Research article

Innovative approach towards early prediction of ovarian cancer: Machine learning- enabled XAI techniques

Sheela Lavanya J M , Subbulakshmi P [*]

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

ARTICLE INFO

ABSTRACT

Globally, ovarian cancer affects women disproportionately, causing significant morbidity and mortality rates. The early diagnosis of ovarian cancer is necessary for enhancing patient health and survival rates. This research article explores the utilization of Machine Learning (ML) techniques alongside eXplainable Artificial Intelligence (XAI) methodologies to aid in the early detection of ovarian cancer. ML techniques have recently gained popularity in developing predictive models to detect early-stage ovarian cancer. These predictions are made using XAI in a transparent and understandable way for healthcare professionals and patients. The primary aim of this study is to evaluate the effectiveness of various ovarian cancer prediction methodologies. This includes assessing K Nearest Neighbors, Support Vector Machines, Decision trees, and ensemble learning techniques such as Max Voting, Boosting, Bagging, and Stacking. A dataset of 349 patients with known ovarian cancer status was collected from Kaggle. The dataset included a comprehensive range of clinical features such as age, family history, tumor markers, and imaging characteristics. Preprocessing techniques were applied to enhance input data, including feature scaling and dimensionality reduction. A Minimum Redundancy Maximum Relevance (MRMR) algorithm was used to select the features in the model. Our experimental results demonstrate that in Support Vector Machines, we found 85 % base model accuracy and 89 % accuracy after stacking several ensemble learning techniques. With the help of XAI, complex ML algorithms can be given more profound insights into their decision-making, improving their applicability. This paper aims to introduce the best practices for integrating ML and artificial intelligence in biomarker evaluation. Building and evaluating Shapley values-based classifiers and visualizing results were the focus of our investigation. The study contributes to the field of oncology and women's health by offering a promising approach to the early diagnosis of ovarian cancer.

1. Introduction

Worldwide, the most lethal form of gynecological malignancy in women is ovarian cancer. It poses a significant challenge due to its asymptomatic nature during the early stages, resulting in a late diagnosis and poor prognosis [1]. As a consequence, accurate and timely predictive models are needed for the early detection and intervention of ovarian cancer.

Ovarian cancer detection faces significant clinical challenges, primarily due to the absence of distinct early symptoms and effective screening tools [2]. Referred to as the "silent killer," ovarian cancer often manifests non-specific symptoms like bloating and abdominal

---

pain, contributing to delayed diagnosis. Unlike some other cancers, there are no routine screening tests for ovarian cancer, and current biomarkers, such as CA-125, lack the needed sensitivity for early-stage detection. The high-grade serous subtype, the most common and aggressive form, is often diagnosed at advanced stages. Additionally, the deep pelvic location of the ovaries and the heterogeneity of ovarian cancer present challenges in achieving early detection and tailoring treatments. Limited patient awareness, prevalence of advanced-stage diagnosis, and the absence of standardized diagnostic pathways further compound the issue. Addressing these challenges requires a comprehensive approach involving advancements in diagnostic technologies, heightened awareness, and research into more effective screening methods and biomarkers to enhance early detection and improve overall outcomes.

Researchers in medicine have increasingly turned to ML techniques in recent years, offering promising avenues for improving cancer prediction and diagnosis. ML models can effectively analyze large volumes of patient data and uncover hidden patterns and relationships, thereby aiding in the development of precise predictive models.

This research article presents an exploratory study on ovarian cancer prediction using intermediate ML techniques [3]. Using intermediate ML methods, this study will explore whether it is possible to identify ovarian cancer patients at high risk accurately. By employing intermediate techniques, we aim to maintain a balance between model complexity and interpretability [4], enabling healthcare professionals to gain valuable research insights regarding the predictive factors that contribute to ovarian cancer while maintaining clinical relevance.

Comprehensive clinical and demographic data were collected in this study using a diverse cohort of ovarian cancer patients. Several intermediate ML algorithms will be applied to train predictive models using the collected dataset, including random forests, gradient boosting, and support vector machines. MRMR feature selection plays a crucial role in ovarian cancer disease prediction by identifying a subset of clinically relevant features that are individually informative about the disease while minimizing redundancy [5]. In the context of ML models for ovarian cancer, MRMR assesses each feature correlation or mutual information with the target variable, ensuring the selection of features that provide maximal information about the disease outcome. By considering both the relevance of individual features and the collective non-redundancy among them, MRMR enhances the accuracy and interpretability of predictive models. This approach aids in constructing more efficient models for early detection, contributing to improved patient outcomes in the context of ovarian cancer diagnosis and prognosis. Through an iterative feature selection process, model optimization, and cross-validation, this study intends to determine the most influential risk factors associated with ovarian cancer development and develop robust and reliable prediction models.

In the landscape of ovarian cancer disease prediction, integrating XAI is paramount to ensure the credibility and acceptance of machine learning models in clinical practice. XAI plays a crucial role in demystifying the decision-making process of these models, providing clinicians with transparent insights into the factors influencing predictions [6]. This interpretability is essential for fostering trust and facilitating collaboration between healthcare professionals and AI systems. For ovarian cancer, an XAI model delivers accurate predictions and elucidates the significance of specific features in driving these predictions. Techniques such as feature importance analysis and interpretable model architectures contribute to this transparency. Understanding the rationale behind AI predictions becomes imperative in a medical context, where decisions impact patient outcomes. Incorporating XAI in ovarian cancer prediction models aligns technological advancements with the needs of healthcare practitioners, paving the way for a more informed and collaborative approach to disease detection and patient care.

The findings of this research can contribute to earlier diagnosis and prevention strategies for ovarian cancer. Our research aims to use intermediate ML techniques to provide clinicians with a tool for identifying high-risk patients, thereby improving the efficiency of screening, facilitating timely interventions, and improving patient outcomes.

As a result of this research article, ensemble ML techniques [7] have been determined to be valuable in predicting ovarian cancer. Besides shedding light on the predictors involved in this malignancy, this study intends to construct accurate models that can assist clinicians in making informed decisions about early detection and intervention. Overall, this research endeavors to make a positive impact on ovarian cancer management and contribute to reducing the burden of this devasting disease.

This research work includes the following main contributions.

- To develop a machine learning-based model for the prediction of ovarian cancer using ensemble methods.
- To address the high dimensionality problem, MRMR, an algorithm for selecting features, has been developed.
- To provide interpretability, the predictions of the model can be interpreted through the SHAP library, which provides interpretable explanations.

The rest of the paper is organized as follows: Section 2 presents a literature review relevant to this research. Section 3 provides a detailed description of the components of the proposed framework. The results of the experiment, as well as an analysis of the results, are presented in Section 4. In Section 5, a description of current research is provided.

2. Related work

This section presents an in-depth review of related work on ovarian cancer prediction using intermediate ML techniques. The studies discussed below highlight the advancements made in the field, the methodologies employed, and the key findings obtained. By examining these studies, we gain valuable insights into the existing approaches and identify research gaps that can be addressed in our own investigation.

