Prediction of Endometrial Carcinoma Recurrence Using a Stacking Ensemble with Meta-Learner

Khushnor Rahman Meem*[1], Anamika Sanyal [1],
Md Wakil Ahmed[1], Shahnewaj Limon[1]
and K.M.Nure Tanvir Siddique [1]

[1]Department of Computer Science and Engineering, Daffodil International University,
Dhaka-1216, Bangladesh

khushnorrahmamemeem@gmail.com*, sanyal15-5788@diu.edu.bd,
ahmedwakil512@gmail.com, mdshahnawajlimon@gmail.com,
tanvir15-5223@diu.edu.bd

Abstract. Endometrial carcinoma is one of the most prevalent gynecological cancer in the world, recurrence of which impacts a lot on the survival of the patient. The conventional clinical markers are not usually able to render the interaction between the genomic and histopathological characteristics. In this research, we provide a state-of-the-art ensemble learning model that combines XGBoost, CatBoost, LightGBM, and multilayer perceptron based on stacking ensembles to predict the recurrence risk. For preprocessing these clinical data we have been used categorical encoding, feature scaling, and KNN imputation. The problem of class imbalance was solved using SMOTE. The performance was compared by use of F1-score, ROC-AUC and sensitivity analysis with the optimization of thresholds. Findings show that the stacking ensemble did better using standalone models, with a macro F1-score of 0.79 and ROC-AUC of 0.82, and better prognostic power than standalone models.

Keywords: Endometrial Carcinoma, XGBoost, LightGBM, CatBoost, Stacking Classifier, SMOTE, disease-free survival.

1 INTRODUCTION

Endometrial carcinoma (EC) is a gynecologic malignancy, which forms in the uterine lining of females [1]. The disease was initially reported to be a major oncologic burden in the mid 20th century which has now become an increasing health issue in the world. Endometrial Carcinoma is a name used to describe a heterogeneous group of tumors that differ significantly in their histological and molecular profile as well as clinical outcome [2]. Throughout history, the EC was considered as an illness that was

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_14

mostly prevalent among postmenopausal women of high-income states. Nevertheless, its rates have grown for the last few decades throughout the world contributed by lifestyle changes and high obesity rate [3]. This increment has not only made endometrial carcinoma a global concern but also reports of increasing cases of mortality in both the developed and the developing countries. There are 2 subtypes: endometrial carcinoma (Usually nonviolent) and serous carcinoma (worse) which are dominating [4]. The patients are the main source of clinical and genomic data whereas the tumor samples and sequencing technologies provide decisive knowledge concerning the underlying molecular mechanisms. In spite of resemblance of EC and other gynecologic malignancies in terms of the presentation, e.g. abnormal bleeding, pain in the pelvis, and weight loss, significant differences are present in aggressiveness, likelihood of recurrence, and responsiveness to treatment [5]. Considering this, early-stage can be treated with surgery, whereas advanced or recurrent disease can be treated with multimodal therapy with mixed results. In some patients, aggressive tumor phenotypes, characterized by a high tumor mutational burden (TMB), a high genomic instability, and poor response to the usual treatment develop [6]. Clinical trials determine recurrence to be occurred for around 15-20% patients who have the disease at the early stages, and higher rates for the more advanced subtypes [7]. Predictive factors of recurrence are commonly determined such as older patients, high level of histological grade, and lymph vascular invasion [8]. Regardless of these findings, predicting recurrence and stratifying the patients based on the disease-free survival is still a low-accuracy area. This diversity doesn't allow early observation and personalized care where resources are limited and genomic testing are difficult. The study will set forth these gaps by creating a stacking ensemble model that predicts recurrence risk among patients with endometrial carcinoma on the basis of clinical, genomic, and demographic features. The study aims to enhance predictive accuracy and assist clinical decision-making by enhancing advanced machine learning algorithms along with the resulting strengths in preprocessing and class balancing to reduce the occurrence of reoccurrence

2 Literature Review

Sabrina et. al [9] developed ML models to identify first rate endometrial cancer and its recurrence. They collected data from eight Canadian centers which include 1237 patients. To predict recurrence rates they were following surgery and tailored adjuvant treatment. Their acquired accuracy is 71.8%.

