An Analysis For Early Diagnosis To Detect Ovarian Cancer

P. Thenmozhi, Dr. T. Parimalam

Assistant Professor, Department of Computer Science, Nandha Arts and Science College, Erode
Associate Professor and Head, Department of Computer Science, Nandha Arts and Science College, Erode

Abstract: Ovarian cancer stands as the primary cause of mortality among gynecological cancers, creating an urgent need for innovative therapies. Despite the established correlation between the presence of tumor-infiltrating lymphocytes and improved prognosis, the integration of immunotherapy into the ovarian cancer treatment arsenal has not yet occurred. Over the past decades, extensive research has been dedicated to identifying early signs of ovarian cancer. Recognized as the "silent killer," ovarian cancer often remains asymptomatic in its early stages, leading to late-stage discoveries and poor prognosis. Early detection is pivotal for timely intervention based on symptomatic manifestations. Various categorization systems have been employed for ovarian tumor diagnosis, encompassing non-molecular factors such as tumor tissue, clinical stage, and pathological characteristics. These factors contribute to a comprehensive understanding of the intricate nature of cancer, guiding the development of effective therapeutic approaches. This review paper briefly outlines the preprocessing, feature extraction, and classification methodologies employed in the analysis of ovarian cancer across various studies from 2018 to 2023, encompassing several research papers.

Keywords: Ovarian Cancer, Deep Learning, CNN, machine Learning

1. Introduction

Ovarian cancer originates in or on an ovary. The growth of abnormal cells makes its incursion through the origin and may subsequently spread to the other parts of the affected person [1]. Initially, cancer doesn't show any significant symptoms. When cancer progresses to further stages, the prevalence of the disease could be uncovered [2]. Unusual abdominal swelling and pain, bloating, pelvic pain, loss of weight, and appetite are some common symptoms. The probability of high-impact attack areas of the cancer is the liver, lungs, abdomen, and lymph nodes [3]. Women who have more than that of higher ovulation than usual in their lifetime are the focal targets of cancer [4].

The following are some aspects proliferating increased risk of this cancer:

1. Young age ovulation
2. Old age menopause
3. Women who don’t have children
4. Hormone treatment alongside the menopause
5. Medication for fertility
6. Stress and obesity

Breastfeeding has a profound impact on decreasing the risk of the disease. Birth control with proper hormone balance and tubal tying are some supplementary aspects to decrease cancer risk.

1.1 Types Of Ovarian Cancer

The historic evidence of ovarian cancer has been published in WHO publication no. 9, 1973, and was subsequently endorsed by the task forces of the FIGO provides the tumor type subdivisions. According to them, the tumor types are as follows: serous, endometrioid, clear cell (mesonephros), mucinous, un-differentiated, and unclassified [5].

1. Serous tumors: Serous tumors are components of the surface epithelial-stromal tumor faction of ovarian neoplasms and originate from the Mullerian epithelium. They are familiar neoplasms with a
strapping tendency to bilaterality and report 50% of all ovarian tumors. 60% are benign (cystadenoma), 10% are borderline and 30% are malignant (cystadenocarcinoma) [6].

2. **Mucinous tumors**: Mucinous tumors are constituent of the outside epithelial-stromal tumor group of ovarian neoplasms, which has nearly 36% ovarian tumors. Nearly 75% are benign staged, 10% are borderline samples, and 15% are malignantly staged. Sometimes, the tumor is seen bilaterally; around 5% of primary mucinous tumors are bilateral [7]. Benign mucinous tumors are usually multilocular. Borderline and malignant mucinous tumors consistently encompass papillae and solid areas. It may also possess hemorrhage and necrosis [8].

3. **Endometrioid tumors**: Endometrioid tumors are a class of tumors formed by a resemblance to endometrium endometrial carcinoma, and over a third of cases have crucial squamous differentiation. They are components of the surface epithelial tumor group of ovarian neoplasms. Benign and norm variants are atypical, as the majority are malignant. There is a connection between endometriosis and concurrent primary endometrial carcinoma (endometrial cancer) [9]. As consolidation in the pathological examination, when the tumor is cystic and solid, it would lead to cystic endometriosis. In 40% of cases, endometrioid tumors are instigated bilaterally [10].

