International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

Explainable Artificial Intelligence based Ensemble Machine Learning for Ovarian Cancer Stratification using Electronic Health Records

Vivekanand Aelgani[1] , Dhanalaxmi Vadlakonda[2]
[1]Department Of Computer Science
Osmania University
Hyderabad, Telangana, India
aelgani.vivekanand@gmail.com
[2]Department of Mathematics
Osmania University
Hyderabad, Telangana, India
dhanalaxmivadlak@gmail.com

Abstract— The purpose of this study is to show how ensemble learning-driven machine learning algorithms outperform individual machine learning algorithms at predicting ovarian cancer on a biomarker dataset. Additionally, this study provides model explanations using explainable Artificial Intelligence methods, The method involved gathering and combining 49 risk factors from 349 patients. We hypothesize that ensemble machine learning systems are superior to individual Machine Learning systems in predicting ovarian cancer. The Machine Learning system consists of five individual Machine Learning and five ensemble Machine Learning systems were trained using K-10 cross validation protocols. These training models were then used to predict the development of benign ovarian tumors and ovarian cancer tumors patients. The AUC and Accuracy metrics for ensemble machine learning increased by 19% and 16%. The MCC and Kappa scores for ensemble Machine Learning also increased over individual machine learning by 29% and 33%, respectively. As a result, we draw the conclusion that ensembled-based algorithms outperform individual machine learning in terms of ovarian carcinoma prediction.

Keywords- Ovarian cancer, Tumor markers, deep-learning, Ensemble Machine Learning, XAI.

I. INTRODUCTION

As per the cancer Surveillance Branch of WHO report of 2022 , there were approximately 313,959 deaths attributed to ovarian cancer globally[1]. Ovarian cancer is a formidable disease that poses significant challenges to accurate diagnosis, treatment planning, and patient outcomes. The complexity of this malignancy, with its diverse histological subtypes and intricate molecular profiles, necessitates sophisticated approaches to classify and understand the disease. Traditional methods of ovarian cancer classification often rely on expert knowledge and subjective interpretation, hindering the ability to uncover the underlying mechanisms driving disease progression and response to therapy. More recently, the Explainable Artificial Intelligence (XAI) has revolutionized the field of machine learning and data-driven decision-making[2]. XAI techniques aim to bridge the gap between complex AI models and human interpretability, enabling insights into the reasoning behind algorithmic predictions. By providing transparency and comprehensibility, XAI methods offer the potential to enhance ovarian cancer classification, facilitating the understanding of disease mechanisms, identification of biomarkers, and improvement of clinical decision-making. This proposed study aims to develop explainable AI (XAI) driven predictive models to stratify ovarian cancer patients using Electronic Health Records (HER). Our goal is to highlight the potential of Explainable AI in addressing critical issues such as model interpretability, trustworthiness, and ethical considerations, while empowering clinicians with actionable insights for personalized patient care. To conduct our study, we evaluated the performance of three sets of feature groups, to evaluate their individual impact on the classifiers' performance.

The paper is organized into 6 sections, beginning with a brief review of existing literature on Ovarian cancer classification in Section 2. In section 3 we described the dataset used for the study and a brief description of predictive models and their performance evaluation metrics. Section 4 presents the experiment results and model predictions using XAI frameworks. In Section 5 we discussed the strengths and weaknesses of our study, and the paper concludes in section 6.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

II. LITERATURE REVIEW

Machine learning (ML) techniques have attracted a lot of attention recently in the healthcare industry, particularly in the creation of prognostic and diagnostic predictive models for cancer[3]. In order to predict the tumor size, Kawakami, Eiryo, et al[4] used supervised machine learning classifiers such as Gradient Boosting Machine, RF, Condition Random Fields, Naive Bayes, and SVM Neural Network, and Elastic Net. However, these models only achieved a 69% accuracy score.

E. S. Paik, et al[5] accurately predicted the cancer stages with a score of about 83% using a four-staged OC, histological data, various primary treatments, and chemotherapy regimen data.