Xiaofeng et. al [10] integrated machine learning model with immune response analysis to forecast disease outcome of endometrial cancer. The prognosis of this cancer was based on the factors such as tumor stage, grade and molecular characteristics. Some of the methods of machine learning that they used include ssGSEA, WGCNA, unsupervised learning, SLC38A3 etc.

Marina et. al [11] acknowledged that Uterine Corpus Endometrial Carcinoma (UCEC) is a popular gynecological asperity that is both morbid and lethal, especially in the late stages. One of the most common cancer in United State and the second popularity in the world is endometrial carcinoma. The conventional diagnostic tests are expensive and time consuming. So they investigated machine learning model on early detection, prognosis prediction, successful treatment and treatment at low cost.

Cheng Zhu et. al [12] used gamma-glutamyl hydrolase role in the UCEC. They applied ML methods of the identification of GGH as a central gene that affects tumor differentiation. They applied some ML model such as LASSO, univariate cox, multivariate cox, ssGSEA, IHC etc. The functional investigations depict that the in vitro depletion of GGH resulted in fewer tumor cells and distorted immune cells. These show that GGH has a major role in UCEC development.

Joshua et. al [13] identified that when the uncontrolled growths take place in the reproductive organs or the genitals of a woman, it is known as a gynecological cancer. One of these cancers is the endometrial cancer. They applied different formats of ML algorithms such as SVM(Support Vector Machine), Decision Tree, Deep learning etc. There is vast heterogeneity within the field, and in the recent past, special ML techniques have been used.

Seyma et. al [14] presented new method of diagnosing and forecast the prognosis of endometrial cancer combining the proteomics and explainable artificial intelligence (XAI). The primary objective of their research was to categorize age, neoplasms grade, myometrium infiltration and tumor volume that are significant in the diagnosis of the endometrial cancer. XAI makes models more transparent, which enables clinicians to understand AI-driven prediction. Proteomics driven studies have come up with important biomarkers with regard to tumor evolution and response to treatment.

Ching-Wae Wang et. al [15] employed deep learning to predict tumor mutational burden(TMB) using histopathology slides and detection of endometrial cancer. It has been demonstrated that Convolutional Neural Networks (CNNs) can be used to accurately classify tumor subtypes and assess genetic changes. They applied some ML model that analyses small tissue pattern and enhance the precision of diagnosis.

3 PROPOSED METHODOLOGY

This paper proposes a customized machine learning pipeline to predict recurring endometrial cancer based on the UCEC TCGA clinical dataset [16] with a specific stacking ensemble configuration [17] and data preprocessing based on subtype. The strategy works on a new base-model-clinical optimization combination to reach a high predictive accuracy, especially on the minority group (Recurred/Progressed). The following Figure 1 demostrates the working procedure to predict the disease free status.

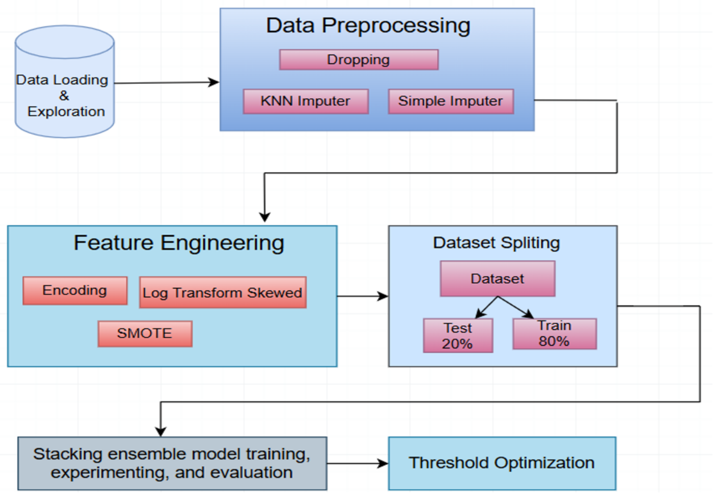

Fig. 1 Work flow diagram for predicting disease-free status

3.1 Dataset Description and Exploration

The clinical dataset from UCEC TCGA [16] consisting of 549 samples and 98 features is used to predict recurrence of the endometrial cancer (DiseaseFree: 398 samples Recurred Progressed: 110 samples) represented in Figure 2. The structure of the dataset is analyzed using the exploratory analysis method which discloses a great imbalance (398 DiseaseFree vs. 110 Recurred/Progressed). Key featured including Diagnosis Age, Mutation Count, Tumor Mutational Burden (TMB) and Neoplasm Histologic Grade are examined statistically.

