Soft Computing Approaches for Ovarian Cancer: A Review

Dilip Kumar Choubey[1,*], Arvind Choubey[2], Shalini Mahto[3], Vaibhav Soni[4]

ARTICLE INFO

Article history:
Received: 25 November 2022
Revised: 11 January 2023
Accepted: 24 Febuary 2023

Keywords:
Ovarian cancer
Soft computing
Classification,
Feature selection
Diagnosis
Medical disease

ABSTRACT

In the current circumstances, accessing the application is benefiting the medical profession in many ways. The classification and clustering algorithms are among the strategies that both researchers and doctors find fascinating. Medical ailments are on the rise for a variety of reasons, and ovarian cancer is one of them. Ovarian cancer is a variant of cancer occurring in the female which spreads all over the body starting from the ovary. Early diagnosis and accurate detection are important because they give better treatment at a lower cost, reduce patient risk. The diagnosis becomes more difficult as the disease progresses. The prime objective of this review article is to find difficulties and advise further research in existing publications, as well as to provide ideas for developing an efficient and successful ovarian cancer categorization technique. The authors of this study looked at a lot of research papers in order to come up with an efficient ovarian cancer prediction model.

1. INTRODUCTION

Ovarian cancer is a variant of cancer occurring in the female which spreads all over the body starting from ovary. The ovaries are a pair of reproductive organs in women that produce eggs and sex hormones. In the early stage, there may be no or vague symptoms. The disease is limited to the ovary, more likely to be successfully diagnosed. It becomes noticeable as the cancer progresses or is in an advanced stage but difficult to treat. Ovarian cancer symptoms include weight loss, swelling, pain or pressure in the pelvis region, variations in bowel habits, such as pain in the back or abdominal pain, constipation, more frequent urination, not regular periods or vaginal bleeding after menopause, loss of appetite, tiredness, and breathlessness. Ovarian cancer is the fifth most prevalent form of cancer among American women. As seventy percent of patients are diagnosed at an advanced stage and thirty percent at an early stage, ovarian cancer is considered a silent murderer. Ovarian cancer is divided into thirty subtypes based on the type of cell from which it originates. Age of the ovarian cancer or tumor plays a crucial role in the formulation of the treatment plan. As the stage level increases, the survival of the patient becomes harder. The stages are defined as:

Stage 1: One or both ovaries are affected with cancer and has not spread elsewhere. Stage 1 has (82-92) % patients.

Stage 2: It extends to the pelvic area but does not migrate to the abdomen. Stage 2 has (51-69) % patients.

Stage 3: Cancer extends beyond the pelvis into the abdominal region. Stage 3 has (17-39) % patients.

Stage 4: It extends into the lung, liver, or to a location beyond the abdominal region. Stage 4 has (11.5) % patients.

Ovarian tumors arise when DNA in a cell stops working correctly, resulting in aberrant cells in the ovary that expand uncontrollably and create a tumor. If left untreated, the tumor can circulate to several parts of the body.

Till date, it is unclear that what are the reasons of ovarian cancer, though physicians have recognized some of the factors that can enhance the chance of ovarian cancer like age (Over 55 years old), inherited gene mutations such as BRCA1, RCA2, Family History (Ovarian, Breast, Gynecological, or colon) cancer, estrogen hormone replacement therapy, age when menstruation begins and ends, BMI more than 30, infertility, endometriosis, never having been on the pill, hormone therapy, obesity and overweight.

Ovarian tumor is dangerous; particularly if it is not found at the early stage. It raises the risk of heart disease and stroke, as well as brittle bones, memory loss, menopausal symptoms, and sexual dysfunction.

[1]*Department of Computer Science & Engineering, Indian Institute of Information Technology Bhagalpur, Bihar, India*.
[2]*Department of Electronics & Communication Engineering, Indian Institute of Information Technology Bhagalpur, Bihar, India*.
[3]*Department of Computer Science & Engineering, Indian Institute of Information Technology Ranchi, Jharkhand, India*.
[4]*Department of Computer Science & Engineering, Maulana Azad National Institute of Technology Bhopal, M.P., India.*
**Corresponding author:* Dilip Kumar Choubey; E-mail: dilipchoubey_1988@yahoo.in.

Early identification, use of oral birth control pills, breastfeeding, discussing your risk factors with your doctor, pregnancy, and surgical procedures on your reproductive organs are methods to reduce your chances of ovarian cancer. Treatment will rely on numerous factors, including the type, stage, age of patient's, grade of the cancer, overall health of patient, patient's personal preferences, accessibility, and affordability of treatment. For certain patients, entering a clinical trial may be the best treatment option.

The following methods are used for the treatment: Surgery, chemotherapy, radiotherapy, immunotherapy (biotherapy), targeted therapy, hormone therapy, a combination of therapy.

The review article is categorized into five sections: Section 1.1 stated Motivation, Section 2 describes Literature Review, Section 3 deals with the Descriptions of Different Techniques and Tools Used, Section 4 stated Discussion and Future Directions and finally Section 5 deals with the Conclusions and Limitations.

Motivation

According to the American Cancer Society, ovarian cancer is the sixth leading cause of cancer-related mortality in women in the United States. In the United States, 14,000 women die annually from cancer, out of 20,000. This condition is most prevalent among women aged 55 to 64. It is challenging to diagnose early because there are few signs or symptoms before the cancer spreads beyond the ovaries, resulting in a 5-year relative survival rate of 46%. The likelihood of survival decreases as the disease progresses to its advanced stages. Nearly eight out of ten patients are diagnosed with ovarian cancer after the disease has progressed to other organs.

2. LITERATURE REVIEW

The literature review section states about the previous researcher's works notably detection of ovarian cancer by majorly soft computing, data mining, machine learning and computational intelligence methods.

Researchers [1] have utilized a combined approach i.e. a Genetic Algorithm (GA) and correlation-based heuristics for data preprocessing and C4.5 Decision Trees (DT), SVM for classification. To improve the classification accuracy, researchers have further implemented Bagging and stacking for the gene expression datasets namely prostate, ovarian, lung cancer for classification. In Ovarian cancer they have used SMO (Sequential Minimal Optimization) for classification whereas in prostate, lung cancer, SVM is used for classification. SMO is not different from SVM but one of the particular simple and fast ways[2] [3] trained method of SVM. The SMO splits a huge QP into a sequence of smaller QP problems that can be addressed analytically. In SMO, standard SVM parameters have been used.

Researchers [4] have proposed a new memetic algorithm notably Cancer Diagnosis with Memetic Fuzzy System (CD-MFS) for an interpretable classifier of gene expression and a compact fuzzy rule base. It is used for the classification of microarray data (14_Tumors data set). The CD-MFS classification system performances have been evaluated through the 14_Tumors cancer data set.

Researchers [5] have applied Complementary Fuzzy Neural Network (CFNN) and Pseudo-associative-Complementary Learning (PACL) FNN as a Diagnostic Decision Support System (DDSS) in ovarian cancer.

Researchers [6] used and evaluated Least Squares Support Vector Machine (LS-SVM) classifiers in order to differentiate between benign and malignant forms of ovarian cancer. Bayesian evidence was utilized in this process. On ovarian tumours, logistic regression models, as well as LS-SVM models with linear and RBF kernels, were constructed. The dataset was analyzed, and several methods, including feature selection, parameter estimation, and ROC analysis, were utilized to assess performance. The LS-SVM that utilized nonlinear RBF kernels performed exceptionally well in comparison to its competitors.

Researchers [7] have performed the classification of ovarian cancers by using Bayesian networks. The key contribution of the researchers here is to combine data, domain literature, and people with specialist skills in medicine and genetics. Researchers have put forth a hybrid Bayesian approach [8] that blends previous knowledge with powerful statistical data. The researchers also explained how the suggested method may offer unclear information about predictions and how it could be applied to rejection tasks in classification activities. Prior to surgical removal, it was put to the test by attempting to distinguish between benign and malignant tumors.

Researchers [9] examined a human ovarian cDNA expression against a database for detecting oncogenes and chose oncogenes that were used to derive pathological phases of ovarian cancer. A novel t-test statistical oncogenes selection and a learning-based classification method are used to validate the discrimination of selected oncogenes. Researchers [10] have employed the fuzzy Yager inference scheme, which is able to mimic human deductive reasoning logically, and has provided a robust, intuitive logical reasoning and decision-making framework with the Hybrid Neural Fuzzy Inference System (HyFIS) approach. Researchers [11] have reviewed many research articles related to recent supervised machine learning (ML) methods utilized in the modeling of cancer progression. Researchers [12] have used Ultrasound Sonography (USG) and Computed Tomography (CT) techniques to evaluate the clinical assessment of suspicious ovarian masses. Researchers [13] have reviewed and summarized many research articles on the prognosis and diagnosis of cancer. It provides an overview of the recent research by using data mining, machine learning techniques on cancer datasets namely prostate cancer, ovarian cancer, and lung cancer. Majorly, all the data mining and machine learning based algorithms such as KNNs, ANN, nearest shrunken centroids,

logistic regression, and SVM have been utilized for cancer classification and gene selection.