In a recent study Akazawa et al [6]. used a variety of models, including Support Vector Machine, Naive Bayes, Boost, Logistic Regression, and Random Forest, to perform machine learning-based analysis. They found that the extreme gradient boost algorithm performed better than the other competing models, with an accuracy score of about 80%. However, this investigation demonstrated a strong correlation between the size of the feature set and its sensitivity. When the number of features decreases, there is a significant decline in accuracy, amounting to approximately 60%. The fact that this work only included 16 different blood parameters is another disadvantage. With biomarkers like blood specimen, general chemistry analytical tests, and OC biomarkers, M. Lu et al[7] achieved impressive accuracy scores during the validation phase. However, their performance on the testing phase showed lower accuracy levels, which suggests the occurrence of over-fitting. Hence, there is an urgent requirement for a robust framework that utilizes machine learning and statistical analysis to categorize individuals with ovarian cancer based on biomarker features. The prediction of ovarian cancer has been the focus of numerous research studies, but the accuracy ratings currently,

are insufficient indicating the need for improvement. Additionally, the data in none of these studies has been categorized using standards like blood specimen, general chemical analysis tests, and OC biological markers. Consequently, we initiate the process by segregating the data. Our method of data analysis combined statistical and machine learning methodologies, unlike earlier studies that only used statistical methods. This innovative method added a fresh viewpoint to the research and improved the validity of clinical testing, which might ultimately help patients and doctors. We hypothesis that due to hierarchical relationships between Ovarian Cancer biomarkers, ensemble machine learning algorithms outperform individual machine learning algorithms in predicting Ovarian cancer. The major contributions of the study are given below.

- Predicting ovarian tumor using tumor biomarkers at initial stage.
- Identifying significant biomarkers that influence predicted variables.
- Training independent and ensembled machine learning models on biomarker dataset and validating on test data using performance metrics and statistical techniques.
- Apply XAI techniques to explain the best performing model predictions.
- different types of cancer, building it an innovative learn with significant promise for the prior diagnosis of cancer types.

III. MATERIAL & METHODS

For this research, we employed a clinically validated dataset that included samples from both benign ovarian tumors and patients with malignant ovarian cancer. Subsequently, we performed statistical analysis to identify the most crucial biomarkers that are strongly linked to malignancy. Furthermore, machine learning classification models were utilized to identify and detect ovarian cancer at its initial stages. **Data Collection** The dataset used for the proposed study was originally published by soo-chow university, China. This dataset has 49 features corresponding to the 349 patients who were diagnosed during the period of July 2011-July 2018. Out of 349 patients 178 were non-ovarian tumor patients and 171 were diagnosed as suffering with Ovarian tumor. The whole dataset is clustered into three groups of 1. Blood sample test features, 2. General chemical Test features 3. Biological tumor markers. The dataset can be accessed publicly using the link https://github.com/martuzaiu/.

Data Preprocessing The procured dataset was preprocessed to eliminate the noise and missing values. The missing values in the feature columns were replaced with the mean of the feature column. To improve the performance of ML strategies all the features are scaled on to a common scale using Min-Max Scalar. The transformation function is described in equation 1.

$$A_{scaled} = \frac{A - A_{min}}{A_{max} - A_{min}} \qquad Eq.\ (1)$$

Where A is the feature column and Amin is the minimum value in the feature column and Amax is the maximum value in the feature column. A_{scaled} is the scaled feature.

Ovarian Carcinoma prediction: Predictive machine learning models can be used to predict ovarian carcinoma (OC) tumors to aid in diagnosis and prediction. This subsection briefly describes various predictive machine learning models used for OC-Tumor prediction. **Support Vector Machines** (SVM) is a predictive machine learning model that that finds the optimal hyperplane that divides instances into different classes by projecting the data to a higher-dimensional feature space. SVM

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

has been applied to ovarian cancer classification using various features such as gene expression profiles or imaging data[8, 9]. **Logistic Regression** is a predictive machine learning model used for binary classification. It computes the probabilities for each target class based on some predictors. Logistic Regression has been applied to ovarian cancer classification using features such as micro array data, clinical data, or proteomic profiles[10, 11]. **k-Nearest Neighbors (k-NN)** predict the class of data points based on feature similarity scores. k-NN has been used for ovarian cancer classification using different features, including gene expression data or clinical variables[12, 13].