To predict ovarian cancer, MRMR and simple decision trees were used to select features for 349 Chinese patients with 49 variables by Mingyang Lu et al. [8]. According to the study, logistic regression was the most effective method, compared with risk of ovarian

malignancy algorithms (ROMA) and risk of ovarian cancer algorithm (ROCA). In the study, Emma L. Barber et al. [9] identified women with OC based on medical case histories and applied natural language processing and ML to CT scan reports of patients who had undergone OC surgery to predict their outcome, and the method was found to be more accurate than traditional methods.

Eiryo Kawakami et al. [10] examined over 300 instances of EOC (Epithelial Ovarian Cancer) alongside over 100 instances of benign ovarian tumors. They utilized various models, including SVM, RF, Naive Bayes, Neural Nets, Elastic Nets, and GB reliant on blood biomarkers, to establish preoperative diagnoses and prognoses for EOC. There is a significant advantage to using ML models over classic regression models. Based on findings by E. Sun Paik et al. [11], GB classification demonstrates enhanced accuracy in delineating prognostic subgroups among individuals with EOC, enabling the prediction of survival outcomes for patients diagnosed with this condition.

To identify genetic variations associated with 35OCa-CR and 162OCa-CS cancers, Chen et al. EOC [12] examined cancer genome atlas data from patients [13]. Laios et al. utilized data from the ovarian database on OC and peritoneal cancer and developed predictive models based on KNNs and logistic regressions for patients with advanced OC. The Cancer Genome Atlas data FPKM sequences and clinical pathologic data of EOC from Shi-yi Liu et al. [14] were used in the establishment of a risk model for the Immune Microenvironment in EOC, using models such as Lasso regression, SVMs, and random forests. High levels of immune cells were found to predict better outcomes and antitumor phenotypes in EOCs.

Chen et al. [15] formulated an initial predictive model for mortality among individuals diagnosed with stage 1 and 2 OC, employing clinical covariates in conjunction with L2-regulated logistic regression techniques. According to Zengqi Yu et al. [16], cancer classification models have been constructed using feature selection, regression, backpropagation neural networks, and machine learning methodologies that rely on data collected from blood plasma samples. As reported by V V P Wibowo et al. [17], KNN and SVM models were used to diagnose ovarian cancer using samples from women with ovarian and no ovarian cancer, with KNN producing a higher level of accuracy than the SVM technique. Jun Ma et al. [18] have investigated 156 patients with EOC and applied a variety of modeling techniques, including Random Forests, SVMs, GBs, and Neural Networks to predict outcomes for blood biomarkers, including CTs, and found that Random Forests performed better than traditional regression models in the prediction of outcomes.

In the study conducted by Sung-Bae Cho and colleagues [19], many different classifiers were used to identify genes that were derived from microarrays which had a great deal of noise, as well as a variety of features used to select genes. The researchers have analyzed three datasets: one associated with leukemia cancer, one associated with colon cancer, and one associated with lymphoma cancer, each containing 72, 62, and 47 samples, respectively, for analysis. In the feature selection process, factors such as Pearson's and Spearman's correlation coefficients, Euclidean distances, mutual information, and signal-to-noise ratios were assessed. Additionally, considerations were given to information gains and mutual information while evaluating signal-to-noise ratios. As part of the classification process, the authors used MLP, kNNs, SVMs, and SOMs. Based on the experiments that were carried out with all of the datasets that were provided, the researchers found that the best accuracy result of 97.1 % was obtained with all of the classifiers presented in this paper on the Leukemia dataset. In a study conducted by Rui Xu et al. [20], gene expression data was used to predict patient outcomes based on PSO, which was performed in order to predict a patient's survival. In PSO, dimensionality is reduced through the implementation of Probabilistic Neural Networks. As a result of testing PSO/PNN on 240 samples of B-cell lymphoma, the results showed that PSO/PNN had a higher degree of accuracy for predicting survival in the case of 240 samples of B-cell lymphoma.

The study by Pirooznia et al. [21] focused on the development of methods for categorizing expression data and for selecting features from microarray data to detect expressed genes. The researchers were able to determine the effectiveness of different classification methods such as SVMs, Radial Basis Functions, Multilayer Perceptrons, DTs, and RFs based on evaluations of the analysis findings. In order to determine the efficiency of the classifier, the accuracy of the 10-fold cross-validation process was applied, including K-means approach, in order to determine how accurate the classifier was. Furthermore, we assessed the efficiency of the different methods of feature selection by using the following methods: SVM-RFE, Chi-Square method as well as Correlation-based feature selection (CFS). With respect to the SVM-RFE feature selection techniques used in the study, the authors concluded that for identifying significant genes with 100 % accuracy using the SVM-RFE feature selection methods, the method SVM-RFE feature selection provided the most efficient results. A novel approach based upon a feature selection method was described by Barnali Sahu et al. [22] to classify high-dimensional cancer microarray data based on a novel methodology reliant on feature selection. A signal-to-noise ratio (SNR) and the Power Spectrum Optimization (PSO) are used in this approach for optimizing the signal. The authors showed that when PSO is implemented alongside SVM, k-NN and PNN, the result is significantly better. There is a dataset that has 72 instances of leukemia and 7129 genes are present in the dataset; 62 instances of colon cancer have 2000 genes present in the dataset; 77 instances of DLBCL have 6817 genes present in the dataset; and it is found in the dataset that 97 instances of breast cancer contain 24481 genes. According to their study, PSO was able to give 100 % accuracy in breast cancer cases when used alongside other classifiers.

Using thousands of microarrays to assess gene expression patterns, C. Gunavathi et al. [23] picked out informative genes from those expression patterns This is due to the fact that many techniques based on swarm intelligence are available for finding informative genes, such as PSO, Cuckoo Search, Shuffled Frog Leaping with Levy Flight (SFLLF), and Shuffled Frog Leaping with Cuckoo Search. The classification method involved categorizing each sample based on its proximity to K-Nearest Neighbors (k-NN) classifiers, utilizing K-Nearest Neighbor (k-NN) classifiers as an integral component of the classification process. As a result of SFLLF feature selection methods, the best classification result is obtained when using the k-NN classifier. As a means of collecting these data, Kent Ridge Biomedical Data Repository was used. The objective of this research is to investigate the impact of swarm intelligence techniques such as PSO, CS, SFL, and SFLLF. The study aims to attain optimal results while analyzing diverse medical conditions including CNS, DLBCL, Lung Cancer, Ovarian Cancer, Prostate Outcomes, AML/ALL, Colon Tumor, among others. These encompass a range of datasets employed for experimenting with swarm intelligence techniques. Compared with the best result available, SFLLF performs much

Table 1
Summarization of related work.