4. **Clear cell tumors**: Clear-cell sarcoma or malignant melanoma is an uncharacter invariant of cancer called sarcoma. It is standard to crop seven normally in the dermis and soft tissues. Nonconforming forms may have their heredity in the gastrointestinal tract earlier than they were discovered to be sundry [11].

5. **Brenner tumor**: They are an extraordinary variant of the float-up epithelial-stromal tumor collection of ovarian neoplasms. The huge numbers are benign; in some cases, it is possible to be malignant. They are most consistently found on pelvic inspection or maybe positioned at laparotomy. Brenner tumors can often arise in further parts, which may take the testes [12].

6. **Undifferentiated carcinomas**: A malignant tumor of epithelial arrangement that is very sensually deviated from being located in any other variant [13]. This type occurs in juvenile cells and doesn’t possess a cell-like constitution. Cancer cells in this abnormal type of structure are dodgier than the well discriminated. These in-discriminated cells are often called anaplastic [14].

7. **Mixed epithelial tumors**: These are the combinatorial type of cancer that may have the instance of two or more types mentioned above [16].

### 1.2 Staging Of Ovarian Cancer

The severity and the extent of the cancer are termed cancer staging. The growth of cancer cells in the originated part and how far it spreads to other parts refer to the staging of the cancer cells [17]. Clinical staging depends on the tests performed before surgery and pathologic staging on tests of tissue removed throughout the operation [18].

1. **Stage I** — it indicates the infection of one ovary or both ovaries
   - IA refers to no tumor on the ovarian surface, malignant cells in ascites, peritoneal clearings, one ovary infection, and capsule intact
   - IB indicates both ovaries; capsule intact; no tumor on the ovarian surface; negative washings
   - IC tumor indicates the ovarian infection of cancer as any one of the following stagings: a tumor on positive clearings, ovarian surface, and capsule ruptured [19]

2. **Stage II** — refers to the cancer infection in the pelvic implant and extension
   - IIA refers the cancer infection in the extension or implants at the region of the fallopian tube or uterus
   - IIB refers to the cancer infection in the region of the extension or implants at pelvic structures
   - IIC indicates the pelvic extension or implants with positive peritoneal clearings [20]

3. **Stage III** — this type refers to the cancer infection at peritoneal implants outside of the pelvis, which may be limited to the pelvis with extension to the small bowel or momentum
   - IIIA refers to the microscopic peritoneal metastases far away from the pelvis
   - IIIB refers to the macroscopic peritoneal metastases ahead of the pelvis less than 2 cm in size
IIIC indicates the peritoneal metastases afar pelvis > 2 cm or lymph node metastases [21]
4. Stage IV — refers to the distant metastases to the liver or exterior than the peritoneal cavity

1.3 Problem Formation
Data analytics creates a crucial role in the health care for early detection of human disease. Interpreting a large amount of data, which is created from different patient records produce a lot of impending information for creating eminence health care at abridged costs. Traditional data analysis methods have become insufficient for handling such huge volume of data. Knowledge discovery of data is a new method that incorporates a range of statistical analysis, pattern recognition, and machine learning methods, which utilize the acquaintance from huge amount of collected data. Data Mining is usually deployed as pattern retrieval technique, which combines and extracts the knowledge from massive information. Data mining methods have officially confirmed to wide range of medicine, like diagnosis, prognosis, and treatment.

2. Background Study
2.1 Ovarian cancer prediction
A. Yaar et al. (2020) The authors demonstrate that learning using privileged information may improve the generalization performance for predicting chemo sensitivity in OvCa patients by using cross-domain knowledge distillation from gene expression profiling to whole slide imaging. The author has also shown that LUPI facilitates cross-domain learning by allowing input space data to be used optimally. Future applications of LUPI knowledge distillation in cross-domain medical image analysis and computational pathology may build on the findings of this work.

Alizadeh, L. et al. (2018) The diagnosis, imaging, and treatment of ovarian cancer are just some of the many applications that have benefited from the use of nanocarriers derived from chatoyant in recent years. These nanocarriers have exciting characteristics such as good biodistribution, biodegradation, biocompatibility, good manipulating capability, and low toxicity. A broad range of pharmaceuticals may be delivered through functionalized chitosan nanoparticles thanks to the presence of reactive groups on the surface of chitosan. Ovarian cancer treatment plans may in the future use chitosan-based nanocarriers for both active and passive targeting because of this property.

