PREDICTION OF OVARIAN CANCER USING EXPLAINABLE AI: A REVIEW

Vidyasre N [1], Manasa T P [2]

[1] *Student, Bangalore Institute of Technology, Karnataka, India*
[2] *Assistant Professor, Dept. of Computer Science & Engineering, Bangalore Institute of Technology, Karnataka, India*

***

Abstract - *The prevention and early detection of ovarian cancer are critical for improving women's health and well-being. Ovarian cancer, often termed the "silent killer," typically presents without apparent symptoms in its early stages, making early detection challenging and significantly reducing the chances of successful treatment and survival. Machine Learning (ML) techniques offer powerful tools for identifying complex patterns in biomarkers, surpassing conventional diagnostic methods in accuracy. However, these advanced algorithms often lack transparency, making it difficult to gain physical insights into the diagnostic process. To address this, integrating eXplainable Artificial Intelligence (XAI) can enhance our understanding of how ML models make decisions, providing valuable insights into the diagnostic process. This paper aims to showcase best practices for combining XAI and ML methods in biomarker validation applications, ultimately improving early detection and treatment outcomes for ovarian cancer.*

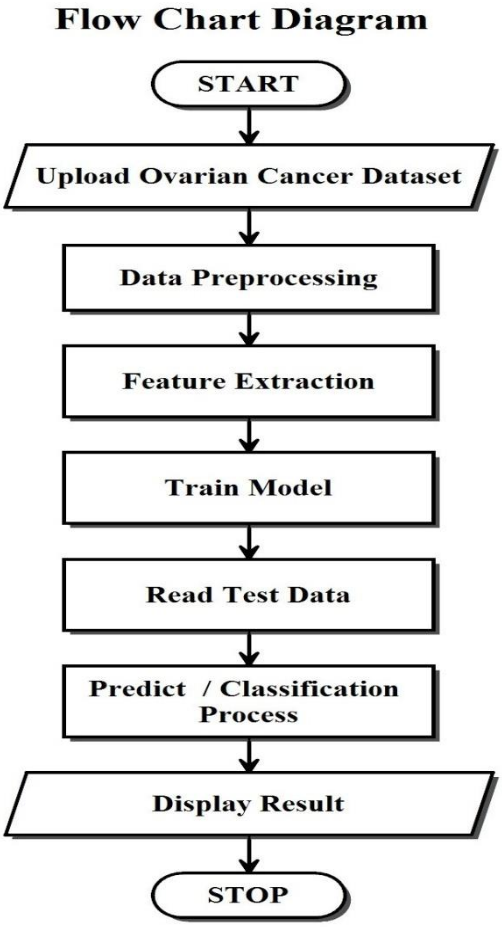
***Key Words*: Ovarian Cancer, Machine Learning Algorithms, eXplainable Artificial Intelligence (XAI), Biomarker Validation, Early Detection.**

1. INTRODUCTION

Ovarian cancer often referred to as the "silent killer," poses a significant threat to women's health due to its typically asymptomatic early stages, which frequently result in delayed diagnosis and decreased treatment effectiveness. The undetected nature of this disease underscores the urgent need for improved detection techniques. Early detection is paramount for enhancing treatment effectiveness and achieving better patient outcomes. To address this critical issue, our study proposes a novel approach for predicting ovarian cancer. This study provides a comprehensive analysis of the development and evaluation process of the proposed model. By leveraging the strengths of both boosting and bagging techniques, our approach overcomes the limitations of individual classifiers. Additionally, the interpretability of the model is enhanced through the application of eXplainable AI (XAI) techniques, such as SHapley Additive Explanations (SHAP), which offer valuable insights into the risk factors for ovarian cancer. This study aims to advance the primary objective of improving patient outcomes and reducing ovarian cancer-related mortality rates. The subsequent sections delve into the methodology, results, and implications of our approach, providing a foundation for future advancements in the detection and management of ovarian cancer.