Decision Trees (DT) uses inverted tree like structure to predict the target class. Decision Trees have been used for ovarian cancer classification using different types of features, such as micro-array data or clinical variables[14-16]. **Naive Bayes** predictive machine model works on the principles of conditional probability to predict the class label. Naive Bayes has been used for ovarian cancer classification using features such as micro array gene expression data or clinical variables[17]. **Ensemble predictive machine learning** works by deriving a new classifier from a set of base classifiers that works better than the constituent base classifiers. In the context of ovarian cancer classification using biomarkers datasets, some ensemble algorithms that have been used in the study includes **Random Forest is an** ensemble predictive machine learning model that combines multiple decision trees predictions into a single predictive value. Random Forest has been applied to ovarian cancer classification using biomarker's dataset to improve accuracy and handle high-dimensional feature spaces[18-20]. **AdaBoost (Adaptive Boosting)** is an ensemble predictive machine learning model that assigns larger weights to samples that are difficult to predict and smaller weights to samples that are easy to classify. AdaBoost has been used for ovarian cancer classification with biomarkers datasets to enhance the classification performance by emphasizing difficult-to-classify sample[21]. **Gradient Boosting** ensemble predictive model that is built by combining weaker predictive models. The result is obtained by combining the output of weaker predictive models. XGBoost (eXtreme Gradient Boosting) and LightGBM have been used for ovarian cancer classification using biomarkers datasets to achieve high accuracy and handle large-scale datasets efficiently[22]. These ensemble algorithms can effectively leverage the biomarkers dataset to improve the, robustness, and generalization of OC-Tumor classification models.

IV. METHODOLOGY

The system architecture diagram of the study is shown in figure 1. It consists of training phase and Testing phase, implemented using Python-based Scikit-learn package. The ML algorithms in Scikit-learn package are divided into two groups: Individual machine learning algorithms(iML) and ensemble machine learning (eML). Conventional predictive models such as Logistic Regression (LR), Support Vector Machine (SVM), K-nearest Neighbors Classifier (KNN), Decision Tree Classifier (DT), and Naive Bayes Classifier (NB) belong to the individual Machine Learning category, while Light Gradient Boosting Machine (LGBM), Random Forest Classifier (RF), Gradient Boosting Classifier (GBC), Extreme Gradient Boosting (Xgboost), and Ada Boost Classifier (AdaBoost) fall under the ensemble Machine Learning category. The iML and eML models were trained and cross validated using k=4, k5 and k=10 cross validation protocols. Machine learning algorithms are tuned using grid search optimization techniques to find best hyper parameters for the ML-models. The iML and eML models were evaluated based accuracy, MCC and kappa scores. The best performing is then explained using LIME and SHAP explainable AI frameworks.