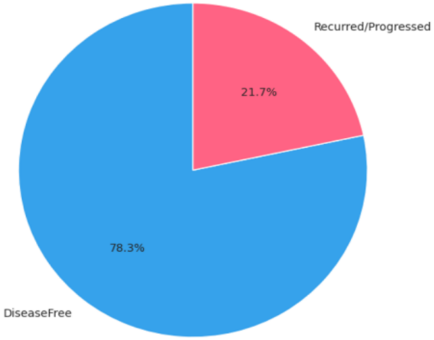

Fig.2. Class distribution of target variable

3.2 Data Preprocessing

In order to maintain the quality of data, the ‘NA’ values are replaced by nan. Columns having more than 80% missing values are eliminated in order to minimize noise. The rows that lacked the target variable (DiseaseFree Status) are dropped and 508 samples remained at end.

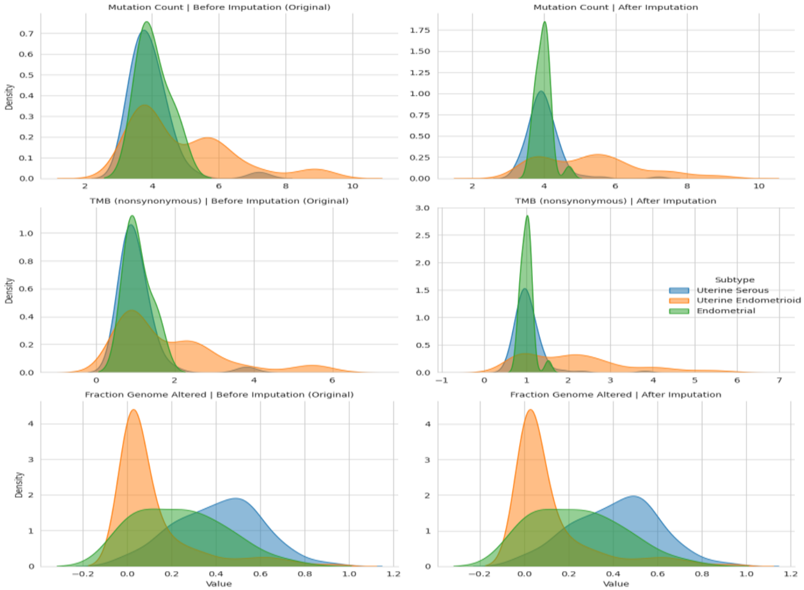

Fig. 3 . Visualization of variables for KNN imputation

Subtype-specific KNN imputation [18] to genomic (Mutation Count, TMB, Fraction Genome Altered) and continuous features is a significant aspect of this study. Imputation is done in the subtypes of cancer (Figure 3) to ensure biological relevance using KNN Imputer with k=5 neighbors. Handling missing values is done by a global KNN imputation of missing values in both continuous and categorical features of mode imputation and mean imputation respectively.

3.3 Feature Engineering

Features are transformed to suitable forms in order to complement machine learning models. OrdinalEncoder is used to code Neoplasm Histologic Grade (G1, G2, G3, High Grade) in order to preserve ordinal relationships of the feature. Menopause Status and AJCC Stage are also ordinal-encoded since they are hierarchical. Cancer Type Detailed, Person Neoplasm Status, Ethnicity, and Surgical Margin Status are categorical features and encoded using OneHotEncoder. Numerical items are standardized with StandardScaler [19]. Skewed variables are transformed using Log Transformer to minimize skewness of dataset (Figure 4). The final dataset includes 21 features (8 numeric, 13 categories) and the target is binarized (DiseaseFree=0, Recurred/Progressed =1).

