



The Open Biomedical Engineering Journal

Content list available at: <https://openbiomedicalengineeringjournal.com>



REVIEW ARTICLE

A Comprehensive Study on the Application of Machine Learning Algorithms in the Prognosis of Ovarian Cancer

S. Sindhu¹, D. Hemavathi¹, K. Sornalakshmi^{1,*}, G. Sujatha² and S. Srividhya¹

¹Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India

²Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India

Abstract:

Ovarian cancer is the third leading type of cancer found in women in India and ranks seventh globally. Several studies have shown that the population affected by ovarian cancer is profound to increase in the future. It is necessary to take steps for identifying cancer at the early stages to avoid mortality and recurrence. This chapter aims to survey the different ways machine learning models have been used in the prognosis of ovarian cancer - to predict the disease progression, recurrence, and mortality rate; analysis of genomic data sets; correlations and pattern analysis, and finding risk factors. The effective analytics on the imaging and other forms of data available from the patient's electronic health records could unveil the possibilities of better or early diagnosis of ovarian cancer. The chapter will summarize the taxonomy of the various ways in which machine learning helps in ovarian cancer diagnosis, early detection, and treatment. In addition to surveying the current state-of-the-art application of machine learning algorithms for ovarian cancer diagnosis, the chapter aims to provide future research directions.

Keywords: Ovarian cancer diagnosis, Machine learning, Pattern analysis, Correlation analysis, Prediction, Cancer.

Article History

Received: December 05, 2022

Revised: March 22, 2023

Accepted: April 17, 2023

1. INTRODUCTION

The female reproductive tract contains two ovaries on the uterus side, which develops eggs for reproduction, and estrogens and progesterone production. Cancer that starts spreading in the ovaries is called ovarian cancer. Overgrowth of non-susceptible cells inside the ovaries leads to ovarian carcinoma. These cells may never die and accumulated to form a tumor. It is one of the most life-threatening cancers among women in the world. As per statistics, women aged 50-65 with high-risk factors are prone to develop ovarian cancer.

The Major key risk in the development of ovarian cancer are as follows (i) Family history of ovarian or colon cancer, (ii) Women in the period of Postmenopausal (iii) Genetic mutations. Compared to other gynaecological cancer, the survival rate of ovarian cancer patients is meagre. Ovarian carcinoma is coined as a "Silent killer" disease because there is no evidence of patients' symptoms until the last stage.

Cancer cells may travel to other parts of the body and multiply there, but the type of cancer is diagnosed with the

name after finding its primary location. The state where cancer cells travel and start growing in some other part of inside the body is called metastasis [1]. Among all gynaecological cancer, ovarian cancer reports 2.5% of cancers in women.

Based on the type of cell it originates from, ovarian cancer is categorized into three types, as shown in Fig. (1). More than 30 sorts of cancer arise from one of these three types.

The epithelial type of ovarian cancer conceives from the cell that covers the lateral surface of the ovaries. Over 90% of ovarian cancer belong to this category. Miserably Epithelial type doesn't show any symptoms till the last stage of cancer.

Germ cell type of malignancy originates from the cells that produce eggs inside the ovary. These types of carcinomas generally affect youths and are easily treated. Over 5% of women are affected by Germ cell-type cancer.

Stromal cell ovarian carcinoma is a rare disorder that evolves from the connective tissues of ovaries. Nearly 5% of women are affected by this category.

Unlike Epithelial and germ cell types of cancer, the Stromal type has many symptoms such as abnormal hormonal production, menstrual cycle change, *etc.* [2].

* Address correspondence to this author at the Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India; E-mail: sornalak@srmist.edu.in

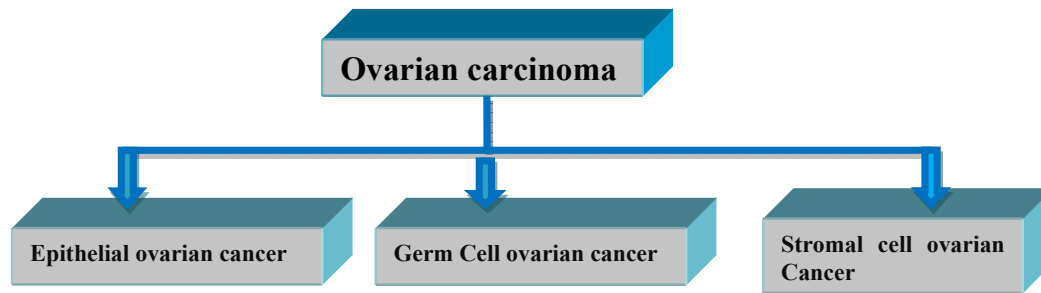


Fig. (1). Types of ovarian carcinoma.

Table 1. Staging in ovarian cancer.

Stage Grouping	Description	Survival Rate (Approximately)	Therapy
Stage I	In this type of cancer, cells are cramped to ovaries and don't spread to any other ovary parts.	90%	Abdominal hysterectomy, Removal of both ovaries and fallopian tubes
Stage II	In this cancer stage, cells may spread around the ovaries, parts such as fallopian tubes, and the uterus may get affected.	70%	In addition to hysterectomy, specimens around the ovaries are tested for infection.
Stage III	The cancer cells may spread beyond the ovaries. It may start laying on the surface of the liver.	39%	Treatment for this stage is the same as stage II. After surgery, the patient might get a chemotherapy procedure if needed.
Stage IV	Cancer has spread to all parts of the body. Fluids taken from other internal organs may contain cancerous cells.	17%	Try to remove the parts where the tumor has spread with the combination of chemotherapy and Radiation therapy.

The percentage of cancer spread inside the patient's body is found by a global system called staging. It helps in deciding the therapy for patients. The categorization of staging with its symptoms and causes is described in Table 1.

The disease can be diagnosed with vaginal ultrasound, CA125 blood test (Biomarkers, usually elevated in ovarian cancer patients), C.T. scans, and biopsy. Therapists generally perform a biopsy, a crucial test that confirms the presence of cancerous cells in ovaries.

Each stage of cancer has different kinds of symptoms. Usually Stage I cancer patient reports very mild or no symptoms at all. In the later stages, symptoms such as frequent urination, constipation, and bloating may occur [3, 4].

Earlier identification of ovarian cancer even before symptoms occur can increase the survival rate of the patients. Various data collected from patients can be stored in analytical sandboxes and analyzed further to predict this deadly disease. Fig. (2) shows the percentage of affected women concerning their age. Middle-aged women are most affected by this disease.

This survey aims at performing a comparison analysis between Machine learning, Deep Learning, and NLP Techniques, which uses various algorithms for the prediction. An accurate forecast for ovarian cancer must satisfy stringent data requirements. The data, such as the patient's profile, images taken from C.T. scan-Ray, and Gene expression is

taken into account. Its required algorithm for analysis is intensely surveyed in this paper.

This review is divided into five sections: Section 2 expresses the Background of existing work in ovarian cancer prognosis. Section 3 briefly explains the Taxonomy of Ovarian cancer prognosis. In section 4, sources available for predicting ovarian cancer are briefly narrated. The various techniques for processing the data are outlined in section 5. The final section concludes the survey with future directions.

