IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

An Intelligent Digital Twin Framework with Biomarker-Aware XGBoost for Ovarian Cancer Prediction

Bhavana B R1*[1], Sindhu A G[2] and Akash Gowda S G3[3]

Department of MCA, Surana College (Autonomous), Bengaluru, India[1]

Department of MCA, Surana College (Autonomous), Bengaluru, India[2]

Department of MCA, Surana College (Autonomous), Bengaluru, India[3]

Abstract: Ovarian cancer is a major killer in women worldwide, and the rate of cancer deaths is especially low because of the lack of good tools to detect it in the early stages and the late diagnosis. Currently, the most used screening tests are CA-125 and transvaginal ultrasound which are not sufficiently precise to make an early diagnosis, so more intelligent prediction methods are needed. In this study, we have proposed a framework for the Digital Twin and XGBoost machine learning to make the personalized prediction of ovarian cancer risk and generation of biomarkers. The clinical data set consisted of 349 patients with 51 biomarkers; 10 were identified as key predictors by Recursive Feature Elimination, including CA125, HE4, ALB and CEA. The accuracy of the XGBoost model is 91.43%. Clinicians can use the Digital Twin component to estimate a patient's current cancer risk and simulate the evolution of cancer risk over time, with the understanding of changes in biomarkers, to obtain interpretable clinical insights. The proposed framework performs well in terms of prediction and is a dynamic and personalized clinical decision support tool for ovarian cancer. .

Keywords: Ovarian cancer, XGBoost, Bio markers, Digital Twin.

INTRODUCTION

Ovarian cancer is one of the most invasive gynaecological cancers and is always among the top ten causes of cancer death in females around the world [1]. The difficulty with this disease is that the majority of patients don't present with any specific symptoms; they present with ambiguous and non-specific symptoms like bloating, abdominal discomfort, discomfort in the legs etc which are easily missed and that's why most of the cases are being diagnosed at late stages [2] [3]. For many years, the most commonly used clinical screening tools for ovarian cancer have been serum markers (e.g., CA125). The sensitivity and specificity of these procedures is, however, well documented, and they have not been able to adequately discriminate between malignant and benign masses with confidence [4] [5]. Algorithms that combine CA-125 with HE4 protein (Risk of Ovarian Malignancy algorithm, ROMA) have proven to be promising but are yet to prove sufficient for the needed diagnostic precision for both the clinician and patient. Such gaps have spurred the development of more complex, data-informed strategies. Machine learning is also a new frontier in research into ovarian cancer. With respect to the various algorithms explored, XGBoost has been shown to offer the best predictive performance in several clinical scenarios, including AUC values reaching 0.87 for survival prediction, 0.843 for metastasis prediction and it performs well with advanced data strategies like ensemble learning, where classification accuracy was high with XGBoost [6][7]. XGBoost has often performed best when compared to Random Forest, Support Vector Machine (SVM), and Logistic Regression, as demonstrated in direct comparisons.XGBoost has consistently outperformed other models, such as Random Forest, Support Vector Machine (SVM), and Logistic Regression, in various direct comparisons. Current models are however, largely static, and make only one prediction at a given time, without considering how the patient's state might change. This constraint is key to its lack of clinical utility. To fill this gap, this study proposes a Digital Twin framework for predicting ovarian cancer. By constructing a virtual representation of a patient from their biomarker profile, A Digital Twin can facilitate real-time prediction of risk, and also simulate future changes in their biomarker profile and risk.

LITERATURE REVIEW

Numverous studies have been conducted to look into the application of Machine Learning techniques for prediction and early detection of ovarian cancer. In this systematic review, Asadi et al.[1] evaluated the performance of AI models for predicting the survival of ovarian cancer patients and found that the ensemble learning models performed better in making predictions than the traditional statistical models. The results are in line with the growing trend in research for

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

data-driven clinical decision support systems in oncology [8]. XGBoost has been an important method in ovarian cancer research because it provides a powerful tool for managing complex clinical data. For the prediction of lung metastasis in ovarian cancer patients, Yuan et al.[9] showed the best performance of XGBoost over traditional logistic regression models. In addition, Hsiao et al.[6] proved that the GA-XGBoost method using gene expression data has high sensitivity in various test sets. Guo et al.[2] were able to demonstrate that