Author	Methodology & Key Findings	Strengths	Weaknesses
Mingyang Lu et al. [8]	MRMR, decision trees, Logistic Regression, ROMA, ROCA; Logistic Regression most effective	- Eight notable features were selected by MRMR, among which two were identified as the top features by the decision tree model: human epididymis protein 4 (HE4) and carcinoembryonic antigen (CEA).	- Potential feature selection bias, and the need for external validation.
Emma L. Barber et al. [9]	NLP and ML on CT scan reports; More accurate than traditional methods	- Integrates NLP and ML	- Medical records from 2 institutions, so model may suffer from overfitting.
Eiryo Kawakami et al. [10]	Various ML models (SVM, RF, Naive Bayes, Neural Nets, Elastic Nets, GB) on blood biomarkers; ML models superior to classic regression	- Diverse set of models used - Achieved high accuracy and AUC values with Random Forest	- lacks explicit external validation and may face challenges in real-world clinical implementation. - Variations in performance across clinical parameters
E. Sun Paik et al. [11]	GB classification on individuals with EOC; Enhanced accuracy in prognostic subgroups	- gradient boosting (GB) for prognostic classification in epithelial ovarian cancer. - Large and diverse dataset with external validation from a separate medical center. - High predictive accuracy (AUC of 0.830 and 0.843) surpasses conventional methods.	- generalizability to other time frames is unclear. - Comparison limited to CoxPHR, lacking a broader assessment against various statistical and machine learning approaches.
Chen et al. [12]	Analysis of cancer genome atlas data for genetic variations; Identifies genetic variations associated with OC	- prediction with an AUC of 0.9658, outperforming individual signature gene-based prediction (AUC 0.6823).	- Limited discussion on the biological interpretation - Lack of external validation to assess the generalizability of the proposed
Laios et al. [13]	Predictive models based on KNNs and logistic regressions on OC and peritoneal cancer data; Predictive models for advanced OC	- k-nearest neighbor (k-NN) classifier, to predict R0 resection in ovarian cancer - mean predictive accuracy of 66 %	- sample size is relatively small (154 patients) - lacks external validation
Shi-yi Liu et al. [14]	Risk model for Immune Microenvironment in EOC using Lasso regression, SVMs, and random forests; High levels of immune cells predict better outcomes	- Addresses immune microenvironment	- lack external validation on independent datasets
Chen et al. [15]	L2-regulated logistic regression for predictive model on mortality in stage 1 and 2 OC; Focuses on mortality prediction	- L2-regularized logistic regression, achieves a proof-of-concept with a 76.1 % accuracy in predicting	- The achieved metrics (AUC of 0.621, sensitivity of 0.130) indicate moderate predictive
Zengqi Yu et al. [16]	Cancer classification models using feature selection, regression, backpropagation neural networks; Construction of cancer classification models	- Laser-induced Breakdown Spectroscopy (LIBS) to record elemental fingerprints - Achieves cancer diagnosis sensitivity and specificity of 71.4 % and 86.5 %,	- spectral features, while informative, introduce complexity and highlight the challenge of selecting appropriate features - need for additional feature selection and optimization.
V V P Wibowo et al. [17]	KNN and SVM models for diagnosing ovarian cancer; KNN produces higher accuracy than SVM	- Utilizes KNN and SVM - Utilizes real-world data from Al Islam Bandung Hospital	- Detailed performance metrics and comparison lacking - limitations in the current study's dataset size. - Does not provide detailed information about the dataset attributes
Jun Ma et al. [18]	Various modeling techniques (Random Forests, SVMs, GBs, Neural Networks) on blood biomarkers; Random Forests outperform traditional regression	- Random Forest, in predicting clinical parameters and providing risk stratification for EOC patients. - Incorporates survival analysis using Kaplan-Meier method	- Limited details on the size of the study - Limited generalizability to a specific cohort of 156 EOC patients. - Lacks explicit details on clinical integration and external validation
Sung-Bae Cho and colleagues [19]	Various classifiers for gene identification from noisy microarrays; Best accuracy obtained with multiple classifiers	- Addresses noisy microarrays - highlighting top performers such as Information gain and Pearson's correlation coefficient.	- Dataset-specific, may lack generalizability - Limited validation of ensemble results
Rui Xu et al. [20]	Gene expression data for predicting patient outcomes using PSO and Probabilistic Neural Networks; PSO/PNN predicts patient survival accurately	- Applies PSO and PNN	- Limited details on B-cell lymphoma - Relies on a specific dataset
Piroznia et al. [21]	Classification methods (SVMs, Radial Basis Functions, Multilayer Perceptrons, DTs, RFs) for expression data; SVM-RFE feature selection efficient for significant gene identification	- Compares multiple classification methods	- Limited details on specific methods used
Barnali Sahu et al. [22]	SVM, k-NN, and PNN with PSO for classifying high-dimensional cancer microarray data; PSO improves classification with SVM, k-NN, and PNN	- Utilizes k-means clustering in the initial stage - Employs well-established classifiers (SVM, k-NN, PNN) and a robust validation approach (leave one out cross-validation). - Successfully applied to Leukemia, DLBCL, and	- Lack of detailed exploration on the biological significance

Table 1 (*continued*)

Author	Methodology & Key Findings	Strengths	Weaknesses
C. Gunavathi et al. [23]	Informative gene selection using swarm intelligence techniques; SFLLF feature selection with k-NN produces best classification	Breast cancer datasets, showcasing its effectiveness across diverse cancer types. - Utilizes multiple swarm intelligence (SI) techniques (PSO, CS, SFL, SFLLF) - Applies the proposed method to 10 diverse benchmark datasets - Results indicate the superiority of SFLLF over PSO, CS, and SFL in conjunction with the k-NN classifier.	- Limited insight into the biological relevance - The impact of the proposed SFLLF algorithm on computational efficiency and scalability is not explicitly discussed.
P. Yasodha et al. [24]	Feature selection and classification for early detection of ovarian cancer; Hybrid particle genetic swarm optimization for feature selection	- Integrates feature selection and classification methods - Utilizes Rough Set Theory for data reduction - Introduces a Hybrid Particle Genetic Swarm Optimization (PGSO) method	- Limited details on the specific feature selection criteria - does not discuss the clinical implications or the impact on patient outcomes.
Hiba Asri et al. [25]	ML algorithms (SVM, C4.5, NB, KNN) on WBCD's dataset for breast cancer modeling; SVM achieves the most accurate accuracy	- Utilizes various ML algorithms - Uses the Wisconsin Breast Cancer dataset - employs the WEKA data mining tool	- lacks a detailed discussion on the specific characteristics of the dataset - SVM is identified as the best-performing algorithm but not justified - does not explicitly address the interpretability of the models
Madeeh Nayer Elgedawy [26]	ML techniques (Naive Bayes, SVM, Regression) for breast cancer classification; RF algorithm performs the best	- Compares multiple ML techniques	- Dataset-specific, may lack generalizability
Min-Wei Huang et al. [27]	SVMs and SVM ensembles for breast cancer classification; Ensemble of RBF kernels and SVMs outperforms other classifiers	- Utilizes SVM ensembles - Utilizes both small and large-scale datasets - Identifies specific SVM ensembles (GA + linear SVM ensembles with bagging and GA + RBF SVM ensembles with boosting) as top-performing models	- Limited details on dataset and comparison - lacks detailed insights into the interpretability of the SVM and SVM ensemble models
Urun Dogan et al. [28]	Multi-class SVMs for multiple datasets; Mathematical explanation of multiclass loss function	-Provides mathematical explanation - Comprehensive unification of multi-class SVMs	- Limited practical guidelines for real-world applications. - Complexity in margin and aggregation operator discussions.
M. Amer et al. [29]	Unsupervised and supervised ML for breast cancer data; Enhanced one-class SVM performs better	- Compares unsupervised and supervised ML -Successful modification of one-class SVMs (Robust and eta variants) to enhance sensitivity to anomalies. - Superior performance on two out of four datasets	- does not discuss potential limitations of SVM-based algorithms. - The computational effort for parameter tuning with Gaussian kernel can be complex
Akazawa et al. [30]	Diverse set of ML models (SVM, Naive Bayes, Boost, Logistic Regression, Random Forests) for dataset analysis; Extreme gradient boost algorithm achieves significant accuracy improvement	- Inclusion of diverse classifiers such as support vector machine, random forest, naive Bayes, logistic regression, and XGBoost for a comprehensive analysis. - Utilization of 16 features commonly available from blood tests, patient background, and imaging tests, enhancing the practicality of the proposed AI model. - Recognition of XGBoost as the machine learning algorithm with the highest accuracy (80 %) - Insightful analysis of the importance of the 16 features, comparing results among correlation coefficient, regression coefficient, and features importance of random forest.	- Correlation between sensitivity and feature set size - Limited dataset size (202 patients) - other performance metrics like precision, recall, or F1-score are not presented
Mintu Pal et al. [32]	Multi-biomarker panels for early detection of ovarian cancer; Combining biomarkers improves sensitivity and specificity	- Recognition of the limitations of CA-125 as a specific biomarker, emphasizing the need for a panel of biomarkers. - including CA-125, CA 19–9, EGFR, G-CSF, Eotaxin, IL-2R, cVCAM, MIF, to improve sensitivity and specificity	- Limited discussion on specific challenges faced in the collaboration between different research units for biomarker panel selection. - The review lacks detailed information on the specific characteristics and performance metrics of the proposed biomarker panels. - While mentioning the challenges, there is a lack of specific recommendations or strategies to address those challenges.

