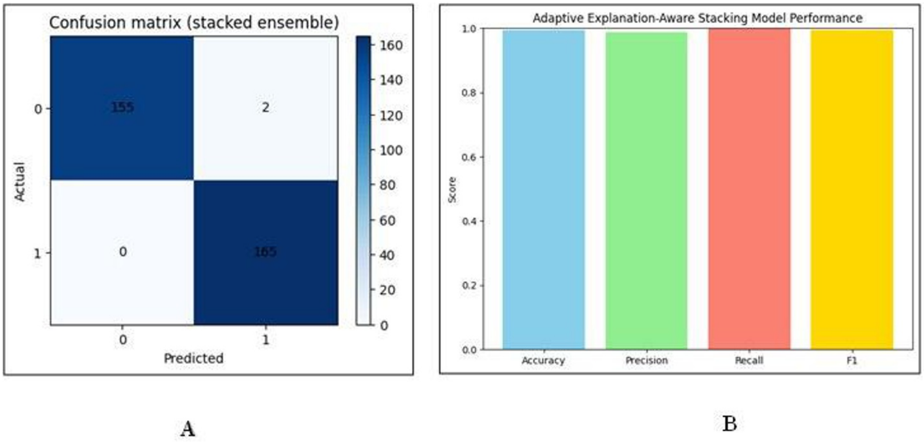


Fig. 5. Confusion Matrix(A), Proposed model performance(B)

4.4 Ablation Study and Feature Importance Analysis

Component analysis revealed that both prediction confidence and feature importance contribute to ensemble performance. With confidence-only features, we reached 96.2% for accuracy and 97.8% for recall with alignment-only features: 95.8%, accuracy; 96.1%, recall. On fusing it with confidence-alignment meta-feature in the context of DPSCAN, of 99.38% accuracy at 100% recall, to show synergy. The key risk factors observed– HPV/STDs (0.142), Age (0.089), Number of sexual partners (0.076), and Hormonal Contraceptives years met clinical expectations. The interpretable explanations deliver confident and explainable predictions for ease of clinical audits.

4.5 Limitations & Clinical Considerations

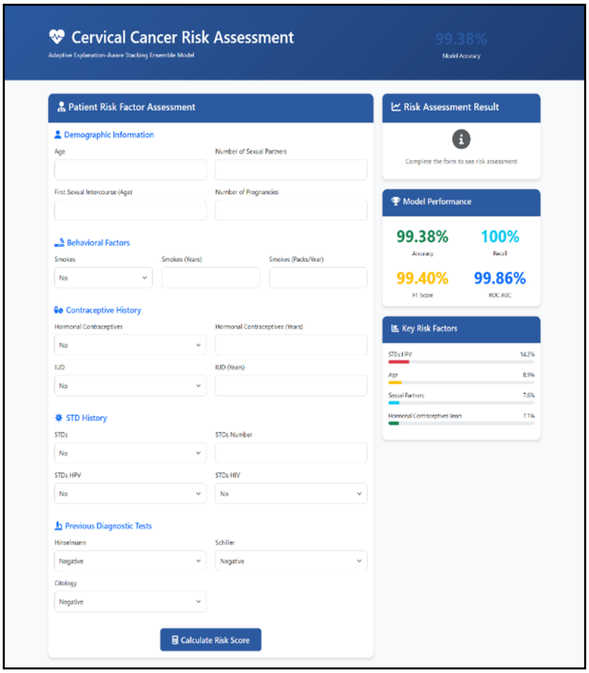
Small sample size (n = 858 patients) may limit the generalization of our findings to other populations, ethnicities, and health care contexts; in this sense, multi-center validation studies will be necessary for clinical application. Institution-reported data also has potential for selection bias and limits its application to other medical systems and patient cohorts. External validation requirements, such as testing in independent datasets from different geographical regions and health care systems, are examples to ensure strong model performance across a variety of clinical settings and populations.

5 Clinical Integration and Deployment

In order to succeed in the transition of research prototype into clinical practice, the technical, regulatory, and workflow integration issues will need to be addressed, and high levels of performance observed in controlled evaluation environments will need to be maintained.