2. BACKGROUND

Cells are an essential element for all living beings, responsible for providing energy to the body. Each cell in the human body is comprised of Deoxyribonucleic acid (DNA). DNA contains four types of chemical bases Adenine (A), Guanine (G), Cytosine (C), and Thymine (T) that form the sequence arrangement for building any living creature. Genes are made from Deoxyribonucleic acid (DNA). Each cell in the Human body is wrapped with a thread-like DNA structure called chromosomes. Flaws may occur when the section divides them, which leads to mutation. In addition to environmental factors, sometimes humans carry these flawed genes from their parents. The cell which drastically divides themselves neither repair nor dies will lead to a deadly disease called cancer. Gene expression patterns reflect the discrepancies in the survival of advanced ovarian carcinoma.

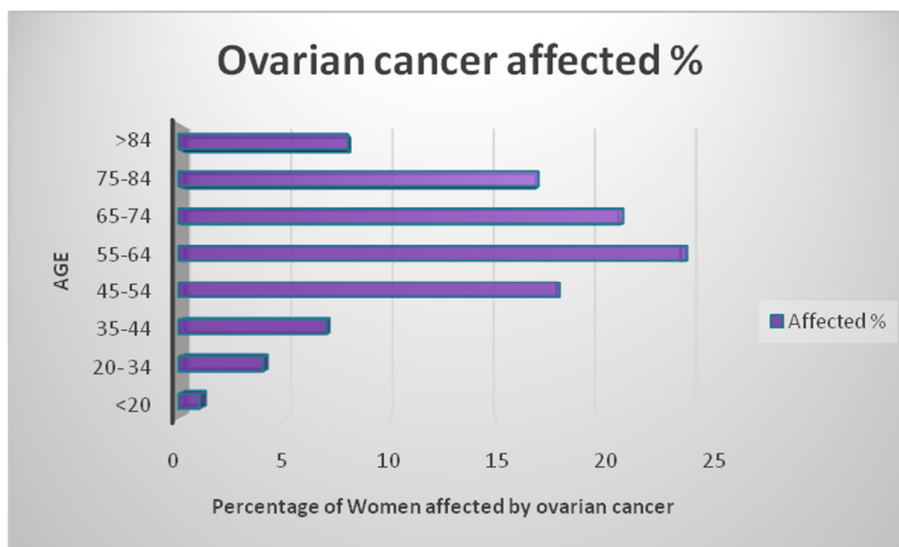


Fig. (2). Percentage of affected women concerning their age.

2.1. Machine Learning in Cancer Prognosis

Identifying ovarian cancer in the early stages is challenging because it doesn't show unusual symptoms in its initial phase. The diagnostic study for various types of ovarian cancer is identified with different kinds of machine learning algorithms. Various classification and clustering algorithms in Machine learning are used in identifying ovarian carcinoma in its early stage. In cancer diagnosis, the focal point is: 1) Predicting the survivability of cancer patients 2) Predicting the reoccurrence of various types of cancer.

2.1.1. Feature Selection Process and Machine Learning Models

The process of feature selection involves automatically choosing the best feature that contributes more to the predictor variable of machine learning models. The feature selection method (FS) aids in improving machine learning model accuracy by decreasing overfitting and simplifies the machine learning model easier to interpret.

The major types of feature selection methods are filter methods, wrapper methods, and embedded methods.

2.1.1.1. Filter Methods

Filter methods capture the unique features assessed by univariate statistics. It performs well with high-dimensional data. Compared to wrapper methods, these techniques are quicker and more computationally efficient.

2.1.1.2. Wrapper Methods

By comparing every possible feature combination to the evaluation criterion, wrapper methods implement a greedy search methodology. It gives better accuracy than filter methods.

2.1.1.3. Embedded Methods

Embedded methods are a combination of filter and wrapper methods.

The authors of a study] used a cervical cancer dataset from UCI Machine Learning Repository collected from Hospital Universitario de Caracas in Caracas, Venezuela [5] This dataset contains the medicinal histories of 858 patients with 36 attributes. By applying Support Vector Machine, K Nearest Neighbor, Random Forest, and C5.0 authors predicted the occurrence of cervical cancer.

Prediction of Ovarian cancer (O.C.) is an essential process in the medical field. The authors suggested that the prognosis of the tumor is in the category of benign or malignant cancer. Cancer value is used to predict the stage of cancer accurately. In one of the papers, a total of 349 patient details were analyzed with the features of demographics, blood routine tests, general chemistry, and tumor markers [6]. In total, 349 attributes were analyzed in this test. Out of 349 patients, 235 patient records were tested. The essential prediction features were extracted using the Machine learning Minimum Redundancy – Maximum Relevance (MRMR) method. Out of which human epididymis protein four and carcinoembryonic antigen were considered to be the most important feature. It has been compared with the Relief F method for a better feature extraction process. One of the required machine learning classification methods, logistic regression, was used to classify the data. Besides, the decision tree and risk of ovarian malignancy algorithm have been compared for better classification. Based on the benignant tumor size, the data was classified in a decision tree. Benign ovarian tumors and ovarian cancers are easily classified with machine learning approaches [6].

Laios *et al.* used the Chi Square test of independence and the Minimum redundancy maximum relevance feature selection method to extract the best features from the categorical data. Various machine learning models including conventional regression analysis were compared For HGSOc prognosis [7].

2.2. Deep Learning in Cancer Prognosis

Deep Learning is a subset of Artificial intelligence that

imitates the human brain structure, which helps us decide with different kinds of data such as text, images, and genomic data. The accuracy of results plays a vital role in ovarian cancer survivability and early prediction. Regarding the voluminous data, deep learning models provide better accuracy with fewer efforts. The classification of neural networks is based on the type of data we are giving as input. The text type of data is handled by Recurrent Neural Network (RNN). The classification of image data is performed by Convolution neural networks (CNN) [8].

The image dataset collected from Guru Gobind Singh Medical College and Hospital Faridkot contained a benign and malignant tumor. The CT images had a record of different stages of ovarian cancer. The initial step began with the pre-processing of raw images. Fuzzy c means the algorithm was used to identify specific regions in the pictures. Various features of the images were retrieved from the scale-invariant feature transform algorithm. Optimized results were obtained using a genetic algorithm. Finally, CNN based classifier was developed for the prediction of various types of cancer [9].

The RNAseq dataset for Lung, Breast, and stomach cancer was collected from TCGA. To avoid the same gene expression data DeSeq method was being used. The neural network was framed with multiple layers to find the optimal prediction [10].

Based on genomic variations, 12 different types of cancer can be predicted using deep neural networks. A deep neural network is built within the Tensor flow model. The Genomic Deep Learning (GDL) method identifies mutation that transforms a normal cell into cancer cells. This GDL identifies cancer in its initial phase [11].

Padideh Danaee *et al.* proposed an in-depth learning approach using a stacked Denoising Autoencoder to extract the relevant feature from gene expression data. Comparison has been made for appropriate feature extraction using principal component analysis and Kernel Principal Component Analysis (KPCA). Further classification of genes related to breast cancer was done using SVM-RBF and deep neural networks [12].

2.3. Natural Language Processing in Cancer Prognosis

Natural Language Processing (NLP) plays a crucial role in making effective decisions in the healthcare industry. It helps in extracting meaningful features from vocal or scripted documents. In cancer treatment, NLP is used for classification, disease recurrence, and various cancer stages. The details collected from the annotated corpus are used for predicting the staging of cancer. After extracting the relevant information, the relation between the entities is made [13].

