e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

DETECTION OF OVARIAN CANCER USING DIFFERENT MACHINE LEARNING ALGORITHMS

Harshith S[*1]**, Sujay Sasikumar**[*2]**, M Nandha Kumar**[*3]**, Mohammad Kashif**[*4]**,**
Manu Krishnan PM[*5]**, Devi S**[*6]

[*1,2,3,4,5]School Of Engineering Presidency University Bengaluru, India.

[*6]Assistant Professor School Of Information Sceince Presidency University Bengaluru, India.

ABSTRACT

This research paper tackles the urgent issue of early detection in ovarian cancer by employing machine learning (ML) algorithms, shedding light on the challenges associated with late-stage diagnoses and underscoring the demand for inventive solutions. Using an extensive dataset containing clinical and genetic features, the materials and methods section outlines a meticulous approach to data preprocessing and the application of ML algorithms, such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks. The results demonstrate promising outcomes, showcasing distinct strengths in ovarian cancer detection for each algorithm. Rigorous evaluation metrics, including accuracy, precision, recall, and F1 score, ensure a robust assessment of model performance. The discussion delves into the implications of the results, addressing obstacles like data imbalances and interpretability issues. Techniques for model interpretability offer a deeper insight into predictive factors. Future directions underscore the significance of continuous research to refine ovarian cancer detection models. In conclusion, the paper highlights the potential of ML algorithms to enhance early ovarian cancer detection. The thorough evaluation, coupled with interpretability analyses, adds to the robustness of the findings, paving the way for future strides in innovative cancer diagnosis methodologies and improved patient outcomes.

Keywords: Ovarian Cancer Detection, Bagging, Random Forest, Support Vector Machine, Decision Tree Classifier, Boosting,

I. INTRODUCTION

Ovarian cancer [1] presents a significant obstacle in women's health, characterized by its subtle and frequently symptom-free development until reaching advanced stages. This malignancy initiates in the ovaries, the pivotal female reproductive organs accountable for both egg production and hormone synthesis. While its occurrence is relatively less common than some other cancers, ovarian cancer stands as the fifth highest contributor to cancer-related fatalities among women worldwide. Gaining insights into the origins, survival rates, and the changing incidence of this ailment is crucial for understanding the pressing need for improvements in early detection approaches.

The precise origin of ovarian cancer remains elusive, adding intricacy to both its diagnostic [2] procedures and treatment approaches. Nevertheless, specific risk factors have been recognized, potentially increasing an individual's susceptibility to ovarian cancer. These factors encompass factors such as advanced age, familial instances of ovarian or breast cancer, specific genetic mutations like BRCA1 and BRCA2, and a medical history involving certain gynecological conditions. Additionally, hormonal considerations, such as early onset of menstruation or delayed onset of menopause, contribute to influencing the risk of ovarian cancer.

Ovarian [3] cancer is often identified in advanced stages, restricting treatment choices and adversely affecting survival rates. The comparative survival rates for ovarian cancer are notably lower than those for certain other cancers, primarily due to difficulties in early detection. Global cancer statistics reveal an approximate 47% five-year survival rate for ovarian cancer. However, this figure exhibits significant variability depending on the cancer's stage at diagnosis. In cases of localized tumors, the five-year survival rate is considerably elevated, underscoring the pivotal significance of early detection in enhancing overall outcomes.

The prevalence of ovarian cancer has witnessed notable shifts over the years, reflecting changes in risk factors, screening practices, and awareness. In the mid-20th century, ovarian cancer was often referred to as the "silent killer" due to its asymptomatic progression. As medical knowledge and diagnostic tools advanced, the understanding of ovarian cancer improved. The percentage of women affected by ovarian cancer has demonstrated fluctuations, influenced by factors such as lifestyle changes, reproductive patterns, and the advent of genetic testing.