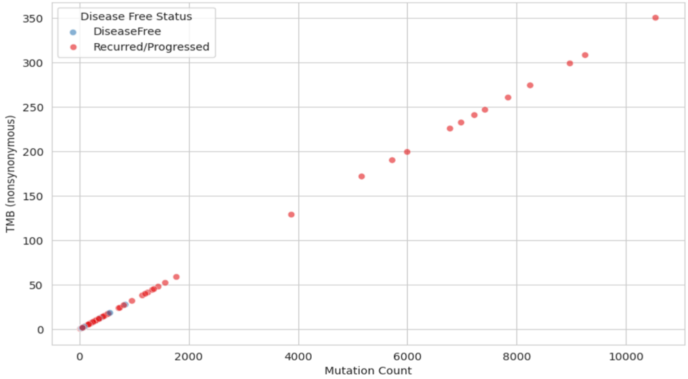

Fig. 4. Visualization of some Skewed Variables

3.4 Dataset Splitting

The data is split into training (406 samples, 80%) and test (102 samples, 20%) sets with the help of stratified sampling in order to maintain the class balance. SMOTE [20] is used to overcome minority class imbalance on the training set with a sampling rate of 0.8 which produce about 574 samples and eventually increased the learning capabilities of the models to learn the minority patterns.

3.5 Stacking Ensemble Design

Stacking ensemble captured the non-linear, genomic-clinical interaction through heterogenous learners like tree based boosting algorithms and neural network and thus outperformed individual models and other ensemble techniques [24]. The baseline models like Logistic Regression or simple ensemble models could not show promising results like the chosen one.

Table 1. BASIC MODEL SPECIFICITY

Base Model	Purpose
XGBoost	Captures complex feature interactions for robust prediction using gradient boosting.
CatBoost	Handles categorical features efficiently to enhance model performance on clinical data.
LightGBM	Provides fast and accurate predictions, optimized for small datasets like UCEC TCGA.
MLP	Models non-linear patterns in clinical features to complement tree-based approaches.

Following stacking ensemble model is developed with four different base models (XGBoost [21], CatBoost [22], LightGBM [23] and Multi-Layer Perceptron (MLP)). All models are chosen due to their specific efficiency showed in Table I The grid search cross-validation technique is applied with 5-fold cross-validation attempting to tune hyperparameters. The best settings are: XGBoost (learning rate=0.05, max depth=3, n estimators= 50), CatBoost (depth=3, iterations=50, learning rate=0.05), LightGBM (learning rate= 0.05, max depth=3, n estimators= 50), and MLP (hidden layer sizes=(50,), max iter=200). The probability output of the base models was added together with an XGBoost meta-learner (learning rate=0.05, max depth=3, n estimators= 50) which is then trained through 5-fold StratifiedKFold to prevent data loss illustrated in Figure 5. This special structure uses the diversity of models to increase the predictive performance.

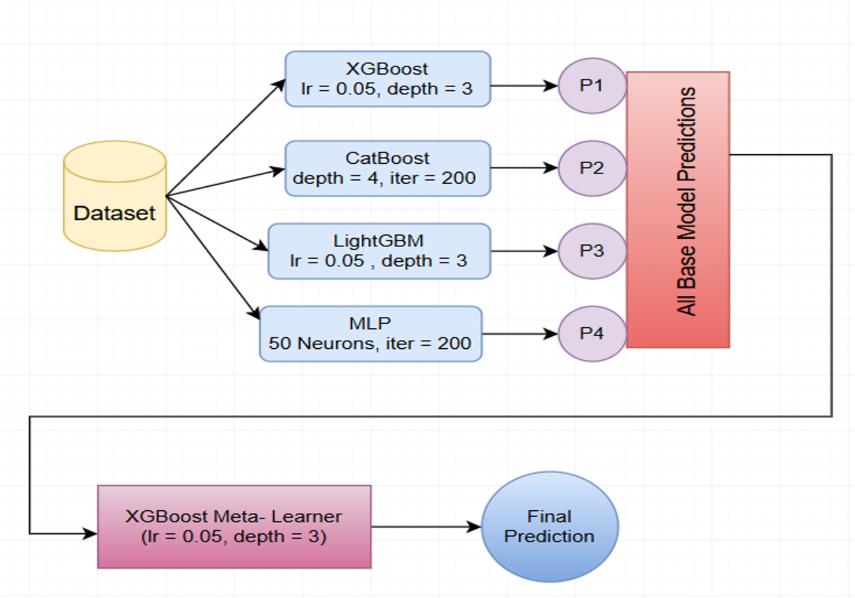

Fig. 5. Stacking Ensemble Model Architecture

3.6 Model Training and Evaluation

Evaluation of models is based on accuracy, precision, recall, F1-score, and AUC (Table. 2) for minority class performance relevant to clinical use. The predictions of the stacking ensemble on the test set are threshold optimized to give an optimal threshold score of 0.3483 based on ROC curves to optimize recall in Recurred/Progressed so that high sensitivity to high-risk patients could be achieved.