Cancer-associated information is retrieved from various sources such as scientific papers, health records, and annotated databases. NLP's first step is to categorize named entity recognition by assigning the word to a predefined category.

The next step is information extraction, which automatically extracts the structured data from unorganized data. With the best text classification algorithm's help, the information received is placed under a systematic group. The final step is to retrieve the required information from the health records and other medical-related data repositories [14].

3. DATA PROCESSING TECHNIQUES

Fig. (3) explains Machine Learning approaches in detail. Input data is received through various sources. Later data pre-processing is done for better optimization through data engineering techniques. In supervised learning, data was split into two parts: Training data and testing data. Training data is used to train the model to obtain better accuracy. Test data is used to validate the model. Validation can be done using various methods. Cross-validation is one of the best techniques during the verification of the model. The model is validated with a small sample of values in the given data set is called cross-validation. In unsupervised learning, the attribute grouping process was implemented for grouping similar kind of features. Dimensionality reduction is also an efficient process in feature engineering. It helps to achieve improved accuracy with less number of attributes. Finally, the efficient machine learning model is used in the deployment of applications.

In Fig. (4), the General flow of the Deep learning process is depicted. Deep learning has the essential characteristics to obtain better accuracy is called automatic feature extraction. Also, it has the multilayer property to do in-depth analytics. Finally, the accuracy of the model prediction is done using improved and deep training techniques.

The in-depth learning image classification process using the Convolution Neural Network (CNN) algorithm is depicted in Fig. (5). With sample tumor images as a reference, the relevant image feature identification process has to be completed. After features were retrieved, the data was fed to the Convolution Neural Network classifier. The trained CNN will predict whether the given input contains traces of ovarian carcinoma or not.

In Fig. (6), the text classification process was explained using an in-depth learning approach. Text data relevant to ovarian cancer was considered as input. Relevant features were extracted using text feature extraction methods. The Recurrent Neural Network (RNN) was used to identify the next word's sequence in the text data. It would be helpful to do the prediction analysis of cancer efficiently.

Fig. (7) depicts the workflow of natural language processing. Input data was split using the parsing technique. The information extraction process was implemented after parsing the input text. Similar word identification and sequence pattern patching were implemented during the Information retrieval process. Useful text classification was evaluated after the information retrieval process.

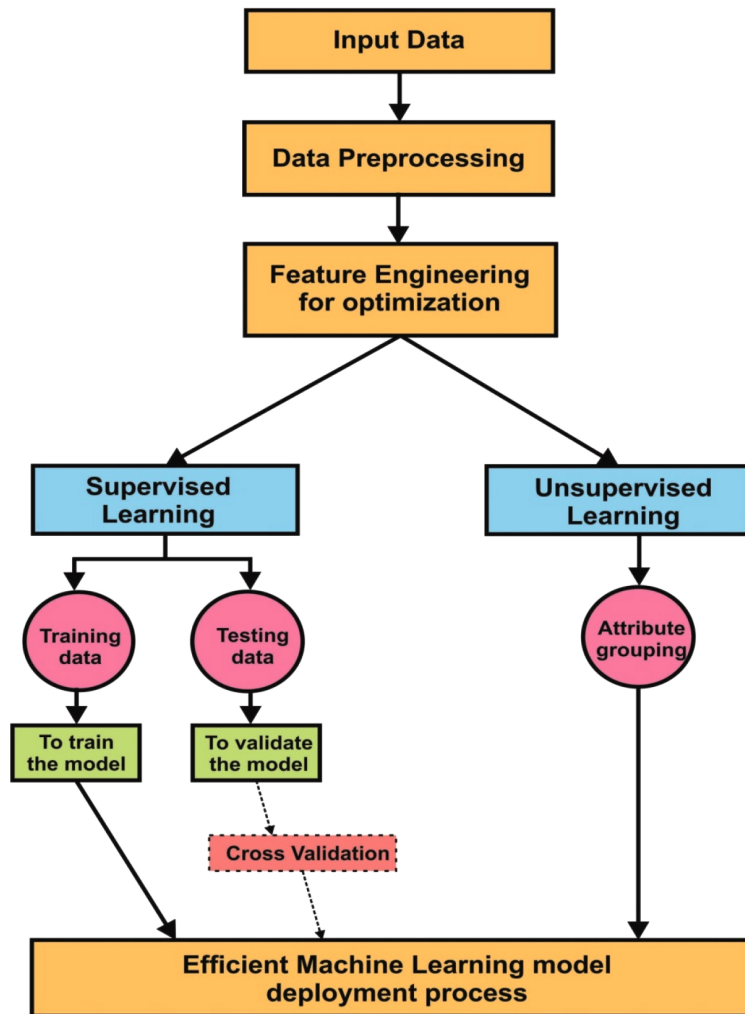


Fig. (3). Machine learning approach.

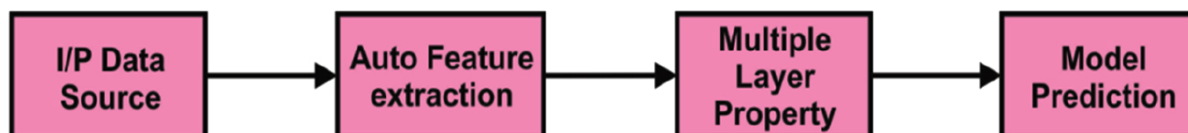


Fig. (4). Flow diagram for deep learning process.

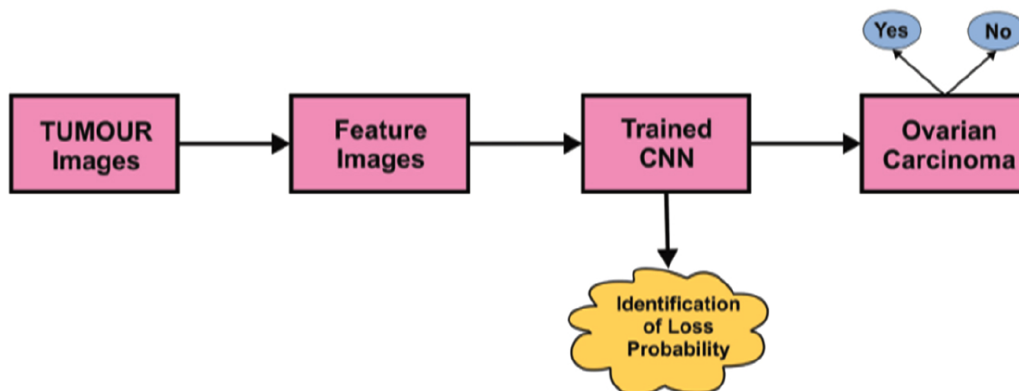


Fig. (5). Image classification process using deep learning.

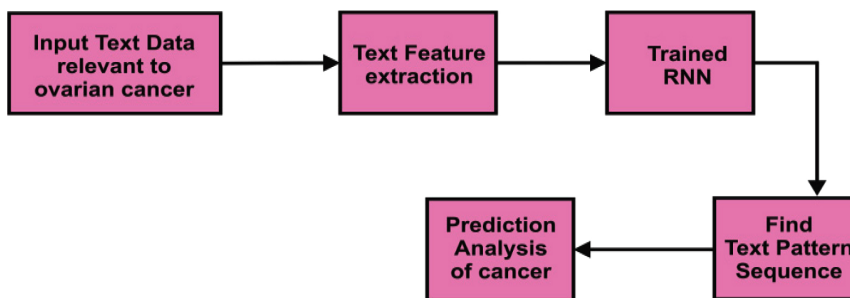


Fig. (6). Deep learning text classification process.