better.

Using data mining techniques on a large amount of data, P. Yasodha et al. [24] were able to discover valuable knowledge from the data. Using a rough set theory for the purpose of detecting the reliance of a data set on certain variables in order to reduce the number

of features in the data set is one way to uncover the reliance of the data set. It has been shown that a hybrid particle genetic swarm optimization technique is suitable for optimizing distinct features identified within ovarian cancer across varying stages, and it has been found that this technique has been applied to optimize selected features of this disease. A multiclass SVM is employed to differentiate between normal ovarian tissue and various stages of ovarian cancer by utilizing an optimized feature set for classification purposes, which uses an optimized feature set to classify. This dataset was retrieved from the TCGA portal (www.tcga-data.nih.gov/) and contains a total of 12042 genes relating to ovarian cancer, of which there are 493 instances for each gene found in the dataset. Based on the results of the experimental studies, Multiclass SVM, ANN and Nave Bayes methods were able to achieve accuracy rates of 96 %, 93 % and 90 %, respectively, according to the results of the experiments. It was also used 172 instances of blood tests from the NUH Singapore (http://www.nuh.com.sg/#) in the study. Each instance had 28 different features attributed to it from the NUH Singapore (http://www.nuh.com.sg/#). Based on their results, three different classes were tested: SVM, ANNs, and Nave Bayes. Their results showed 98 %, 95 %, and 93 % accuracy depending on the classifier.

According to Hiba Asri et al. [25], they used WBCD's dataset which contains 699 instances and 11 integer-valued attributes in order to model this dataset. For this dataset, they used four different The utilized ML algorithms comprise SVM (Support Vector Machines), C4.5, NB (Naive Bayes), and KNN (K-Nearest Neighbors). The WEKA data mining tool is a popular and widely used software suite in the field of ML and data mining. By using this tool, the results obtained by using SVM are the algorithm that is capable of obtaining the most accurate accuracy (97.13 %) and the least error rate in comparison to all alternative algorithms based on their accuracy and error rates. There were three types of ML techniques used by Madeeh Nayer Elgedawy [26]. These techniques included Nash Bayes, SVM and Regression-Fast regression techniques. As a result of the comparison of these algorithms, it is clear that the RF algorithm is the most effective and appropriate algorithm, as it can achieve a peak accuracy of 99.42 %, whereas SVM and NB yield lower rates at 98.8 % and 98.24 %, respectively. A study conducted by Min-Wei Huang et al. [27] used SVMs and SVM ensembles to classify breast cancer data from a large dataset. Furthermore, the team applied a kernel-based SVM based on RBF kernels to help predict the accuracy from training data to build a model, which will be able to evaluate the accuracy as a result of the F-measure, ROC curve, etc. on each of the training observations. It further applied the boosting technique as well as a kernel-based SVM based on RBF kernels in order to achieve the same objective. The results of the experiments revealed that, based on the boosting method, the ensemble of RBF kernels and SVMs outperformed other types of classifiers as far as accuracy was concerned.

The research demonstrated by Urun Dogan et al. [28] the use of multi-class SVMs for multiple datasets in one unified view. The mathematicians assisted with a mathematical explanation of all components of a multiclass loss function, such as the margin function, the aggregator operator, and the fisher consistency when applied to a multiclass loss function. As an additional benefit, the data from linear SVM as well as non-linear SVM is also tabulated so that you can evaluate their accuracy as well. A study published by M. Amer et al. [29] demonstrated that either unsupervised or supervised ML techniques could produce better classifiers when applied to breast cancer data. Unsupervised ML is not dependent on the use of a training dataset for the training of the models; according to their findings, supervised ML requires a training dataset to be used for the training of the models [30]. To detect anomalies in the dataset, one-class SVMs with enhanced attributes are applied as well as one-class SVMs that are robustly applied. As a result of our analysis, we find that enhanced one-class SVM performs better than robust one-class SVM in terms of performance.

In their study involving the analysis of a specific dataset using ML techniques, Akazawa et al. [31] employed a diverse set of ML models. These models encompassed a range of methodologies such as Support Vector Machines, Naive Bayes, Boost, Logistic Regression, and Random Forests to conduct a comprehensive analysis of the dataset. Compared with the other competing algorithms, there was a significant increment in accuracy scores achieved by the extreme gradient boost algorithm, with a score of approximately 80 %, when it was compared with the other similar algorithms. This study found, however, that the sensitivity of the model was strongly correlated with the size of the feature set, and that this correlation was statistically significant. There is a noticeable decrease in accuracy as the number of features decreases, amounting to approximately 60 % of the accuracy reduces as the number of features decreases. In addition, the small number of blood parameters included in this study was another disadvantage of the study.