Table 2. BASE MODEL PERFORMANCE

Model	Accuracy	Precision		Recall		F1-Score		AUC
		Disease Free	**Recurred**	**Disease Free**	**Recurred**	**Disease Free**	**Recurred**	
XGBoost	0.84	0.86	0.70	0.94	0.45	0.90	0.55	0.80
CatBoost	0.85	0.87	0.75	0.95	0.48	0.91	0.59	0.82
LightGBM	0.84	0.86	0.72	0.94	0.46	0.90	0.56	0.81
MLP	0.80	0.83	0.65	0.92	0.40	0.87	0.50	0.78
Logistic regression	0.79	0.82	0.68	0.90	0.40	0.88	0.51	0.77

4 RESULT ANALYSIS

In this work, model performance is measured with the help of accuracy, AUC, recall of minority classes, and F1-score. Accuracy is the measure of overall prediction accuracy and AUC is the measure of the ability of the model to distinguish between classes which is an important measure for imbalance datasets. The recall of minority classes focuses more on the recurrence detection, which is crucial to clinical sensitivity, and F1-score balances precision and recall to ensure a healthy evaluation.

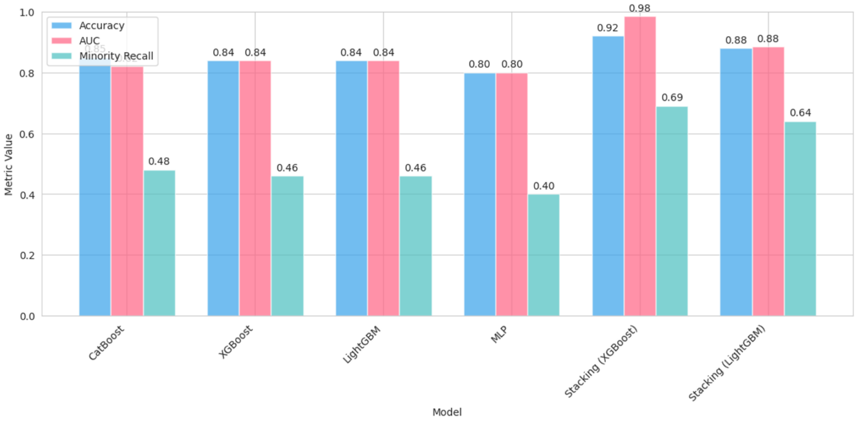

Fig. 6. Model Performance Comparison Bar Chart

Accuracies of base models vary from 0.80 to 0.85 (Table 2). CatBoost is the best with 0.85 accuracy and 0.82 AUC whereas XGBoost and LightGBM (0.84 each) come next, and MLP (0.80) showed in Figure 6. SMOTE did not increase minority class (Recurred/Progressed) recall because of imbalance in the classes.

Table 3. ENSEMBLED MODEL EVALUATION SUMMARY

Model Configuration	**Data Set (Sample size)**	**Accuracy**	**Precision**		**Recall**		**F1-Score**		**AUC**
			Disease Free	**Recurred**	**Disease Free**	**Recurred**	**Disease Free**	**Recurred**	
Stacking (XGBoost Meta Learner)	Full (508)	0.92	0.92	0.94	0.99	0.69	0.95	0.80	0.984
Stacking (LightGBM Meta-Learner)	Test (102)	0.88	0.90	0.78	0.95	0.64	0.93	0.70	0.884
Stacking (LightGBM Threshold 0.3483)	Test (102)	0.86	0.88	0.82	0.90	0.82	0.90	0.81	0.884

The stacking ensemble performance improves with increase in the memory size (Table 3). The XGBoost stacking ensemble having an XGBoost meta-learner had the highest accuracy and AUC (0.92 and 0.9844) on the entire dataset (Figure 7), with a recall of 0.69 with the minority classes (F1-score: 0.80) and DiseaseFree recalling 0.99. It is robust on the test set using a LightGBM meta-learner with an accuracy of 0.88, AUC of 0.8847, and minority recall of 0.64.