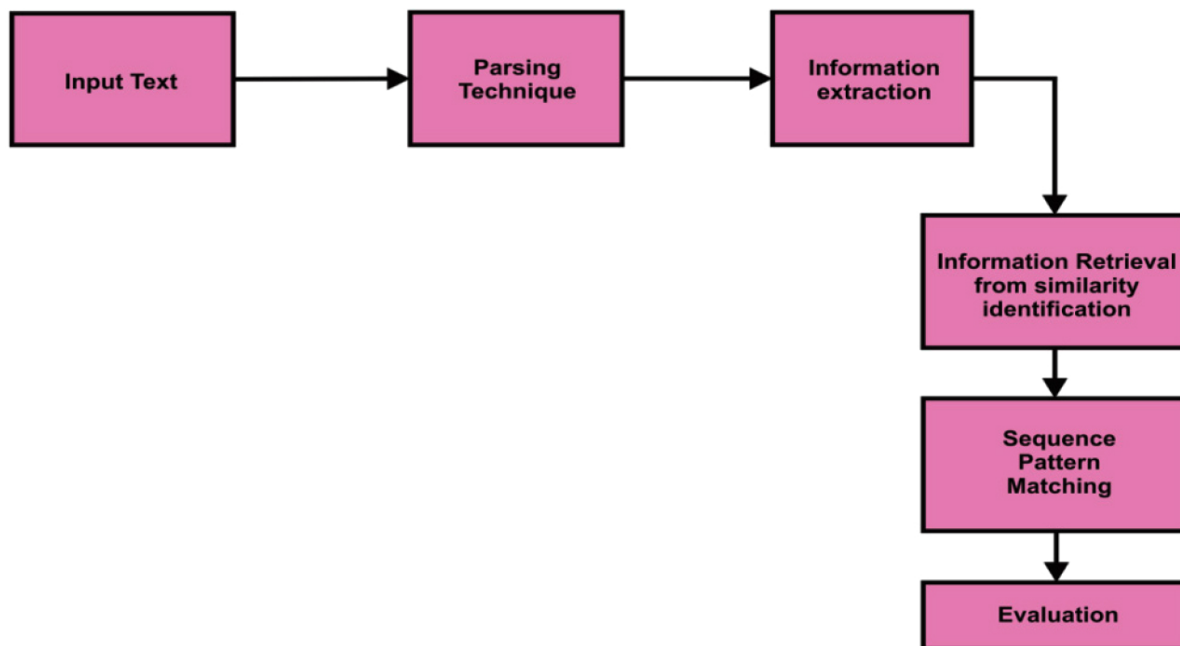


Fig. (7). Workflow of natural language processing.

The general architecture of the deep learning process with multiple hidden layers is shown in Fig. (8). Deep learning always provides better accuracy due to the multilayer protocol. Input data is collected from various sources such as image, text, or genomic representation. It will be classified using multiple layers. In detail, general learning consists of multiple hidden layers for analyzing the data in depth. Fig. (6) shows that there are five layers that act as hidden layers. In that hidden layer, the auto feature extraction, and classification process was implemented efficiently. Finally, the output was obtained with improved classification accuracy.

4. TAXONOMY OF OVARIAN CANCER PROGNOSIS

Ovarian cancer prognosis can be done in three ways: image analytics, text analytics, or Genomic data analysis. The data was received from various sources, and then it was pre-processed in the data staging technique. Pre-processing of information is very important for any kind of analytics. Null

values, irrelevant features, and missing values always reduce model prediction accuracy. After the pre-processing of data, it was analyzed using various machine learning algorithms. Machine learning, Deep Learning, and Natural Language Processing are considered in our taxonomy. In Machine Learning, two major categories of the learning environment were considered, supervised, and unsupervised learning (Fig. 9).

In supervised learning, the classification of data is handled by various classification algorithms like Naive Bayes, decision tree, support vector machine, K nearest neighbour (KNN), and logistic regression. An efficient clustering process is done in the unsupervised environment by various clustering algorithms like k means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), Gaussian mixture, and hierarchical clustering model. In the Deep learning model, the classification process is handled by various algorithms like Recurrent Neural Networks (RNN), Convolution Neural Networks (CNN), Autoencoder, Deep Belief Networks (DBF), and Radial Basis Function, etc.

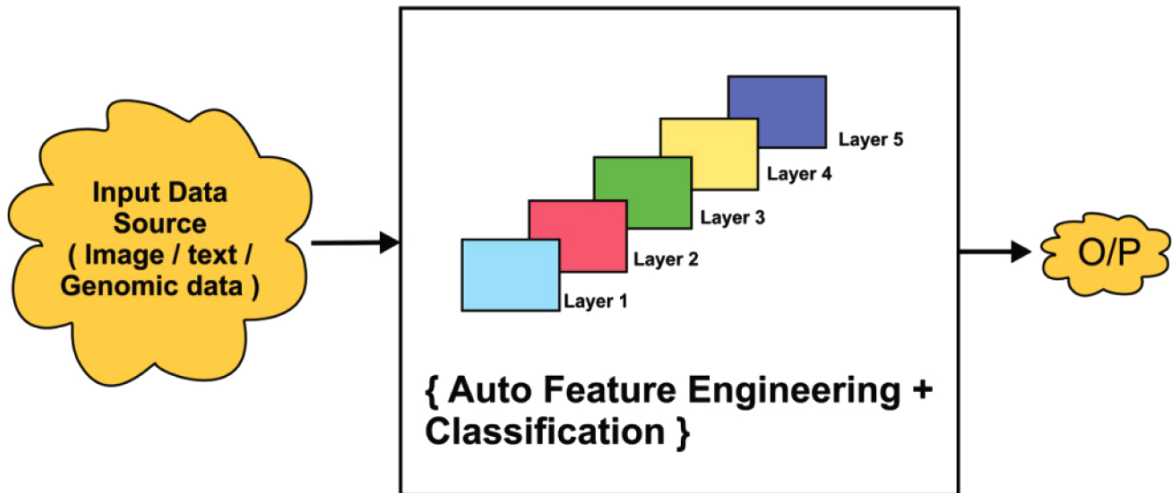


Fig. (8). General deep learning architecture with multiple hidden layers.

