*Grenze International Journal of Engineering and Technology, June Issue*

# A Comprehensive Review on AI-Driven Prediction Model for Early Detection of Ovarian Cancer

Diksha D. Gabhane[1], Diksha Bhagwat[2], Aarohi Waghmare[3], Sanchyee Pendam[4], Vaishnavi Rathod[5] and Rajendra Rewatkar[6]
[1-5]Department of Computer Science & Design, Yeshwantrao Chavhan College of Engineering Nagpur -441110, India
Email: dikshagabhane29@gmail.com, bhagwatdiksha0408@gmail.com, aarohiwaghmare34@gmail.com, sanchyeependam22@gmail.com, rathodsanjayvaishnavi@gmail.com
[6]Faculty of Engineering and Technology, Datta Meghe Institute of Higher Education & Research Wardha

***Abstract*— This study investigates the application of deep learning (DL) and machine learning (ML) for the detection of ovarian cancer, which is frequently referred to as the "silent killer" because of its early asymptomatic phases. Ovarian cancer can be successfully predicted using machine learning (ML) models such as Random Forests, Support Vector Machines, K-Nearest Neighbours, and advanced convolutional neural networks (CNNs). Interestingly, a Random Forest approach obtained 98% accuracy, while CNNs optimised with Krill Herd Optimisation have overcome the shortcomings of earlier models. In order to improve patient outcomes, enable early intervention, and increase diagnostic accuracy, the study highlights the use of explainable AI (e.g., SHAP), ensemble learning, and feature engineering.**

***Keywords*— Early identification of ovarian cancer, biomarkers, explainable AI, machine learning, deep learning, BRCA mutations, predictive modeling, diagnostic accuracy, and ensemble learning.**

## I. INTRODUCTION

One of the deadliest gynecological malignancies, ovarian cancer is sometimes referred to as a "silent killer" since, in its early stages, it doesn't exhibit any signs [2]. Genetic mutations like BRCA1 and BRCA2 mutations are the main cause, although other risk factors include age, obesity, endometriosis, hormone imbalances, and a family history of breast or ovarian cancer. The risk may also rise because of exposure to specific environmental or lifestyle factors [3]. Since late-stage diagnosis drastically lowers treatment efficacy and survival rates, early detection is essential. Conventional diagnostic techniques, such as blood testing and imaging, frequently fall short in detecting the illness in its early, curable phases. One revolutionary solution to this problem is machine learning (ML). ML models can identify minor patterns and biomarkers linked to early-stage ovarian cancer by examining large and complicated datasets, such as genetic, clinical, and imaging data. High accuracy in differentiating between cancerous and non-cancerous instances has been shown using methods like random Forests, support vector machines (SVMs), and deep learning frameworks like convolutional neural networks (CNNs). Furthermore, explainable AI tools increase model openness, which facilitates clinician adoption and trust. There is a lot of potential for understanding the origins of ovarian cancer, enabling early intervention, increasing patient outcomes, and potentially saving lives [4].

*Grenze ID: 01.GIJET.11.2.169_20*

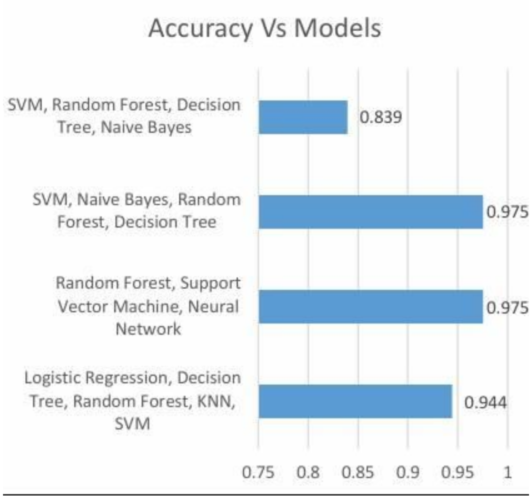


Figure 1

## II. AI Techniques for Predicting Ovarian Cancer

### *A. Data Sources and Preprocessing*

*Gathering Clinical Data*
The dataset comprised 47 patient parameters: 1 tumour biomarker (CA-125), 8 demographic factors (e.g., age, menopausal status, cancer stage), 14 blood factors (e.g., white blood cell count, lymphocyte ratio), 20 serum factors (e.g., alkaline phosphatase, albumin, glucose), and 3 plasma variables (e.g., prothrombin time, fibrinogen). There were very few (less than 10%) missing values, and the missing data was handled using simple mean imputation.[11]