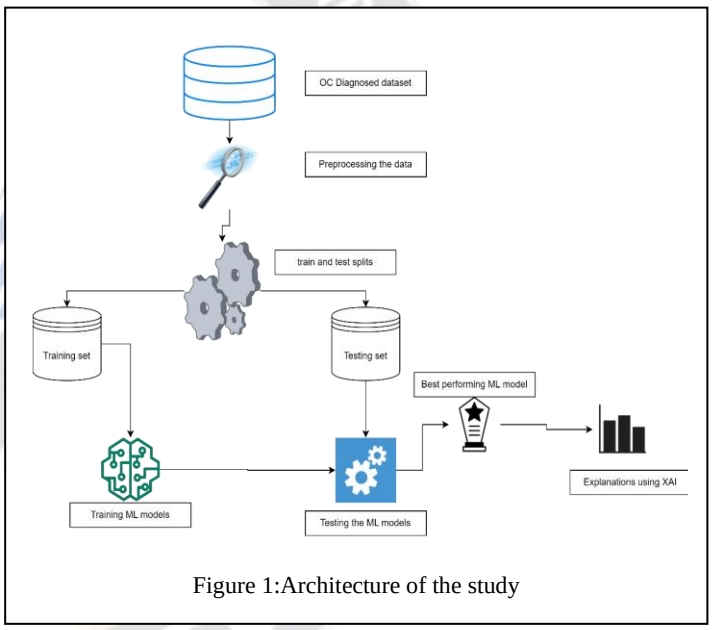

Figure 1:Architecture of the study

Handling Class Imbalance Skewed Classes is a major concern in healthcare datasets due to the uneven distribution of patients with or without ovarian cancer (OC) tumors. This imbalance can lead to bias in the trained predictive models which can further hamper the performance of the predictive machine learning models[23]. To address this issue, a widely used technique called -Synthetic Minority Oversampling technique (SMOTE)" was employed. SMOTE utilizes the nearest neighbor method to generate synthetic minority samples, ensuring they are distinct from existing samples[24]. To ensure unbiased training, SMOTE was exclusively applied to the training dataset and not on the testing dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

Model Evaluation Metrics The predictive model evaluation metrics are computed using the parameters **TP**, **TN**, **FP**, **FN**. The parameter True Positive (TP) is defined as the number of OC-Tumor(positive) samples that are correctly classified as OC-Tumors(positive). True Negative is defined as the number of benign samples (negative)that are correctly classified as BOT (Benign Ovarian Tumors). False Positive is defined as number of BOT (negative)samples that are misclassified as OC-Tumor (positive)samples. False Negative is defined as the number OC-Tumor (positive) samples that are misclassified as BOT sample(negative). ***Table 1*** lists the formula to compute the model evaluation metrics. The AUC-ROC represents the area under the ROC curve. The curve is drawn by taking false positive rate (FPR) on abscissa-axis and true positive rate on ordinate-axis. The use of Mathew Correlation Coefficient (MCC) in machine learning is significant because it can handle class imbalanced datasets. It gives a balanced assessment of the model's performance, considering both the true positive rate and true negative rate. In machine learning, the kappa score is employed to evaluate the performance of a classification model by comparing its predictions with the true labels. The kappa score lies in the interval -1 to +1, where 1 indicates perfect classified, 0 indicates classification is by random, and -1 indicates complete misclassification i.e model failed to classify correctly.

Table 1:Performance Evaluation Metrics

Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
F1-Score	$\frac{2\,P\times R}{P+R}$
Mathew Correlation Coefficient (MCC)	$\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Kappa score (KS)	$\frac{2\times(TP\times TN-FP\times FN)}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$

Explainable AI (XAI) plays a crucial role in healthcare by providing transparency and interpretability to AI-driven decision-making process. XAI is instrumental in ovarian cancer prediction by providing insights and explanations regarding the factors and patterns contributing to the predicted outcome[25, 26]. Ovarian cancer is a complex and challenging disease, and the ability to understand and interpret the predictions made by AI models is of utmost importance in the healthcare domain. XAI in ovarian cancer prediction facilitates the interpretation of AI models, enables identification of risk factors, ensures transparency and trust, aids in error detection, and empowers patients. By integrating explainable AI techniques into the prediction process, healthcare professionals can harness the power of AI while maintaining interpretability and insights that are crucial for effective clinical decision-making and patient care. The popular XAI techniques are LIME and SHAP. **LIME (Local Interpretable Model-agnostic Explanations)** approximates the behavior of complex models locally by creating interpretable surrogate models around specific instances. It generates local explanations that highlight the features and their contributions to the predictions[27]. **SHAP (Shapley Additive Explanations)** is based on cooperative game theory and assigns importance values to each feature by considering all possible feature subsets. It provides a unified framework for explaining the output of any model by decomposing the prediction into the contributions of individual features[28]. In this study we employed **LIME** and **SHAP** techniques to explain the feature contribution to the predicted outcome.