Researchers [14] have examined Gene Expression Microarray (GEM) profiles of epithelial malignancy using oligonucleotide microarrays containing 1200 genes. Mamma globin 2 (MGB2) is a novel biomarker that is significantly overexpressed in essential and metastatic ovarian cancer compared to normal ovarian tissue, as discovered by researchers. Researchers [15] have developed Rank Gene software for feature selection and evolutionary method for leukemia microarray data classification which gave accurate and robust outcome. Polymerase Chain Reaction (PCR) analysis for Epidermal Growth Factor Receptor (EGFR) type III variant (EGFRvIII) gene expression and Western blot protein analysis have been investigated by researchers [16]. In the EGFRvIII mutation, neither the mRNA nor the protein levels of tissue samples were positive. Using a DNA microarray containing 9121 genes, researchers [17] analyzed the gene expression of nine ovarian tumors. Researchers have discovered 115 genes that are distinct between the two types of tumors. Researchers [18] have evaluated the significance of Tissue polypeptide specific antigen (TPS) in conjunction with Cancer antigens 125 (CA125) and Carciniembryonic antigens (CEA) in colorectal cancer patients. TPS is a more precise proliferative marker that can pinpoint the activity of tumor cells. In some instances, it has been discovered to be more sensitive than CEA. In this investigation, patients with metastatic ovarian cancer had higher levels of TPS and CA125 than those with stable disease.

Researchers [19] demonstrated the efficiency of the Risk of Malignancy Index, also referred as (RMI-3) in distinguishing malignant and benign tumors prior to surgery and also revealed the appropriate cut-off value. SVMs have been created by [20] for the examination of both tissue categorization and data exploration for mislabeled or dubious tissue results. The CART method has demonstrated efficacy in distinguishing ovarian cancer. from benign and malignant [21]. The CART classification method has been implemented by using Biomarker Patterns Software (BPS). The proposed method with the help of this tool has been implemented and analyzed on serum protein mass spectrum profiles.

Researchers [22] have utilized logistic regression, MLPs, and Generalized Regression Neural Networks (GRNNs) classification methods for discriminations between malignant and benign ovarian tumors. Researchers have assessed the performance of the models by using the ROC curve. The experimental study indicates that neural network classifier has the potential of a more reliable prediction.
Researchers [23] have used Multi-Layer Perceptron (MLP) networks for ovarian cancer classification. The use of standardized blood-test results for the classification of benign and malignant disorders has been investigated by researchers. The results of the experiment show that it is possible to distinguish between preceding and late-stage tumors. In addition, MLP also deals with missing data and is decided as a suitable technique to deal with missing data. Support Vector Regression (SVR) is a new feature reduction technique devised by researchers [24] for ovarian cancer. Using the proposed method, variable gene selection for up- and down-regulated genes in ovarian cancer may be feasible. The block effect on ovarian cancer data is overcome with this method. The experimental investigation demonstrates the validity and application of the method, as well as the possibility of gene class identification and prediction.

Researchers [25] developed a novel intelligent method called SVR for variable gene selection or feature selection, as well as similar analysis for ovarian cancer categorization. The proposed approach can also overcome multi-class problems. Researchers [26] have proposed a fuzzy ANN technique which is authorized by the GARSC for ovarian cancer diagnosis. The proposed approach is competent to address problems in the actual world with great interpretability. An ovarian tumor diagnosis is used, and the hybrid method proves to be dependable or superior in capabilities. Finally, the proposed method is capable of discovering the application's inherent knowledge without the use of prior information or human interaction.

On high throughput SELDI-TOF mass spectroscopy data for the classification of ovarian tumors, researchers [27]developed a novel combinational attribute selection method. The proposed method consists of three steps: dataset normalization, dimensionality reduction through feature filtration, and binary PSO selection of pertinent features. It was successfully validated using a proteomic dataset of ovarian cancer. It reduces proteomic data's high dimensionality to 3-dimensional, linear separable data, resulting in high-accuracy, high-speed diagnosis methods. Researchers [28] have employed GA and PSO for gene selection and ANFIS, SVM, KNN, and CART for cancer classification based on Microarray Gene Expression data pertaining to Breast, ALL-AML, Colon, Prostate, Lung, and Lymphoma. ANFIS is compared to the three other classifiers mentioned above, and it is shown that ANFIS produces superior results, is best for a smaller number of genes, and can provide TSK-like fuzzy if-then rules that are easy to understand. Researchers [29] evaluated the levels of CA125 and HE4 in the serum of healthy people and ovarian cancer patients. CA125 and HE4 have been measured in serum using the sandwich ELSA technique. In women with epithelial ovarian cancer, the HE4 test is used in conjunction with the CA125 test. Researchers [30] have summarized data mining techniques which include classification, clustering, association methods with advantages, drawbacks, and applications for the healthcare industry and biomedical. They have presented discrimination between statistics and data mining and also many real-world examples of data mining in the healthcare industry.

Researchers [31] have used Redundancy-based Filter, ReliefF, Correlation Feature Subset-Sequential Forward (CFS-SF) for the selection of relevant genes or removing the redundant genes on microarray datasets namely Colon cancer, Leukemia, Lung cancer, and Breast cancer. Researchers [32] have emphasized on ovarian cancer, its generating process, signs, symptoms, and the major causes. Researchers have laid emphasis on diagnosis techniques that assist to discover the cancer cells and treatment of the patient. In multicenter data, Researchers [33] used a LR model to differentiate prior to surgery between benign and malignant overt adnexal tumors. The new method appears to be more resilient than the present one. Researchers [34] have outlined a place for the use of MR imaging in ovarian cancer. (Claus E.B., 2001) [35] have discussed risk definitions used in clinical oncology, as well as risk calculation methodologies for breast and ovarian cancer. Researchers [36] have used the RF method for the classification of microarray data and have also used a new method based on random forest attribute selection or relevant gene selection.

Researchers [37] have utilized ANN and logistic regression methods for the classification of ovarian cancer. ANN achieved good performance comparatively with logistic regression. The effectiveness of the proposed method may improve with increasing dataset size and thereby needs further exploration. Researchers [38] have extracted the features in 4 phases: LL, HL, LH, and HH followed by classification using ML along with spatial domain algorithm. The proposed approach has been implemented on MRI medical images where accuracy 94%, specificity 0.99, and sensitivity value 0.9978 have been achieved. Researchers [39] proposed a data mining technique including classification and clustering for ovarian cancer prediction. The proposed technology is easy, cost-effective, and time-saving. To rate the relevance of risk factors and identify ovarian tumor recurrence, [40] combined ensemble learning, C5.0, SVM, MARS, ELM and Random Forest (RF) are used.

Researchers [41] have utilized ANN to classify ovarian cysts. The primary objective of the proposed research is the classification of follicular and dermoid cysts. It is common for women's fertility to be affected by fibroids. The ovarian cyst images in the database were collected from hospitals, physicians, and the internet. Researchers [42] have created a novel hybrid filter wrapper gene subset selection method for gene selection and ML algorithms for the classification of 10 public dataset of cancer microarray data sets.

Researchers [43] developed a bioinformatics algorithm that can distinguish between neoplastic and non-neoplastic illness in the ovary based on proteomic patterns in serum. Researchers [44] suggested a hybrid PSO using a GA approach for gene selection and SVM for classification. Cancer patients with colon cancer, leukemia, and breast cancer are testing the proposed method. According to the results of the experiment, the proposed methodology can determine the dataset's dimensionality, locate the relevant gene subset, and improve accuracy.

K-TSP (K-Top Scoring Pairs) has been proposed for cancer categorization using microarray gene expression data [5]. On nineteen binary and multi-class gene expression datasets, the proposed method for class prediction is compared to existing ML algorithms. As a prediction analysis, the proposed approach works well. It generates simple and precise decision-making rules. Researchers [45] have utilized the CLFNN method to aid existing diagnosis. The proposed method is accurate and consistent. Here CLFNN has proved as a promising tool for clinical decision support. For the classification of microarray gene expression datasets, researchers [46] have used SVM, RBFNN, MLPNN, Bayesian, Decision Tree, and RF techniques. Clustering techniques such as K-means, DBC, and EM have been applied to microarray gene expression datasets. As a feature selection technique, SVM-RFE, Chi-Squared, and CSF have been used. These techniques were executed to eight two-class microarray datasets. They compared how well different approaches performed in classifying test datasets. The findings of the experiments demonstrate the importance of feature selection in identifying important genes and determining classification accuracy. Researchers [47] have built the profiles of patient serum proteins using mass spectrometry, and integrated with advanced data mining was found to be a great technique to achieve the mortality. The ovarian dataset has been obtained from the clinical proteomics program databank website.