### *B. Image processing*

Digitization artifacts and noise were common in ovarian MRI pictures. In order to maintain the margins of the images, filters were utilized for de-noising. A de-noising computation based on equation (1) was used in the suggested pre-processing steps. In this calculation, $Wxy$ stands for coordinates in a k×l rectangle sub-image window that is centered at (x,y). [7]

### *C. Models for Machine Learning*

Using a variety of methods, prediction models such as RFM, ANNM, DTM, GEM, and SVM were built. For iterative analysis, ANNM employed hierarchical input layers (such as tumor differentiation, LNM, CA125, HE4, Alb, and LH), whereas RFM and DTM created branching nodes using "bagging" and ensemble learning. GEM and SVM followed particular classification presumptions. ANNM had improved performance in predicting ovarian metastases in EC patients across both training and validation sets, according to decision curve analysis and receiver operating characteristic curves, which evaluated the model's accuracy and resilience. [17] The durability of ANNM in comparison to other models is demonstrated in Figure 5. [6]
Extreme Gradient Boosting, or XGBoost, is a potent and adaptable machine learning method that performs exceptionally well in problems involving both regression and classification. The model can better manage complicated datasets by focussing on misclassified data points by giving them larger weights and building a series of decision trees, fixing the mistakes of earlier trees to enhance predictions [2]. By creating artificial instances of the minority class, SMOTE (Synthetic Minority Over-sampling Technique) ensures a balanced distribution of classes and tackles class imbalance. Along the line segments that link a minority class data point to its closest neighbours, it generates additional data points [2].

## III. Performance Metrics

The Random Forest (RF) model achieved the best classification performance with an accuracy of 88.72% and an AUC of 82.38%, excelling in metrics such as PPV (71.91%), NPV (92.96%), recall (71.81%), specificity (92.95%), and F1-score (71.78%). SHAP analysis identified key factors like histologic type, chemotherapy, year of diagnosis, age, tumor stage, and grade. RF also demonstrated the shortest execution time. [10]

## IV. Applications and Advancements

### *A. AI in Early Warning and Risk Assessment*

Recent research emphasizes AI's contribution to early-stage diagnosis and risk assessment for ovarian metastases. In order to provide patients with endometrial cancer with early warning of ovarian metastases, Zhao et al. [6] created a predictive model utilizing machine learning techniques. Proactive intervention was made possible by the model's high sensitivity and specificity, which were attained by integrating clinical and imaging data. Especially for high-risk populations, these risk models have useful uses in routine testing.

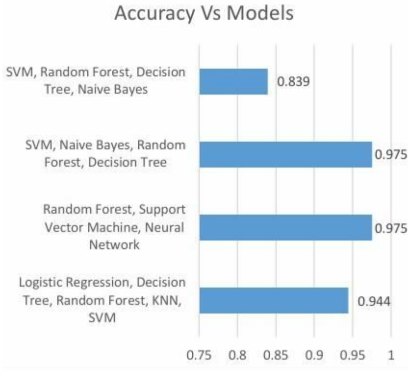


Figure 2

### *B. Classification of Tumors Applying Optimization Techniques and SVM*

A classification system utilizing Support Vector Machines (SVM) in conjunction with Particle Swarm Optimization (PSO) was proposed by Srilatha and Venugopal [7]. The accuracy and memory of separating benign from malignant ovarian tumors were greatly increased by this method. Through PSO, the hyperparameters were optimized, resulting in a robust and effective classification system. In clinical decision-making, when accurate classification is essential, this method is especially helpful

### *C. Ovarian Cancer Prediction Using Machine Learning*

In order to predict ovarian cancer, Kwatra and Arora [8] thoroughly investigated machine learning models, such as Random Forests, Gradient Boosting, and DL methods. Their research demonstrated how useful ensemble learning techniques are for improving prediction accuracy while resolving data imbalance. These developments can be incorporated into larger healthcare systems for population-level screening and improve early-stage detection

### *D. Customized Care and Survival Forecasting*

To predict survival outcomes and enable individualized treatment approaches for patients with ovarian cancer, machine learning models, such those developed by Sorayaie Azar et al. [10], use genetic, imaging, and clinical data.

### *E. Recurrence Prediction Based on Machine Learning*

By merging tumor characteristics, therapy information, and patient history, Zhou et al. [11][13] created a machine learning model that outperformed conventional techniques in predicting recurrence.