V. EXPERIMENTAL RESULTS

Performance metrics of individual machine learning algorithms (iML) and ensemble Machine learning algorithms (eML) are listed in *Table 2* and *Table 3* respectively. In the iML category, Logistic regression was the best performing. model, with an individual accuracy of 74.15% and AUC score equal to 0.8229, whereas in the eML category the best performing model was Light Gradient Boosting Machine with an accuracy of 91% and AUC score equal to 0.9529. *Table 4* presents the comparison between mean accuracy and mean AUC score of iML and eML. *Table-5* shows that, ensemble machine algorithms are superior to individual machine learning algorithms and thus validating our hypothesis that due to hierarchical relationships between OC biomarkers, ensemble machine learning algorithms outperforms individual predictive machine learning algorithms in predicting Ovarian carcinoma tumors based on biomarkers.

Table 2:Performance Evaluation Metrics of individual Machine Learning algorithms

iML	ACC	AUC	R	P	F1	KS	MCC
LR	0.74	0.82	0.75	0.74	0.74	0.48	0.49
SVM	0.71	0.60	0.74	0.75	0.70	0.41	0.43
KNN	0.72	0.79	0.77	0.71	0.74	0.44	0.45

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

DT	0.71	0.71	0.75	0.71	0.73	0.43	0.44
NB	0.70	0.75	0.81	0.67	0.73	0.39	**0.57**

Table 3:Performance Evaluation Metrics of Ensemble Machine Learning Algorithms

eML	ACC	AUC	R	P	F1	KS	MCC
LGBM	0.91	0.95	0.94	0.89	0.91	0.82	0.83
RF	0.90	0.95	0.94	0.88	0.90	0.80	0.81
GBM	0.89	0.95	0.93	0.88	0.90	0.79	0.79
XgBoost	0.88	0.94	0.90	0.87	0.88	0.75	0.76
AdaBoost	0.85	0.9	0.86	0.84	0.85	0.7	0.70

Table 4:ACC and AUC comparison

Metric	iML(a)	eML(b)	%Increase= (b-a)X100
Accuracy	0.72	0.88	16%
AUC	0.74	0.93	19%

Table 5 MCC and Kappa Score Comparison

Metric	iML(c)	eML(d)	% Increase= (d-c) X100
Mean MCC	0.47	0.76	29%
Mean KS	0.43	0.76	33%

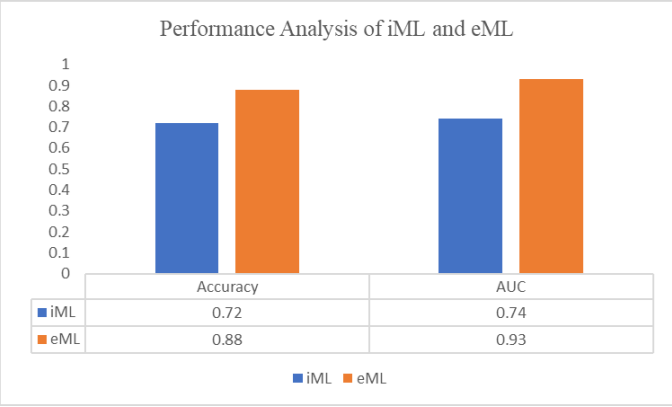

Figure 2: Bar chart showing the performance of iML and eML.

Model Explanations using LIME-The best performing ensemble ML was Light Gradient Boosting (LGBM) model. LGBM is not an inherently interpretable model and based on the number of estimators, depth of the tree or other hyper parameters the complexity of the algorithm might vary. LGBM model predictions are explained using XAI framework LIME and shown in figure 3.

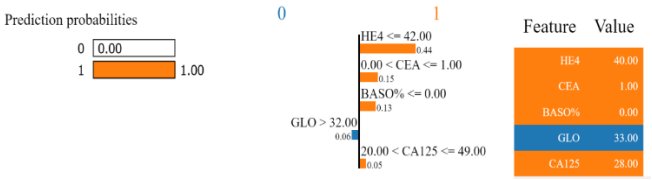

Figure 3:LIME Explanations

The left-most visualization from Figure 3 shows a range of possible values and the position of the model's predicted outcome. Intuitively speaking, all model predictions should lie within the minimum and the maximum possible value as this indicates to the user to compare the current forecast with the best-case and the worst-case values. The middle part of the visualization shows which features contribute to the prediction being on the higher side or the lower side. Considering our prior knowledge of Ovarian, a higher HE4, as well as CEA and BASO levels, do indicate increasing progression of the Ovarian tumor. The right-most visualization in Figure 3 shows us the actual local data values for the most important features identified, arranged in descending order of their relevance.