Mintu Pal et al. [32] CA-125, the gold standard tumor marker, exhibits high levels in ovarian cancer but lacks specificity for early-stage screening due to high levels in non-cancerous conditions. HE4, with FDA approval, is a prospective single serum biomarker but with low sensitivity and specificity. Combining CA-125 with other biomarkers like CA 19–9, EGFR, G-CSF, Eotaxin, IL-2R, cVCAM, and MIF has shown improved sensitivity (98.2 %) and specificity (98.7 %) in early-stage detection. A panel of biomarkers has been proposed to improve sensitivity and specificity for early detection, outperforming single biomarker assays. Current trends suggest the potential of multi-biomarker panels for prototype development and advanced approaches in the early diagnosis of ovarian cancer, aiming to avoid false diagnoses and excessive cost. Table 1 summarizes the related work systematically compared and represented.

Therefore, there is a need for an integrated framework that combines statistical analysis with ML in order to be able to identify individual patients with ovarian cancer in accordance with the biomarker features they exhibit and classify them accordingly. It has been well documented that many studies have been conducted on the diagnosis of ovarian cancer, but the accuracy of the methods are currently insufficient, indicating that further improvement would be beneficial. In addition, there is no standard method for classifying the data in any of these studies, regardless of the type of data in question, such as blood specimens, general chemical analysis, or biological markers found in OCs. As a first step, the data needs to be separated so that the process can begin. The analysis of data in our study was conducted using both statistical methods as well as ML methods, in contrast to prior studies that only used statistical methods. Through the use of this innovative methodology, new perspectives were added to the research, and clinical testing gained greater validity, which may ultimately benefit both patients and physicians. A hypothesized reason for the better performance of ensemble ML strategies in the diagnosis of ovarian cancer over individual ML algorithms is the hierarchical relationship between ovarian cancer biomarkers that lies at this algorithm's core. Following is a brief summary of the study's major contributions.

Overall, the related work showcases the effectiveness of intermediate ML techniques in ovarian cancer prediction. Additionally, the

studies emphasize the significance of feature selection and the identification of key risk factors in developing robust prediction models for ovarian cancer.

By building upon the findings of these studies and addressing their limitations, our research aims to contribute further to the field of ovarian cancer prediction using intermediate ML techniques. While the related studies consider reducing the number of features, the explainability of these results has not been adequately addressed to the best of our knowledge. Therefore,our research developed a novel approach to classification of OC based on the analysis of the explainability of the prediction model developed. ML models should be transparent and understandable, which is the primary motivation behind this research. Through the use of XAI, the research aims to facilitate informed choices about patient health for both healthcare professionals and patients.

3. Methodology

As part of this section, describe the methodology for the proposed system. This process begins with the pre-processing of the data from the datasets, which is the first step in this process, and filling in null values where necessary, as well as encoding the data in order to transform categorical variables into numerical values to be used in the analysis. It is then necessary to divide the data after pre-processing into two parts in order to separate both the test data and the train data [33]. In order to predict the outcome of an ovarian cancer case and perform the prediction, the training data is passed through various ML approaches on the basis of a prediction model. In order to evaluate the overall performance of ML classifiers, cross-validation will be carried out on the classifiers, and the results will be evaluated again. An illustration of the proposed methodology for the prediction of Ovarian Cancer can be found in Fig. 1.

3.1. System model

Algorithm 1: The Working Procedure for Predicting Ovarian Cancer.
Input: Ovarian Cancer Dataset from Kaggle Output: Predicted Ovarian Cancer (Positive or Negative) with Model Interpretability.

1 Begin
2. data ← Load the input dataset from Kaggle containing Ovarian Cancer data.
3. **if** data.value **is** NaN or empty:
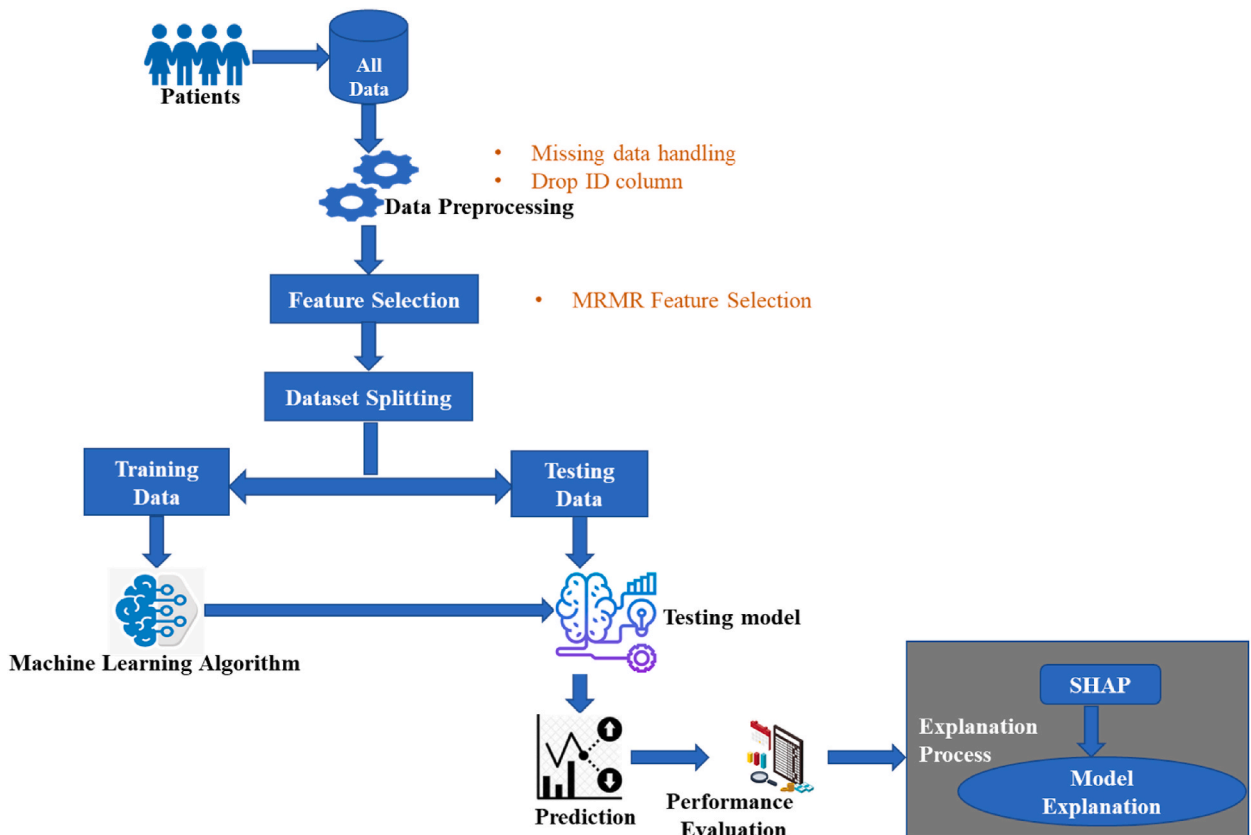4. Replace NaN or missing values in the dataset.

Fig. 1. Proposed Ovarian Cancer prediction.