2. FLOW DIAGRAM

A flowchart visually represents a process, using standardized symbols to illustrate steps, decisions, and flow of information. It serves to clarify processes, improve understanding, and facilitate communication in various fields, from engineering to business and healthcare.



The flow chart diagram outlines the process of predicting and classifying ovarian cancer using machine learning techniques, starting with the initiation of the workflow. The first step involves uploading the ovarian cancer dataset, followed by data preprocessing, where the raw data is cleaned and prepared for analysis. Next, feature extraction identifies significant variables that contribute to the prediction of ovarian cancer. The machine learning model is then trained using the extracted features. After training, the model is evaluated using a separate test dataset to assess its performance. The trained model then makes predictions or classifications based on the test data, determining the likelihood of ovarian cancer. Finally, the results of the prediction or classification process are displayed, and the workflow concludes.

3. LITERATURE REVIEW

Research in the field of ovarian cancer detection has investigated a collection of approaches and strategies with the objective of enhancing the outcomes of early diagnosis and treatment.

This research focuses on improving tumor classification as malignant or benign, minimizing false positives through various machine learning methods, including logistic regression, Random Forests, decision trees, SVM, KNN, and deep learning techniques [1]. The goal is to predict Critical Care Unit (CCU) admission more accurately for ovarian cancer patients undergoing surgery, with a graphical user interface for real-time analysis [2].

The study compares premenopausal and postmenopausal women using HE4, CA125, and ROMA markers to differentiate between ovarian cancer and benign masses, also exploring the impact of age and cutoff values on marker performance [3]. It proposes a hybrid approach combining CNN and ReliefF algorithms for microarray data analysis to enhance cancer detection accuracy [4]. Challenges such as high dimensionality and small sample sizes are addressed.

The research employs Raman spectroscopy and machine learning to differentiate ovarian cancer from cysts and normal cases, aiming to improve diagnostic accuracy [5]. It also develops a predictive model for mortality risk using gene expression data, combining genetic algorithms and extreme gradient boosting [6]. The study uses statistical analysis and ensemble models to identify important biomarkers for distinguishing cancer from benign tumors [7].

Machine learning models developed on the SEER database predict survival rates, with techniques including SVM, KNN, DT, RF, AdaBoost, and XGBoost [8]. The study suggests optimizing feature weights and classifier parameters with LASSO regularization and Asynchronous Differential Evolution (ADE) to enhance detection efficiency [9]. It also creates a model to forecast progression-free survival (PFS) at 12 months, emphasizing the need for larger sample sizes for validation [10].

The research aims to build a predictive model for ovarian cancer detection, evaluating biomarkers and algorithms to compare with conventional methods and provide clinical insights [11]. It explores dimensionality reduction and classification approaches to improve diagnostic accuracy, using techniques like AdaBoost, Decision Trees, and t-SNE [12]. Despite a limited sample size and potential biases, the study recommends further research using heuristic algorithms [13].

SVM outperforms KNN for diagnosis based on medical histories, with future research recommended to integrate advanced techniques like fuzzification and transfer learning [14]. To enhance diagnostic accuracy, the study develops Ocys-Net, a deep learning network for categorizing ovarian cysts from ultrasound images [15]. It also assesses various machine learning methods and compares Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) for diagnosing ovarian cancer from colposcopy images [18].

Table 1: Summary of Research Work on Ovarian Cancer Detection

Sl.No.	Methodology	Advantages	Limitations
1	Logistic Regression, Random Forests, Decision Trees, SVM, KNN, Deep Learning	Reduces false positives, improved categorization	Possible misclassifications, limited sample size
2	Machine Learning for CCU admission prediction	Greater accuracy for predicting CCU admission	Focused on specific cancer type (HGSOC)
3	Comparison of HE4, CA125, and ROMA markers	Insights into clinical relevance of markers	Potential publication bias, age and cutoff value impact
4	Hybrid approach using CNN and ReliefF algorithms for microarray data	Decreases mistakes, increases precision	High dimensionality, small sample size may affect validity
5	Raman spectroscopy and machine learning	Increases diagnostic accuracy	Challenges in data pretreatment and feature extraction
6	GA-XGBoost models for mortality risk	Reliable mortality risk estimation based on gene	Small sample size, limited generalizability