To accurately detect malignancy, researchers [48] have proposed a two-step feature selection method with a 15-neural network. Utilising the feature selection method to deduce features. The proposed methods have been applied to the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and compared to existing methods, demonstrating a 99.4% improvement in classification accuracy. The utilized methods are more promising and significant. Researchers have performed [49] the work in two stages: Stages one deals with J48graft decision tree for classification whereas stages two deals with the GA as a feature selection followed by the same J48graft decision tree classification method. Afterwards, they compared the performance of classification to the performance of feature selection followed by classification. In the same manner [50] modified RBF NN and GA followed by RBF NN were used to classify diabetes. The authors then contrasted the classification and feature selection methods to the classification method. Feature selection followed by classification methods has yielded positive results in both of the aforementioned works. For the classification of the Pima Indian Diabetes Dataset, researchers [51] have implemented Naive Bayes with Genetic Algorithm. First, GA selected the pertinent features, then Bayesian classification was performed on the relevant features. They [52] have utilized a functional link convolutional neural network to classify

diabetes mellitus. They have presented [53] a hybrid intelligent system for diagnosing diabetes disease. They [54] have demonstrated the implementation and analysis of diabetes classification algorithms. They have [55] provided a comparative analysis of diabetes classification methods using PCA and LDA. They have [19] evaluated the efficacy of PCA and PSO classification methods for diabetes. Various soft computing and computational intelligence techniques have been used to predict diabetes in the preceding works. Using both public and collected real-world data, researchers contrasted and analysed their proposed algorithms with those already in use. Similarly, [56] discussed soft computing approaches for breast cancer detection; Researchers [57] conducted a comparative analysis for leukaemia using machine learning and data mining methods, and [58] conducted a comparative analysis for leukaemia using soft computing approaches. Researchers [59] have summarised the classification techniques of soft computing, data mining, machine learning, and deep learning for the classification of diabetes. In addition, they have proposed benefits, issues, the applicability of techniques and tools, and prospective directions for existing articles. Researchers [60], [52] employed Multi-channel FLANN and Cat Swarm Optimization-based FLANN for noise removal in ultrasound images; Researchers [61], [62], [63] analyzed and compared the categorization of heart disease using ML techniques. ; Researchers [64], [65] studied and compared computational intelligence and data mining for thunderstorm and lightning prediction. Researchers [66] suggested and tested a new statistical strategy for ovarian cancer feature reduction using high-resolution SELDI-TOF data. For feature reduction, the t-test and the four statistical moments technique are used. Kernel PLS models are used to improve accuracy and classification performance. The proposed model is appropriate for analyzing proteomics data with a high throughput.

In Table 1, the overall summary of the existing works for ovarian cancer diagnosis has been discussed.

Table 1. Summary of the existing works for ovarian cancer diagnosis

Ref.	Dataset Used	Techniques and Tool Used	Purpose	Significance	Accuracy
[1]	Ovarian, Prostate, Lung Cancer.	A Combined approach i.e. GA and correlation-based heuristics, Decision trees (C4.5), SVM, Bagging and Stacking	GA and correlation-based data preprocessing heuristics. C4.5, SVM for Classification. To improve the accuracy, researchers have further used Bagging and stacking for the classification.	Identification of the most significant genes. Due to partitioning of data, it does not need any earlier information of the genes thus it allows for flexibility. The mapping of genotype information to phenotype will reduce the complexity of cancer prediction. Bagging and stacking algorithms were used to enhance accuracy, and GA was employed for gene selection.	Ovarian DT-(94.07%), SMO-(97.46%), Bagging-(97.46%), Stacking-(97.46%). Prostate DT-(55.88%), SVM-(67.65%), Bagging-(26.47%), Stacking-(67.65%). Lung DT-(81.88%), SVM-(98.66%), Bagging-(94.63%), Stacking-(80.54%).
[8]	Ovarian Cancer	Bayesian Networks, Logistic Regression, ANN.	Bayesian Networks, Logistic Regression and ANN have been used for the classification of ovarian cancer.	Bayesian networks are a useful tool for gaining medical knowledge, making judgments, and learning. They are mostly efficient in a combined sub-model. A Bayesian network has been proposed as a feasible solution for combining background knowledge and observations efficiently.	Bayesian Network ROC: (0.952), Sensitivity: (94.7), Logistic Regression ROC: (0.904), Sensitivity: (85.4), ANN ROC: (0.951), Sensitivity: (87.5).

[4]	14_Tumors Data set.	CD-MFS.	CD-MFS used for the microarray data cancer tumor detection (14_ Tumors dataset)	Compatibility, fewer rules, fewer genes used, and fuzzy rule capable of discriminating any type of tumor.	69.43%.
[5]	Ovarian Cancer	CLFNN, PACL FNN	CLFNN and PACL are utilized in ovarian cancer diagnosis.	Interoperability, improved accuracy and training time.	CLFNN (FALCON-AART) -(94%), PACL (FCMAC-FALCON) -(86%).
[6]	Ovarian Tumors Ultra-Sonographer	LS-SVM models with linear RBF kernels and logistic regression.	To distinguish between benign and malignant ovarian cancer, the Bayesian evidence framework was used to implement and evaluate LS-SVM with linear, Radial Basis Function (RBF) kernels and logistic regression models. The LS-SVM with nonlinear RBF kernels has the greatest performance.	Optimal input variable selection, support for statistical learning theory, reliability, and assistance to doctors in generating a correct diagnosis, thereby assisting those investigators with limited expertise in accurately predicting preoperative outcomes.	LS-SVM with RBF Kernels ROC: (0.94), Sensitivity: (90.0%), Specificity: (80.6%).
[67]	Ovarian Cancer, Colon and Breast.	Regression analysis, GA, PSO, SVM, Analysis of Variance (ANOVA) and improved fuzzy model.	To locate target genes, regression analysis was performed. To identify gene markers from target genes, the GA, PSO, SVM, and ANOVA procedures have also been combined. The gathered gene markers are used to classify cancerous tissues using the updated fuzzy model.	The suggested methodology can be used to evaluate gene expressions in ovarian cancer microarray data in this investigation. It performs well in terms of gene selection and classification.	ovarian cancer …… Colon (99.13%), Breast (98.55%).
[68]	Lymphoma, Lung, Ovarian Cancer Dataset,	Genetic Programming (GP)	Common gene expression datasets, including lymphoma cancer dataset, lung cancer dataset, and ovarian cancer dataset, are classified using GP.	Enhances classification effectiveness and diversity. The proposed method is independent of training data, allowing it to be effective in cancer classification.	Lymphoma (91.3%), Lung (98.2%), Ovarian (96.9%).
[45]	Ovarian Cancer Dataset, blood tests, Proteomic Spectra of Ovarian Cancer	CLFNN	CLFNN is a tool that aids in the diagnosis of existing conditions.	CLFNN is a promising instrument for clinical decision support due to the precision and consistency of its diagnostic decisions. It also outperforms most conventional methods and surpasses them.	ovarian cancer dataset (90.70%), Based on blood tests (78.90%), Proteomic spectra (94.70%).
[69]	Leukemia, Colon, Lung and Ovarian Cancer.	Correlation analysis, Fisher ratio and Estimation of Distribution Algorithm (EDA)	EDA is used to design the classifier committee for classification, while correlation analysis and Fisher ratio are utilized to extract valuable gene features and reduce dimensionality while maintaining informative features.	Best recognition rates.	Leukemia (98.6%), Colon (95.2%), Lung cancer (100%), Ovarian (99.6%).

[70]	Pima Indian Diabetes Dataset (PIDD), Wisconsin Breast Cancer Dataset, Ovarian Cancer.	GA and Rough Set Incorporated Neural Fuzzy Inference System (GARSINFIS), GA and Rough Set Clustering (GARSC).	GARSINFIS employs the foundation of inference rules automatically generated by our proposed GARSC method. GARSC is a clustering method that combines GA & RS theory together. GA is used to find the best solutions. The RS theory is used to solve the problem of dimensionality curse. The original knowledge base is considerably deduced via RS approximations without losing crucial information.	Achieves majorly correct diagnosis, automatically derives rules which are consistent with expert knowledge and has high level of interpretability. No human interference is required as well as a smaller number of parameter and constrain required.	PIDD (74.36%), Wisconsin breast cancer dataset (95.33%), Ovarian cancer diagnosis-1 dataset (98.95%), Ovarian cancer diagnosis-2 dataset (71.26%), Ovarian cancer diagnosis-3 dataset (81.58%), Ovarian cancer diagnosis-4 dataset (83.31%).
[71]	Eleven Cancer Datasets.	PSO C4.5.	PSO integrated with a C4.5 DT technique is used for gene selection. It selects a minimal number of significant genes from the vast number of genes that can play a role in cancer detection.	Superior performance over all other existing techniques.	…….
[72]	Breast, Colon DLBC, Leukemia, Lung Ovarian Cancer.	SVM-RFE, Binary Dragon Fly (BDF).	SVM-RFE is utilized to find 60% of candidate genes, and BDF is employed to identify the optimal subsets of candidate genes. This is optimized with the help of an objective function.	The experimental study demonstrates that the methodology is efficient and achieves a good accuracy using much smaller genes. The employed methodology can effectively address the problem of gene selection and cancer diagnosis. Is stable, and increases the complexity and quality of the classification model.	phase 1, phase2 Breast cancer (64.71%, 86.22%), Colon cancer (81.08%, 97.46%), DLBC (80.62%, 89.44%), Leukemia (81.48%, 95.81%), Lung cancer (80.73%, 99.14%), Ovarian cancer (93.80%, 98.19%).
[73]	Wisconsin Diagnostic Breast Cancer, Parkinson's Disease, Heart-Statlog, Colon Tumor, Central Nervous System, All-AML (Leukemia), Breast cancer, Ovarian cancer.	BBHA, RF, Bagging, C5.0, C4.5, Boosted C5.0 and CART, Naïve Bayes.	BBHA have been used for feature selection whereas RF, Bagging, C5.0, C4.5, Boosted C5.0, CART, Naïve Bayes have been used for classification.	Removes the problem stuck in local optima. Computationally cheap, with higher efficiency in terms of CPU time. Reduces the number of model configuration parameters. The number of selected optimized characteristics and performance parameters (accuracy, specificity, and sensitivity).	BBHA_RF Wisconsin diagnostic breast cancer (97.38%), parkinson's disease (93.91%), Heart-statlog, (85.75%), Colon tumor (91.41%), Central nervous system (91.85%), All-AML (Leukemia) (98.61%), Breast cancer (87.77%),