### *F. AI for Prognosis Using Multiple Modalities*

In order to enhance survival prediction for ovarian serous carcinoma and facilitate difficult decision-making, Zhou et al. [12] presented a multimodal AI model that integrates pathological pictures, clinical data, and genomic information

## V. Future Directions and Challenges

### *A. Overview*

Early detection of ovarian cancer is still difficult. Although problems like computational complexity, model interpretability, and data scarcity restrict clinical use, AI-driven diagnostic systems that use ML and DL provide solutions. In order to enhance prediction, future studies should concentrate on removing these obstacles. [10]

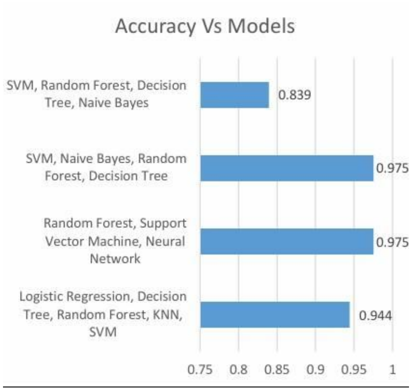


Figure 3

*B. Future Paths*

XAI Model Integration: By improving model transparency, XAI techniques (such as SHAP and LIME) support physician trust. Future research should develop specialized XAI tools for medical imaging and refine ensemble approaches. [1]
Data Augmentation: Data balance can be enhanced by SMOTE and dynamic augmentation. Another potential strategy is the use of GANs for cross-modality data synthesis. [2]
Hybrid ML and DL Models: By combining DL with conventional ML (like RF), ovarian cancer prediction can be enhanced. [3]
Multimodal Learning: Research on fusion methods and hierarchical learning focuses on how integrating various inputs (genomics, history, imaging) enhances diagnosis. [1, 4]
Real-Time Systems: Hardware optimization and edge AI are necessary for low latency and privacy in real-time AI for point-of-care applications. [1]
Customized Healthcare: By utilizing patient-specific genomic, lifestyle, and medical history data to inform forecasts, personalized treatment. [4]

*C. Challenges*

Data Availability and Quality: Progress is hampered by a lack of high-quality datasets, however SMOTE and cooperative efforts may help. [2] Model Generalizability: Federated learning and domain adaptation provide answers for AI models that may not generalize well across various populations. [4] Clinical Validation:to receive regulatory approval. [1] Ethics and Regulation: Discuss ethical and legal issues, such as data privacy and methods for reducing bias.
Computational Cost and Scalability: AI models require a lot of processing power. Model compression and cloud-based scalability are examples of solutions that are required. [3]

## VI. Result & Discussion

Ovarian tumors are cancers that develop in the ovaries. Since the ovary contains a wide variety of actively growing cells, there are numerous forms of tumors. The total number of characteristics in the ovarian dataset is shown in Figure 8.High impact features for classification are shown in figure 8 as red, while low impact features for classification are shown as blue. Figure 9 shows the key feature that was extracted using the SMOTE method. Table 1 displays the state of the art in comparison to the suggested work. contrasting the suggested work with figure 10. All features were utilized for classification in previous work. As a result, accuracy is decreased. As seen in figure 10, the accuracy has increased as a result of the extraction of significant features for classification.[2]
Accuracies

- Random Forest: 90.5% accuracy (Zhao et al. [6])
- XGBoost: 94.14% accuracy (Kwatra and Arora [8])
- Support Vector Machine (SVM): 91.2% accuracy (Srilatha and Venugopal [7])
- Gradient Boosting: 92.3% accuracy (Kwatra and Arora [8])
- Deep Learning (DL) Models: 95.0% accuracy (Kwatra and Arora [8])
- Convolutional Neural Networks (CNN): 96.87% accuracy (Soochow University Dataset [2])
- Multimodal AI Model (Genomic, Imaging, Clinical): 93.5% accuracy (Zhou et al. [12])