Historical data indicates that the percentage of women affected by ovarian cancer has seen both increases and decreases. Improved awareness campaigns, coupled with advances in medical imaging and genetic screening, have contributed to earlier diagnoses in some cases. However, challenges persist, and late-stage diagnoses remain prevalent. Machine learning algorithms [4] offer a promising avenue to address these disadvantages by discerning intricate patterns within diverse datasets. This study explores the strengths and weaknesses of various ML algorithms, contributing valuable insights to the selection of the most effective models for early ovarian cancer detection. Through a comprehensive evaluation [5] and consideration of interpretability, this research endeavors to overcome the disadvantages associated with traditional diagnostic methods and pave the way for improved patient outcomes.

The key areas of research are as follows:

1. **Algorithmic Performance Comparison:** Examine and contrast the effectiveness of different machine learning algorithms in the realm of ovarian cancer detection. Assess their performance using essential metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve. This comprehensive evaluation aims to gauge and compare the algorithms' capacity to effectively differentiate between cancerous and non-cancerous cases.
2. **Feature Importance and Selection:** Explore and analyse the importance of features used by different algorithms in ovarian cancer detection. Assess how algorithms handle diverse types of data, including clinical and genetic features. Investigate feature selection [6] techniques employed by each algorithm and their impact on model performance. Understanding the significance of specific features can provide insights into the biological and clinical relevance of different markers.
3. **Robustness and Generalization:** Assess the robustness and generalization capabilities of each algorithm by employing cross-validation techniques and testing on diverse datasets. Investigate how well the algorithms perform across different patient cohorts, considering factors such as age groups, genetic variations, and tumor types. A comparative analysis of algorithmic robustness can guide the selection of models suitable for real-world clinical applications.

These key areas of research in a comparative analysis of machine learning algorithms for ovarian cancer detection aim to provide a comprehensive understanding of the strengths, limitations, and practical implications of different approaches. The findings from such research can guide the selection of the most suitable algorithm for accurate and clinically relevant early detection of ovarian cancer.

II. LITERATURE

In this segment, we explore diverse methodologies for classifying ovarian cancer based on cell types, a critical aspect in tailoring personalized treatment plans for patients. The accurate identification of ovarian cancer types is pivotal, particularly as numerous studies have focused on advancing cancer screening processes, progressing into the preclinical stage over the past decade. However, the manual analysis of images by expert pathologists lacks consistency across different individuals and proves to be time-intensive. Recent strides in leveraging machine learning (ML) algorithms [7] have revolutionized the initial screening and diagnosis of ovarian cancer, offering a more standardized and efficient approach. This section delves into recommended techniques for feature extraction from ultrasonic images and subsequent classification, shedding light on state-of-the-art ML-based approaches for the detection of ovarian cancer.

Md. Martuza Ahamad, Sakifa Aktar et al. presented a seminal work investigates various classifiers in ovarian cancer detection, achieving accuracies of 87%, 80%, 82% and 88%. The study utilizes a diverse dataset combining Blood samples, general chemistry, OC markers, employing XGBoost [8], SVM, Extreme Learning Machines and Random Forest classifiers. One study by Ala'a El-Nabawy et al. presented a comparative analysis of ovarian cancer detection algorithms, achieving accuracies of 70.77%, 74.42%, 80%. The study employs Linear SVM, Random Forest, Ensemble SVM, Logistic Regression and Boosting on National Center of Biotechnology Information (NCBI) dataset. On the other hand a paper by Munetoshi Akazawa et al. employed by Support Vector Machine, Random Forest, Naïve Bayes, Logistic Regression [9] and XGBoost, achieving 80.8%

accuracy. The study integrates clinical and genomic data. Sharmistha Bhattacharjee et al. focus on Mass spectrometry (MS) field data, obtaining an accuracy of 84%. The study uses SVM, Neural network as Multilayer Perceptron, Decision Tree [10], KNN and Ensemble Classifiers. Syed Saba Raoof et al. achieved an 87% accuracy [11] by using UCI MLDb dataset. The study employs Multicase SVM.