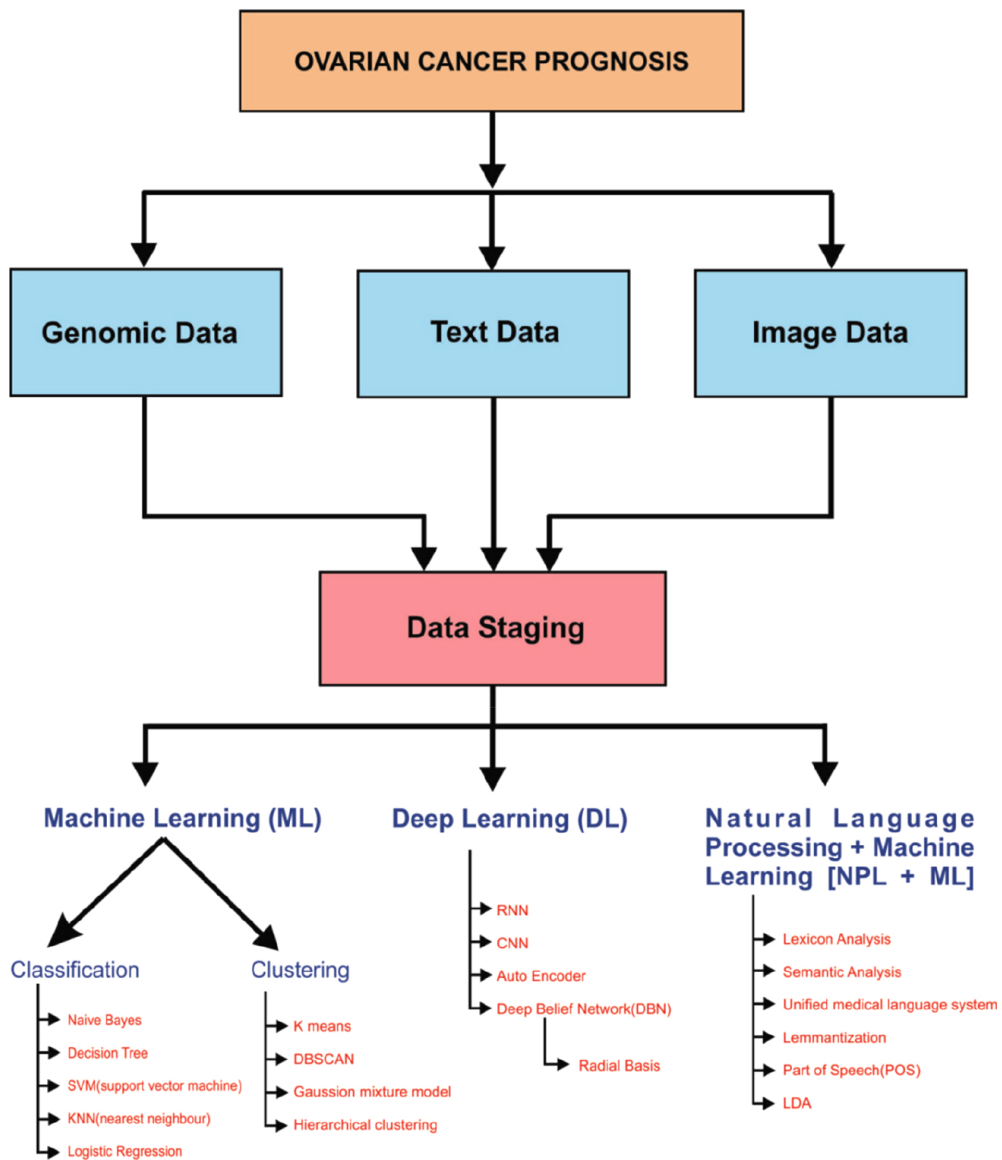


Fig. (9). Taxonomy of ovarian cancer based on ML, DL, and NLP.

Natural language processing handles classification by Lexicon approaches semantic analysis, Unified Medical Language system, speech, and lemmatization. The correlation between gene expression and phenotype data such as tumor morphology, a consequence of chemotherapy, and a related blood test from the laboratory should be studied. NLP is used to extract cancer-related information automatically from electronic medical records.

5. VARIETY OF SOURCES: TEXT, IMAGES, AND GENOMIC DATA

The sources include personal health data, and CT images; MRI images are used in the prediction and recurrence analysis of ovarian cancer. Ovarian cancer can be detected with ultrasound, high CA125 values, Positron emission tomography (PET), and biopsy. The sources of images are helpful in the detection, portrayal, and detecting the stages of ovarian cancer. The benign and malignant tumor is easily distinguished with image sources. Various types of ovarian carcinoma, such as Epithelial, Stromal, and Germ cells are easily detected with the radiologic appearance of Topographic images. The mutation in a specific type of gene is a primary reason for causing ovarian cancer. Data related to gene mutations and personal health data can be integrated to predict cancer early.

An ensemble approach for the prognosis of ovarian cancer is developed by Kawakami *et al.* [15]. Based on the blood biomarkers dataset. Clinical parameters based on ovarian carcinoma and ovarian tumors are differentiated using EOC blood markers. Multiple logistic regression analysis is made to choose the numerous variables as predictors in bio-blood markers. With the help of machine learning approaches, risky patients are identified prior and preliminary treatment for them can be provided.

An ensemble-based approach is proposed by Tarek *et al.* [16] for the classification of different types of gene data associated with various types of cancer. Each ensemble approach chooses the best feature selection method for improving the accuracy in cancer classification.

A study on Genetic profiles with methylation data plays a vital role in the prediction of ovarian cancer. Methylation is a procedure that includes methyl groups in the existing DNA sequence. Main feature selection methods are used to select the useful feature from the combination of gene and methylation profiles. Gene profiles collected from the TCGA repository and methylation beta values play a vital role in ovarian cancer [17].

Li *et al.* developed an integrated framework that predicts recovery chances in the chemotherapy process. The support vector machine algorithm is applied as a base model to classify the data with necessary profile information and merge it with gene expression. The Leave-One-Out Cross Validation method was used to evaluate the performance measure of both models. The final results showed that the fusion of clinical information and gene expression data provided better accuracy [18].

Hybrid machine learning algorithms such as SVM, C5.0, MARS (Multivariate Adaptive Regression Spines), and Random forest were implemented over the patient's profile to predict persistent ovarian cancer [19].

Certain types of genes are responsible for causing the deadly disease of ovarian cancer. The mutations in BRCA1, BRCA2, and TP53 Genes are linked to causing hereditary ovarian cancer. Screening based on gene expression prevents this type of hereditary cancer. Genetic testing can be done by taking the sample from blood, saliva, hair, tissue, *etc.*; integration of Genetic based testing in ovarian cancer reduces the risk and identification of this disease in its early stage. It is estimated that 64% of women will develop ovarian cancer by BRCA1 and BRCA2 gene mutations [20].

Ahn *et al.* performs prediction of ovarian cancer based on an RNAseq dataset collected from The Cancer Genome Atlas (TCGA). Genes were marked as 'good' and 'bad' based on Genes coefficient calculation. The study shows the classification of high-risk and low-risk of patients [21].

With the K-means clustering technique's help, the ovarian cancer Gene dataset from TCGA (The Cancer Genome Atlas) is subdivided into three types. The Dunn index metric is used to evaluate the performance of the clusters. Based on the survival rate of the cancer, patient's morphological type is defined. The ensemble classifier predicts each subtype. The endurance of patients in all three subtypes is significantly different risks. They used two different data sets for their experimental purpose. One was from TCGA, and the is an independent dataset collected from previous work Gyorffy *et al.*, 2012 [22]. They combined and utilized the drug target information from the following three different types of databases. They were the Therapeutic Target Database, Drug Bank, and Drug-Gene Interaction Database. Using the Pearson correlation coefficient and rank-based method, they constructed each subtype's co-expression network. They compared their work with two different controls. The first control applied a t-test to select the features and the second control used Euclidean distance of the vector to assign the patient with a similar subtype. Subsequently, they predicted the prognosis. They have done a survival analysis of cancer patients in all three types. The resultant model showed better significance than the existing methods in classifying the patients in different survival risks [23].