5. Preprocessing:
6. For each column in the dataset:
7. **if** data.dtypes **is** equal to 'object':
8. Encode categorical data in the column.
9 Split data into features and target
10. **x** ← data.drop 'Ovarian Cancer' column from the dataset.
11. **y** ← data.'Ovarian Cancer' Target variable, specifically.
12 Split data for training and testing
13. **x_train, x_test, y_train, y_test** ← Split data (features and target) into training and testing sets.
14 Train model
15. **model** ← Train a model using **x_train** (training features) and **y_train** (training target).
16 Test model
17. **predict** ← Test the trained model using **x_test** (testing features).
18 Explainable AI (XAI) - SHAP values

Apply XAI techniques (e.g., SHAP values) to interpret the model predictions.

19. **explainer** ← Create a SHAP explainer for the trained model.
20. **shap_values** ← Compute SHAP values for **x_test.**
21. **Visualize** SHAP values or generate explanations to understand feature importance and model predictions.
22 Apply various classifiers
23 model ← Apply different classifiers using the trained model.
24 Compute performance evaluation metrics
25 Calculate and **analyze** performance metrics
26 End.

The detailed algorithm of the proposed work is elaborated in the above algorithm 1. Initially, the algorithm loads the Kaggle dataset associated with ovarian cancer, addresses missing values through data cleaning, and preprocesses categorical data. In order to train and evaluate the model, the dataset is divided into training and testing datasets, and the target variable is 'Ovarian Cancer.' An artificial intelligence model is trained based on the training set of features, and its predictions are then tested against a set of testing features. An XAI technique is used to visualize the importance of features and generate explanations for the model predictions in order to enhance interpretability. Besides this, the algorithm explores the use of different classifiers based on the trained model. The performance evaluation metrics are then computed to evaluate the model accuracy, precision, recall, and other relevant measures, thus providing a comprehensive workflow for predicting ovarian cancer.

3.2. Dataset description

In this study, data from Kaggle was collected, comprising 349 instances and 51 features [8]. There are two types of outcomes in the type column, 0 indicates that there was no ovarian cancer detected, and 1 indicates that ovarian cancer was found. The data in this dataset consists of 171 cases of non-ovarian cancer and 176 cases of ovarian cancer. The data inclusion criteria were defined based on age range, histological types, and the availability of relevant features. Only high-quality and well-annotated datasets were included to ensure the reliability of the analysis. An age range of 23–68 was selected for the collection of data. There were a total of 234 post-menopausal patients among the 349 patients studied. Various blood routine examinations, protein analyses, and general chemistry tests were included in the dataset. These attributes encompassed a range of factors, including Alpha Feto Protein (AFP), Albumin Globulin Ratio (AG), Age (Years), Albumin being a gene that codes for a protein (ALB), Alkaline Phosphate (ALP), Alkaline Transminase, which is an enzyme that can be found in the liver (ALT), Asparate Aminotransferase, an enzyme primarily found in the liver but also present in all other organs (AST), Basophils, a specific type of white blood cell (BASO), Blood Urea Nitrogen (BUN), Level of Calcium in Blood (Ca), and more.

3.3. Data wrangling

In the context of data wrangling, also known as data munging or data cleansing, it is the process of transforming and preparing raw data into a structured format that is suitable for analysis in the context of a database. There are several tasks involved in this process, including cleaning, transforming, combining, and aggregating data to ensure its quality and usability. Data wrangling plays an essential role in preparing data for interpretation and is often considered one of the most time-consuming and challenging steps in the data analysis workflow [34]. In this process, column types are changed to the desired type, the ratio of missing data in each column is calculated, and the missing data is imputed with a median value. Finally, the ID column is eliminated by using Drop ID.

3.4. Feature engineering

In the process of feature engineering, new features are created or existing features are transformed so that ML models can perform

better on datasets in the future. An essential part of this process is obtaining relevant information from raw data and presenting it in a more meaningful and useful manner for model training purposes. An important aspect of feature engineering is that it captures the underlying patterns, relationships, and characteristics of data that have not been explicitly encoded in the original features so that these patterns are discovered as well as improved. Then, by enabling ML models to uncover informative features, they will be able to make better predictions, and they will be able to classify data more accurately through a much better understanding of the data. As a result of the specific dataset, the problem at hand, and the algorithms used, the choice of feature engineering techniques may differ from dataset to dataset. It often requires a deep understanding of the data and domain knowledge. Feature engineering is iterative, and experiments and validation should be done to assess how effective engineered features are. In this research, the MRMR feature selection algorithm is used. A feature selection algorithm called MRMR is one of the most widely used in data mining. MRMR is the ideal algorithm for practical ML applications where feature selection is performed frequently and automatically in a short period of time due to its high efficiency. As a result of the MRMR feature selection algorithm, 18 features are selected. The selected features are 'Age', 'CEA', 'IBIL', 'NEU', 'Menopause', 'CA125', 'ALB', 'HE4', 'GLO', 'LYM%', 'AST', 'PLT', 'HGB', 'ALP', 'LYM#', 'PCT', 'Ca', 'CA19-9'.
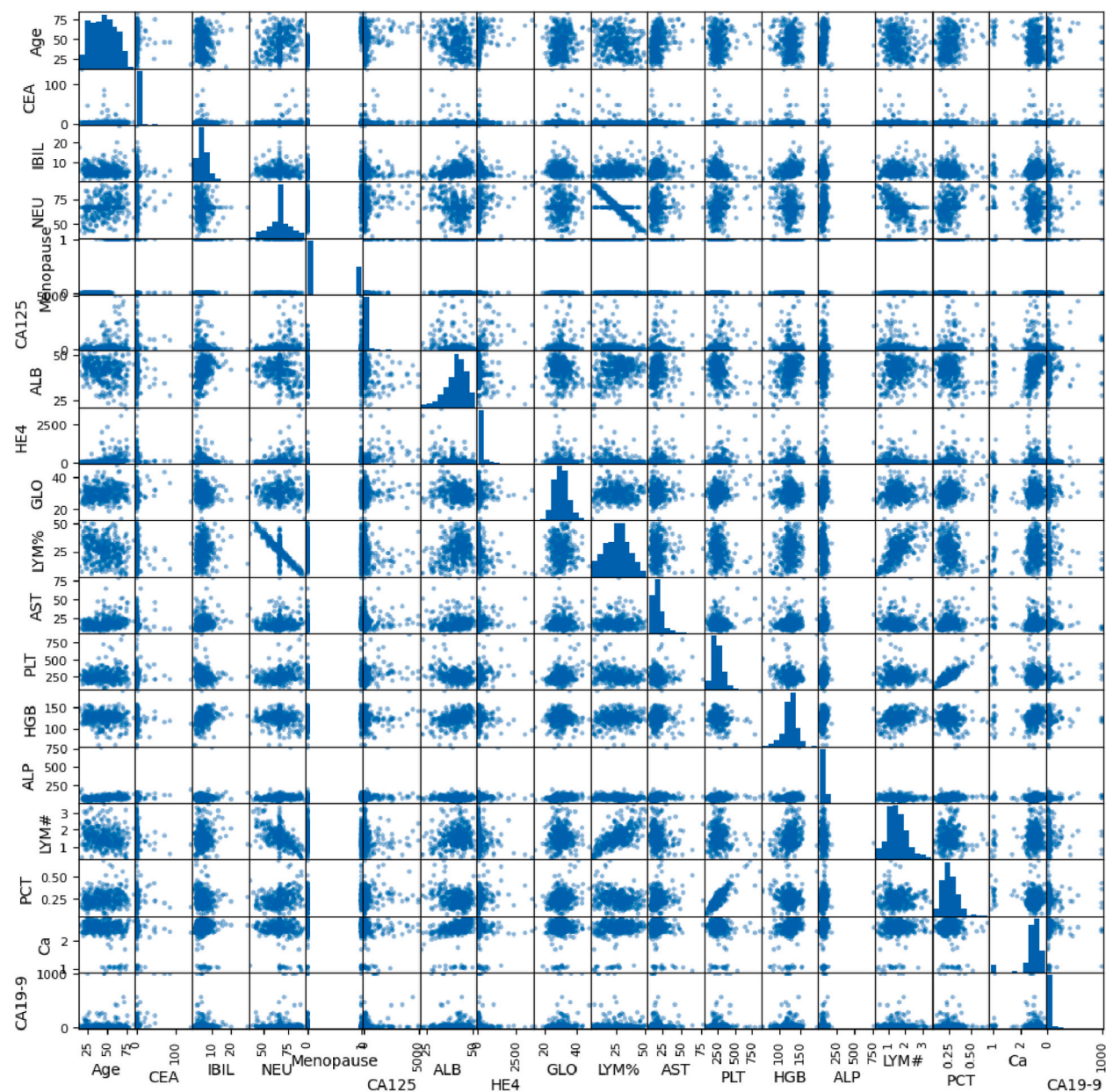