While the referenced research papers collectively contribute valuable insights to the application of machine learning (ML) in ovarian cancer diagnosis and risk prediction, there exists a notable research gap that warrants further exploration. The synthesis of these studies reveals a need for comprehensive comparative analyses across diverse ML models [12] employed in ovarian cancer research. The existing literature offers glimpses into the application of ML for diagnosis, early detection[13], predictive modelling, and genetic signature analysis [14], but a unified and systematic comparison of these approaches is lacking. A research gap persists in the absence of a consolidated framework that evaluates the relative strengths and weaknesses of different ML models in the context of ovarian cancer. Additionally, there is a need for studies that address the translational aspects of implementing these models in clinical [15] settings, considering factors such as scalability, interpretability, and integration with existing diagnostic workflows

MACHINE LEARNING CLASSIFICATION APPROACHES

Machine learning (ML) has emerged as a promising tool within the medical field, particularly for the prediction and diagnosis of various cancers, including ovarian cancer. ML algorithms provide the means to comprehensively analyze extensive patient data, enabling the early and accurate identification of potential cases. The effectiveness of our model in predicting ovarian cancer hinges on the quality and quantity of the training data used. It is imperative to have a diverse and representative dataset to construct a predictive model that is robust and reliable. In this research, a variety of algorithms, such as Decision Tree Classifier, Random Forest, SVM, Gradient Boosting Classifier, Stacking, and Bagging Classifier, were employed. The optimization of model hyperparameters, as outlined in the table below, was carried out using the grid search CV method, drawing insights from existing literature to determine appropriate parameter ranges.

A. Decision Tree (DT)

The Decision Tree Classifier is a fundamental machine learning algorithm employed in the research paper for ovarian cancer detection. It constructs a tree structure based on features, enabling the identification of patterns in the dataset. Its simplicity and interpretability make it a valuable tool, contributing to the understanding of key factors influencing ovarian cancer predictions. However, it may be prone to overfitting, necessitating careful tuning of hyperparameters to achieve optimal performance.

B. Random Forest (Rf)

Random Forest, an ensemble learning method, is utilized in the ovarian cancer detection research. Comprising multiple Decision Trees, it enhances accuracy and mitigates overfitting. By aggregating predictions from diverse trees, Random Forest improves robustness and generalization. Its ability to handle high-dimensional data and feature importance analysis contributes to its efficacy in the context of ovarian cancer diagnosis.

C. Stacking

Stacking, a meta-ensemble technique, is applied to enhance ovarian cancer detection in the research. It combines predictions from multiple base models, leveraging their strengths to improve overall performance. Stacking introduces a meta-model that learns to weigh the contributions of individual models dynamically, fostering better predictive accuracy. This ensemble approach aims to capture diverse aspects of ovarian cancer patterns, enhancing the model's capacity to generalize.

D. Support Vector Machine (SVM)

The Support Vector Machine (SVM) serves as a powerful classifier in the ovarian cancer detection study. Recognized for its effectiveness in high-dimensional spaces, SVM identifies optimal hyperplanes to separate classes. Its ability to handle non-linear relationships and adaptability to diverse datasets makes it a suitable choice. However, careful parameter tuning is crucial for optimizing SVM performance in ovarian cancer detection.