Barber, E.et al. (2021) In this work, the author assessed the feasibility of combining machine learning and natural language processing to enhance the accuracy with which postoperative complications and readmissions can be predicted for women with ovarian cancer. The author discovered that the discrimination of the models using machine learning models like random forest or XGBoost was only slightly better than using the more standard method of logistic regression for predicting what would happen after surgery.

C. Nayak et al. (2023) late diagnosis was the leading cause of mortality from ovarian cancer. This study recommends using machine learning methods to aid medical science in the fight against this fatal disease. Although these techniques were no substitute for medical professionals, they may be helpful in areas where modern medical care was difficult to get. It may also help a doctor make a more informed choice. The suggested methods were limited in that they can only be used to determine whether or not a person has cancer. It will be blind to the extent of the problem. Further development of this system will allow for more precise predictions and the identification of appropriate feature selection strategies for the straightforward diagnosis of ovarian cancer.

2.2 Ovarian cancer prediction using machine learning
D. Lin et al. (2021) previous research has shown that these three genes play a crucial role in the progression of ovarian cancer. Earlier genes in these authors shortlist of candidates, such as ACTB, EEF1A1, and CD74, have already been validated as good ovarian cancer reference genes in earlier research. These findings corroborate these authors study's hypothesized relationship between highly expressed genes in tumors and cancer development.

Guo, L.-Y et al. (2020) In order to choose appropriate treatment plans and predict clinical outcomes, a deeper understanding of the heterogeneity of ovarian cancer across individuals was necessary. In this research,
the author aimed to identify the subtypes of ovarian cancer by proposing a unique deep learning framework for integrating multi-omics data with denoising auto encoder. Ovarian cancer was broken down into two molecular subgroups, proving the efficacy and viability of these authors suggested approach. The author found that these authors strategy performed better than both the conventional and the deep learning-based alternatives. In addition, three GEO test datasets were used to validate the categorization model. All the p-values for the differences between the various cancer groupings were less than 0.05, indicating that they were statistically significant.

Icki, M. G. et al. (2021) in this group of women with ovarian cancer, the author discovered that the Frailty Index was significantly correlated with both postoperative complications and overall survival. This highlights the value of preoperative screening, especially for the elderly and other at-risk patients. The Frailty Index is a powerful tool that is also manageable during busy office routine and may be useful to doctors in identifying patients at increased risk and aiding in treatment planning.

Jiang, X. et al. (2019) The existence of homologous recombination deficit, which may be investigated by mutational analysis of HRD gene panels, genomic scar analysis, and functional testing, remained a robust predictor of clinical success from PARP medicines. The HRD biomarker has not been shown to effectively identify the fraction of patients with wild-type BRCA who would benefit much from PARP inhibitor treatment in terms of progression-free survival. This incapacity might be the consequence of other justifications for PARP inhibitor sensitivity or additional HRD processes not discovered by present techniques. Deficits in the HR pathway may also be brought on by a wide variety of therapy combinations. In particular for BRCA proficient ovarian cancer, the response to platinum based chemotherapy remained a powerful predictor of the response to PARP inhibitor treatment.

Laios, A. et al. (2020) The author looked at the challenge of predicting CCR in aEOC patients using pre- and intra-operative clinical factors, and the author compared AI with traditional regression models under the identical resembling settings. The results showed that the k-NN method, which was particularly reflective of "previous clinical experience," may be used to accurately forecast R0 resection during aEOC surgery. Traditional logistic regression was marginally outperformed by the model.

M. E. Frésard et al. (2020) here, the author provide a fresh method of using multi-objective ML to build clinical prediction models. Although other methods have been used to predict VTE/DVT in patients with ovarian cancer, this was the first time that data asymmetry was taken into account. The proposed models outperform prior models, but more external validation was required before they can be recommended for use in the decision-making process.