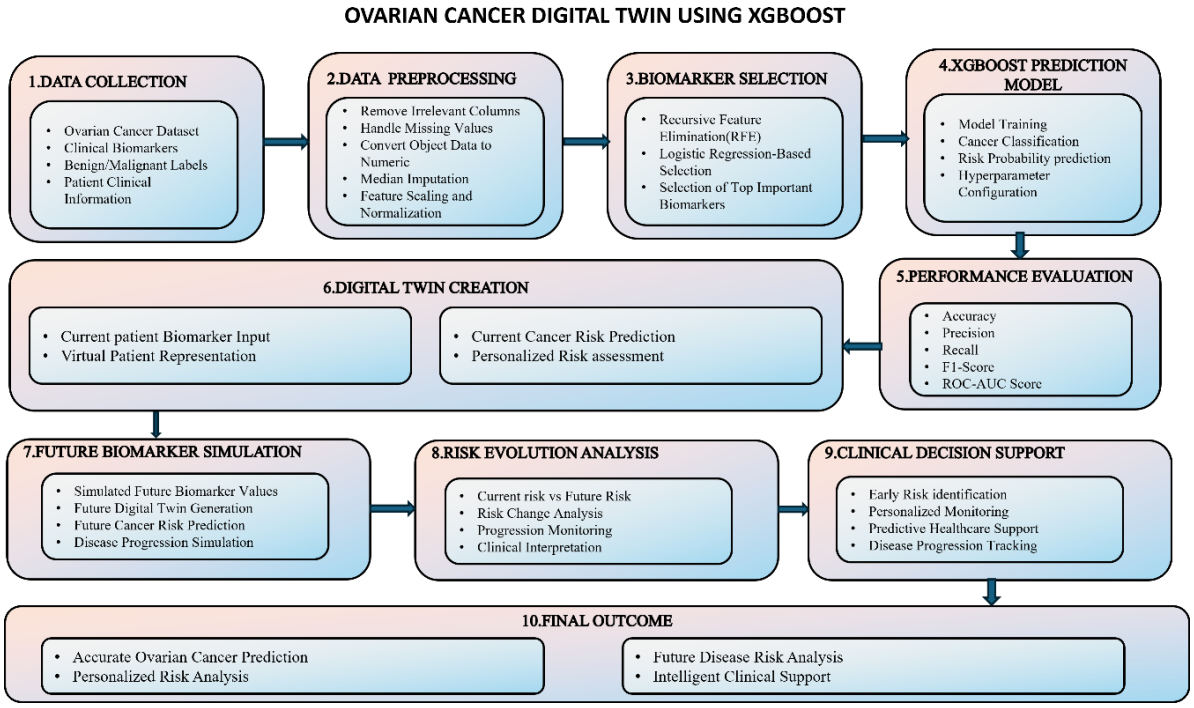

Fig. 1: XGBoost-Driven Digital Twin Framework for Ovarian Cancer Prediction

machine learning models using SHAP analysis could be used in an interpretable way to correctly identify key factors for the metastasis of ovarian cancer patients. Huang[3] showed that the systematic selection of biomarkers using Recursive Feature Elimination could achieve better model performance than the existing screening algorithms in diagnosis of early ovarian cancer. Recent research has also demonstrated that several new biomarkers, such as HE4 and plasma metabolites, can be used in addition to CA-125 to increase the classification accuracy when integrated into machine learningbased models[10][11]. Digital Twin technology is a new and exciting development in the realm of personalized healthcare, alongside machine learning. As pointed out by Zhang and Liu[12], digital twin simulation can be used for patient-specific disease trajectory modelling and treatment planning. In oncology, dynamic patient modeling using AI is closely related to Digital Twin concepts, as highlighted by Woitek et al. [13]. This is the first study to use a Digital Twin framework specifically for the prediction of ovarian cancer risk and the simulation of biomarkers.

METHODOLOGY

This study aims to present a Digital Twin structure for ovarian cancer risk prediction with XGBoost as a tool for personalized clinical decision making, both for risk assessment and for simulation of future biomarkers, as illustrated in Fig 1.