The authors in a study identified diffusion-weighted image value using an MRI scan method [24]. Epithelial ovarian cancer patient tumor load is calculated and can be analyzed whether the complete cytoreduction is achieved. In this paper, the peritoneal cancer index value has been compared in 2 stages. Stage one is the MRI scan period and step 2 is surgery time. The entire cytoreduction is evaluated in both locations using ROC. The correlation coefficient has been identified in both phases. As a result, it was observed that the stage of MRI provides better accuracy than the surgery time to predict the cytoreduction of ovarian cancer.

Gopichand *et al.* clarifies in real-time investigation predicting patient response to chemotherapy for ovarian cancer is analyzed using C.T. scan images. The patient response was calculated according to the shape, density, texture, and wavelet of the image feature. At the time of wavelet transmission was categorized in 4 categories Low-Low, Low -high, high -low, and high-high. (L-H: applying short scale to high scale filter). The authors have identified the feature selection process in two different stages: pre-treatment and post-treatment. In the pre-treatment stage, some of the image features are recognized as the best features. They are skewness, contrast, uniformity,

entropy, density, kurtosis, and compactness. In the post-treatment stage, in addition to the characteristics of pre treatment stage, some new features were also considered like surface area, energy, media, volume, etc. [25].

The recurrence of ovarian cancer causes significant risk in the survival of cancer patients [26]. Wang *et al.* proposed an unsupervised deep learning method for extracting required features from the CT images. The raw images are converted

into 16-dimensional features. The decoder is used to reconstruct the original image of the tumor. The radiologist has been used to identify the tumor part from the CT images. The Cox proportional-hazards regression model is used to find the recurrence of cancer in individual patients. The feature selected is combined with personal clinical characteristics for better predictions [27].

The following Table 2 summarizes the various models and their accuracy for predicting ovarian carcinoma.

Table 2. Summary of various models and their accuracy for predicting ovarian carcinoma.

References	Model	Cancer Type	Type of Data	Validation Method
[6]	Machine learning Minimum Redundancy - Maximum Relevance (MRMR) feature selection method & Decision tree	Ovarian cancer	Demographics, blood routine test, general chemistry, and tumour markers dataset.	Compared with the risk of ovarian malignancy algorithm (ROMA) and logistic regression model.
[8]	Feature Extraction--Scale-invariant feature transform algorithm Convolution Neural network	Stages in ovarian cancer	Ovarian cancer image	Genetic algorithm
[11]	Deep Neural network	Breast cancer, colon cancer, glioblastoma multiform, kidney renal clear cell cancer, low-grade glioma, lung squamous cell cancer, Ovarian cancer, prostate cancer, Skin cutaneous melanoma, Thyroid cancer, and Uterine corpus endometrial cancer.	Gene mutation data	Sensitivity and specificity formula to evaluate the classifiers' performance.
[15]	1. Gradient Boosting Machine 2. Support Vector Machine 3. Random Forest 4. Conditional Random Forest 5. Naive Bayes 6. Neural Network 7. Elastic Net	Epithelial ovarian cancer	Blood biomarkers	10-fold cross-validation
[17]	1. Supervised Principal component and the 2.Elastic net method	Ovarian cancer	Gene expression and methylation profiles	2-fold cross validation.
[18]	Support vector machine	Chemo-response for ovarian cancer	Gene expression data	Leave-one-out cross-validation
[21]	Univariate Cox PH model	LASSO	Microarray and RNASeq data	N/A
[23]	Clustering analysis based on subtypes	Ovarian cancer	Genomic	Leave-one-out and independent validation
[24]	MRI PCI and diffusion-weighted imaging	Epithelial ovarian cancer	MRI Image data	Receiver Operating Characteristic (ROC) Curve
[25]	Equal-weighted fusion method and CAD	Ovarian cancer	Computed tomography (CT) images	RECIST

Table 3. Summary of deep learning models and their accuracy for predicting ovarian carcinoma.

References	Model	Type of Data	Accuracy	Significant Algorithm Settings Discussed
[28]	Deep interactive learning with a pretrained segmentation model	Image	86	Stochastic gradient descent
[29]	Convolutional Neural networks	MRI sequence	81	CNN with Xception architecture
[30]	Hybrid evolutionary deep learning model, using multi-modal data	Gene expression	98.8	Ant Lion Optimizer
[31]	Supervised Deep learning approaches	Histopathology images	88.2	5-fold cross-validation
[32]	Cox Proportional Hazards Model	Histopathological images	88.2	-
[33]	Deep convolutional neural network	Image	86.9	DenseNet-121
[34]	DCNN	Images	81.38	-

(Table 3) contd....

References	Model	Type of Data	Accuracy	Significant Algorithm Settings Discussed
[35]	Deep Convolutional Neural Networks (DCNN)	Histopathological images.	84.64	Visual Geometry Group – 16
[36]	Variational Autoencoder	Gene Expression	81.44	Min-max, SMOTE, ReLU activation function
[37]	Multi-layer perceptron and Logistic Regression	Numerical	99	PCA, Stochastic Gradient
[38]	Improved RCNN	Images	93.2	Attention-based activation function
[39]	GoogleNet+Random Forest Classifier	Ultrasound images	99.79	Softmax activation function, Reverse conduction optimization
[40]	Residual Encoder Decoder CNN	CT images	95.2	Adam optimizer

In addition to the above work related to applying machine learning algorithms, the recent work in the prognosis of ovarian carcinoma involving deep learning techniques is summarized below. The class of deep learning algorithms has considerable advantages - automating feature extraction, efficiency of understanding unstructured data, appreciable self-improvement capability with feedback, and ease of parallelizing to name a few. Deep learning algorithms use different layers of neural networks allowing them to learn from huge volumes of data. The role of deep learning in diagnosing ovarian cancer is summarized in Table 3.