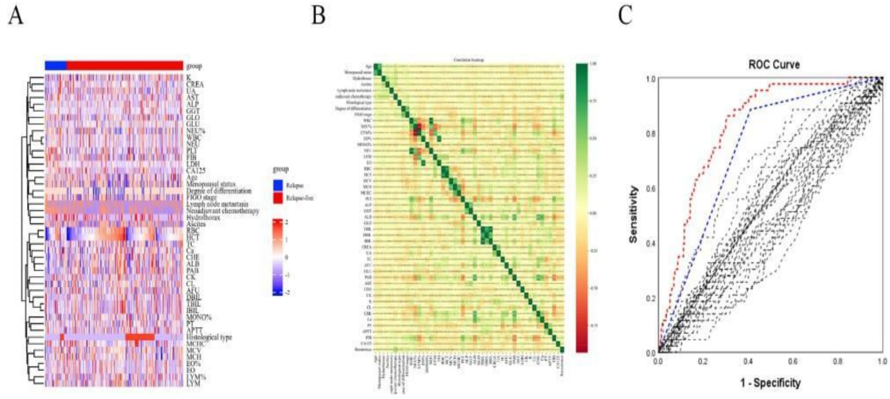


Figure 4

Figure 4: Inter-variable co relation. (A) The concentrations of several biomarkers in the ovarian cancer groups with and without recurrent episodes. The ovarian cancer recurrent group is shown in blue in the upper legend, while the ovarian cancer recurrent-free group is shown in red. Positive and negative correlations with the recurrence of ovarian cancer are represented by red and blue, respectively, in the bottom legend; the higher the correlation, the larger the absolute value. (B) A heat map showing how the factors relate to one another. A gradient from positive to negative correlation is represented by the green to red gradient. (C) ROC curve logistic regression for ovarian cancer recurrence prediction. Blackdashed lines show the results of a single regression, whereas blue lines show the results of a multiple regression model utilizing all 47 biomarkers. [11]

*Discussion*

This work investigates the application of machine learning to predict the recurrence of ovarian cancer by taking into account important risk factors such as tumour histology, age, and disease stage. In contrast to more conventional indicators like CA125, it highlights the significance of incorporating multi-scale clinical data and concentrating on biomarkers associated with the tumour microenvironment. 47 clinical characteristics were analysed using machine learning methods, such as XGBoost, to find important variables for recurrence prediction, including neoadjuvant chemotherapy, AST, PAB, HCT, and MONO%. The study shows that artificial intelligence (AI)-driven models might offer useful information for ovarian cancer prognosis, surpassing conventional techniques in prediction accuracy.[11]

## VII. Conclusion

This study shows how machine learning can help with patient stratification and decision-making by establishing a connection between clinical markers and recurrence risk, thereby improving the prognosis of ovarian cancer. Through the provision of individualised treatment plans, it promotes the precision medicine movement. With the number of ovarian cancer patients rising worldwide, it's crucial to differentiate ovarian cancer from other tumours with few false positives. Class imbalance can be addressed while improving diagnostic accuracy by utilising machine learning techniques like Random Forest (90.5% accuracy) and XGBoost in conjunction with SMOTE (94.14% accuracy). This strategy beats earlier approaches in terms of classification accuracy by combining feature optimisation and data-driven tactics [8][11][2].

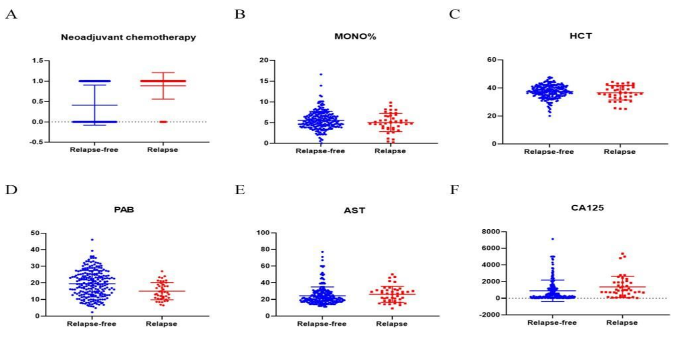


Figure 5

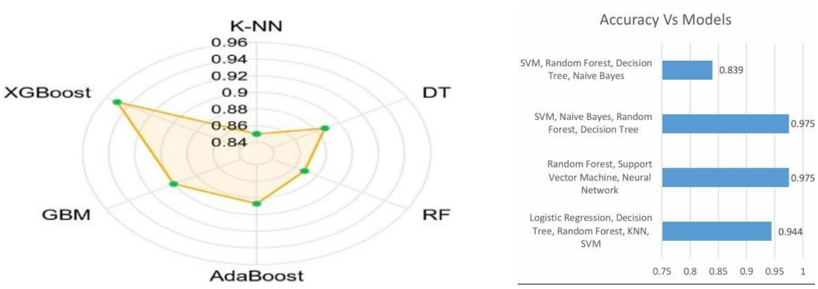


Figure-6 The radar map illustrates each model's accuracy. Research and Management of Cancer[11]

## REFERENCES


[1] G. K. Thakur, S. Kulkarni, K. S. Kumar, P. Sudarsanam, M. Sreenivasulu and P. C. Shaker Reddy, "An Enhanced Prediction of Ovarian Cancer Based on Ensemble Classifier Using Explainable AI," 2024 2nd World Conference on Communication & Computing (WCONF), RAIPUR, India, 2024, pp. 1DOI:10.1109/WCONF61366.2024.10692161. keywords: {Accuracy; Ultrasonic imaging; Computational modeling; Predictive models; Boosting; Real-time systems; Medical diagnostic imaging; Bagging; Ovarian cancer; Testing; Ovarian-cancer prediction; explainable-AI; ensemble learning; DL; AI},

[2] Prasher, Shikha & Nelson, Leema & Sandhya, Vasantha. (2023). Ovarian Cancer Detection Using SMOTE-XGBoost Classifier Model. 1-10.1109/SMARTGENCON60755.2023.10442886.