Octaviani, T. L. et al. (2019) Classifying ovarian cancer data using this strategy may be a useful resource for doctors making a diagnosis. However, more work has to be done on it before it can be used accurately. In the future, it will be put to use in other areas, such as the economy.

prabhakar, S. K., & Lee, S.-W. (2020) Ovarian cancer was the most frequent kind of gynecological cancer. Computer-Aided Diagnosis was crucial for making an accurate diagnosis. Micro array technology allows the expression levels of thousands of genes to be monitored under controlled circumstances all at once. The examination of gene expressions was made feasible by microarray technology, and a massive quantity of data was produced as a result. Therefore, additional processing was challenging owing to the curse of dimensionality issue and a limited sample space. This work proposes a two-tiered feature selection procedure, the first involving the use of conventional gene selection methods, and the second including the use of optimization strategies, both of which must precede any kind of categorization. Classification accuracies of 98.96% and 98.69% were achieved, respectively, when the results of T-static tests were further improved using CFO and LAPO and then categorized using MLP and LDA.
Table 1: Comparison Table for Existing Work

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Methodology</th>
<th>Limitation</th>
<th>Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Nayak et al.</td>
<td>2023</td>
<td>Machine learning</td>
<td>While the reported accuracy was high, it's essential to assess how well the model generalizes to unseen data.</td>
<td>ovarian cancer dataset was collected from Kaggel</td>
<td>91.00%</td>
</tr>
<tr>
<td>Yang, X. et al.</td>
<td>2020</td>
<td>Machine Learning</td>
<td>The study should ideally validate its findings on external datasets to demonstrate the generalizability of the model.</td>
<td>prostate, Lung, Colorectal and Ovarian dataset</td>
<td>90.41%</td>
</tr>
<tr>
<td>Rahman, M. A. et al.</td>
<td>2019</td>
<td>Artificial Neural Network</td>
<td>The high accuracy achieved by the proposed model might raise concerns about overfitting, especially if the dataset used for training was small or unrepresentative.</td>
<td>benchmarked on ovarian cancer dataset</td>
<td>98.70%</td>
</tr>
<tr>
<td>Yeganeh, P. N., &amp; Mostafavi, M. T.</td>
<td>2018</td>
<td>Microarray Analysis</td>
<td>The study mentions the use of an integrated dataset with 530 ovarian tissues. The size of the dataset may have an impact on the generalizability of the results.</td>
<td>Seven ovarian cancer-related datasets were retrieved by accessing NCBI gene expression omnibus (GEO) portal.</td>
<td>81.00%</td>
</tr>
</tbody>
</table>

Rahman, M. et al. (2019) This research analyzes the efficacy of a 15-neuron ANN model for classifying ovarian cancer. Accurate categorization with a high classification rate has been accomplished. Using the ovarian cancer dataset, the suggested 15-Neuron ANN model obtained 98.7% classification accuracy, making it a major step forward in the field. It was clear from the results that the suggested cancer categorization model was a significant step forward in assisting doctors in making informed treatment decisions for their patients. While this study examined ovarian cancer data, feature selection was not a part of the process. With this improvement, precision should increase.

2.3 Ovarian cancer prediction using deep learning

Rizzo, S. et al. (2018) this research concluded that the association between specific radiomic traits and prognosis suggests that radiomics may provide useful supplementary information for OC patients. Texture, mass size, and homogeneity were all associated with disease progression in this cohort of women during the first 12 months of diagnosis, whereas mass size, randomness, and homogeneity were all connected to residual tumour.

Shannon, N. B. et al. (2021) To wrap things off, the author emphasize how molecular traits, in addition to established clinicopathological variables, may play a crucial role in predicting the risk of platinum-based chemoresistance. It was quite simple to include combined clinical and genetic models into regular histopathological examination of tumor specimens, and they may be useful in identifying individuals at high risk of acquiring platinum resistance. If these individuals could be identified, their care following first cytoreductive
surgery and the choice of adjuvant treatment may be optimized. Future research may investigate the efficacy of increased monitoring or the use of second-line chemotherapy drugs other than the conventional platinum and taxane combination for patients at high risk of developing treatment resistance. As a result, the author could tailor each patient's chemotherapy treatment plan to more effectively attack the unique biology of each tumor.

Sharbatoghli, M. et al. (2020) The ability to identify ctDNA before to therapy not only aids in early diagnosis, but also leads to more informed treatment decisions thanks to better patient stratification. Keeping an eye on any traces of the tumor may help cut down on the risk of it returning. Drug resistance due to genetic mutations was always present but cannot be detected using traditional methods; however, this may be revealed with routine monitoring of ctDNA throughout therapy. Therefore, new doors may be opened in oncology via genomic-based medication response prediction, improving care for cancer patients.