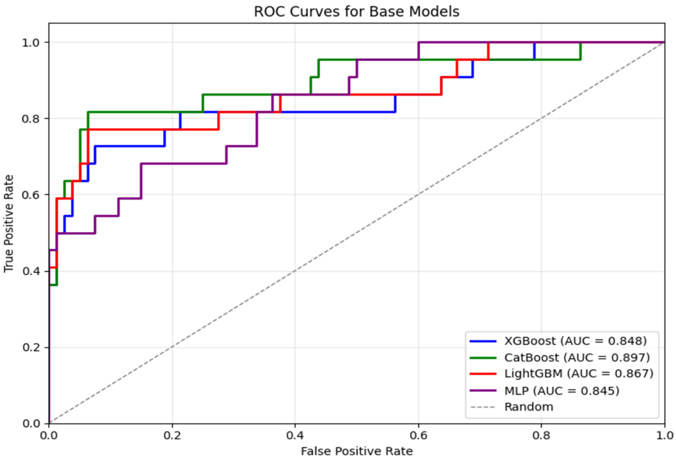

Fig.7. ROC Curve for base model

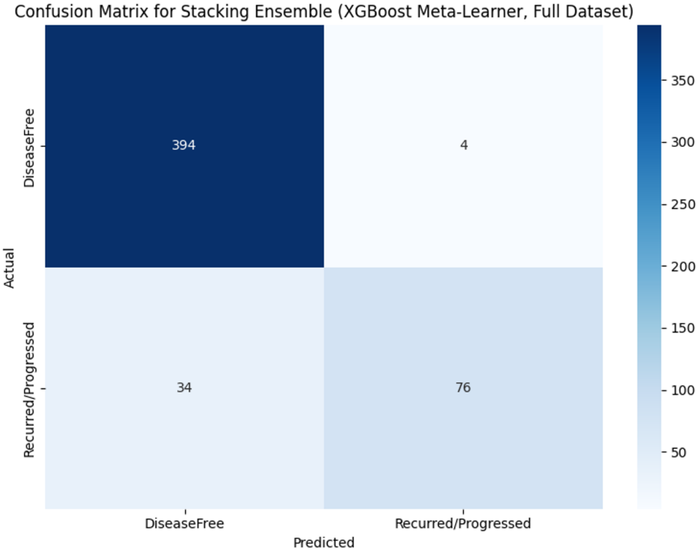

Fig. 8. Confusion Matrix for the best performancing model

A more optimal threshold of 0.3483 (LightGBM metalearner, test set visualized in Figure 9) increased minority class recall to 0.82 (28 percentage points higher), with an accuracy of 0.86, a F1-score of 0.81 and no change in the AUC (0.8847), increasing clinical sensitivity to recurrence detection.

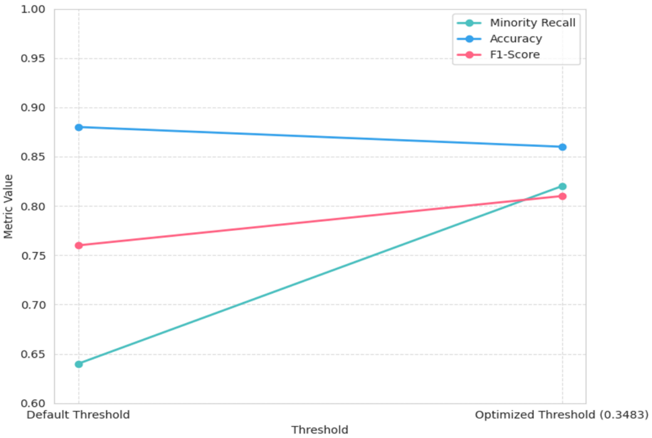

Fig. 9. Threshold Optimization plot of LightGBM meta-learner

The stacking ensemble, which relies on different base models and subtype-specific imputation, has a 7-12% higher accuracy than individual models, and high minority class performance (0.82 recall). These results support to integrate this model in practical world for early intercession because the high-risk patients have been identified.