Dataset

The study is conducted using a clinical ovarian cancer dataset of 349 patients and 51 biomarker features. Pathological diagnosis (the target variable) is represented by the number of malignant cases (178) and benign cases (171), and there is a near balance in the classes. The dataset consists of several serum biomarkers and clinical parameters that are frequently used in the diagnosis and monitoring of ovarian cancer[14].Ovarian cancer is a major killer in women worldwide, and the rate of cancer deaths is especially low because of the lack of good tools to detect it in the early stages and the late diagnosis. Currently, the most used screening test is CA-125 and transvaginal ultrasound which are not sufficiently precise to make an early diagnosis, so more intelligent prediction methods are needed. In this study, we have proposed a framework for the Digital Twin and XGBoost machine learning to make the personalized prediction of

DOI: 10.17148/IJARCCE.2026.15816

ovarian cancer risk and generation of biomarkers. The clinical data set consisted of 349 patients with 51 biomarkers; 10 were identified as key predictors by Recursive Feature Elimination, including CA125, HE4, ALB and CEA. The accuracy of the XGBoost model is 91.43%. Clinicians can use the Digital Twin component to estimate a patient's current cancer risk and simulate the evolution of cancer risk over time, with the understanding of changes in biomarkers, to obtain interpretable clinical insights. The proposed framework performs well in terms of prediction and is a dynamic and personalized clinical decision support tool for ovarian cancer. monitoring [15].

Data Preprocessing

The raw data was preprocessed in multiple steps. The subject identifier column was dropped because it did not provide any predictive value. All non-numeric attributes were converted to numeric and features with more than 50% of data missing were removed. CA72-4 was the only feature that was found to be above this threshold and was thus dropped. Median imputation was used to fill in the missing values, as this is a robust method for missing values that are typically observed in clinical biomarker data[3][16].

Feature Selection

To select the ten most diagnostically relevant biomarkers from the preprocessed data, Recursive Feature Elimination (RFE) with Logistic Regression as the base estimator was used.To select the ten most diagnostically relevant biomarkers out of the preprocessed data, Recursive Feature Elimination (RFE) with Logistic Regression as the base estimator was used.

XGBoost Architecture

Inspired by gradient boosted decision trees, Extreme Gradient Boosting (XGBoost) is a cutting-edge ensemble machine learning technique. It is highly used in medical prediction systems because of its good prediction capability, regularization and resistance to overfitting. This study utilizes XGBoost in the Digital Twin framework to investigate the biomarker profiles and predict the risk of ovarian cancer. The architecture sequentially builds several decision trees, with each new tree trying to reduce the errors left by the previous trees. This is a recursive boosting process which allows the model to learn complex non-linear interactions between clinical markers and disease measures. The prediction made by XGBoost can be mathematically expressed as shown in the Eq. 1:

$$\hat{y_i} = \sum_{k=1}^{K} f_k(x_i), f_k \in F \quad (1)$$

Here, y^i is the predicted output for the i th patient sample, K is the number of trees and fk is a single regression tree in the functional space F. The function to be minimized when training the model is given by the Eq. 2:

$$L(\phi) = \sum_{i=1}^{n} l(\hat{y_i}, y_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

The definition of the regularisation component is given in the Eq. 3:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda |w|^2 \quad (3)$$

where T is the number of leaves in the tree, w represents leaf weights, and γ and λ are regularization parameters. The XGBoost architecture used in this framework as shown in the Fig. 2 consists of biomarker input features, feature preprocessing, sequential boosted tree generation, risk score estimation, and final ovarian cancer prediction. The generated prediction is further integrated into the Digital Twin system for future biomarker simulation and personalized risk analysis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

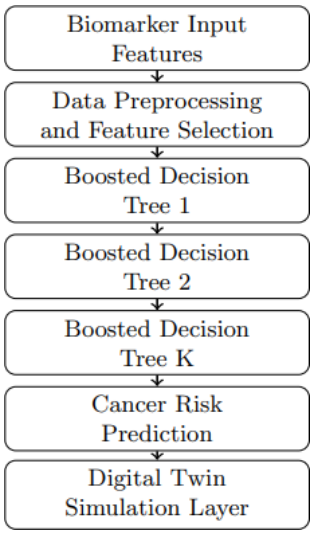

Fig. 2: Architecture of XGBoost Integrated with the Digital Twin Framework for Ovarian Cancer Prediction

Results and Discussion

This section will involve all the necessary evaluations of the proposed A Digital Twin assisted XGBoost framework for ovarian cancer risk. prediction. The analysis is broken down in three parts: (i) dataset characteristics and preprocessing outcomes, (ii) quantitative To assess the model classification performance for the XGBoost model, and (iii) the personalized risk simulation enabled by the Digital Twin architecture. All experiments were done in a CUDA equipped GPU environment with Multiple machine learning libraries for Python, guaranteeing computational Efficiency in both the training and inference process.