Fig. 2. Visualization of different data.

3.5. Preliminary data analysis

This section of the code involves an initial examination of the dataset to understand its structure, contents, and quality. This includes checking for the extent of missing or erroneous data, exploring the distribution of the variables, identifying any outliers, and computing summary statistics. The purpose of preliminary data analysis is to gain insights into the dataset and inform subsequent data processing steps. It also involves visualizing the data using various plotting techniques to reveal patterns or relationships between the variables. Fig. 2 represents the visualization of data in the dataset. In Fig. 3, the target variable distributions are visualized based on the training data in relation to the target variable.

3.6. Correlation matrix with heatmap

Correlation matrices with heatmaps provide a visual representation of the pair-wise relationships between variables in a dataset with respect to a pairwise correlation matrix. The relationship between variables in the model provides insights into how strongly and in what direction they are related. Fig. 4 displays the heat map and correlation matrix for the dataset utilized in the research. The correlation value in this case is open to any number between −1 and 1. The relationships between variables can be plainly seen through the color-coding of the cells.

3.7. Model selection and training

A number of machine learning approaches were considered, in particular support vector machines, Decision Tree models, and K nearest neighbor algorithms as base models in this study, and also Max voting, Stacking, Bagging, Boosting, and Stacking of various ensemble learning techniques were used. The models were trained using appropriate training algorithms and produced optimal performance.

3.8. Model evaluation

As a means of evaluating whether the models are capable of predicting events in the future, multiple performance metrics were applied. A number of metrics were calculated to assess different aspects of how well the model performed, such as AUC-ROC, F1 score, recall, accuracy, precision, and specificity are used to determine the model performance [35].

3.9. Interpretation and analysis

The implications of the model analyses and evaluations were carefully analyzed and interpreted. Key findings, including performance metrics and statistical analyses, were presented. The implications of the findings for ovarian cancer prediction were discussed, highlighting the potential clinical or biological significance of the identified features and the models' predictive capabilities.

3.10. Machine learning algorithms

3.10.1. SVM (support vector machine)

Statistical learning algorithms such as SVM have been widely applied in the healthcare field for several years for making predictions about results as well as for problems related to classification and regression [36,37]. By using this classification method, the dataset is
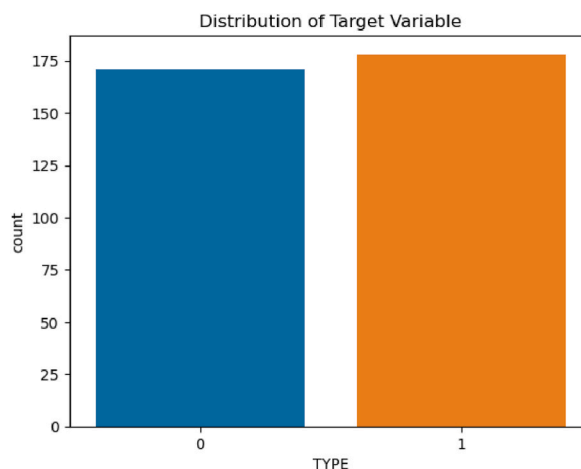

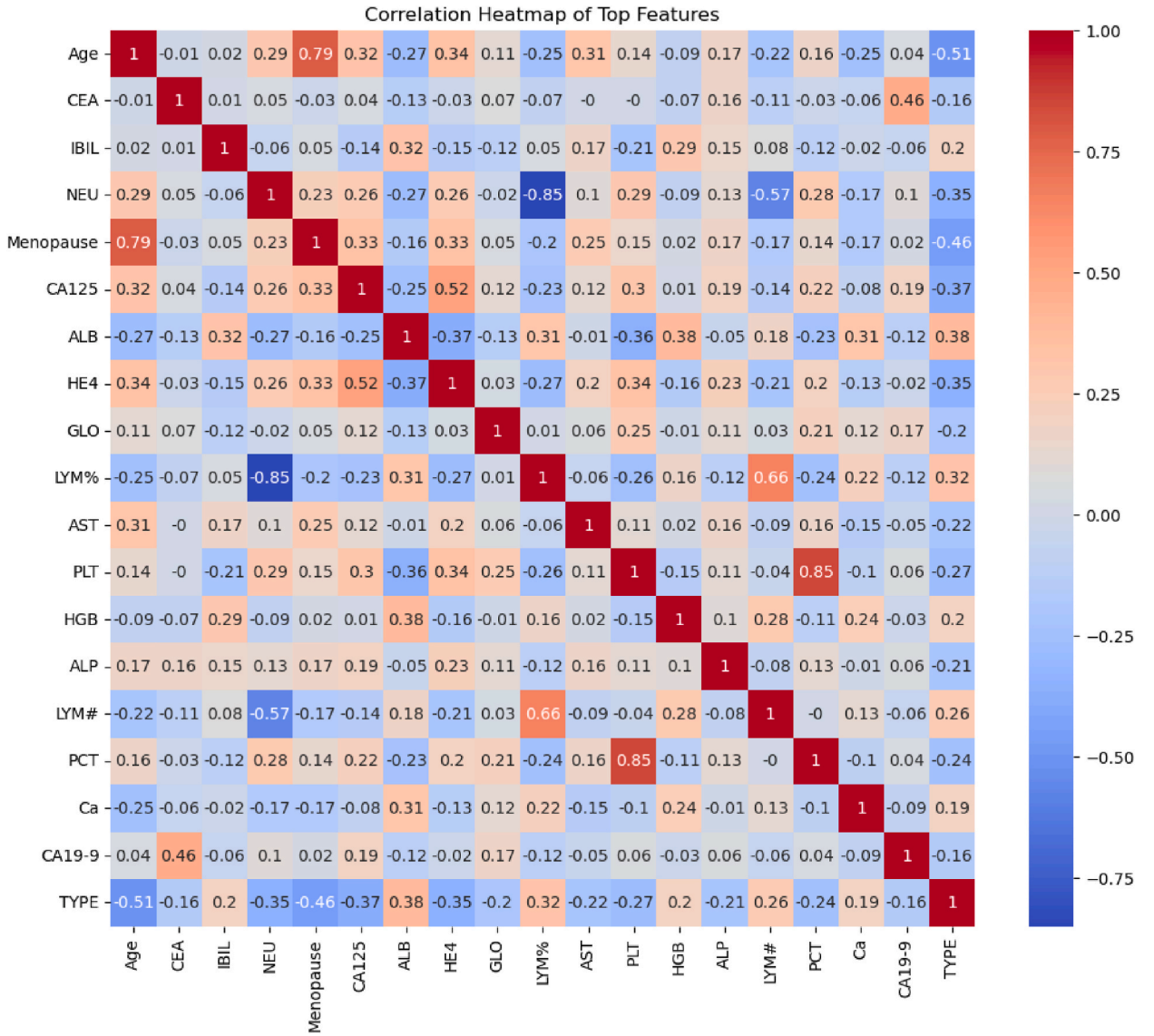**Fig. 3.** Distribution of target variable.