Model Explanations using SHAP- We also used SHAP force plots to explain local inference data (record #=1). With force plots, we can see the model prediction, which is denoted by f(x), as shown in *Figure 4*. The base value in the diagram represents the average predicted outcome of the model. The features colored pink tries to increase the model prediction and the features colored with BLUE try to decrease the predicted value. The HE4 and NEU biomarkers of this patient are responsible for predicting the outcome as negative (Benign Tumor).

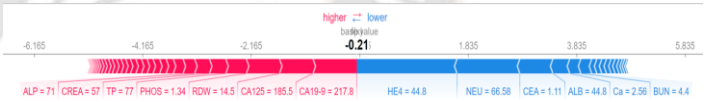

Figure 4:SHAP-Force Plot

Global interpretability with summary plots- A summary plot is a visualization method in SHAP for providing global explain-ability of black-box models. *Figure 5* shows the SHAP summary plot which displays the most significant feature that impact LGBM model predictions. Low values of HE4 and high ALB values are responsible for positive prediction by the LGBM model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

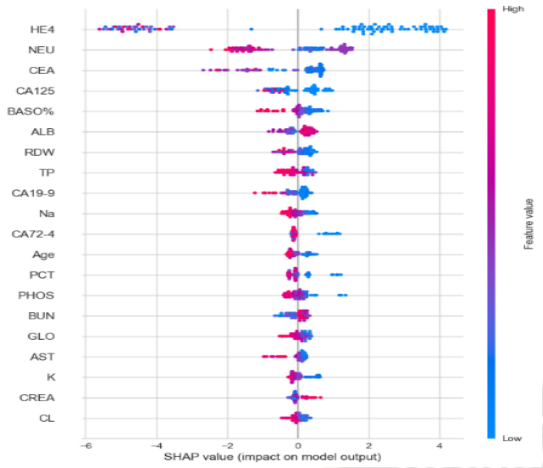

Figure 5:SHAP Summary Plot

VI. DISCUSSION

Detecting ovarian cancer early can decrease the mortality rate by prolonging survival. We proposed an explainable approach for early-stage detection of ovarian carcinomas, identifying varying groups of biomarkers associated with the occurrence of the disease. We followed the traditional approach of building Machine Learning models and integrated with XAI to explain the model prediction. At first, we performed exploratory data analysis (EDA) to clean and understand the data spread. Data imputation is done by replacing missing values with the mean value of the feature. We normalized all features by applying Minimax scaling technique and. Class imbalance issue is addressed by applying SMOTE technique. We then partitioned the dataset in to train and test in the ratio of 90:10. Afterwards we trained 5 individual Machine learning models (iML) and 5 ensemble Machine learning models(eML) and compared on accuracy, AUC, MCC and kappa scores. We found that eML outperformed iML models in prediction ovarian carcinoma with an increase of 19% in accuracy and 16% AUC value of eML. The increase in MCC and kappa scores of eML (Table-8) justifies that fact eML models are better at classifying OC-tumors than iML. . Lastly, we applied LIME and SHAP explainable-AI techniques to explain the best performing eML model LGBM. Our investigation suggests that XAI driven ensemble machine learning models are superior to solo machine models and to the best of our knowledge the proposed study is a novel approach. We benchmark our work with previous studies in ***Table 7*** to prove the superiority of our study.

Table 6:Comparative Study

Reference	Classifiers	ACC	AUC	MCC	KS	XAI?
M.Lu	DT	0.87	0.80	-	-	NO
Akazawa.	XGBoost	0.80	0.80	-	-	NO
Martínez	SVM, ELM	0.87	0.85	-	-	NO
Md.Ahamad	RF, GBM	0.88	0.85	-	-	NO
Proposed	eML	**0.91**	0.95	0.83	0.82	YES

The major hurdle in our study is the availability of experimental data. The study was carried out with only 349 patients EHR data. Thus, it may not be feasible to generalize the conclusions of our study. However, the proposed system is a good interpretable predictive model for ovarian carcinoma which can gain the trust of healthcare professionals and increases the chance of clinical deployment of the AI-models.