					Ovarian cancer (99.82%).
[74]	Ovarian cancer.	GA and ANFIS.	GA is used for feature selection, whereas ANFIS is used for classification.	Minimize costs, maximize precision, execute quickly, improve classification accuracy, and ease in implementation.	ANFIS (RMSE-1.15153) GA_ANFIS (RMSE-1.15153)
[75]	Ovarian cancer, Pancreatic cancer.	GA, Support Vector Machine (SVM).	GA has been used as a technique for feature selection and SVM for classification.	In context of feature selection technique, economic with respect to time, reduce datasets, remove irrelevant data, improved accuracy of Classifier and overall reliable.	Ovarian cancer (99.07%), Pancreatic cancer ……
[76]	Leukemia, Breast, Colon, Ovarian Prostate, Lung cancer datasets.	GA, PSO and SVM.	The GA and PSO methods are used to select informative genes from thousands of candidates, while the SVM algorithm is used for classification.	Acceptable solutions.	Leukemia (97.38%), Breast (86.35%), Colon (100%), Ovarian (99.44%), Prostate (98.66%), Lung (90.00%).
[77]	Serum Samples	SVMRFE, 1SVM, SVMRFE_NL, VMRW, SVM, SVM_NL.	SVMRFE with linear SVM, L1SVM, VMRFE_NL, SVMRW have been utilized as a feature selection technique whereas SVM, SVM_NL have been used as the classification methods.	Accurate, reliable, most promising metabolomic-based approach for detection of ovarian cancer.	…….
[20]	Ovarian, Acute Myeloid Leukemia (AML)/Acute Lymphoblastic Leukemia (ALL), Colon.	SVMs	SVMs have been utilized for the classification of microarray expression dataset namely Ovarian, AML/ALL, Colon.	Robustness, SVMs are able to successfully classify the microarray expression data and give good performance. It may be utilized to find mis-labeled data.	……..
[78]	Breast Cancer, Lung Cancer, Colon Tumor, Ovarian Cancer, Leukemia.	K-S test, CFS, K-S test-CFS, mRMR, ReliefF, and SVM.	K-S test, CFS, K-S test-CFS, mRMR, ReliefF algorithm have been utilized for gene subset selection and SVM for classification.	K-S test-CFS demonstrates the effectiveness of the combined of the K-S test and CFS. Overall, the proposed approaches are very effective and promising.	K-S test-CFS (Best Perf.) Breast cancer (87.4%), Lung cancer (91.6%) Colon tumor (90.1 %), Ovarian cancer (98.5%), Leukemia (79.6%).
[79]	Lymphoma Data sets, Leukemia data set.	PCA, Class-separability analysis, Fisher-ratio, t-test, SVM.	PCA, Class-separability analysis, Fisher-ratio, t-test are utilized for feature reduction and SVM are utilized for microarray cancer datasets classification.	Dimensionality reduction eases the task because significant genes establish a new input space in which the instances are more likely to be correctly classified, SVM provides High classification accuracy	…….

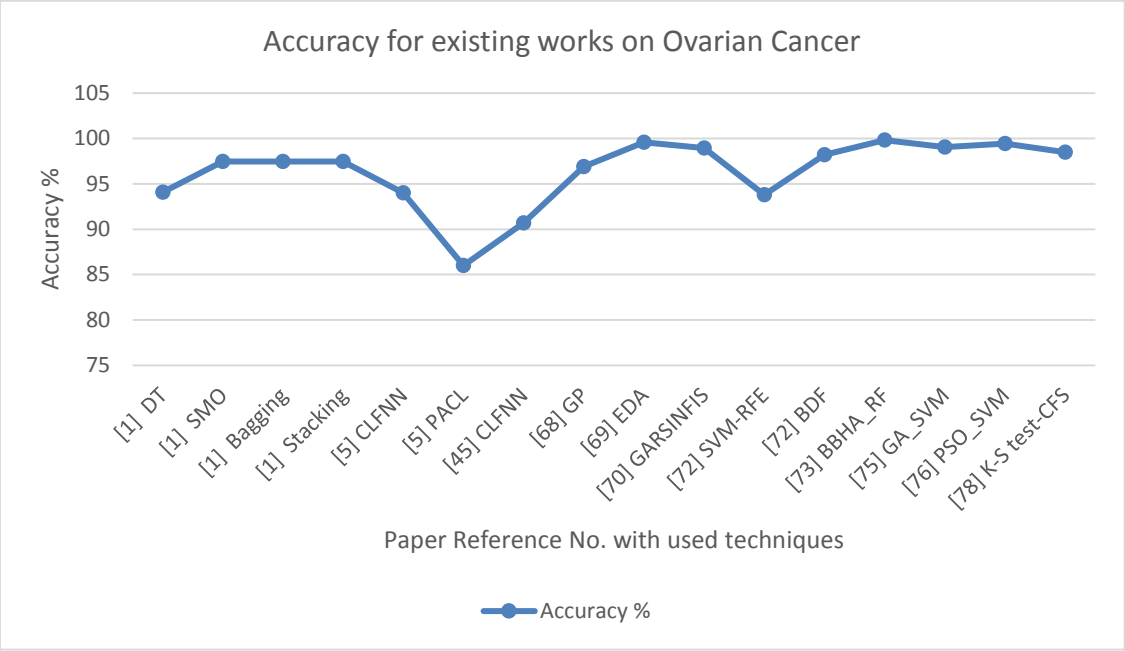

Fig. 1. Accuracy for existing works on Ovarian Cancer.

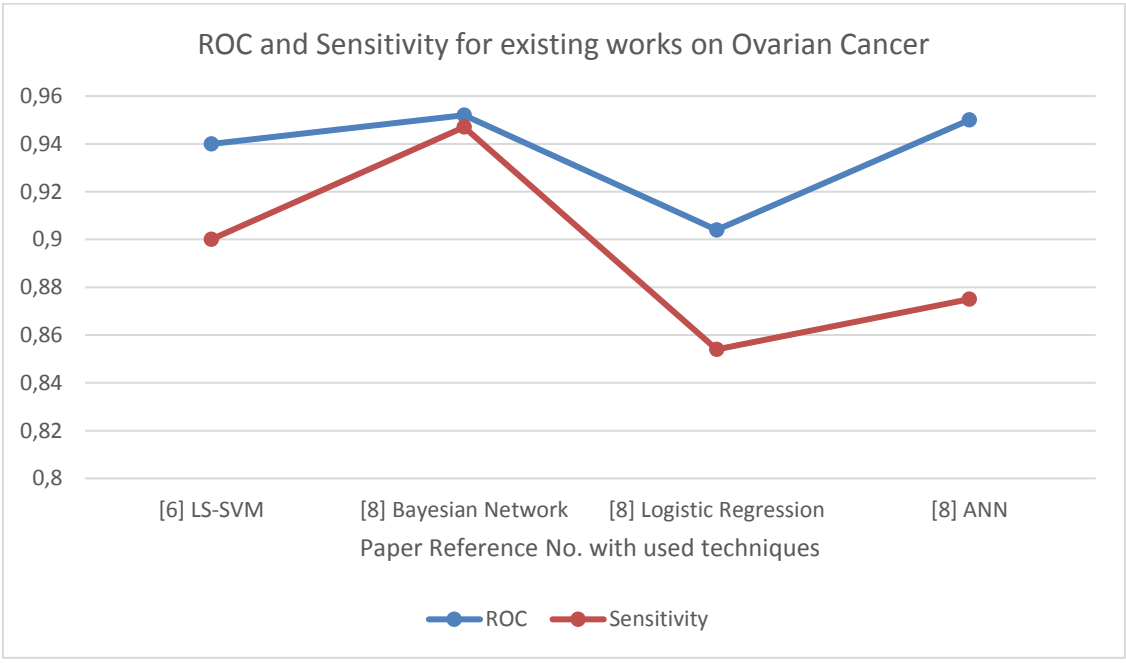

Fig. 2. ROC and Sensitivity for existing works on Ovarian Cancer.

Fig. 1 represents the accuracy for existing works of ovarian cancer. This figure illustrates the utilized techniques performance in context of accuracy for the existing articles of ovarian cancer. Table 1 is graphically represented in Figure 1 which illustrates the accuracy comparison for utilized techniques.