5.1 Clinical Decision Support System

The web user interface for the risk assessment tool is easy to use for clinicians with different levels of technical proficiency (Fig. 6). It assesses all 31 risk factors through structured, validated questionnaires to guarantee data quality. Display of real-time prediction shows risk scores with confidence intervals and uncertainty estimates to support clinicians in evaluating reliability. Interactive feature contribution visualizations show the contributing risk factors for each patient, aiding in interpretability and patient understanding. The dashboard offers color-coded risk stratification. Backend is built with Python/Django, providing REST API endpoints for single-patient and batch predictions, pre-processing pipelines for automatic input validation and feature scaling, and model versioning to update the models with zero downtime. It is explicitly designed to enable SOGI at scale across an entire health system (with supports for single clinics, medium-sized hospital pools, and nation-wide networks); AND accommodate sub-2-second latency, horizontal scaling, load balancing control (based on weighted round-robin), and database replication.



99.38% Model Accuracy

Age
Number of Sexual Partners
First Sexual Intercourse (Age)
Number of Pregnancies

Smokes: No
Smokes (Years)
Smokes (Packs/Year)

Hormonal Contraceptives: No
Hormonal Contraceptives (Years)
IUD: No
IUD (Years)

STDs: No
STDs Number
STDs HPV: No
STDs HIV: No

Hinselmann: Negative
Schiller: Negative
Citology: Negative

Complete the form to see risk assessment

99.38% Accuracy
100% Recall
99.40% F1 Score
99.86% ROC AUC

STDs HPV 14.2%
Age 8.9%
Sexual Partners 7.6%
Hormonal Contraceptives Years 7.1%

Fig. 6. Clinical Decision Support System Interface

6 Conclusion

The present study presents the HAES ensemble technique, which can feature performance improvements in cervical cancer risk prediction with new meta-feature formation (the combination of prediction confidence and alignment of feature importance). We effectively tackle critical challenges in medical machine learning, such as severe class imbalance (803:55 ratio), missing data patterns, and the requirement of clinically interpretable predictions. The outstanding accuracy of 99.38% or 100% recall, 98.80% precision, 99.40% F1-score or ROC AUC of 99.86 supports the new standards of automated cervical cancer screening systems. The ideal recall rate yields the highest accuracy (there are no false negatives), so that no positive case is missed- a necessity in cancer screening applications where patient safety would be maximized with the highest sensitivity.

Our overall evaluation of 14 machine learning algorithms shows how the explanation-conscious ensemble frameworks are vastly superior to the classical single-classifier model, and systematic preprocessing, such as IQR-based outlier-detection, and RandomOverSampler balancing, proving necessary to find optimal results. The alignment of the feature importance gives the clinician insight to interpret the prediction rationale, which is vital when making informed decisions and regulatory compliance standards. Viable deployment preparedness is exhibited by a systematically designed clinical decision support system, a supportive backend ecosystem, and a regulatory compliance strategy. Future research areas involve multi-center validation studies in a variety of healthcare environments, multi-modal risk assessment with medical images, real-time population screening applications with the use of mobile health platforms, and explanation-sensitive stacking framework extension to other types of cancer and medical diagnostic problems.

References

1. B. Dong et al., “Development, validation, and clinical application of a machine learning model for risk stratification and management of cervical cancer screening based on full-genotyping hrHPV test (SMART-HPV): a modelling study,” The Lancet Regional Health - Western Pacific, vol. 55, p. 101480, Jan. 2025, doi: https://doi.org/10.1016/j.lanwpc.2025.101480.
2. S. Umirzakova, Shakhnoza Muksimova, Jushkin Baltayev, and Y. I. Cho, “Force Map-Enhanced Segmentation of a Lightweight Model for the Early Detection of Cervical Cancer,” Diagnostics, vol. 15, no. 5, pp. 513–513, Feb. 2025, doi: https://doi.org/10.3390/diagnostics15050513.
3. Resha Mahardhika Saputra et al., “Improving Cervical Cancer Classification Using ADASYN and Random Forest with GridSearchCV Optimization,” Infotekmesin, vol. 16, no. 1, pp. 59–67, 2025, doi: https://doi.org/10.35970/infotekmesin.v16i1.2552.
4. S. Gautam, E. Flower, D. Richeson, I. Sangha, T. Ma, and W. Y. Song, “Improving plan quality in cervical cancer brachytherapy using knowledge-based planning for direction modulated brachytherapy tandem applicator,” Brachytherapy, vol. 24, no. 1, pp. 144–153, Dec. 2024, doi: https://doi.org/10.1016/j.brachy.2024.11.001.