Dataset Characteristics and Preprocessing

The clinical data was obtained from 349 patients with diagnosis of benign and malignant ovarian disease. Hematological, biochemical and oncological biomarkers were obtained as part of regular clinical evaluation and included in each record. The predictor variables were normalized using feature-wise mean substitution and Standard Scaling to standardize all the predictor variables across various measurement ranges for consistent model learning. Recursive Feature Elimination (RFE) was used to feature select to an optimal set of ten biomarkers: Albumin (ALB), Cancer Antigen 125 (CA125), Carcinoembryonic Antigen (CEA), Chloride (CL), Globulin (GLO), Human Epididymis Protein 4 (HE4), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), Menopause status, and Neutrophil count (NEU). This diminished feature space incorporates proven tumor markers and hematological and biochemical markers linked to ovarian cancer with reduced dimensionality and overfitting risk. A stratified train-test split was used to ensure class balance in the dataset. An additional test set of 70 patients (about 20% of the patients) was included in a held out set exclusively for final performance evaluation, and it was neither used for model training nor used for hyperparameter tuning.

XGBoost Classifier Performance

The training set was preprocessed and then used to train the XGBoost classifier. partitioned and then tested on the held-out test set of 70 samples. XGBoost builds an ensemble of gradient-boosted. The nonlinear, high-order interactions between the variables are modeled by The complex, multi-faceted nature of the field made this apparent, and it would be particularly suitable for biomarketing. Complex nonmonotonic relationship landscape of oncological biomarker data. Hyperparameter optimization was done using the ”Grid” algorithm and cross-validated. search over critical parameters such as tree depth, learning rate, Using the optimal subsampling ratio, and regularization terms, the best Model selected from the validation fold by the AUC criteria. Table 1 summarizes the classification performance. The optimized model was evaluated on the five critical metrics and compared to the original model. The model’s accuracy is shown to be 91.43% and The results for recall, precision and f1-score were at 91.67%, 91.67% and 91.67% respectively. The Similar values of these three metrics mean a balanced classifier that has little systematic error for either benign or malignant. malignant class — a key characteristic in the clinical screening situation, where both false positives and false negative have different significance and 6

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

Table 1: XGBoost Model Performance

Metric	Score
ROC-AUC	98.45%
Accuracy	91.43%
Recall	91.67%
Precision	91.67%
F1 Score	91.67%

significant clinical consequences. A false negative (malignant case) diagnosis (also considered benign) delays life-saving treatment unnecessarily when the case is actually positive (benign). invasive procedures and patient anxiety. The precise balancing and - recall confirm that the proposed framework deals with this trade-off. They can be practically done without any unnecessary manipulation of the threshold.

The most clinically relevant performance measure is the best result achieved with the ROC-AUC score, which is 98.45% A classification accuracy of close to one means that the classifier is consistently performing higher ranking of a randomly selected malignant patient. More likely to have a risk probability than a random benign patient — property that makes the model easy to be threshold-adjustable on. At the population level or in a resource-limited setting. This model's high AUC value also indicates its good performance. The discriminative ability is not a result of threshold tuning; rather, it is a characteristic of the instruments that represents a basic, well-calibrated probabilistic output.

The confusion matrix of Fig. 3 also indicates the classification accuracy of the model. The number of misclassified instances that lie between the benign and malignant classes is few, which is in accordance with the high precision and recall that have been reported in Table 1. The low non-diagonal elements suggest that the tenbiomarker panel chosen using the RFE approach has a good discriminative power that is useful for accurate class separation. Compared to other traditional methods like logistic regression and single decision trees, which generally produce classification accuracy ranging from 75–85%, the proposed XGBoost model demonstrates excellent classification performance by capturing nonlinear feature interactions and avoiding overfitting through gradient-boosted ensemble learning.