Fig. 2 represents the ROC and sensitivity for existing works of ovarian cancer. This figure illustrates the utilized techniques performance in context of ROC and sensitivity for the existing some articles of ovarian cancer. Table 1 is graphically represented in Figure 2 which illustrates the ROC and Sensitivity comparison for utilized techniques.

3. DESCRIPTIONS OF DIFFERENT TECHNIQUES AND TOOLS USED

There are various used existing techniques and tools which are used for classification algorithms for ovarian cancer diagnosis. Table 2 provides a concise summary of Techniques' benefits, issues, and applicability.

Table 2. Concise summarization of existing techniques/tools for Ovarian Cancer classification

Techniques/Tools Used	Advantages	Issues	Applicability
Analysis of Variance (ANOVA)	It is simple to deduce the experimental error to a large extent, and it is possible to deduce or increase the number of treatments. Design suitable for laboratory experiments; increased statistical power.	Unequal sample sizes influence the robustness of the assumption of equal variance, local control is completely neglected. Design is inefficient and sensitive as compared to others and difficult to determine which group is different.	Manufacturing, Healthcare service, Power reactors, Chemical plants, Recommendation of fertilizer against others for the improvement of a crop yield.
Regression Analysis	Simple implementation, easier to interpret and can be performed on linearly separable datasets. Over fitting can be reduced by regularization.	Prone to under fitting and prone to outliers. It assumes that the data is independent. Poor data, software limitations, human error and linear relationship are among the variables so not recommended for most practical applications. Includes a protracted and intricate calculation and analysis procedure. Cannot be utilized in qualitative phenomenon viz honesty, criminality, etc.	Operation efficiency, Predictive analytics, Supporting decisions, New insights and Correcting errors.
Correlation Analysis	Determine the direction and strength of two variables' relationship. Study behavior like Gain quantitative information that can be readily analyzed. Possesses a favorable starting position. Research studies are simple to categorize.	Cannot show cause and effect. Ha no control of third variable that might affect the correlation. Research can be a time-consuming process.	Insurance companies, Government through census, Item analysis, Factor analysis, Graphical form.
Binary Dragon Fly (BDF)	Is scalable, flexible and robust.	Is non-optimal, uncontrollable, unpredictable, non-understandable and non-immediate.	Image Processing, Machine Learning, Engineering (Mechanical, electrical, etc.), Wireless and Network.
Binary version of Black Hole Algorithm (BBHA)	Individual modes may dominate the time evolution of some perturbation and an entire set of them could be used to totally depict this time evolution.	When a black hole evaporates information is gone. Due to this there is difficulty in energy conservation and invariance in time predictability.	Optimization Problems (Travelling Salesman Problem), Set Covering, Image Processing, Data Mining, Computer Vision, Science and Engineering.
Kolmogorov-Smirnov (K-S) test	More powerful test, easier to compute. The test statistic is independent of the expected frequency distribution. Can be used for goodness of an exact test. Small sample could work as well. Is distribution free, K-S tables are easily available.	More sensitive to deviations near the middle of the distribution than at the tails. Only applies to continuous distribution. Distribution to be compared which must be fully specified.	Continuous distributions, Pseudo random numbers, Uniform distributions, Cumulative Probability Distribution (CDF), Normality assumptions.
GA, Decision Trees, SVM, Bayesian Networks, Logistic Regression, ANN, PSO, Fuzzy Logic, Rough Sets, ANFIS, PCA, KNN, Likelihood, Naive Bayes, CART, Random Forest, Adaboost	[59]	[59]	[59]
WEKA, MATLAB, C++ Builder, Java, Oracle, SPSS.	[59]	[59]	[59]
GDA, Ranker Search, K-	[59]	[59]	[59]

means clustering, Bijective Soft Set, ACO, LDA, Association rules, K-fold cross validation, Gausian mixture model, Validation.			

Table 3. Summary of the Future works over existing works for ovarian cancer

Ref.	Existing Works	Future Works
[1]	For data preprocessing, genetic algorithms (GA) and correlation-based heuristics were used. C4.5 decision trees and SVM have been used to classify gene expression data sets, such as ovarian, prostate, and lung cancer. To improve the accuracy, researchers have further used Bagging and stacking for the classification.	To attain good accuracy, the proposed approach can be used in conjunction with other models. The incorporated gene-search algorithm will aid in the selection of treatment options and the development of medications by detecting categorization algorithms in a timely and accurate manner.
[8]	Bayesian Networks, Logistic Regression, ANN have been used for the classification of ovarian cancer. The Bayesian networks have been suggested the best among all for the classification of ovarian tumor.	The applicability of the proposed techniques has been summarized. A possible hybrid usage of the technique has been outlined to design an algorithm which will predict the correct class i.e., benign or malignant.
[4]	A new memetic algorithm called CD-MFS was used to develop a fuzzy rule base system for explainable gene expression classifier. It is used for cancer tumor detection from 14_Tumors data set. The performance has been evaluated through this data set and also compared with other existing classification system.	Improvement in accuracy can be obtained using CD-MFS as well as concise rule set for medical conditions can be developed.
[5]	In ovarian cancer, CLFNN (FALCON-AART) and PACL (FCMAC-FALCON) are used as diagnostic decision support systems. Both networks are capable of automatically inferring and formulating a rule-base without previous knowledge of the problem domain.	They can rely on intuitive fuzzy rules to support their thinking on their own, which is critical for building client trust.
[6]	To use and assess LS-SVM classifiers in a Bayesian evidence framework to distinguish between malignant and benign ovarian cancer. The dataset was examined, relevant features were chosen, parameters were estimated, and performance was evaluated using a ROC chart.	Designing a hybrid strategy that combines the black-box learning capabilities with the expert knowledge of white-box models to increase model performance even more.
[67]	In this study, gene selection and ovarian cancer classification were combined. Using regression analysis, target genes were identified. Selection of gene marker from target genes was done using GA, PSO, SVM and ANOVA Additionally, the updated fuzzy model uses the collected gene markers to categorize cancer tissues. The researcher's method has also been tested on datasets related to colon and breast cancer.	The researcher's method has shown to be effective in the treatment of ovarian cancer, and it may also be effective in other diseases. Biologists can use the results of this study as a guide.
[68]	Classification using GP was done using gene expression datasets lymphoma cancer dataset, lung cancer dataset, and ovarian cancer dataset. Diversity was estimated by using a distance measurement among classification rules.	For assessing accurate diversity, a more complex distance measurement will be developed for ensemble GP. A non-pair-wise method for assessing diversity will be investigated as well.
[45]	CLFNN model was proposed to help in diagnosis. Compared to CI approaches, the CLFNN takes advantage of lateral inhibition between positive and negative events. It also has a feature that allows it to generate rules on its own. The FALCON-AART is a good example of CLFNN's performance.	The proposed methodology could be applied to verify different ovarian cancer hypotheses or theories.
[69]	A new ensemble method has been proposed by researchers. To begin, correlation analysis and Fisher ratio utilized to extract useful gene features, decrease features, and keep the important ones. The EDA is then used to create a classification classifier committee. Finally, the outcomes of the suggested techniques are compared with several advanced artificial techniques to find that the proposed methods had a higher recognition rate.	The proposed novel approach can be used for other medical diseases.

[70]	GARSINFIS is a novel self-organizing model proposed by researchers that employ the inference rules base automatically derived by our proposed GARSC method. GARSC is a clustering method that combines GA & RS theory together. GA is used to determine optimal solutions. The RS theory is used to solve the problem of dimensionality curse. RS approximations deduce a substantial portion of the original knowledge base without sacrificing vital information. Simple convincing fuzzy inference rules are systematically devised to diagnose both published medical data sets and hospital-collected real-world ovarian cancer data. The proposed method can be used to diagnose ovarian cancer as a trustworthy decision support system.	The proposed approach may increase accuracy and become more reliable for other medical diseases.
[71]	Using a PSO combined with a C4.5 decision tree, researchers have suggested a unique method for relevant gene selection. From the millions of genes in the data, a small number of significant genes that can aid in the identification of tumors are identified. The proposed PSO C4.5 technique was presented, and the results were compared to those of other well-known classifiers, including the support vector machine, self-organizing map, BP neural networks, and C4.5 decision tree. Eleven cancer datasets were used to evaluate the effectiveness of the proposed method. The proposed technique achieved a good accuracy compared with existing techniques.	It is necessary to conduct further research into the adjustment of PSO parameters as well as the local optima trapping problem.
[72]	For the process of diagnosing malignancy, researchers have proposed kernel-based learning and attribute selection. The proposed strategy consists of two phases: In the first phase, the SVM-RFE algorithm is used to select sixty percent of candidate genes, and in the second phase, the BDF algorithm is used to identify the optimal subsets of candidate genes. This is optimized by using objective function. Colon cancer, breast cancer, leukemia, ovarian cancer, lung cancer and DLBCL cancer were among the six microarray datasets used in this trial.	The objective function may be derived from distances between classes or gene correlation.
[73]	Researchers have proposed a BBHA for feature selection and have applied 6 well known DT classifier notably RF, Bagging, C5.0, C4.5, Boosted C5.0 and CART for biological data. The proposed methodology is applied to eight publicly accessible datasets, and BBHA is compared to PSO, GA, SA, and CFS. Further BBHA is combined with Naive Bayes and executed on datasets along with also image domains. The study shows that the RF outperforms and BBHA is the better feature selection technique compared to Binary PSO, GA, SA, CFS.	The proposed technique can be used in a variety of fields, including pattern recognition and bioinformatics. Also, researchers may apply a new strategy for the implementation of BBHA for the improvement than existing.
[74]	Researchers have proposed a GA for feature selection and ANFIS for classification. The primary objective of this study is to simultaneously optimize the parameters and feature subset without degrading ANFIS classification performance. The features were chosen from a total of 15154 using feature selection algorithms. It reduces the cost of computation and boosts the efficacy of classification methods.	The proposed methodology is applicable to additional medical conditions.
[75]	SVM and GA have been used for attribute selection and classification, respectively. Through GA technique, 20 relevant attributes have been selected among 15154 for ovarian cancer, whereas 16 relevant attributes have been selected among 6771 for pancreatic cancer. By reducing datasets, it decreases the computation cost of CPU.	The proposed approach may be used in another biological dataset.
[76]	Researchers have discussed and compared the use of genetic algorithms (GA) and particle swarm optimization (PSO) with support vector machines (SVM) for the classification of high-dimensional microarray data.. The GA and PSO methods are used to choose a smaller number of relevant genes from thousands of candidates, whilst SVM is used for classification.	In an effort to identify the optimal set of genes, amalgamation of other meta-heuristics with classification was used. In this context, the use of multi objective approaches may contribute in gene subset selection.
[77]	Researchers have looked into new methods for automatically classifying metabolic data derived from ovarian cancer and	The advanced machine learning algorithm may be developed for the mentioned evaluation procedures.