The reduction in the accuracy by 0.92 to 0.88 between full and test set is indicative of slight overfitting, which is mitigated through test-set validation

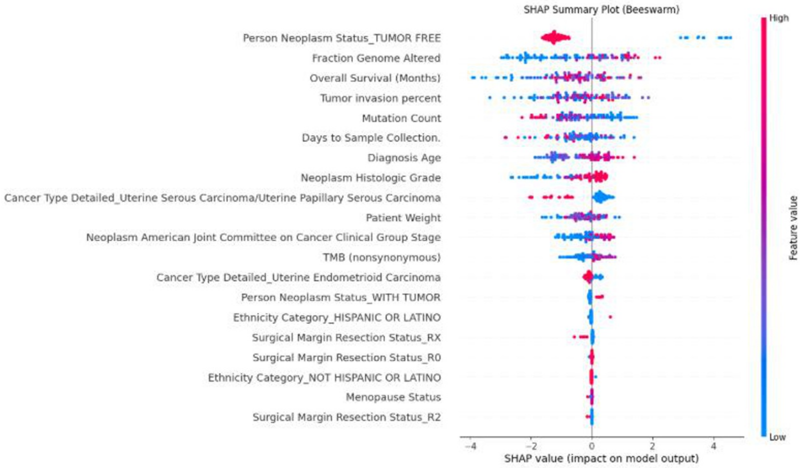

Fig. 10. SHAP Summary Plot

According to SHAP analysis from Figure 10, the most influential factors in the model predictions are tumor status, genomic alterations, and some clinical factors like number of months survived, and tumor invasion. Large genomic alteration and tumor invasion rates raise the risk and less survival decreases risks. This proves the fact that the model is working good based on the combination of biological and clinical data.

5 LIMITATIONS

Although the methodology is effective, it is limited because of the small sample which limits its ability to generalize to different populations. Subtype-specific imputation of KNN requires proper subtype classification which can be unsuccessful with clinical data of low quality. The computational complexity of the stacking ensemble, including a number of base models and hyperparameter optimization makes real-time deployment difficult. The imbalance of classes irrespective of SMOTE can bring in artificial sample biases which can influence real world performance. External validation and more cross-clinical data are required to make it applicable for clinical applications.

6 FUTURE WORK

In order to overcome the mentioned weaknesses future work may include increasing the dataset through more sources in order to increase sample size and diversity. In order to mitigate the synthetic biases of SMOTE, one can use generative adversarial networks (GANs) to oversample data. Real-time deployment of the stacking ensemble might be made possible by model compression or pruning. The strength would be enhanced by external validation on other independent clinical datasets across different populations. The use of multimodal data (e.g., imaging, proteomics) and deep survival model like cox-net, random survival forests etc. may improve the clinical validation.

7 CONCLUSION

The paper is a strong stacked ensemble model which combines XGBoost, CatBoost, LightGBM, and MLP on the UCEC TCGA to predict the recurrence of endometrial carcinoma with an XGBoost meta-learner. The model provided a macro F1-score of 0.79, ROC-AUC of 0.82, and minor class recall of 0.82 at a threshold of 0.3483 on the test set with the aid of subtype-specific KNN imputation, SMOTE to deal with class imbalance, and optimum preprocessing methods. Such findings portray high results in comparison to the personal base models, which illustrates the capacity of the ensemble to measure complex patterns of clinical and genomic information. The sensitivity of the model to recurrence detection was very high, which would facilitate clinical decision-making, as it is possible to identify the high-risk patients early and develop a personalized approach to treatment. Regardless of the drawbacks (small sample size and computational complexity), the suggested method provides a strong framework on how to enhance the disease-free survival prediction. Its use can be improved in clinical practice in the future by adding more data, more sophisticated methods of oversampling, and multimodal data integration.