It is noteworthy that these results were obtained from a relatively A small clinical group of 349 cases. The strong generalization supported by the test metrics held out, as well as the high AUC, proposing the elimination of the RFE-derived biomarker panel was effective. Features that are redundant or noisy and could cause overfitting. In the small sample setting of medical data sets. In addition, the use of Standard Scaling before training helped to achieve numerical stability and guaranteed that none of a single biomarker dominated the update of gradients when the model is fitted.

Digital Twin-Assisted Personalized Risk

Simulation Going beyond one-time predictions, the fundamental scientific key idea of this work is to integrate the XGBoost model into a Digital Twin (DT) framework to facilitate 7

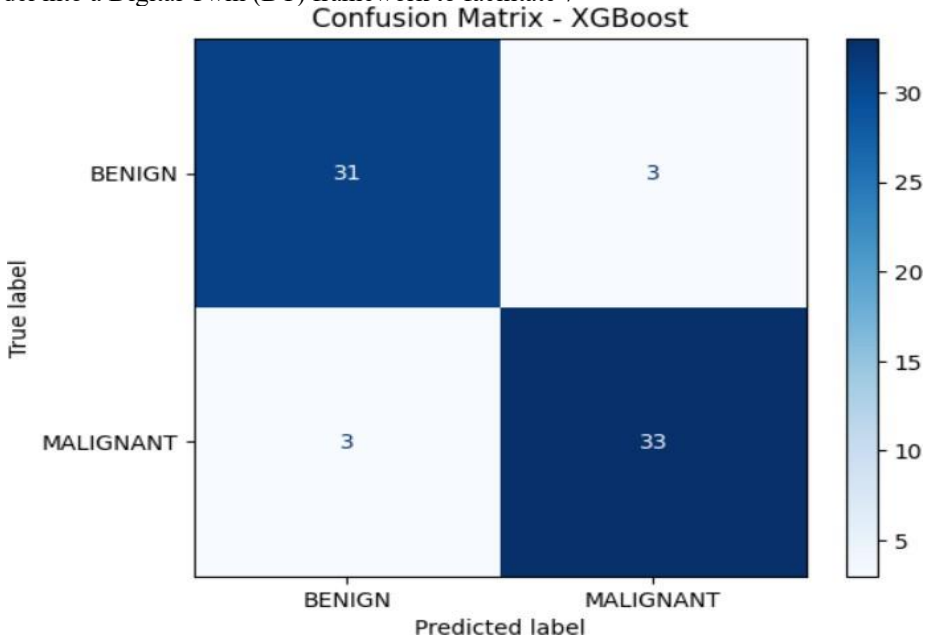

Fig. 3: Confusion Matrix: Distribution of correctly and incorrectly classified benign and malignant ovarian cancer cases on the 70-sample held-out test set.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

individual, longitudinal data Modelling of cancer risk pathways. Unlike conventional machine-learning-based diagnostic systems, which produce a single result derived from a given set of biomarker readings, and proposed Digital Twin continually ingests a patient's changing probiological risk assessments of projected future states. This is an important feature for proactive Ovarian cancer biomarker, dynamics are complex, making clinical decisions challenging. Time-varying, treatment-responsive, and disease-dependent, Physiological comorbidities.

The Digital Twin results are shown in Table 2. Patient simulations for a typical person in the evaluation cohort. The column to the left represents the patient's current measured biomarker. values; the right column shows biomarker values projected for the simulated future clinical state.