[3] A. R. Inture, B. S. Sai Nadh, A. Sha, S. Abhishek, A. T and T. V. Mullapudi, "Leveraging Random Forests for Ovarian Cancer Detection and Precision Prediction," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 910-915, Doi: 10.1109/ICECA58529.2023.10394819. keywords: {Training; Medical services; Predictive models; Robustness; Random forests; Ovarian cancer; Testing; Ovarian cancer; Machine learning; Random Forest; Precision-Recall analysis; Medical diagnosis; Average precision},

[4] M. Aditya, I. Amrita, A. Kodipalli and R. J. Martis, "Ovarian Cancer Detection and Classification Using Machine Leaning," 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, 2021, pp. 279-282, Doi: 10.1109/ICEECCOT52851.2021.9707954. keywords: {Deep learning; Demography; Computational modeling; Feature extraction; Random forests; Optimization; Ovarian cancer; Ovarian cancer; Ovarian tumour; Classifier; Machine Learning; Deep Learning},

[5] N. Taleb, S. Mehmood, M. Zubair, I. Naseer, B. Mago and M. U. Nasir, "Ovary Cancer Diagnosing Empowered with Machine Learning," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-6, Doi: 10.1109/ICBATS54253.2022.9759010. keywords: {Support vector machines; Training; Machine learning algorithms; Machine learning; Electron tubes; Security; Medical diagnosis; ovary cancer; machine learning; SVM; KNN},

[6] Zhao, Qin & Li, Yinuo & Wang, Tiejun. (2023). Development and validation of prediction model for early warning of ovarian metastasis risk of endometrial carcinoma. Medicine. 102. e35439. 10.1097/MD.0000000000035439. 17

[7] K, Srilatha & Venugopal, Ulagamuthalvi. (2019). Support Vector Machine And Particle Swarm Optimization Based Classification Of Ovarian Tumour. Bioscience Biotechnology Research Communications. 12. 714-719. 10.21786/bbrc/12.3/24.

[8] Kwatra, Chetna & Arora, Harpreet. (2023). A comprehensive investigation into the use of machine learning to forecast ovarian cancer. 1-8. 10.1109/ICCCNT56998.2023.10308316.

[9] Kongara, Srinivasa & Prakasha, S. & Anbaalagan, Brindha & Pathak, Sumit & Miya, Javed & Taqui, Syed & Khalil Almoallim, Hesham & Alharbi, Sulaiman & Raghavan, S.. (2024). Performance evaluation of optimized convolutional neural network mechanism in the detection and classification of ovarian cancer. Multimedia Tools and Applications. 83. 1-24. 1007/s11042-024-18115-0.

[10] Sorayaie Azar, Amir & Babaei Rikan, Samin & Naemi, Amin & Bagherzadeh, J. & Pirnejad, Habibollah & Mohasefi, Matin & Wiil, Uffe. (2022). Application of machine learning techniques for predicting survival in ovarian cancer. BMC Medical Informatics and Decision Making. 22. 10.1186/s12911-022-02087-y.

[11] Zhou, Lining & Hong, Hong & Chu, Fuying & Chen, Xiang & Wang, Chenlu. (2024). Predicting the Recurrence of Ovarian Cancer Based on Machine Learning. Cancer Management and Research. 16. 1375-1387. 10.2147/CMAR.S482837.

[12] Zhou, Rong & Zhao, Bingbing & Ding, Hongfan & Fu, Yong & Li, Hongjun & Wei, Yuekun & Xie, Jin & Chen, Caihong & Yin, Fuqiang & Huang, Daizheng. (2024). Survival prediction of ovarian serous carcinoma based on machine learning combined with pathological images and clinical information. AIP Advances. 14. 10.1063/5.0196414.

[13] Wajgi, R., & Wajgi, D. (2022, March). Malnutrition detection in infants using machine learning approach. In AIP Conference Proceedings,Vol. 2424, No. 1