ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Prediction Of Ovarian Cancer Using Machine Learning

[1] Bhumika G, [2] Dr. G. Sasikala
[1] Student,School Of CSA, bhumikagangadhar30@gmail.com.
[2] Associate Professor,School Of CSA, REVA University, Bengaluru, India.
sasikala.g@reva.edu.in

Abstract: ***Women's morality has grown recently as a result of gynecological cancer diagnoses. It is mostly the outcome of a delayed diagnosis caused by a delayed discovery of the illness. Ovarian cancer is typically discovered late and frequently has peritoneal and distant metastases at diagnosis since there aren't many distinct signs and symptoms in the early stages of the illness. Detecting this type of cancer early is crucial for improving survival rates. However, it poses a significant challenge. Early disease detection may be feasible by the use of biomarker data, such as proteomic, genomic, and other molecular data. To anticipate the risk of ovarian cancer and analyze biomarker data, machine learning techniques have become a prominent tool. These methods are gaining greater recognition for their ability to effectively assess biomarker data and support ovarian cancer risk prediction. In this work, we employed various algorithms like logistic regression, support vector machine (SVM), k-nearest neighbor (KNN), random forest, and decision trees. Among these approaches, the random forest classifier has shown promising results, demonstrating a high level of accuracy in predicting ovarian cancer risk. This advancement in predictive modeling holds significant promise for improving patient outcomes by enabling early detection and the implementation of personalized treatment strategies. Ultimately, such efforts contribute to enhancing overall healthcare practices and the well-being of women affected by gynecological cancer.***

I. INTRODUCTION

Ovarian cancer, a formidable adversary in the realm of oncology, is notorious for its asymptomatic nature and often late-stage detection, contributing to the challenges associated with effective intervention. The quest for improved diagnostic tools has led to the integration of machine learning techniques, presenting a promising avenue for enhancing the accuracy and timeliness of ovarian cancer detection.This project endeavours to develop a predictive model for ovarian cancer using machine learning, leveraging the wealth of data available in medical records, genetic profiles, and patient histories. The overarching goal is to contribute to the advancement of early detection methods, ultimately leading to improved patient outcomes in the face of this complex and insidious disease.The significance of early detection in ovarian cancer cannot be overstated, as the disease typically manifests subtle symptoms or remains asymptomatic until advanced stages. By harnessing the power of machine learning, this project seeks to identify patterns and relationships within diverse datasets that may elude traditional diagnostic approaches. The ability to predict the likelihood of ovarian cancer in individuals can potentially revolutionize clinical practices, enabling healthcare professionals to intervene proactively and tailor treatment plans more effectively.The project is structured to follow a systematic and rigorous methodology, encompassing data collection, preprocessing, feature selection, model training, and evaluation. The choice of machine learning algorithm is pivotal, considering the nuances of ovarian cancer prediction. The algorithms selected for this project will undergo meticulous training and refinement through hyperparameter tuning to optimize their predictive capabilities.Feature selection plays a crucial role in the project, aiming to distil the most relevant information from the comprehensive datasets. By identifying key features associated with ovarian cancer, the model's interpretability is enhanced, and the potential for overfitting is mitigated. Furthermore, the project recognizes the importance of ethical considerations and privacy in handling sensitive medical data, adhering to regulatory standards and collaborating closely with healthcare professionals throughout the development process.Validation and cross-validation techniques will be employed to ensure the model's robustness across diverse datasets, simulating real-world conditions and bolstering confidence in its generalizability. Additionally, the project emphasizes the interpretability of the model's predictions, acknowledging the importance of fostering trust among healthcare professionals who will utilize these tools in clinical settings.

II. Existing System

In the existing system, researchers and healthcare professionals are exploring various machine learning algorithms for predicting ovarian cancer. Logistic Regression, Support Vector Machines, Random Forest, and Gradient Boosting are among the commonly employed algorithms for binary classification tasks. These algorithms undergo rigorous training on diverse

datasets, learning to identify subtle patterns indicative of ovarian cancer presence. The process involves preprocessing steps such as handling missing data, encoding categorical variables, and scaling numerical features to ensure the input data is conducive to effective model training.