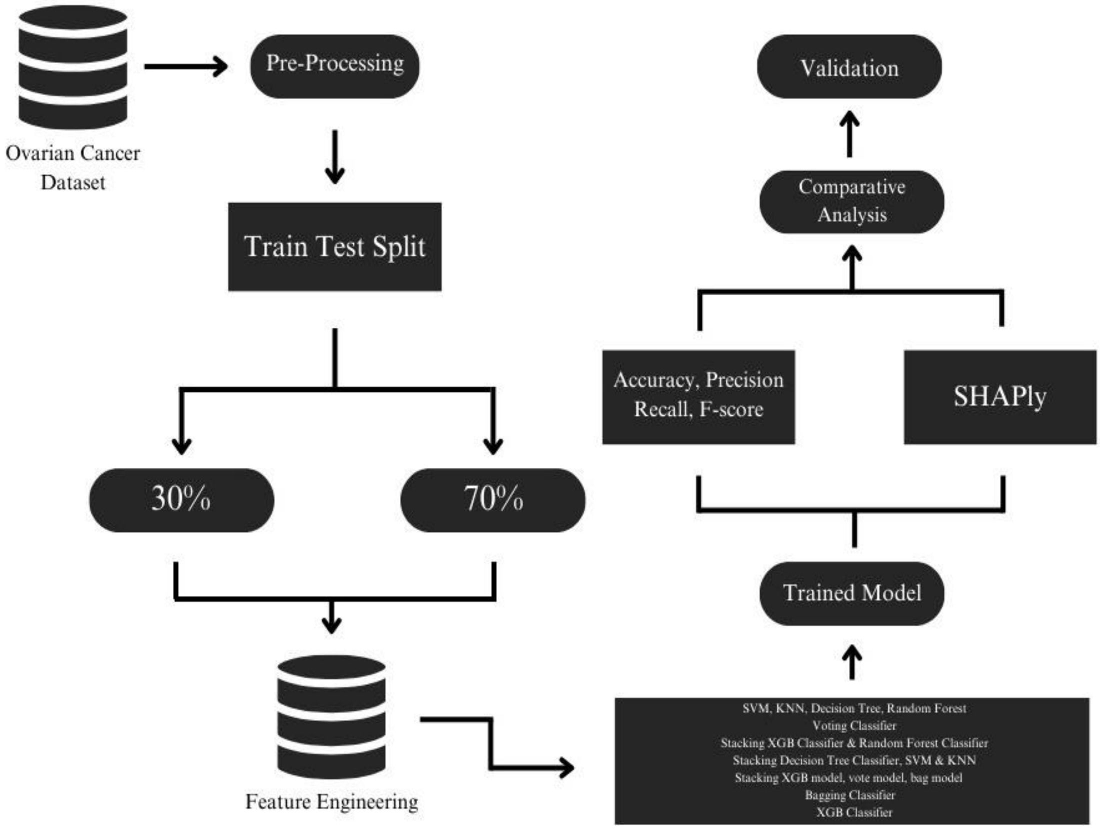
E. Bagging Classifier

The Bagging Classifier, often employed with Decision Trees, is instrumental in the ovarian cancer detection study. By training multiple models on random subsets of the dataset, Bagging reduces variance and improves overall stability. This ensemble technique mitigates overfitting and contributes to enhanced accuracy in identifying ovarian cancer cases. Its effectiveness lies in the diversity introduced through bootstrap sampling, aiding in capturing the complexity of the underlying data.

F. Gradient Boosting Classifier

The Gradient Boosting Classifier is a powerful ensemble algorithm utilized in the ovarian cancer detection research. Building sequentially on weak learners, it minimizes errors and enhances accuracy. Gradient boosting combines the strengths of multiple models, refining predictions iteratively. Its adaptability to various data types and robustness against overfitting make it suitable for complex tasks like ovarian cancer detection. However, parameter tuning is crucial to optimize its performance and prevent potential overfitting.

III. PROPOSED METHOD


Ovarian Cancer Dataset
Pre-Processing
Train Test Split
30%
70%
Feature Engineering
SVM, KNN, Decision Tree, Random Forest
Voting Classifier
Stacking XGB Classifier & Random Forest Classifier
Stacking Decision Tree Classifier, SVM & KNN
Stacking XGB model, vote model, bag model
Bagging Classifier
XGB Classifier
Trained Model
Accuracy, Precision Recall, F-score
SHAPly
Comparative Analysis
Validation

The following steps were used:

1 **Acquisition of Dataset:** Obtain a comprehensive dataset containing clinical and genetic information of ovarian cancer patients. Ensure diversity in the dataset to capture various patient profiles. Acquire data from reputable sources such as medical records, research databases, or collaborative institutions.