CONCLUSION AND RESEARCH DIRECTIONS

This review endeavored to elucidate and assess the performance of various machine learning and deep learning methods in the prediction of ovarian cancer. The taxonomy for ovarian cancer prognosis is presented concerning text, image, and genomic data. In particular, we compared the pattern observed concerning the type of model being used, the nature of cancer diagnosed, the kind of data incorporated in the prediction and overall accuracy of the various methods with its validation techniques, and the effective analytics in ovarian cancer prediction. The survey emphasized a strong bias towards Machine learning, deep learning, and NLP models in the earlier forecast of this deadly disease. In this review, we outlined the performance of deep learning models with similar or improved accuracy as compared to machine learning models. Future research should extend by optimizing the existing algorithm and diagnosing ovarian carcinoma in its early stage.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflicts of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] E. Lengyel, "Ovarian cancer development and metastasis", *Am. J. Pathol.*, vol. 177, no. 3, pp. 1053-1064, 2010. [http://dx.doi.org/10.2353/ajpath.2010.100105] [PMID: 20651229]
- [2] N. Auersperg, M.M. Woo, and C.B. Gilks, "The origin of ovarian carcino-mas: A developmental view", *Gynecol. Oncol.*, vol. 110, no. 3, pp. 452-454, 2008. [http://dx.doi.org/10.1016/j.ygyno.2008.05.031]
- [3] M. Hermens, A.M. van Altena, J. Bulten, H.A. van Vliet, A.G. Siebers, and R.L. Bekkers, "Ovarian cancer prognosis in women with endometriosis. A retrospective nationwide cohort study of 32,419 women", *Am J. Obstet. Gynecol.*, vol. 224, no. 3, pp. 284.e1-284.e10, 2020.
- [4] L.A. Torre, B. Trabert, C.E. DeSantis, K.D. Miller, G. Samimi, C.D. Runowicz, M.M. Gaudet, A. Jemal, and R.L. Siegel, "Ovarian cancer statistics, 2018", *CA Cancer J. Clin.*, vol. 68, no. 4, pp. 284-296, 2018. [http://dx.doi.org/10.3322/caac.21456] [PMID: 29809280]
- [5] B. Nithya, and V. Ilango, "Evaluation of machine learning based optimized feature selection approaches and classification methods for cervical cancer prediction", *SN Appl Sci.* vol. 1, no. 6, p. 641, 2019. [http://dx.doi.org/10.1007/s42452-019-0645-7]
- [6] M. Lu, Z. Fan, B. Xu, L. Chen, X. Zheng, J. Li, T. Znati, Q. Mi, and J. Jiang, "Using machine learning to predict ovarian cancer", *Int. J. Med. Inform.*, vol. 141, p. 104195, 2020. [http://dx.doi.org/10.1016/j.ijmedinf.2020.104195] [PMID: 32485554]
- [7] A. Laios, A. Katsenou, Y.S. Tan, R. Johnson, M. Otify, A. Kaufmann, S. Munot, A. Thangavelu, R. Hutson, T. Broadhead, G. Theophilou, D. Nugent, and D. De Jong, "Feature selection is critical for 2-year prognosis in advanced stage high grade serous ovarian cancer by using machine learning", *Cancer Contr.*, vol. 28, 2021. [http://dx.doi.org/10.1177/10732748211044678] [PMID: 34693730]
- [8] B. Kaur, K.S. Mann, and M.K. Grewal, "Ovarian cancer stage based detection on convolutional neural network", *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, pp. 855-859, 2017. [http://dx.doi.org/10.1109/CESYS.2017.8321206]
- [9] W. Zhu, L. Xie, J. Han, and X. Guo, "The application of deep learning in cancer prognosis prediction", *Cancers*, vol. 12, no. 3, p. 603, 2020. [http://dx.doi.org/10.3390/cancers12030603] [PMID: 32150991]
- [10] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction", *Comput. Methods Programs Biomed.*, vol. 153, pp. 1-9, 2018. [http://dx.doi.org/10.1016/j.cmpb.2017.09.005] [PMID: 29157442]
- [11] Y. Sun, S. Zhu, K. Ma, W. Liu, Y. Yue, G. Hu, H. Lu, and W. Chen, "Identification of 12 cancer types through genome deep learning", *Sci. Rep.*, vol. 9, no. 1, p. 17256, 2019. [http://dx.doi.org/10.1038/s41598-019-53989-3] [PMID: 31754222]
- [12] P. Danaee, R. Ghaeini, and D.A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification", *Pac. Symp. Biocomput.*, vol. 22, pp. 219-229, 2017. [http://dx.doi.org/10.1142/9789813207813_0022] [PMID: 27896977]
- [13] W. Yim, M. Yetisgen, W.P. Harris, and S.W. Kwan, "Natural language processing in oncology: A review", *JAMA Oncol.*, vol. 2, no. 6, pp. 797-804, 2016. [http://dx.doi.org/10.1001/jamaoncol.2016.0213] [PMID: 27124593]
- [14] I. Spasić, J. Livsey, J.A. Keane, and G. Nenadić, "Text mining of cancer-related information: Review of current status and future directions", *Int. J. Med. Inform.*, vol. 83, no. 9, pp. 605-623, 2014. [http://dx.doi.org/10.1016/j.ijmedinf.2014.06.009] [PMID: 25008281]
- [15] E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, K. Koseki, Y. Iida, M. Saito, H. Komazaki, J.S. Shapiro, C. Goto, Y. Akiyama, R. Saito, M. Saito, H. Takano, K. Yamada, and A. Okamoto, "Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers", *Clin. Cancer Res.*, vol. 25, no. 10, pp. 3006-3015, 2019. [http://dx.doi.org/10.1158/1078-0432.CCR-18-3378] [PMID: 30979733]
- [16] S. Tarek, R. Abd Elwahab, and M. Shoman, "Gene expression based cancer classification", *Egyptian Inform J.*, vol. 18, no. 3, pp. 151-159, 2017. [http://dx.doi.org/10.1016/j.eij.2016.12.001]