Table 2: Digital Twin-Assisted Ovarian Cancer Risk Prediction and Future Biomarker Simulation for a Representative Patient

Biomarker / Parameter	Current Patient State	Future Simulated State
ALB	12	17
CA125	22	67
CEA	45	78
CL	1.0	4.40
GLO	2.4	6.8
HE4	5.89	56.0
MCH	22	0.08
MCV	12	0.45
Menopause Status	89	0.112
NEU	0.03	11
Predicted Condition	**MALIGNANT**	**MALIGNANT**
Cancer Risk Probability	**70.82%**	**53.09%**
Risk Change	**−17.73%**	
Clinical Insight	**Risk Stable / Reduced**	

The Digital Twin is expected to play a crucial role in the future biomarker scenario. Re-calculated the probability of cancer risk as 53.09%, and The group was termed 'Malignant' but did show a statistically significant reflection of the term. That is, a risk reduction of 17.73 percentage points. There are significant changes in multiple biomarkers in the simulated future state: The CA125 level rises above 67 U/mL (clinical cut-off) When 35 U/mL, the level of CEA increases to 78 ng/mL and HE4 significantly increases. ¿56.0 pmol/L — changes that would normally indicate disease progression. The simultaneous improvement of ALB (from 12 to 14) was not observed. to 17 g/dL), CL (from 1.0 to 4.40 mmol/L), GLO (from 2.4 to 6.8 g/dL), and NEU (from 0.03 to 11 ×109/L) reflects An anticipated gain in the system physiological reserves, it May have been a treatment effect or nature given host response.

The Digital Twin is capable of combining (synthesizing) those. opposing biomarker trends and translate them into a net reduced single-marker or single-choice assessments. Its benefit over single-marker or single-choice assessments can be seen in the probability of risk. "threshold-based interpretation": it quantifies the aggregate Moving beyond reacting to single biomarker to physiological state exceedances in isolation.

The clinical interpretation that comes out of the framework is Risk Stable / Reduced — translates this 17.73% risk reduction into an Useful message to the clinician that is meant to lead to a decision. Such information can inform decisions about treatment escalation, Time of follow-up biomarker evaluation, or "watchful waiting." This is an important point as the classification continues to be Malignant under the future simulated state properly alerts. risk reduction does not equal remission, and preserves clinical vigilance. This is a subtle, probabilistic risk reporting that's a hallmark of it. The Digital Twin paradigm has been demonstrated to be stronger than binary, threshold-based. clinical decision tools.

Comparative Analysis and Clinical Implications

The proposed framework facilitates two major steps in the diagnosis of ovarian cancer using computational techniques. First, the XGBoost classifier and the RFE-based biomarker selection method demonstrate comparable or better results on ovarian cancer sets compared with previous machine learning methods published.First, on the 9 ovarian cancer sets, the performance of the XGBoost classifier and the RFE-based biomarker selection method is similar or better than

DOI: 10.17148/IJARCCE.2026.15816

some recent machine learning methods. The AUC of 98.45% is an important improvement from the usual range of 0.88–0.95, which would lead to misclassification of borderline-risk cases.

Secondly, the Digital Twin methodology goes beyond simple classification by modelling the patient risk as iterative, and not just a single-time forecast. In contrast to static AI systems, it allows for modeling potential progression of risks based on future states of biomarkers. This functionality includes identification of increasing risk trends, longitudinal disease surveillance and 'what if' scenario analysis for personalized clinical decision making in order to consider possible treatment outcomes in the early stages.

The framework also has good clinical interpretability with the 10 selected biomarkers. Established markers include CA125, HE4 and CEA, along with markers for nutritional status (ALB, GLO), hematological alterations (MCH, MCV), metabolic balance (CL), inflammatory response (NEU) and menopause status. Combination of clinically meaningful features and high predictive accuracy increases transparency and increases clinician confidence in the model recommendations.

CONCLUSION

In this study, a novel Digital Twin framework for ovarian cancer risk prediction is presented which integrates the XGBoost classification with dynamic biomarker simulation. The model was validated on a clinical dataset of 349 patients, with an accuracy of 91.43% and an ROC-AUC of 98.45%, showing good discriminative ability between malignant and benign ovarian conditions. Recursive Feature Elimination (RFE) was used to identify ten key biomarkers with a clinically interpretable and reduced feature set for prediction. The Digital Twin architecture went beyond static prediction and enabled the creation of personalized risk simulation—where clinicians could estimate the impact of anticipated shifts in a patient's biomarker profile over time on cancer risk. This work makes a useful step toward dynamic, personalized and explainable clinical decision support in managing ovarian cancer. Further studies will investigate external validation, the integration of imaging and genomic information and deployment within clinical workflow systems.