	benign control sera. SVMRFE with linear SVM, L1SVM, SVMRFE_NL, SVMRW have been used as a method for selecting features and SVM SVM_NL have been used as the classification methods. The proposed approaches were used to examine based on the three evaluation procedures. The validation was done using: (a)50 trials of 52-20-split validation (b) leave-one-out-cross validation (LOOCV), (c) 12-fold cross validation (CV). The LOOCV is the most reliable among all three evaluation procedures. SVMRFE_NL is the best feature selection among all three and SVM_NL is best classification method. Overall SVMRFE_NL with SVM_NL performs the best performance.	
[20]	SVMs have been employed in the classification of tissue instances as well as the investigation of data for mislabeled or dubious tissue outcomes.	There is an assumption that forecasts of a treatment's success or failure will be feasible, but so far, the findings of these types of experiment studies have been indeterminate.
[78]	Researchers have presented CFS, K-S test-CFS, Minimum Redundancy Maximum Relevance (mRMR), K-S test and ReliefF algorithm for gene subset selection and SVM for classification. Initially, the K-S test algorithm select distinguish genes or candidate gene set, and then, CFS was used to identify genes among those selected by the K-S test.	Compared to other gene selection method, K-S test-CFS outperforms, but it has no advantages over the other algorithms in terms of runtime. As a result, the study's future focus will be on how to improve the K-S test-CFS algorithm's performance.
[79]	Researchers have presented a Fisher-ratio, PCA, t-test, class-separability analysis, utilized for feature selection and SVM are utilized for Gene expression data sets, lymphoma data sets, Leukemia data set classification. Due to the fact that significant genes designate a new input space in which the data set is anticipated to be accurately classified, dimension reduction can simplify the task. SVM provides results that are effective.	Linear SVMs can be implemented with nonlinear problems.

4. DISCUSSION AND FUTURE DIRECTIONS

The future works over the existing works for the ovarian cancer are discussed in Table 3.

The primary goal of this review article is how to achieve early diagnosis, enhance the procedure and accurate detection for ovarian cancer.

Fig. 3 illustrates the summarization of this review article.

In this review article, authors have collected information from various sources such as journals, books, book chapters, conference papers, internet, etc. They have majorly found research articles on classification techniques, and feature selection followed by classification techniques for the prediction of ovarian cancer. Classification is one of the most extensively used and efficient techniques for a variety of applications, such as the medical diagnosis of ovarian cancer patients. The ovarian cancer datasets were not the same in each article as it was from different sources. Thereafter, authors have briefly summarized the existing research articles in context of name of dataset, techniques and tools used for the purpose of techniques, advantages, issues and applications of particular tool, and techniques. The performance of each technique has been summarized in context of accuracy. The review article is suggesting us that researchers may directly start his work with the help of future work and issues given in the article.

5. CONCLUSIONS AND LIMITATIONS

It is commonly known that up until a few years ago, doctors only used their clinical judgement to diagnose the ovarian cancer disease and relied heavily on the patient's unprocessed clinical data, which mostly consisted of laboratory test results. These laboratory test results differed according to meals, exercise, illness, stress, slight temperature variations, various pieces of equipment, and sample handling techniques. Therefore, this form of disease diagnosis is not only time-consuming, but also completely dependent on the availability and expertise of the physician, who must deal with ambiguous and imprecise patient clinical data. The necessity for an effective (intelligent) diagnosis system is therefore critical for enhancing and accelerating the decision-making process utilizing only readily available clinical data. The significance of the research article has been summarized considering the performance of many existing articles and also outlined the optimally designed hybrid techniques with a medical point of view to design a model to predict the correct class. The limitation of these works is that, it doesn’t guarantee a good accuracy for the application of classification techniques on medical datasets like ovarian cancer. Also, accuracy will differ for different databases. The same thing when applied

with feature selection methods has proven to be better for Ovarian Cancer Dataset but its efficacy w.r.t. other databases need to be ascertained.

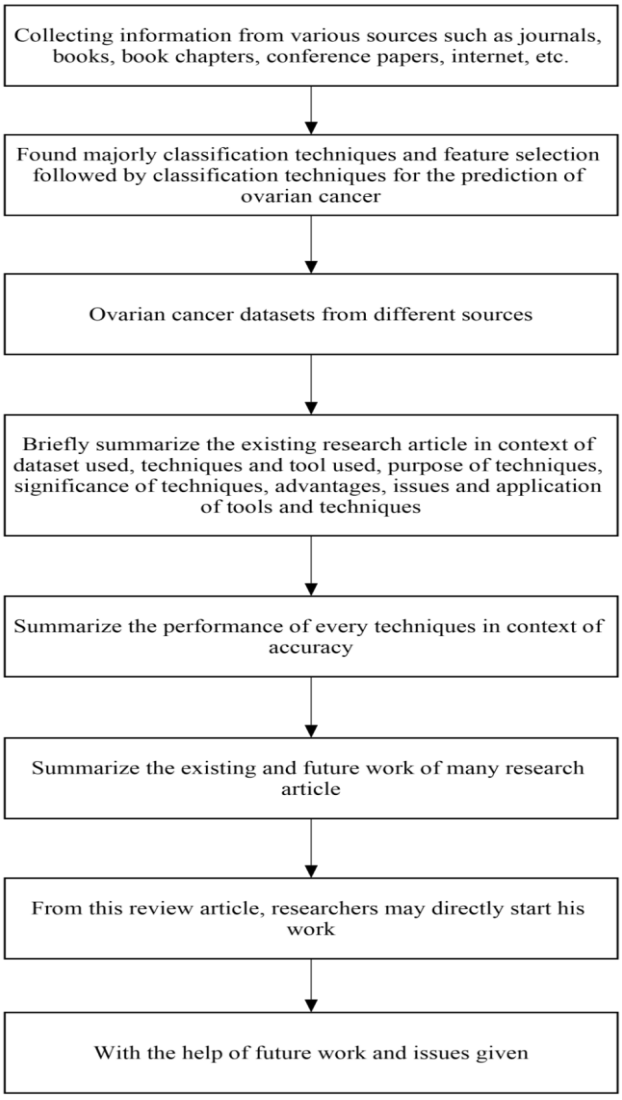

Fig. 3. Summarization of Review Article

REFERENCES

[1] S. Shah and A. Kusiak, "Cancer gene search with data-mining and genetic algorithms," *Comput. Biol. Med.*, vol. 37, no. 2, pp. 251–261, 2007, doi: 10.1016/j.compbiomed.2006.01.007.

[2] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. their Appl.*, vol. 13, no. 4, pp. 18–28, 1998.

[3] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Comput.*, vol. 13, no. 3, pp. 637–649, 2001.

[4] A. Zibakhsh and M. S. Abadeh, "Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function," *Eng. Appl. Artif. Intell.*, vol. 26, no. 4, pp. 1274–1281, 2013, doi: 10.1016/j.engappai.2012.12.009.

[5] T. Z. Tan, C. Quek, and G. S. Ng, "Ovarian cancer diagnosis by hippocampus and neocortex-inspired learning memory structures," *Neural Networks*, vol. 18, no. 5–6, pp. 818–825, 2005, doi: 10.1016/j.neunet.2005.06.027.

[6] C. Lu, T. Van Gestel, J. A. K. Suykens, S. Van Huffel, I. Vergote, and D. Timmerman, "Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines," *Artif. Intell. Med.*, vol. 28, no. 3, pp. 281–306, 2003, doi: 10.1016/S0933-3657(03)00051-4.