- [17] L. Zhang, H. Liu, T.H. Hsiao, Y. Chen, and Y. Huang, "An investigation of clinical outcome prediction from integrative genomic profiles in ovarian cancer. *Proceedings 2012 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, 2012, pp. 103-106.
[http://dx.doi.org/10.1109/GENSIPS.2012.6507739]
- [18] L. Li, L. Chen, D. Goldof, F. George, Z. Chen, A. Rao, J. Cragun, R. Sutphen, and J.M. Lancaster, "Integration of clinical information and gene expression profiles for prediction of chemo-response for ovarian cancer", *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 4818-4821, 2006.
- [19] Y.C. Lu, C.J. Lu, C.C. Chang, and Y.W. Lin, "A hybrid of data mining and ensemble learning forecasting for recurrent ovarian cancer", *2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIBMS)*, pp. 216-216, 2017.
[http://dx.doi.org/10.1109/ICIBMS.2017.8279723]
- [20] B. Rahman, L. Side, S. Gibbon, S.F. Meisel, L. Fraser, S. Gessler, J. Wardle, and A. Lanceley, "Moving towards population-based genetic risk prediction for ovarian cancer", *BJOG*, vol. 124, no. 6, pp. 855-858, 2017.
[http://dx.doi.org/10.1111/1471-0528.14603] [PMID: 28217902]
- [21] T. Ahn, N. Kang, Y. Kim, and T. Park, "Gene expression based prediction of prognostic outcome in ovarian cancer", *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018pp. 1753-1757
[http://dx.doi.org/10.1109/BIBM.2018.8621205]
- [22] B. Györfy, A. Lániczky, and Z. Szállási, "Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients", *Endocr. Relat. Cancer*, vol. 19, no. 2, pp. 197-208, 2012.
[http://dx.doi.org/10.1530/ERC-11-0329] [PMID: 22277193]
- [23] Y.C. Gao, X.H. Zhou, and W. Zhang, "An ensemble strategy to predict prognosis in ovarian cancer based on gene modules", *Front. Genet.*, vol. 10, p. 366, 2019.
[http://dx.doi.org/10.3389/fgene.2019.00366] [PMID: 31068972]
- [24] M.P. Engbersen, I. van 't Sant, C. Lok, D.M.J. Lambregts, G.S. Sonke, R.G.H. Beets-Tan, W.J. van Driel, and M.J. Lahaye, "MRI with diffusion-weighted imaging to predict feasibility of complete cytoreduction with the peritoneal cancer index (PCI) in advanced stage ovarian cancer patients", *Eur. J. Radiol.*, vol. 114, pp. 146-151, 2019.
[http://dx.doi.org/10.1016/j.ejrad.2019.03.007] [PMID: 31005166]
- [25] G. Danala, T. Thai, C.C. Gunderson, K.M. Moxley, K. Moore, R.S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Applying quantitative C.T. image feature analysis to predict response of ovarian cancer patients to chemotherapy", *Acad. Radiol.*, vol. 24, no. 10, pp. 1233-1239, 2017.
[http://dx.doi.org/10.1016/j.acra.2017.04.014] [PMID: 28554551]
- [26] S. Wang, Z. Liu, Y. Rong, B. Zhou, Y. Bai, W. Wei, W. Wei, M. Wang, Y. Guo, and J. Tian, "Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer", *Radiother. Oncol.*, vol. 132, pp. 171-177, 2019.
[http://dx.doi.org/10.1016/j.radonc.2018.10.019] [PMID: 30392780]
- [27] M. Akazawa, and K. Hashimoto, "Artificial intelligence in ovarian cancer diagnosis", *Anticancer Res.*, vol. 40, no. 8, pp. 4795-4800, 2020.
[http://dx.doi.org/10.21873/anticancer.14482] [PMID: 32727807]
- [28] D.J. Ho, M.H. Chui, and C.M. Vanderbilt, "Deep Interactive Learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation", *J. Pathol. Inform.*, vol. 14, p. 100160, 2022.
[http://dx.doi.org/10.1016/j.jpi.2022.100160]
- [29] T. Saida, K. Mori, S. Hoshiai, M. Sakai, A. Urushibara, T. Ishiguro, M. Minami, T. Satoh, and T. Nakajima, "Diagnosing ovarian cancer on mri: a preliminary study comparing deep learning and radiologist assessments", *Cancers*, vol. 14, no. 4, p. 987, 2022.
[http://dx.doi.org/10.3390/cancers14040987] [PMID: 35205735]
- [30] R.M. Ghoniem, A.D. Algarni, B. Refky, and A.A. Ewees, "Multi-modal evolutionary deep learning model for ovarian cancer diagnosis", *Symmetry*, vol. 13, no. 4, p. 643, 2021.
[http://dx.doi.org/10.3390/sym13040643]
- [31] C.W. Wang, C.C. Chang, Y.C. Lee, Y.J. Lin, S.C. Lo, P.C. Hsu, Y.A. Liou, C.H. Wang, and T.K. Chao, "Weakly supervised deep learning for prediction of treatment effectiveness on ovarian cancer from histopathology images", *Comput. Med. Imaging Graph.*, vol. 99, p. 102093, 2022.
[http://dx.doi.org/10.1016/j.compmedimag.2022.102093] [PMID: 35752000]
- [32] Y. Gao, S. Zeng, X. Xu, H. Li, S. Yao, K. Song, X. Li, L. Chen, J. Tang, H. Xing, Z. Yu, Q. Zhang, S. Zeng, C. Yi, H. Xie, X. Xiong, G. Cai, Z. Wang, Y. Wu, J. Chi, X. Jiao, Y. Qin, X. Mao, Y. Chen, X. Jin, Q. Mo, P. Chen, Y. Huang, Y. Shi, J. Wang, Y. Zhou, S. Ding, S. Zhu, X. Liu, X. Dong, L. Cheng, L. Zhu, H. Cheng, L. Cha, Y. Hao, C. Jin, S. Zhang, P. Zhou, M. Sun, Q. Xu, K. Chen, Z. Gao, X. Zhang, Y. Ma, Y. Liu, L. Xiao, L. Xu, L. Peng, Z. Hao, M. Yang, Y. Wang, H. Ou, Y. Jia, L. Tian, W. Zhang, P. Jin, X. Tian, L. Huang, Z. Wang, J. Liu, T. Fang, D. Yan, H. Cao, J. Ma, X. Li, X. Zheng, H. Lou, C. Song, R. Li, S. Wang, W. Li, X. Zheng, J. Chen, G. Li, R. Chen, C. Xu, R. Yu, J. Wang, S. Xu, B. Kong, X. Xie, D. Ma, and Q. Gao, "Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: A retrospective, multicentre, diagnostic study", *Lancet Digit. Health*, vol. 4, no. 3, pp. e179-e187, 2022.
[http://dx.doi.org/10.1016/S2589-7500(21)00278-8] [PMID: 35216752]
- [33] H. Farahani, J. Boschman, D. Farnell, A. Darbandsari, A. Zhang, P. Ahmadvand, S.J.M. Jones, D. Huntsman, M. Köbel, C.B. Gilks, N. Singh, and A. Bashashati, "Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images", *Mod. Pathol.*, vol. 35, no. 12, pp. 1983-1990, 2022.
[http://dx.doi.org/10.1038/s41379-022-01146-z] [PMID: 36065012]
- [34] K.R. Kasture, B.B. Sayankar, and P.N. Matte, "Multi-class classification of ovarian cancer from histopathological images using deep learning - VGG-16. *2nd Global Conference for Advancement in Technology (GCAT)*, 2021, pp. 1-6.
[http://dx.doi.org/10.1109/GCAT52182.2021.9587760]
- [35] X. Meng, X. Li, and X. Wang, "A computationally virtual histological staining method to ovarian cancer tissue by deep generative adversarial networks", *Comput. Math. Methods Med.*, vol. 2021, pp. 1-12, 2021.
[http://dx.doi.org/10.1155/2021/4244157] [PMID: 34306174]
- [36] M.T. Hira, M.A. Razzaque, C. Angione, J. Scrivens, S. Sawan, and M. Sarker, "Integrated multi-omics analysis of ovarian cancer using variational autoencoders", *Sci. Rep.*, vol. 11, no. 1, p. 6265, 2021.
[http://dx.doi.org/10.1038/s41598-021-85285-4] [PMID: 33737557]
- [37] B. Yesilkaya, M. Perc, and Y. Isler, "Manifold learning methods for the diagnosis of ovarian cancer", *J. Comput. Sci.*, vol. 63, p. 101775, 2022.
[http://dx.doi.org/10.1016/j.jocs.2022.101775]
- [38] S. Wang, L. Wang, L. Wang, Z. Yu, X. Zhao, and X. Wang, "AT-Cascade R-CNN: A novel attention-based cascade R-CNN model for ovarian cancer lesion identification", *Int. J. Innov. Technol. Adapt. Manag.*, vol. 3, no. 1, pp. 74-86, 2021.
[http://dx.doi.org/10.1504/IJAIS.2021.117913]
- [39] L. Zhang, J. Huang, and L. Liu, "RETRACTED ARTICLE: Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system", *J. Med. Syst.*, vol. 43, no. 8, p. 251, 2019.
[http://dx.doi.org/10.1007/s10916-019-1356-8] [PMID: 31254110]
- [40] Y. Urase, M. Nishio, Y. Ueno, A.K. Kono, K. Sofue, T. Kanda, T. Maeda, M. Nogami, M. Hori, and T. Murakami, "Simulation study of low-dose sparse-sampling ct with deep learning-based reconstruction: Usefulness for evaluation of ovarian cancer metastasis", *Appl. Sci.*, vol. 10, no. 13, p. 4446, 2020.
[http://dx.doi.org/10.3390/app10134446]