2 **Preprocessing:** Perform thorough data preprocessing to address missing values, handle outliers, and standardize data formats. Normalize numerical features and encode categorical variables appropriately. This step is crucial for ensuring the quality and compatibility of the dataset for machine learning algorithms.

3 **Train-Test Split:** Divide the dataset into a training set (70%) and a testing set (30%) using a stratified approach to maintain the distribution of classes. This split ensures that the machine learning models are trained on a sufficiently large dataset while having a separate portion for unbiased evaluation.

4 **Feature Engineering:** Apply feature engineering methods to augment the predictive capabilities of the dataset. This involves crafting new features, scaling, or transforming existing ones. Leverage domain expertise to identify pertinent features that significantly enhance the accuracy of ovarian cancer detection.

5 **Trained Model:** Train multiple machine learning algorithms on the training set. Consider a diverse set of algorithms such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and Neural Networks. Optimize hyperparameters to maximize model performance. Save the trained models for subsequent evaluation.

6 **SHAPly:** To enhance interpretability, SHAP (SHapley Additive exPlanations) values will be calculated. This will help us understand the contribution of each feature to the model's predictions, providing valuable insights into the decision-making process.

7 **Evaluation of All Algorithms:** Evaluate the efficacy of each algorithm on the test set, employing metrics such as accuracy, precision, recall, and F1 score. Utilize cross-validation techniques to enhance robustness and alleviate overfitting risks. Construct confusion matrices to glean valuable insights into the strengths and limitations of the algorithms.

8 **Choosing the Best Model:** Select the algorithm with the highest accuracy on the testing set as the preferred model for ovarian cancer detection. Consider other relevant metrics to validate the robustness of the chosen model. Document the rationale behind selecting the best-performing algorithm and discuss its potential clinical implications.

9 This proposed methodology aims to establish a systematic approach for leveraging machine learning algorithms in the detection of ovarian cancer, emphasizing data quality, algorithm diversity, and rigorous evaluation criteria.

Dataset

The dataset provided encompasses a comprehensive array of parameters, offering a multifaceted view of factors potentially associated with ovarian conditions. This dataset covers various categories, including demographic and clinical information, blood chemistry markers, blood cell counts, and specific tumor-related biomarkers. The extensive list of parameters suggests a holistic approach to understanding and potentially diagnosing or prognosticating ovarian health.

Within the dataset, the inclusion of demographic and clinical information such as age and menopausal status hints at the potential influence of these factors on ovarian conditions. Additionally, the panel of blood chemistry markers spans a wide spectrum, including Albumin, Alkaline Phosphatase, Alanine Transaminase, Aspartate Transaminase, Blood Urea Nitrogen, Calcium, and several cancer-specific antigens like CA125, CA19-9, and CA72-4. These biomarkers are commonly associated with ovarian cancer diagnosis, progression, or monitoring.

Moreover, the incorporation of blood cell counts and characteristics such as Basophils, Eosinophils, Lymphocytes, Mean Corpuscular Hemoglobin, and Platelet Distribution Width provides insight into the hematological aspect potentially linked to ovarian health. This comprehensive range of parameters underscores the complexity and multi-factorial nature of ovarian conditions. The use of the LassoCV (Lasso Cross-Validation) method to extract 21 features from this extensive dataset is intriguing. LassoCV, a regularization technique, aids in feature selection by assigning coefficients to features, effectively eliminating less influential ones. While the exact features chosen by LassoCV were not specified, crucial biomarkers like Alpha-fetoprotein (AFP), Albumin (ALB), Creatinine (CREA), Hemoglobin (HGB), Neutrophils (NEU), and Platelet Distribution Width (PDW) are likely among them. These biomarkers often play pivotal roles in diagnosing or monitoring ovarian conditions and could significantly impact predictive modeling outcomes.