VII. CONCLUSION

This study introduces a new approach to classify ovarian carcinoma using ensemble explainable AI algorithms. We pre-processed a combined dataset and trained machine learning models to identify crucial biomarkers in the initial diagnosis of ovarian carcinoma patients. The most significant biomarkers associated with ovarian cancer are HE4, CEA, CA125, NEU, PHOS, GLO, CA19-9, ALB, and BASO ratio. The study demonstrates that ensemble Machine Learning classifiers exhibit high accuracy and model stability, suggesting their potential for computer-aided clinical diagnostics as a cost-effective method to assist healthcare professional in analyzing ovarian cancer. An important benefit of this research is that it enhances trust among clinicians and physicians through model interpretability. However, the main constraint of the study is the scarcity of data. In future research, we plan to include a larger dataset, including a validation cohort of patients, to further explore ovarian cancer.

VIII.ACKNOWLEDGEMENTS

We would like to thank Osmania University for supporting this research. We would also like to thank our colleagues from CMR College of Engineering & Technology who gave insight and knowledge that considerably aided the research. We appreciate Dr.V.A.Narayana, for assistance with the research and Mr. Chandrashekhar Reddy for reviewing the paper which really improved the manuscript.

REFERENCES

[1] H. Sung et al., "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] W. Guan et al., "Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines," vol. 10, no. 1, pp. 1-15, 2009.

[3] A. M. Alqudah, "Ovarian cancer classification using serum proteomic profiling and wavelet features a comparison of machine learning and features selection algorithms," Journal of Clinical Engineering, vol. 44, no. 4, pp. 165-173, 2019.

[4] E. Kawakami et al., "Application of Artificial Intelligence for Preoperative Diagnostic and Prognostic Prediction in Epithelial Ovarian Cancer Based on Blood BiomarkersArtificial Intelligence in Epithelial Ovarian Cancer," Clinical cancer research, vol. 25, no. 10, pp. 3006-3015, 2019.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v11i7.7832
***Article Received**: 20 April 2023 **Revised**: 15 June 2023 **Accepted**: 03 July 2023*

[5] E. S. Paik et al., "Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods," Journal of gynecologic oncology, vol. 30, no. 4, 2019.

[6] M. Akazawa and K. Hashimoto, "Artificial intelligence in ovarian cancer diagnosis," Anticancer research, vol. 40, no. 8, pp. 4795-4800, 2020.

[7] Rastogi, A. K. ., Taterh , S. ., & Kumar, B. S. . (2023). Dimensionality Reduction Approach for High Dimensional Data using HGA based Bio Inspired Algorithm. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 227 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2621

[8] M. Lu et al., "Using machine learning to predict ovarian cancer," International journal of medical informatics, vol. 141, p. 104195, 2020.

[9] C. Lu, T. Van Gestel, J. A. Suykens, S. Van Huffel, I. Vergote, and D. Timmerman, "Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines," Artificial Intelligence in Medicine, vol. 28, no. 3, pp. 281-306, 2003.

[10] J. Dong and M. Xu, "A 19-miRNA Support Vector Machine classifier and a 6-miRNA risk score system designed for ovarian cancer patients Corrigendum in/10.3892/or. 2019.7385," Oncology reports, vol. 41, no. 6, pp. 3233-3243, 2019.

[11] Z. Zhang and Y. Han, "Detection of ovarian tumors in obstetric ultrasound imaging using logistic regression classifier with an advanced machine learning approach," IEEE Access, vol. 8, pp. 44999-45008, 2020.

[12] J. Kaijser et al., "Improving strategies for diagnosing ovarian cancer: a summary of the International Ovarian Tumor Analysis (IOTA) studies," Ultrasound in obstetrics & gynecology, vol. 41, no. 1, pp. 9-20, 2013.