Sl.No.	Methodology	Advantages	Limitations
	prediction	expression	
7	Statistical analysis and ensemble models for biomarker identification	Identifies important biomarkers	Limited sample size, no control group, possible misclassifications
8	SEER database, SVM, KNN, DT, RF, AdaBoost, XGBoost for survival prediction	Collaborative design with clinician, data preparation methods like SMOTE	Small sample size, need for larger datasets
9	LASSO regularization, cross-validation error, ADE for ovarian cancer detection	Improves model efficiency, generalization	Limited to SVM and KNN classifiers
10	RFE, K-Nearest Neighbors, Random Forest, Logistic Regression for PFS prediction	Encouraging outcomes despite small sample size	Needs confirmation on larger patient groups
11	ReliefF, MRMR for feature selection, machine learning for tumor differentiation	Evaluates biomarkers and algorithms for detection	Small sample size, potential biases in feature selection
12	Evaluation of dimensionality reduction and classification algorithms	Identifies effective combinations for diagnosis	Complexity in implementing multiple techniques
13	Feature selection and extraction for microarray gene data	Improves identification accuracy	Limited by sample size, data quality issues
14	SVM vs KNN for ovarian cancer diagnosis	SVM shown to be superior	Future research needed for integrating advanced techniques
15	Ocys-Net for categorizing ovarian cysts from ultrasound images	Better diagnosis accuracy, reduces doctors' burden	Limited to ultrasound images, needs further validation
16	Comparison of machine learning methods for ovarian cancer identification	Assesses classifier accuracy, improves early diagnosis	Focus on cross-validation and holdout techniques
17	SHAP algorithm for gene identification, Kaplan-Meier plots for validation	Identifies potential biomarkers	Needs further investigation for practical application
18	MLP and LSTM for colposcopy image diagnosis	Evaluates algorithm effectiveness for stage prediction	Limited to colposcopy images, needs broader validation

4. CONCLUSIONS

This study presents a comprehensive evaluation of ovarian cancer research, focusing on articles from 2021 to 2023. The survey covers various machine learning and deep learning approaches for improving the detection and classification of ovarian cancer. By comparing different biomarkers, feature selection strategies, and algorithmic techniques, this research highlights the advancements and challenges in the field. The findings emphasize the potential of machine learning models to enhance diagnostic accuracy and patient outcomes. This study serves as an initial step towards developing a more robust and explainable AI model for predicting ovarian cancer, ultimately aiming to improve early detection and treatment strategies.

REFERENCES

[1] Aditya, Ms, I. Amrita, Ashwini Kodipalli, and Roshan Joy Martis. "Ovarian cancer detection and classification using machine leaning." In 2021 5th international conference on electrical, electronics, communication, computer technologies and optimization techniques (ICEECCOT), pp. 279282. Ieee, 2021.

[2] Laios, Alexandros, Raissa Vanessa De Oliveira Silva, Daniel Lucas Dantas De Freitas, Yong Sheng Tan, Gwendolyn Saalmink, Albina Zubayraeva, Racheal Johnson et al. "Machine learning based risk prediction of critical care unit admission for advanced stage high grade serous ovarian cancer patients undergoing cytoreductive surgery: The LeedsNatal score." Journal of Clinical Medicine 11, no. 1 (2021): 87.

[3] Suri, Arpita, Vanamail Perumal, Prajwal Ammalli, Varsha Suryan, and Sanjiv Kumar Bansal. "Diagnostic measures comparison for ovarian malignancy risk in Epithelial ovarian cancer patients: A metaanalysis." Scientific reports 11, no. 1 (2021): 17308.

[4] Kilicarslan, Serhat, Kemal Adem, and Mete Celik. "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network." Medical hypotheses 137 (2020): 109577.