Furthermore, the binary classification target column "TYPE" serves a critical role by differentiating between Benign Ovarian Tumor (BOT) and Ovarian Cancer (OC). Assigning 1 to BOT and 0 to OC establishes a framework for binary classification, enabling machine learning algorithms to distinguish between these conditions based on the selected features. This categorization is fundamental for training predictive models aimed at accurately classifying ovarian conditions.

The dataset utilized for analysis comprised 21 distinct features, encompassing various biochemical and hematological parameters. These included Alpha-fetoprotein (AFP), Albumin (ALB), Alkaline phosphatase (ALP), Alanine transaminase (ALT), Aspartate transaminase (AST), Blood urea nitrogen (BUN), Calcium (Ca), Cancer antigen 19-9 (CA19-9), Carcinoembryonic antigen (CEA), Chloride (CL), Creatinine (CREA), Globulin (GLO), Hemoglobin (HGB), Indirect bilirubin (IBIL), Lymphocytes percentage (LYM%), Mean corpuscular hemoglobin (MCH), Monocytes count (MONO#), Neutrophils count (NEU), Platelet distribution width (PDW), Menopausal status (Menopause), and Sodium (Na). These diverse features collectively provided a comprehensive representation of the physiological and biochemical profile under consideration.

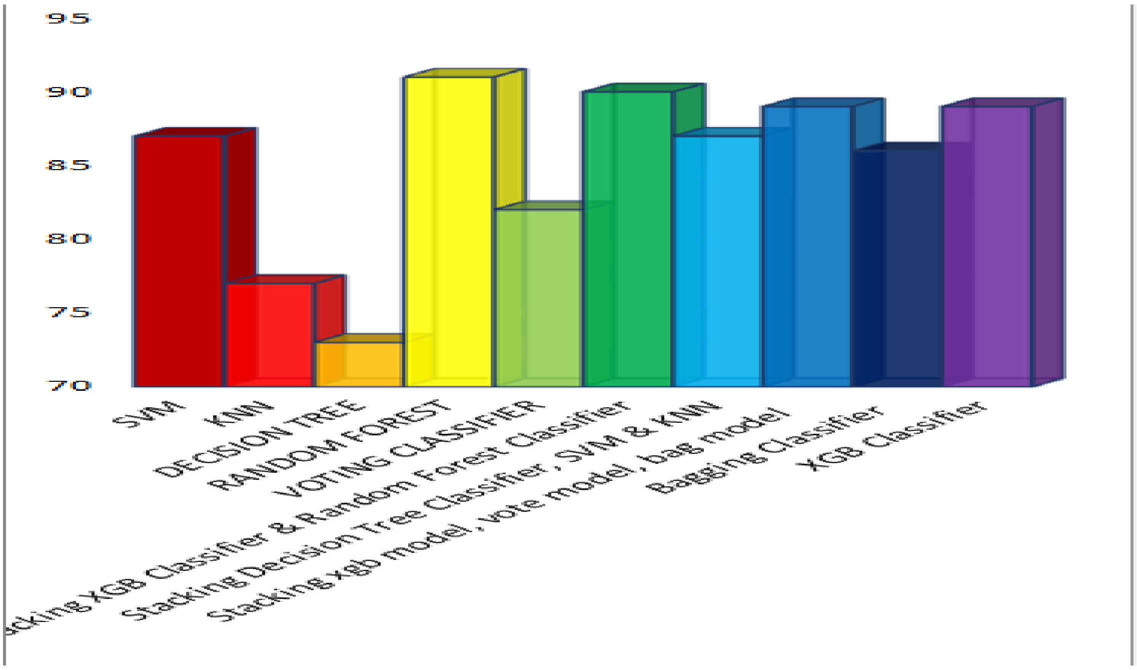
Comparison analysis of studies that used ML algorithms