[13] V. Wibowo, Z. Rustam, S. Hartini, F. Maulidina, I. Wirasati, and W. Sadewo, "Ovarian cancer classification using K-Nearest Neighbor and Support Vector Machine," in Journal of Physics: Conference Series, 2021, vol. 1821, no. 1: IOP Publishing, p. 012007.

[14] B. Wu et al., "Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data," Bioinformatics, vol. 19, no. 13, pp. 1636-1643, 2003.

[15] A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, "Diagnosis of ovarian cancer using decision tree classification of mass spectral data," Journal of Biomedicine and Biotechnology, vol. 2003, no. 5, pp. 308-314, 2003.

[16] A. Osmanović, L. Abdel-Ilah, A. Hodžić, J. Kevric, and A. Fojnica, "Ovary cancer detection using decision tree classifiers based on historical data of ovary cancer patients," in CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017, 2017: Springer, pp. 503-510.

[17] M.-H. Tsai, H.-C. Wang, G.-W. Lee, Y.-C. Lin, and S.-H. Chiu, "A decision tree based classifier to analyze human ovarian cancer cDNA microarray datasets," Journal of medical systems, vol. 40, pp. 1-8, 2016.

[18] S. K. Maliha, R. R. Ema, S. K. Ghosh, H. Ahmed, M. R. J. Mollick, and T. Islam, "Cancer disease prediction using naive bayes, K-nearest neighbor and J48 algorithm," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019: IEEE, pp. 1-7.

[19] H. Zhang et al., "A random forest-based metabolic risk model to assess the prognosis and metabolism-related drug targets in ovarian cancer," Computers in Biology and Medicine, vol. 153, p. 106432, 2023.

[20] L. Cheng, L. Li, L. Wang, X. Li, H. Xing, and J. Zhou, "A random forest classifier predicts recurrence risk in patients with ovarian cancer," Molecular Medicine Reports, vol. 18, no. 3, pp. 3289-3297, 2018.

[21] A. Arfiani and Z. Rustam, "Ovarian cancer data classification using bagging and random forest," in AIP Conference Proceedings, 2019, vol. 2168, no. 1: AIP Publishing.

[22] B. Yesilkaya, M. Perc, and Y. Isler, "Manifold learning methods for the diagnosis of ovarian cancer," Journal of Computational Science, vol. 63, p. 101775, 2022.

[23] Y.-W. Hsiao, C.-L. Tao, E. Y. Chuang, and T.-P. Lu, "A risk prediction model of gene signatures in ovarian cancer through bagging of GA-XGBoost models," Journal of advanced research, vol. 30, pp. 113-122, 2021.

[24] Mr. Rahul Sharma. (2018). Monitoring of Drainage System in Urban Using Device Free Localization Neural Networks and Cloud computing. International Journal of New Practices in Management and Engineering, 7(04), 08 - 14. https://doi.org/10.17762/ijnpme.v7i04.69

[25] X. Yang, M. Khushi, and K. Shaukat, "Biomarker CA125 feature engineering and class imbalance learning improves ovarian cancer prediction," in 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2020: IEEE, pp. 1-6.

[26] V. Aelgani, D. Vadlakonda, and V. Lendale, "Performance analysis of predictive models on class balanced datasets using oversampling techniques," in Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020, Volume 1, 2021: Springer, pp. 375-383.

[27] Omondi, P., Ji-hoon, P., Cohen, D., Silva, C., & Tanaka, A. Deep Learning-Based Object Detection for Autonomous Vehicles. Kuwait Journal of Machine Learning, 1(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/149

[28] A. Abdollahi and B. Pradhan, "Urban vegetation mapping from aerial imagery using explainable AI (XAI)," Sensors, vol. 21, no. 14, p. 4738, 2021.

[29] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Incorporating Explainable Artificial Intelligence (XAI) to aid the Understanding of Machine Learning in the Healthcare Domain," in AICS, 2020, pp. 169-180.

[30] M. T. Ribeiro, S. Singh, and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135-1144.

[31] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in neural information processing systems, vol. 30, 2017.