[5] Chen, Fengye, Chen Sun, Zengqi Yue, Yuqing Zhang, Weijie Xu, Sahar Shabbir, Long Zou et al. "Screening ovarian cancers with Raman spectroscopy of blood plasma coupled with machine learning data processing." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 265 (2022): 120355.

[6] Hsiao, YiWen, ChunLiang Tao, Eric Y. Chuang, and TzuPin Lu. "A risk prediction model of gene signatures in ovarian cancer through bagging of GAXGBoost models." Journal of advanced research 30 (2021): 113122.

[7] Ahamad, Md Martuza, Sakifa Aktar, Md Jamal Uddin, Tasnia Rahman, Salem A. Alyami, Samer AlAshhab, Hanan Fawaz Akhdar, A. K. M. Azad, and Mohammad Ali Moni. "Earlystage detection of ovarian cancer based on clinical data using machine learning approaches." Journal of personalized medicine 12, no. 8 (2022): 1211.

[8] Sorayaie Azar, Amir, Samin Babaei Rikan, Amin Naemi, Jamshid Bagherzadeh Mohasefi, Habibollah Pirnejad, Matin Bagherzadeh Mohasefi, and Uffe Kock Wiil. "Application of machine learning techniques for predicting survival in ovarian cancer." BMC medical informatics and decision making 22, no. 1 (2022): 345.

[9] Juwono, Filbert H., Wei Kitt Wong, Hui Ting Pek, Saaveethya Sivakumar, and Donata D. Acula. "Ovarian cancer detection using optimized machine learning models with adaptive differential evolution." Biomedical Signal Processing and Control 77 (2022): 103785.

[10] Arezzo, Francesca, Gennaro Cormio, Daniele La Forgia, Carla Mariaflavia Santarsiero, Michele Mongelli, Claudio Lombardi, Gerardo Cazzato, Ettore Cicinelli, and Vera Loizzi. "A machine learning approach applied to gynecological ultrasound to predict progressionfree survival in ovarian cancer patients." Archives of Gynecology and Obstetrics 306, no. 6 (2022): 21432154.

[11] Lu, Mingyang, Zhenjiang Fan, Bin Xu, Lujun Chen, Xiao Zheng, Jundong Li, Taieb Znati, Qi Mi, and Jingting Jiang. "Using machine learning to predict ovarian cancer." International journal of medical informatics 141 (2020): 104195.

[12] Yesilkaya, Bartu, Matjaž Perc, and Yalcin Isler. "Manifold learning methods for the diagnosis of ovarian cancer." Journal of Computational Science 63 (2022): 101775.

[13] Kalaiyarasi, M., and Harikumar Rajaguru. "Performance analysis of ovarian cancer detection and classification for microarray gene data." BioMed Research International 2022 (2022).

[14] Taleb, Nasser, Shahid Mehmood, Muhammad Zubair, Iftikhar Naseer, Beenu Mago, and Muhammad Umar Nasir. "Ovary cancer diagnosing empowered with machine learning." In 2022 International Conference on Business Analytics for Technology and Security (ICBATS), pp. 16. IEEE, 2022.

[15] Fan, Junfang, Juanqin Liu, Qili Chen, Wei Wang, and Yanhui Wu. "Accurate Ovarian Cyst Classification with a Lightweight Deep Learning Model for Ultrasound Images." IEEE Access (2023).

[16]Singh, Samridhi, Malti Kumari Maurya, and Nagendra Pratap Singh. "STRAMPN: Histopathological image dataset for ovarian cancer detection incorporating AI-based methods." Multimedia Tools and Applications 83, no. 9 (2024): 28175-28196.

[17] Schilling, Vincent, Peter Beyerlein, and Jeremy Chien. "A Bioinformatics Analysis of Ovarian Cancer Data Using Machine Learning." Algorithms 16, no. 7 (2023): 330.

[18] Ch, Chandana, and N. P. G. Bhavani. "Innovative Long ShortTerm Memory is a more accurate and efficient method of predicting the stage of ovarian cancer than MultiLayer Perceptron." In 2023 9th International Conference on Smart Structures and Systems (ICSSS), pp. 15. IEEE, 2023.