In conclusion, while the existing system for ovarian cancer prediction using machine learning demonstrates significant promise, it is essential to acknowledge and address these limitations. Ongoing research and advancements in data collection, model interpretability, and ethical considerations are vital to overcoming these challenges and realizing the full potential of machine learning in improving ovarian cancer diagnostics

Limitation Of Exixting System

The limitations of the current approaches highlight the need for more advanced and precise diagnostic tools. Machine learning models, with their ability to analyse complex datasets, offer a promising avenue for improving the accuracy and early detection of ovarian cancer. Integrating machine learning into the existing diagnostic framework could potentially enhance sensitivity, specificity, and overall diagnostic performance. However, it is crucial to address ethical considerations, data privacy issues, and ensure the interpretability of these models before widespread clinical adoption.

Proposed System

A user-friendly interface will be developed, catering to both healthcare professionals and data scientists.Interactive dashboards and visualizations will enhance user experience and facilitate effective interaction with the system.In summary, the proposed system for ovarian cancer prediction harnesses the power of machine learning, with a particular focus on leveraging the Kaggle dataset. By incorporating this diverse and well-structured data source, the system aims to overcome limitations of traditional datasets, ultimately contributing to more accurate and reliable predictions for early ovarian cancer detection.

The proposed system combines the advantages of utilizing the Kaggle dataset, incorporating advanced machine learning techniques, and prioritizing user accessibility and ethical considerations. These features collectively position the system as a promising tool for improving ovarian cancer prediction, ultimately contributing to more effective and timely diagnoses in clinical settings.

Results

Table 2. Classifier Performance: Accuracy Comparison

Classifiers	**Accuracy (%)**
Logistic Regression	82
SVM	81
Navie Bayes	84
Random forest	93
KNN	89

The performance evaluation of different classifiers reveals varying degrees of accuracy in predicting the target variable. Random Forest emerges as the most accurate classifier among the tested models, achieving an impressive accuracy of 93%. Navie Bayes also demonstrates strong performance, with an accuracy of 84%, showcasing its effectiveness in the given context. KNN follows closely with an accuracy of 89%, while Logistic Regression and SVM exhibit respectable accuracies of 82% and 81%, respectively. These findings underscore the importance of selecting the appropriate classifier based on the specific characteristics and requirements of the dataset, with Random Forest proving to be the standout performer in this comparative analysis.

Table 3. Performance Evaluation Metrics for Various Classifiers

Classifiers		**Precision**	**Recall**	F1-**Score**
Logistic Regression	0	0.81	0.86	0.83
	1	0.85	0.80	0.82
SVM	0	0.83	0.54	0.66
	1	0.66	0.89	0.76

Navie Bayes	0	0.82	0.80	0.81
	1	0.81	0.83	0.82
Random forest	0	0.88	0.86	0.87
	1	0.86	0.87	0.89
KNN	0	0.81	0.60	0.69
	1	0.68	0.89	0.76

The evaluation of precision, recall, and F1-Score for each classifier provides a nuanced understanding of their performance. In precision, Random Forest stands out with high values for both classes (0 and 1), indicating its ability to correctly identify positive instances while minimizing false positives. This is further emphasized by its elevated recall values, demonstrating a balanced capacity to capture true positives across both classes. Navie Bayes exhibits commendable precision and recall scores, showcasing a well-rounded performance. Logistic Regression and SVM, while maintaining decent overall scores, exhibit trade-offs between precision and recall, highlighting the challenge of balancing the two metrics. KNN, while achieving competitive results, demonstrates a noticeable trade-off between precision and recall, suggesting potential sensitivity to parameter tuning. Overall, the detailed metrics underscore the importance of considering both precision and recall in classifier selection, with Random Forest emerging as a robust choice for achieving a harmonious balance between these metrics in the given context.