Song, H.-J. et al. (2018) High performance in cancer detection was seen in combinations including TTR and Prolactin. Her-ELISA, PDGF-AA, Prolactin, and TTR were the most reliable and accurate biomarkers for differentiating between healthy and cancerous tissue samples. High death rates from ovarian cancer may be mitigated by early diagnosis. It was crucial to find many biomarkers to use in diagnostic tests that have high sensitivity and specificity.

Wibowo, V. et al. (2021) This research tested the performance of the supervised machine learning techniques of KNN and SVM with the RBF kernel on ovarian cancer classification data. In all, there were 203 cases and 5 characteristics taken from the Al Islam Bandung Hospital dataset. There were 130 cases of ovarian cancer and 73 other types of cancer. The KNN concept was developed to assign a sample's class based on the sample's majority class. The closest distance was used to determine this. Meanwhile, SVM's goal was to find the best possible hyperplane for clustering the data. To do this, the author determined the largest possible gap between the nearest vectors and the hyperplane.

Yang, X. et al. (2020) This research demonstrates that the rate of change of CA125, together with six additional characteristics chosen by feature engineering, enhances the prediction of ovarian cancer in individuals who were cancer-positive. Class-imbalanced approaches were explored since there was a large disparity between the amount of malignant tumour samples and benign samples in this data set. Using the nine traditional classifiers, the author find that applying class imbalance learning to the best-performing decision tree boosts the model's prediction power for positive situations.

Yeganeh, P. N., & Mostafavi, M. T. (2018) The results of this research demonstrate that the 26-mRNA panel may be used to effectively stratify ovarian cancer tissues with high sensitivity and specificity. The superset offers sufficient sample heterogeneity across tissues relevant to HGSOC, which was necessary since false positive predictions of the present biomarkers emerge from both healthy persons and patients with benign tumors. Random Forest and Support Vector Machine classification pipelines were shown to be effective in separating HGSOC from benign tumors, normal ovarian tissue, and fallopian tube tissue.

3. Discussion

The confluence of radiomics and machine learning in the context of ovarian cancer introduces a transformative paradigm aimed at bolstering diagnostics and tailoring treatments. Statistical insights, though evolving, underscore the potential impact of this integrated approach on patient outcomes. Firstly, early diagnosis, a critical determinant of survival rates, stands to benefit significantly. Statistical trends indicate that ovarian cancer often presents at advanced stages, contributing to its formidable mortality rates. The fusion of radiomics and machine learning, as observed in current studies, shows a statistical correlation with improved early detection rates. This is pivotal as early interventions correlate with better prognoses and increased chances of successful treatment. Additionally, statistical analyses emphasize the potential for personalized treatment strategies. Ovarian cancer is known for its molecular and pathological heterogeneity, and statistical models leveraging radiomic data can discern patterns that guide more tailored therapeutic interventions. Such personalized approaches align with the evolving paradigm of precision medicine, aiming to improve treatment efficacy and reduce adverse effects. However, statistical considerations also bring attention to the challenges at hand. The need for diverse and comprehensive datasets is statistically evident, emphasizing that the robustness of machine learning models is contingent on the inclusivity and representativeness of the data they are trained
on. Furthermore, statistical assessments underscore the significance of careful feature selection to ensure that machine learning models generalize well beyond the training set.

Looking forward, statistical records reflect the dynamic nature of ongoing research trends. Increasingly sophisticated machine learning algorithms, statistically proven for heightened predictive accuracy, are becoming pivotal. The statistical integration of novel biomarkers, including genomic data, enhances the comprehensiveness of the approach, promising a more holistic understanding of ovarian cancer.

4. Conclusion

In this review underscores the critical importance of early detection in ovarian cancer, often termed the "silent killer." While existing methodologies have shown promise, achieving a reported accuracy of up to 98.7%, there exists a pressing need to address limitations. Recognizing the drawbacks in current approaches, particularly in achieving optimal accuracy, the imperative lies in exploring diverse neural network architectures and hybrid algorithms. Embracing innovative technologies and collaborative research efforts can aspire to surpass existing limitations and refine detection strategies. The global impact of ovarian cancer demands a concerted push for advancements that enhance diagnostic precision. Navigating forward, the overarching goal is to significantly improve patient outcomes by pushing the boundaries of accuracy in the early detection of this formidable gynecologic malignancy.

Reference


5113