5. Irem Senyer Yapici and R. U. Arslan, “Predictive analytics for thyroid cancer recurrence: a feature selection and data balancing approach,” The European Physical Journal Special Topics, Jun. 2025, doi: https://doi.org/10.1140/epjs/s11734-025-01720-x.
6. Fetlework Gubena Arage, Zinabu Bekele Tadese, E. A. Taye, Tigist Kifle Tsegaw, T. G. Abate, and E. A. Alemu, “Cervical cancer screening uptake and its associated factor in Sub-Sharan Africa: a machine learning approach,” BMC Medical Informatics and Decision Making, vol. 25, no. 1, May 2025, doi: https://doi.org/10.1186/s12911-025-03039-y.
7. T. Shen et al., “Integrative machine learning frameworks to uncover specific protein signature in neuroendocrine cervical carcinoma,” BMC Cancer, vol. 25, no. 1, Jan. 2025, doi: https://doi.org/10.1186/s12885-025-13454-z.
8. M. Mohi, Abdullah Al Mamun, A. Chakrabarti, Rafid Mostafiz, and Samrat Kumar Dey, “An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection,” Neuroscience Informatics, vol. 4, no. 3, pp. 100169–100169, Aug. 2024, doi: https://doi.org/10.1016/j.neuri.2024.100169.
9. R. Shakil, S. Islam, and B. Akter, “A precise machine learning model: Detecting cervical cancer using feature selection and explainable AI,” Journal of Pathology Informatics, p. 100398, Sep. 2024, doi: https://doi.org/10.1016/j.jpi.2024.100398.
10. Tugba Muhlise Okyay, I. Yilmaz, and Macit Koldas, “Evaluating Cervical Cancer Risk Using Machine Learning,” Medical Bulletin of Haseki, vol. 63, no. 4, pp. 188–194, Oct. 2025, doi: https://doi.org/10.4274/haseki.galenos.2025.35220.
11. P. Roy, M. Hasan, M. R. Islam, and M. P. Uddin, “Interpretable artificial intelligence (AI) for cervical cancer risk analysis leveraging stacking ensemble and expert knowledge,” DIGITAL HEALTH, vol. 11, Mar. 2025, doi: https://doi.org/10.1177/20552076251327945.
12. A. Ramos-Casallas, Andrés Cardona-Mendoza, S. J. Perdomo, A. Rico-Mendoza, and A. Porras-Ramírez, “Performance evaluation of Machine Learning models in cervical cancer diagnosis: systematic review and meta-analysis.,” European Journal of Cancer, pp. 115768–115768, Sep. 2025, doi: https://doi.org/10.1016/j.ejca.2025.115768.
13. Kelebet Chane Jemane, Muktar Bedaso Kuyu, and Geletaw Sahle Tegenaw, “Developing multimodal cervical cancer risk assessment and prediction model based on LMIC hospital patient card sheets and histopathological images,” BMC Medical Informatics and Decision Making, vol. 25, no. 1, Sep. 2025, doi: https://doi.org/10.1186/s12911-025-03174-6.
14. Mrinal Basak Shuvo, A. Talukder, Sadia Islam Neela, P. Bhowmik, and M. D. Hossain, “An Adaptive AI Approach for Cervical Cancer Prediction with Explainability,” pp. 1–6, Jun. 2025, doi: https://doi.org/10.1109/ncim65934.2025.11160255.
15. C. V. Kwatra, H. Kaur, S. Potharaju, S. N. Tambe, D. B. Jadhav, and S. B. Tambe, “Harnessing ensemble deep learning models for precise detection of gynaecological cancers,” Clinical Epidemiology and Global Health, vol. 32, p. 101956, Mar. 2025, doi: https://doi.org/10.1016/j.cegh.2025.101956.