[7] P. Antal, G. Fannes, D. Timmerman, Y. Moreau, and B. De Moor, "Using literature and data to learn Bayesian networks as clinical models of ovarian tumors," *Artif. Intell. Med.*, vol. 30, no. 3, pp. 257–281, 2004, doi: 10.1016/j.artmed.2003.11.007.

[8] P. Antal, G. Fannes, D. Timmerman, Y. Moreau, and B. De Moor, "Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection," *Artif. Intell. Med.*, vol. 29, no. 1–2, pp. 39–60, 2003, doi: 10.1016/S0933-3657(03)00053-8.

[9] M. H. Tsai, C. H. Lai, and S. S. Yu, "A statistical and learning based oncogene detection and classification scheme using human cDNA expressions for ovarian carcinoma," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10066–10074, 2011, doi: 10.1016/j.eswa.2011.02.010.

[10] S. W. Tung, C. Quek, and C. Guan, "SoHyFIS-Yager: A self-organizing Yager based Hybrid neural Fuzzy Inference System," *Expert Syst. Appl.*, vol. 39, no. 17, pp. 12759–12771, 2012, doi: 10.1016/j.eswa.2012.02.056.

[11] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015, doi: 10.1016/j.csbj.2014.11.005.

[12] S. A. Hussain, "Assessment of suspicious ovarian masses by using USG & CT Techniques," *Assess. suspicious ovarian masses by using USG CT Tech.*, vol. 5, no. 10, pp. 36–40, 2019.

[13] S. Khanna and S. Agarwal, "An integrated approach towards the prediction of likelihood of diabetes," *Proc. - 2013 Int. Conf. Mach. Intell. Res. Adv. ICMIRA 2013*, pp. 294–298, 2014, doi: 10.1109/ICMIRA.2013.62.

[14] T. R. Adib *et al.*, "Predicting biomarkers for ovarian cancer using gene-expression microarrays," *Br. J. Cancer*, vol. 90, no. 3, pp. 686–692, 2004, doi: 10.1038/sj.bjc.6601603.

[15] T. Jirapech-Umpai and S. Aitken, "Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes," *BMC Bioinformatics*, vol. 6, pp. 1–11, 2005, doi: 10.1186/1471-2105-6-148.

[16] K. D. Steffensen *et al.*, "Mutant epidermal growth factor receptor in benign, borderline, and malignant ovarian tumors," *Clin. Cancer Res.*, vol. 14, no. 11, pp. 3278–3282, 2008, doi: 10.1158/1078-0432.CCR-07-4171.

[17] K. Ono *et al.*, "Identification by cDNA microarray of genes involved in ovarian carcinogenesis," *Cancer Res.*, vol. 60, no. 18, pp. 5007–5011, 2000.

[18] S. Upadhya, S. Upadhya, and D. M. Vasudevan, "Tissue polypeptide specific antigen in the post therapeutic evaluation of patients with ovarian and colorectal cancer," *Indian J. Clin. Biochem.*, vol. 18, no. 1, pp. 46–51, 2003, doi: 10.1007/BF02867664.

[19] S. K. Dora, A. B. Dandapat, B. Pande, and J. P. Hota, "A

prospective study to evaluate the risk malignancy index and its diagnostic implication in patients with suspected ovarian mass," *J. Ovarian Res.*, vol. 10, no. 1, pp. 1–9, 2017, doi: 10.1186/s13048-017-0351-2.

[20] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000, doi: 10.1093/bioinformatics/16.10.906.

[21] A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, "Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data," *J. Biomed. Biotechnol.*, vol. 2003, no. 5, pp. 308–314, 2003, doi: 10.1155/S1110724303210032.

[22] C. Lu, J. De Brabanter, S. Van Huffel, I. Vergote, and D. Timmennan, "To Predict Malignancy of Ovarian Tumors," pp. 1637–1640, 2001.

[23] C. Renz, J. C. Rajapakse, K. Razvi, and S. K. C. Liang, "Ovarian cancer classification with missing data," *ICONIP 2002 - Proc. 9th Int. Conf. Neural Inf. Process. Comput. Intell. E-Age*, vol. 2, pp. 809–813, 2002, doi: 10.1109/ICONIP.2002.1198171.

[24] C. C. Chuang, S. F. Su, and J. T. Jeng, "Dimension reduction with support vector regression for ovarian cancer microarray data," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2, pp. 1048–1052, 2005, doi: 10.1109/icsmc.2005.1571284.

[25] J. T. Jeng, T. T. Lee, and Y. C. Lee, "Classification of ovarian cancer based on intelligent systems with microarray data," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2, pp. 1053–1058, 2005, doi: 10.1109/icsmc.2005.1571285.

[26] D. Wang, G. S. Ng, and C. Quek, "Ovarian cancer diagnosis using fuzzy neural networks empowered by evolutionary clustering technique," *2006 IEEE Congr. Evol. Comput. CEC 2006*, pp. 2764–2770, 2006, doi: 10.1109/CEC.2006.1688655.

[27] M. Alipoor, M. K. Parashkoh, and J. Haddadnia, "A novel biomarker discovery method on protemic data for ovarian cancer classification," *Proc. - 2010 18th Iran. Conf. Electr. Eng. ICEE 2010*, pp. 1–6, 2010, doi: 10.1109/IRANIANCEE.2010.5507114.

[28] S. Mahmoudi, B. S. Lahijan, and H. R. Kanan, "ANFIS-based wrapper model gene selection for cancer classification on microarray gene expression data," *13th Iran. Conf. Fuzzy Syst. IFSC 2013*, pp. 1–6, 2013, doi: 10.1109/IFSC.2013.6675687.

[29] M. Bakrani, K. Poor, V. Mehrzad, and N. Razmi, "Comparison of the Serum Level of Cancer Antigen 125 and Human Epididymis Protein 4 in Ovarian Cancer Patients and Healthy Groups in Isfahan City," *Adv. Biomed. Res.*, vol. 6, no. 1, p. 124, 2017, doi: 10.4103/2277-9175.216778.

[30] I. Yoo *et al.*, "Data mining in healthcare and biomedicine: A survey of the literature," *J. Med. Syst.*, vol. 36, no. 4, pp. 2431–2448, 2012, doi: 10.1007/s10916-011-9710-5.

[31] L. Yu and H. Liu, "Redundancy based feature selection for microarray data," *KDD-2004 - Proc. Tenth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, no. 2, pp. 737–742, 2004, doi: 10.1145/1014052.1014149.

[32] U. Shafi, "Ovarian cancer sign , symptoms and detection techniques," vol. 4, no. 6, pp. 257–261, 2018.

[33] D. Timmerman *et al.*, "Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: A multicenter study by the International Ovarian Tumor Analysis Group," *J. Clin. Oncol.*, vol. 23, no. 34, pp. 8794–8801, 2005, doi: 10.1200/JCO.2005.01.7632.

[34] S. A. A. Sohaib and R. H. Reznek, "MR imaging in ovarian cancer," *Cancer Imaging*, vol. 7, no. SPEC. ISS. A, pp. 119–129, 2007, doi: 10.1102/1470-7330.2007.9046.

[35] Claus E.B., "Risk models used to counsel women for breast and ovarian cancer: A guide for clinicians.," *Fam. Cancer*, pp. 197–206, 2001.

[36] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, pp. 1–13, 2006, doi: 10.1186/1471-2105-7-3.

[37] A. Enshaei, C. N. Robson, and R. J. Edmondson, "Artificial Intelligence Systems as Prognostic and Predictive Tools in Ovarian Cancer," *Ann. Surg. Oncol.*, vol. 22, no. 12, pp. 3970–3975, 2015, doi: 10.1245/s10434-015-4475-6.

[38] S. Rashid and R. Kaur, "Automatic detection of ovarian cancer based on improved DWT transformation," *Int. J. Eng. Technol.*, vol. 7, no. 2.27 Special Issue 27, pp. 104–108, 2018, doi: 10.14419/ijet.v7i3.12572.

[39] V. Thulasiraman and S. Kavitha, "RISK PREDICTION SYSTEM USING DATA MINING TECHNIQUES IN GYNECOLOGICAL OVARIAN CANCER," vol. 6956, no. July, pp. 1993–1998, 2019, doi: 10.21917/ijsc.2019.0278.

[40] C. J. Tseng, C. J. Lu, C. C. Chang, G. Den Chen, and C. Cheewakriangkrai, "Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence," *Artif. Intell. Med.*, vol. 78, pp. 47–54, 2017, doi: 10.1016/j.artmed.2017.06.003.

[41] D. T. John and M. Suvarna, "Classification of Ovarian Cysts Using Artificial Neural Network," *Int. Res. J. Eng. Technol.*, vol. 3, no. 6, pp. 2509–2514, 2016, [Online]. Available: http://www.irjet.net/archives/V3/i6/IRJET-V3I6469.pdf.

[42] S. Nakariyakul, "A hybrid gene selection algorithm based on interaction information for microarray-based cancer classification," *PLoS One*, vol. 14, no. 2, pp. 1–17, 2019, doi: 10.1371/journal.pone.0212333.

[43] E. F. Petricoin *et al.*, "Use of proteomic patterns in serum to identify ovarian cancer," *Lancet*, vol. 359, no. 9306, pp. 572–577, 2002, doi: 10.1016/S0140-6736(02)07746-2.