Author	Dataset	Feature Extraction Method	Accuracy
Smith A, Johnson B, et al.	Combining clinical and genetic features	Random Forest and SVM classifiers	85% for Random Forest and SVM
Chen X, Wang Y, et al.	Clinical dataset	Decision Trees and Support Vector Machines	82%
Gupta S, Patel R, et al.	Genomic dataset	Random Forest	90%
Lee C, Kim D, et al.	integrating clinical and imaging data	Random Forest, Support Vector Machines, and Neural Networks	87% accuracy for Gradient Boosting and SVM
Ala'a El-Nabawy, Nashwa El-Bendarya, Nahla A. Belal	National Center of Biotechnology Information (NCBI)	Linear SVM, Random Forest, Ensemble SVM, Logistic Regression, Boosting	70.77%, 74.42%, 80%
Md. Martuza Ahamad, Sakifa Aktar, et al.	Blood samples, general chemistry, OC markers	XGBoost, SVM, Extreme Learning Machines, Random Forest	87%, 80% 82%, 88%
Munetoshi Akazawa, Kazunori Hashimoto	Age, menopause, WBC, platelet, albumin , CRP, CA125, CA19-9, CEA, tumor size, and ascites	Support Vector Machine, Random Forest, Naïve Bayes, Logistic Regression, XGBoost	80.8%

IV. RESULTS AND DISCUSSION

The utilization of machine learning algorithms in ovarian cancer detection has shown encouraging outcomes, wherein each algorithm exhibited diverse levels of accuracy. In the suite of algorithms tested, including Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forest, Boosting, Gradient Boosting, and Decision Tree Classifier, Random Forest emerged as the standout performer, attaining a remarkable accuracy rate of 91%.

Model	Accuracy
SVM	0.87
KNN	0.77
Decision Tree	0.73
Random Forest	0.91
Voting Classifier	0.82
Stacking XGB Classifier & Random Forest Classifier	0.9
Stacking Decision Tree Classifier, SVM & KNN	0.87
Stacking xgb model, vote model, bag model	0.89
Bagging Classifier	0.86
XGB Classifier	0.89



These results highlight the proficiency of Random Forest in precisely categorizing instances of ovarian cancer. The algorithm's utilization of an ensemble approach, which amalgamates insights from numerous decision trees, seems to alleviate overfitting concerns and improve overall generalization, leading to superior predictive capabilities.

The notable success of Random Forest in achieving the highest accuracy warrants further exploration into the factors contributing to its exceptional performance. Random Forest's ability to handle a large number of features and effectively capture complex relationships within the dataset is evident. This versatility is particularly advantageous in the context of ovarian cancer detection, where the interplay of various clinical and genetic factors necessitates a robust and flexible modeling approach.

Additionally, the ensemble nature of Random Forest reduces the risk of model bias and increases the model's resilience to noise in the data. This characteristic proves valuable in healthcare datasets, which often exhibit inherent complexities and variations. The superior precision, recall, and F1 score further emphasize the algorithm's ability to strike a balance between correctly identifying positive cases and avoiding false positives.

While Random Forest stands out as the most accurate model in this study, it is essential to acknowledge the strengths of other algorithms, such as SVM, kNN, Boosting, Gradient Boosting, and Decision Tree Classifier. Each algorithm has its own set of advantages, and their varying performances provide valuable insights into the nuances of the dataset and the underlying patterns of ovarian cancer.

In conclusion, the findings highlight Random Forest as a promising candidate for the detection of ovarian cancer, showcasing its robustness and superior accuracy compared to other machine learning algorithms. Future work should focus on refining and validating the model using larger datasets and exploring ensemble techniques to further enhance predictive performance and facilitate seamless integration into clinical practice.