III. Conclusion

The conclusion highlights the growing prevalence of ovarian cancer globally and emphasizes the critical need for precise classification between ovarian cancer and tumours. Recognizing the significance of minimizing false classifications, the conclusion suggests leveraging computer-assisted and machine learning methods to enhance accuracy while eliminating human errors in the diagnostic process. By employing these advanced technologies, the conclusion suggests that more accurate disease classification can be achieved, facilitating the delivery of appropriate treatments.
Specifically, the conclusion points out that among the evaluated methods, Random Forest with median imputation stands out with an impressive accuracy of 93%. This high accuracy makes it a strong candidate for the task of distinguishing between ovarian cancer and tumours. The implication is that utilizing Random Forest, along with the specified imputation technique, could contribute significantly to the accurate identification of the disease, thereby aiding in the provision of timely and targeted medical interventions for individuals affected by ovarian cancer.

Reference

[1] Ullas Pandey1, Tushar Deshpande1, Amit Agrawal2, and Shiv Govind Singh// Graphene-Oxide-Assisted Biosensor With Optimum Response Selection Algorithm for Detecting and Quantifying Vimentin, a Potential Biomarker for Ovarian Cancer, 2023.
[2] Hye-Jeong Song, Eun-Suk Yang, Jong-Dae Kim, Chan-Young Park, Yu-Seop Kim and Min-Sun Kyung Bio-IT Research Center, Hallym University// Improving Performance for Classifying Ovarian Cancer with Menopause Information, 2018.
[3] Feng Yan , Ji-Hee Ha, Yuyang Yan, Sam B. Ton, Chen Wang, Bornface Mutembei , Zaid A. Alhajeri , Aubrey F. McNiel, Andrew J. Keddissi, Qinghao Zhang, Muralidharan Jayaraman, Danny N. Dhanasekaran, and Qinggong Tang// Optical Coherence Tomography of Tumor Spheroids Identifies Candidates for Drug Repurposing in Ovarian Cancer, 2023.
[4] Laboni Akter and Nasrin Akhter Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh// Ovarian Cancer Classification from Pathophysiological Complications using Machine Learning Techniques, 2021.
[5] Aditya MS, Amrita I, Dr. Ashwini Kodipalli, Dr. Roshan Joy Martis, Global Academy of Technology Bengaluru, India// Ovarian Cancer Detection and Classification Using Machine Leaning, 2021.
[6] Rinki Singh, Anup Som, Centre of Bioinformatics University of Allahabad Allahabad // Profiling of Ovarian Cancer Reveals Common Features Shared by Sub-types of Ovarian Cancer: A network biology approach towards finding therapeutic drug targets, 2023.

[7] Filbert H. Juwono a,* , W.K. Wong a , Hui Ting Pek b , Saaveethya Sivakumar a , Donata D. Acula c, a Department of Electrical and Computer Engineering, Curtin University Malaysia, 98009 Miri, Sarawak, Malaysia// Ovarian cancer detection using optimized machine learning models with adaptive differential evolution, 2022.
[8] Md. Ali Hossaina,b,1 , Sheikh Muhammad Saiful Islamc,1 , Julian M.W. Quinnd , Fazlul Huqe , Mohammad Ali Monid,e,∗ a Dept of CSE, Manarat International University, Dhaka 1212, Bangladesh b Dept of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh c Dept. of Pharmacy, Manarat International University, Dhaka 1212, Bangladesh d Bone Biology Divisions, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia e The University of Sydney, School of Medical Sciences, Faculty of Medicine & Health, NSW 2006, Australia// Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality, 2019.
[9] Sumit Das1*, Dipansu Mondal2, Priyanka Roy1, Tanusree Das1, Risha Roy1, Diprajyoti Majumdar1 1JIS College of Engineering, Information Technology, Kalyani, 741235, India 2 University of Kalyani, Centre for Information Resource Management, 741235, India// A Comparative Analysis and Prediction of Ovarian Cancer using AI Approach.
[10] Donna Badgwell and Robert C. Bast Jr.∗ Department of Experimental Therapeutics, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA// Early detection of ovarian cancer.