[44] S. Li, X. Wu, and M. Tan, "Gene selection using hybrid particle swarm optimization and genetic algorithm," *Soft Comput.*, vol. 12, no. 11, pp. 1039–1048, 2008, doi: 10.1007/s00500-007-0272-x.

[45] T. Z. Tan, C. Quek, G. S. Ng, and K. Razvi, "Ovarian cancer diagnosis with complementary learning fuzzy neural network," *Artif. Intell. Med.*, vol. 43, no. 3, pp. 207–222, 2008, doi: 10.1016/j.artmed.2008.04.003.

[46] M. Pirooznia, J. Y. Yang, M. Q. Qu, and Y. Deng, "A comparative study of different machine learning methods on microarray gene expression data," *BMC Genomics*, vol. 9, no. SUPPL. 1, pp. 1–13, 2008, doi: 10.1186/1471-2164-9-S1-S13.

[47] J. M. Sorace and M. Zhan, "A data review and re-assessment of ovarian cancer serum proteomic profiling," *BMC Bioinformatics*, vol. 4, pp. 1–13, 2003, doi: 10.1186/1471-2105-4-24.

[48] M. A. Rahman and R. C. Muniyandi, "An enhancement in cancer classification accuracy using a two-step feature selection method based on artificial neural networks with 15

neurons," *Symmetry (Basel).*, vol. 12, no. 2, 2020, doi: 10.3390/sym12020271.

[49] D. K. Choubey and S. Paul, "GA _ J48graft DT : A Hybrid Intelligent System for Diabetes Disease Diagnosis," vol. 7, no. 5, pp. 135–150, 2015.

[50] D. K. Choubey and S. Paul, "GA-RBF NN: A classification system for diabetes," *Int. J. Biomed. Eng. Technol.*, vol. 23, no. 1, pp. 71–93, 2017, doi: 10.1504/IJBET.2017.082229.

[51] D. K. Choubey, S. Paul, S. Kumar, and S. Kumar, "Classification of pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection," *Commun. Comput. Syst. - Proc. Int. Conf. Commun. Comput. Syst. ICCCS 2016*, pp. 451–455, 2017, doi: 10.1201/9781315364094-82.

[52] M. Kumar, S. K. Jangir, S. K. Mishra, S. K. Choubey, and D. K. Choubey, "Multi-Channel FLANN Adaptive Filter for Speckle & Impulse Noise Elimination from Color Doppler Ultrasound Images," pp. 1–4, 2020, doi: 10.1109/iconc345789.2020.9117288.

[53] D. K. Choubey and S. Paul, "GA _ MLP NN : A Hybrid Intelligent System for Diabetes Disease Diagnosis," no. January, pp. 49–59, 2016, doi: 10.5815/ijisa.2016.01.06.

[54] D. K. Choubey, S. Paul, S. Shandilya, and V. K. Dhandhania, "Implementation and Analysis of Classification Algorithms for Diabetes," *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.*, vol. 16, no. 4, pp. 340–354, 2018, doi: 10.2174/1573405614666180828115813.

[55] D. K. Choubey, M. Kumar, V. Shukla, S. Tripathi, and V. K. Dhandhania, "Comparative Analysis of Classification Methods with PCA and LDA for Diabetes," *Curr. Diabetes Rev.*, vol. 16, no. 8, pp. 833–850, 2020, doi: 10.2174/1573399816666200123124008.

[56] D. Sharma, P. Jain, and D. K. Choubey, "A Comparative Study of Computational Intelligence for Identification of Breast Cancer," in *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, 2020, pp. 209–216.

[57] A. Parthvi, K. Rawal, and D. K. Choubey, "A Comparative study using Machine Learning and Data Mining Approach for Leukemia," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 672–677, 2020, doi: 10.1109/ICCSP48568.2020.9182142.

[58] S. Pahari and D. K. Choubey, "Analysis of Liver Disorder Using Classification Techniques: A Survey," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–4, 2020, doi: 10.1109/ic-ETITE47903.2020.300.

[59] D. K. Choubey and S. Paul, "Classification techniques for diagnosis of diabetes: A review," *Int. J. Biomed. Eng. Technol.*, vol. 21, no. 1, 2016, doi: 10.1504/IJBET.2016.076730.

[60] M. Kumar, S. K. Mishra, S. K. Choubey, S. S. Tripathy, D. K. Choubey, and D. Das, "Cat Swarm Optimization based Functional Link Multilayer Perceptron for Suppression of Gaussian and Impulse Noise from Computed Tomography Images," *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.*, vol. 16, no. 4, pp. 329–339, 2018, doi: 10.2174/1573405614666180903115336.

[61] K. Srivastava and D. K. Choubey, "Soft Computing, Data Mining, and Machine Learning Approaches in Detection of Heart Disease: A Review," in *International Conference on Hybrid Intelligent Systems*, 2019, pp. 165–175.

[62] K. Srivastava and D. K. Choubey, "Heart Disease Prediction using Machine Learning and Data Mining," *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 21–219, 2020, doi: 10.35940/ijrte.f9199.059120.

[63] S. Kumar, U. M. Mohapatra, D. Singh, and D. K. Choubey, "EAC: Efficient associative classifier for classification," *Proc. - 2019 Int. Conf. Appl. Mach. Learn. ICAML 2019*, pp. 15–20, 2019, doi: 10.1109/ICAML48257.2019.00011.

[64] K. Bala, D. K. Choubey, and S. Paul, "Soft computing and data mining techniques for thunderstorms and lightning prediction: A survey," in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8203729.

[65] K. Bala, D. K. Choubey, S. Paul, and M. G. N. Lala, "Classification Techniques for Thunderstorms and Lightning Prediction: A Survey," in *Soft-Computing-Based Nonlinear Control Systems Design*, IGI Global, 2018, pp. 1–17.

[66] K. L. Tang, T. H. Li, W. W. Xiong, and K. Chen, "Ovarian cancer classification based on dimensionality reduction for SELDI-TOF data," *BMC Bioinformatics*, vol. 11, 2010, doi: 10.1186/1471-2105-11-109.

[67] Z. J. Lee, "An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer," *Artif. Intell. Med.*, vol. 42, no. 1, pp. 81–93, 2008, doi: 10.1016/j.artmed.2007.09.004.

[68] J. H. Hong and S. B. Cho, "The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming," *Artif. Intell. Med.*, vol. 36, no. 1, pp. 43–58, 2006, doi: 10.1016/j.artmed.2005.06.002.

[69] Y. Chen and Y. Zhao, "A novel ensemble of classifiers for microarray data classification," *Appl. Soft Comput. J.*, vol. 8, no. 4, pp. 1664–1669, 2008, doi: 10.1016/j.asoc.2008.01.006.

[70] D. Wang, C. Quek, and G. See Ng, "Ovarian cancer diagnosis using a hybrid intelligent system with simple yet convincing rules," *Appl. Soft Comput. J.*, vol. 20, pp. 25–39, 2014, doi: 10.1016/j.asoc.2013.12.018.

[71] K. H. Chen, K. J. Wang, K. M. Wang, and M. A. Angelia, "Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data," *Appl. Soft Comput. J.*, vol. 24, pp. 773–780, 2014, doi: 10.1016/j.asoc.2014.08.032.

[72] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Kernel-based learning and feature selection analysis for cancer diagnosis," *Appl. Soft Comput. J.*, vol. 51, pp. 39–48, 2017, doi: 10.1016/j.asoc.2016.12.010.

[73] E. Pashaei and N. Aydin, "Binary black hole algorithm for feature selection and classification on biological data," *Appl. Soft Comput. J.*, vol. 56, pp. 94–106, 2017, doi: 10.1016/j.asoc.2017.03.002.

[74] K. Rawat and K. Burse, "A Soft Computing Genetic-Neuro fuzzy Approach for Data Mining and Its Application to Medical Diagnosis 410 representation and an adequate evaluation function," no. 1, pp. 409–411, 2013, [Online]. Available: http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp.

[75] T. K. Mansoori, S. Amrit, and S. K. Mishra, "Feature selection by genetic algorithm and SVM classification for cancer detection," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 4, no. 9, 2014.

[76] E. Alba, J. García-Nieto, L. Jourdan, and E. G. Talbi, "Gene selection in cancer classification using PSO/SVM and

GA/SVM hybrid algorithms," *2007 IEEE Congr. Evol. Comput. CEC 2007*, pp. 284–290, 2007, doi: 10.1109/CEC.2007.4424483.

[77] W. Guan *et al.*, "Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines," *BMC Bioinformatics*, vol. 10, p. 259, 2009, doi: 10.1186/1471-2105-10-259.

[78] Q. Su, Y. Wang, X. Jiang, F. Chen, and W. C. Lu, "A Cancer Gene Selection Algorithm Based on the K-S Test and CFS," *Biomed Res. Int.*, vol. 2017, pp. 1–7, 2017, doi: 10.1155/2017/1645619.

[79] F. Chu and L. Wang, "Applications of support vector machines to cancer classification with microarray data," *Int. J. Neural Syst.*, vol. 15, no. 6, pp. 475–484, 2005, doi: 10.1142/S0129065705000396.