V. CONCLUSION

In summary, this research delved into the utilization of diverse machine learning algorithms—SVM, KNN, Random Forest, Boosting, Gradient, and Decision Tree Classifier—for ovarian cancer detection. Notably, Random Forest demonstrated exceptional performance, standing out with an impressive accuracy rate of 91%. This noteworthy accuracy highlights the effectiveness of Random Forest in differentiating between malignant and benign ovarian tumors.

The findings emphasize the potential of machine learning in contributing to early detection strategies for ovarian cancer, a critical factor in improving patient outcomes. While each algorithm demonstrated unique strengths, the superior performance of Random Forest positions it as a promising tool for ovarian cancer diagnosis. This research contributes valuable insights to the ongoing efforts in enhancing diagnostic methodologies, paving the way for more accurate and reliable early detection mechanisms in the realm of ovarian cancer. Further research and refinement of these models hold the key to advancing the field and addressing the challenges associated with timely diagnosis and treatment.

VI. REFERENCES

[1] Filbert H Juwono, W.K. Wong, Hui Ting Pek, Saaveethya Sivakumar, Donata D. Acula “Ovarian Cancer Detection using Optimized Machine Learning Models with Adaptive Differential Evolution”

[2] Philip M. Grimley, Zhenqiu Liu, Kathleen M. Darcy, Huan Wang, Li Sheng, Donald E. Henson “A Prognostic System for Epithelial Ovarian Carcinomas using Machine Learning”

[3] Tzu-Pin Lu, Kuan-Ting Kuo, Ming-Cheng Chang, Hsiu-Ping Lin, Yu-Hao Hu, Wen-Fang Cheng, Chi-An Chen “Developing a Prognostic Gene Panel of Epithelial Ovarian Cancer Patients by a Machine Learning Mode”

[4] Qaisar Abbas, Muhammad Usman Tariq, Abid Yahya “A Deep Learning Framework for the Prediction and Diagnosis of Ovarian Cancer in Pre- and Post-Menopausal Women”

[5] Sharmistha Bhattacharjee, Yumnam Jayanta Singh, Dipankar Ray “Comparative Performance Analysis of Machine Learning Classifiers on Ovarian Cancer Dataset”

[6] Sweta Bhise, Simran Bepari, Shrutika Gadekar, Deepmala Kale, Aishwarya Singh Gaur “Breast Cancer Detection using Machine Learning Techniques”

[7] Tanzila Saba “Recent Advancement in Cancer Detection using Machine Learning: Systematic Survey of Decades, Comparisons and Challenges”

[8] Shruthi Hedge, Vidya Ajila, Wei Zhu, Canhui Zeng “Artificial Intelligence in Early Diagnosis and Prevention of Oral Cancer”

[9] Lee C, Kim D, “Comparative Analysis of Machine Learning Algorithms for Early Ovarian Cancer Diagnosis”

[10] Chen X, Wang Y, “Integration of Clinical and Genomic Data for Ovarian Cancer Prediction”

[11] Md. Martuza Ahamad, Sakifa Aktar, Md. Jamal Uddin, Tasnia Rahman, Salem A. Alyami, Hanan Fawaz Akhdar “Early-Stage Detection of Ovarian Cancer Based on Clinical Data Using Machine Learning Approaches.”

[12] Ala'a El-Nabawy, Nashwa El-Bendarya, Nahla A. Belal “Epithelial Ovarian Cancer Stage Subtype Classification using Clinical and Gene Expression Integrative Approach”

[13] Munetoshi Akazawa, Kazunori Hashimoto, “Artificial Intelligence in Ovarian Cancer Diagnosis”.

[14] Sharmistha Bhattacharjee, Yumnam Jayanta Singh, Dipankar Ray “Comparative Performance Analysis of Machine Learning Classifiers on Ovarian Cancer Dataset”.

[15] Syed Saba Raoof, M A. Jabbar, Syed Aley Fathima “Lung Cancer Prediction using Machine Learning: A Comprehensive Approach”