References

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: a cancer journal for clinicians, vol. 71, no. 3, pp. 209–249, 2021.
2. Corr, B., Cosgrove, C., Spinosa, D., & Guntupalli, S. (2022). Endometrial cancer: molecular classification and future treatments. BMJ medicine, 1(1), e000152.
3. Smrz, S. A., Calo, C., Fisher, J. L., & Salani, R. (2021). An ecological evaluation of the increasing incidence of endometrial cancer and the obesity epidemic. American Journal of Obstetrics and Gynecology, 224(5), 506-e1.
4. Bokhman, J. V. (1983). Two pathogenetic types of endometrial carcinoma. Gynecologic oncology, 15(1), 10-17.
5. Nagel, J., Paschoalini, R. B., Barreto, P. S. D., Credidio, C. H., Paulino, E., & Del Pilar Estevez-Diz, M. (2024). Predictive biomarkers in endometrial carcinomas: a review of their relevance in daily anatomic pathology. Surgical and Experimental Pathology, 7(1), 21
6. Horvat, T. (2025). Evaluacija zadebljanog endometrija u postmenopauzalnih asimptomatskih pacijentica (Doctoral dissertation, University of Rijeka. Faculty of Medicine).
7. Nwachukwu, C., Baskovic, M., Von Eyben, R., Fujimoto, D., Giaretta, S., English, D., & Kidd, E. (2021). Recurrence risk factors in stage IA grade 1 endometrial cancer. Journal of gynecologic oncology, 32(2), e22.

8. Altın, D., Akıncı, T., Taşkın, S., & Ortaç, F. (2023). Predictors of recurrence and survival in lymphovascular space invasion negative early-stage endometrioid endometrial cancer patients. Journal of the Turkish German Gynecological Association, 24(4), 261.
9. Piedimonte, S., Feigenberg, T., Drysdale, E., Kwon, J., Gotlieb, W. H., Cormier, B., ... & Vicus, D. (2022). Predicting recurrence and recurrence-free survival in high-grade endometrial cancer using machine learning. Journal of Surgical Oncology, 126(6), 1096-1103.
10. Wang, X., Guan, J., Feng, L., Li, Q., Zhao, L., Li, Y., ... & Wang, X. (2024). A machine learning-based immune response signature to facilitate prognosis prediction in patients with endometrial cancer. Scientific Reports, 14(1), 30801.
11. Mercioni, M. A., & Holban, S. (2024, May). Uterine Corpus Endometrial Carcinoma Prediction from Genomic Analysis with Machine Learning. In 2024 International Conference on Development and Application Systems (DAS) (pp. 121-126). IEEE.
12. Liu, X., Wang, W., Zhang, X., Liang, J., Feng, D., Li, Y., ... & Ling, B. (2024). Metabolism pathway-based subtyping in endometrial cancer: An integrated study by multi-omics analysis and machine learning algorithms. Molecular therapy Nucleic acids, 35(2).
13. Zhu, C., Du, Y., Huai, Q., Fang, N., Xu, W., Yang, J., ... & Dai, Y. (2024). The identification of gamma-glutamyl hydrolase in uterine corpus endometrial carcinoma: a predictive model and machine learning. Reproductive Sciences, 31(2), 532-549.
14. Suman, S., & Kulshrestha, A. (2024). Integrative approach: Advancing endometrial cancer biomarkers through machine learning and network-derived RNA-Seq screening. Human Gene, 39, 201257.
15. Sheehy, J., Rutledge, H., Acharya, U. R., Loh, H. W., Gururajan, R., Tao, X., ... & Kondalsamy-Chennakesavan, S. (2023). Gynecological cancer prognosis using machine learning techniques: A systematic review of the last three decades (1990–2022). Artificial Intelligence in Medicine, 139, 102536.
16. cBioPortal: UCEC TCGA Clinical Dataset. cBioPortal for Cancer Genomics. https://www.cbioportal.org/study/summary?id=ucec_tcga (accessed: 09 Oct 2025)
17. Wolpert, D. H. (1992). Stacked generalization. Neural networks, 5(2), 241-259.
18. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., ... & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. Bioinformatics, 17(6), 520-525.
19. Fabian, P. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research 12, 2825.
20. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
21. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
22. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.

23. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.
24. Ahmed, F., Ahmed, F., Mubin, A. S., Rahman, H., & Ayon, R. (2024, June). Drug Addiction Analysis in Bangladesh Using Machine Learning. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.