REFERENCES

[1] Farkhondeh Asadi et al. "Comparing the Effectiveness of Artificial Intelligence Models in Predicting Ovarian Cancer Survival: A Systematic Review". In: Cancer Reports 8 (2025), e70138. doi: 10.1002/cnr2.70138.

[2] Qin-Hua Guo et al. "Application of Interpretable Machine Learning Algorithms to Predict Distant Metastasis in Ovarian Clear Cell Carcinoma". In: Cancer Medicine 13 (2024), e7161. doi: 10.1002/cam4.7161.

[3] Weitong Huang. "Machine Learning for Early Detection of Ovarian Cancer: Improving Evaluation and Explanation from Technical and Domain Perspectives". MA thesis. Australian National University, 2024. 10

[4] Chuhan Zhang and Muyu Liu. "Integrating Ovarian Reserve Biomarkers with Machine Learning for Gonadotoxicity Risk Prediction in Young Female Cancer Patients: A Scoping Review". In: Computing Innovations and Applications (2026). doi: 10.63575/CIA.2026.40111.

[5] Yue Gao et al. "Artificial Intelligence-Based Models Enabling Accurate Diagnosis of Ovarian Cancer Using Laboratory Tests in China: A Multicentre, Retrospective Cohort Study". In: Lancet Digital Health 6.3 (2024), e176–e186. doi: 10.1016/S2589-7500(23)00245-5.

[6] Yi-Wen Hsiao et al. "A Risk Prediction Model of Gene Signatures in Ovarian Cancer Through Bagging of GA-XGBoost Models". In: Journal of Advanced Research 30 (2021), pp. 113–122. doi: 10.1016/j.jare.2020.11.006.

[7] Ching-Ju Lin et al. "An Integrated Algorithm with Feature Selection, Data Augmentation, and XGBoost for Ovarian Cancer". In: Mathematics 12.24 (2024), p. 4041. doi: 10.3390/math12244041.

[8] Seyed Mohammad Ayyoubzadeh et al. "Prediction of Ovarian Cancer Using Artificial Intelligence Tools". In: Health Science Reports 7 (2024), e2203. doi: 10.1002/hsr2.2203.

[9] Yufei Yuan et al. "A Machine Learning Approach Using XGBoost Predicts Lung Metastasis in Patients with Ovarian Cancer". In: BioMed Research International 2022 (2022), p. 8501819. doi: 10.1155/2022/8501819.

[10] Alina Tokareva et al. "Machine Learning Framework for Ovarian Cancer Diagnostics Using Plasma Lipidomics and Metabolomics". In: International Journal of Molecular Sciences 26.14 (2025), p. 6630. doi: 10.3390/ijms26146630.

[11] Munetoshi Akazawa and Kazunori Hashimoto. "Artificial Intelligence in Ovarian Cancer Diagnosis". In: Anticancer Research 40.8 (2020), pp. 4795–4800. doi: 10.21873/anticanres.14482.

[12] J. M. Sheela Lavanya and P. Subbulakshmi. "Innovative Approach Towards Early Prediction of Ovarian Cancer: Machine Learning-Enabled XAI Techniques". In: Heliyon 10 (2024), e29197. doi: 10 . 1016 / j . heliyon . 2024 . e29197.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15816

[13] Ramona Woitek, Ferdia A. Gallagher, and Evis Sala. "Ovarian Cancer Beyond Imaging: Integration of AI and Multiomics Biomarkers". In: European Radiology Experimental 7 (2023), p. 47. doi: 10.1186/s41747-023-00364-7.

[14] and others. "Early Diagnosis of Ovarian Cancer: A Comprehensive Review of the Advances, Challenges, and Future Directions". In: Diagnostics 15.4 (2025), p. 406. doi: 10.3390/diagnostics15040406.

[15] Imran Ashraf et al. "Improved Prediction of Ovarian Cancer Using Ensemble Classifier and SHAP Explainable AI". In: Cancers 15.24 (2023), p. 5793. doi: 10.3390/cancers15245793.

[16] Diederick De Jong et al. "Explainable Artificial Intelligence for Prediction of Complete Surgical Cytoreduction in Advanced-Stage Epithelial Ovarian Cancer". In: Journal of Personalized Medicine 12.4 (2022), p. 607. doi: 10.3390/ jpm12040607.