Fig. 4. Correlation heatmap of top features.

classified using the most appropriate hyperplane that distinguishes it as one of two classifications [38]. Eq. (1) represents a linear combination of a kernel with a SVM, where pj represents patterns of the training set, S is the SVM, and $qj\ \varepsilon\{+1,-1\}$ is the corresponding class label.

$$f(p) = \sum pj\ \varepsilon S\ \boldsymbol{\alpha} j\ qjK(pj,p) + b \tag{1}$$

3.10.1.1. K-nearest neighbor. One of the simplest and fastest algorithms is the K-Nearest Neighbor algorithm for finding similar features in the training dataset based on the test data [39]. There are various distance algorithms that are commonly used to determine the distance between new data and all of the previous data. They are the Manhattan, Euclidian, and Minkowski distance algorithms [40]. The formulas outlined in Eqs. (2)–(4) represent the formulas to calculate the Euclidian, Manhattan, and Minkowski distances, respectively.

$$Euclidean = \sqrt{\sum_{i=1}^{k} (x_i - y_i)2} \tag{2}$$
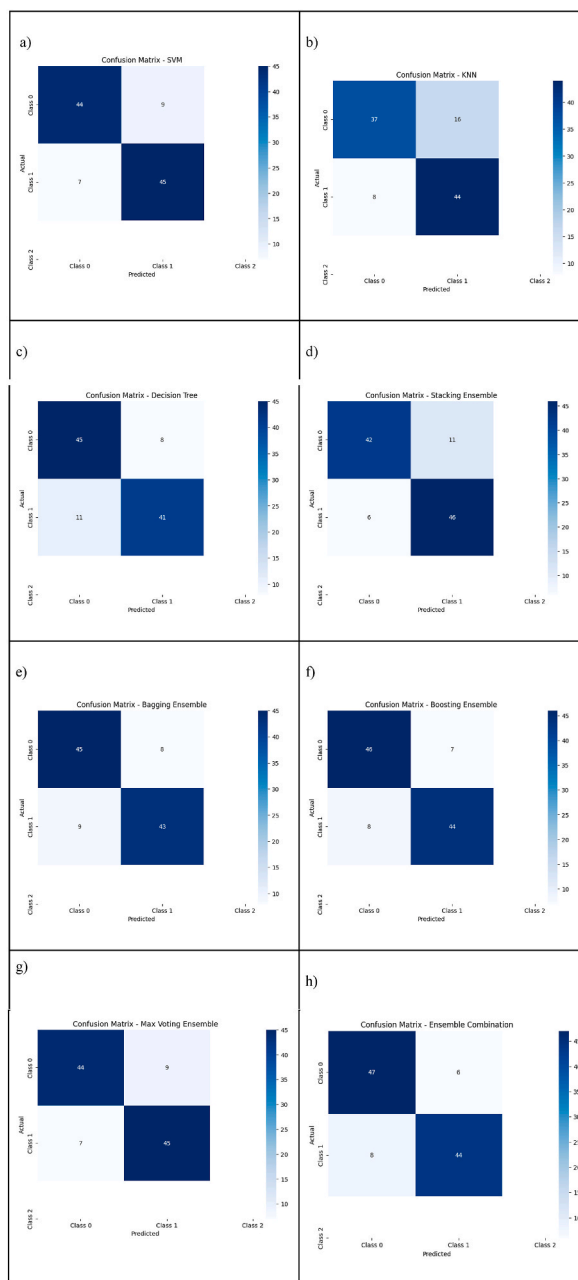
Fig. 5. Confusion matrix. a) Confusion matrix of SVM; b) Confusion Matrix of KNN; c) Confusion Matrix of DT; d) Confusion Matrix of Stacking Ensemble; e) Confusion Matrix of Bagging Ensemble; f) Confusion Matrix of Boosting Ensemble; g) Confusion Matrix of Max Voting Ensemble; h) Confusion Matrix of Ensemble Combination.

$$Manhattan = \sum_{i=1}^{k} |x_i - y_i| \tag{3}$$

$$Minkowski = \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{4}$$

3) DT (Decision tree)

DT is not a parametric supervised learning classifier and supervised learning classifier. Decision trees are constructed of a collection of internal nodes that are necessary for making the decisions, and leaf nodes are used to represent the final outcome of the decision. In order to design a node of the DT, we use Eq. (5) [41].

$$E^{''}(d_i) = [T_0(di) - T(d_i)] + K[\varepsilon_0(d_i) - \epsilon(d_i)] + \sum_{j=1}^{c_i} p_{i+j} E_0 (d_{i+j}) \tag{5}$$

A heuristic search function is applied to a node in order to get the correct decision for this node based on the values provided by $E^{''}$. It is the evaluation function for the node.

4. Results and discussion

As part of the testing and training process, the proposed system had been trained and tested on data from 20 % to 80 %, respectively. A variety of ML classifiers have been used to evaluate the best model, including: DT, SVM, K-NN, MVC, GB, XGB, and Stacking. The best ML model is found by combining the performance of these classifiers, including such measures as accuracy, recall, precision, and the F1-score, which are all used to determine the best. An approach of k-fold straitified cross-validation was used to evaluate the performance of each classifier. Five-fold stratified cross-validation involves dividing the dataset into five subsets and maintaining the original class distribution at each fold. As a consequence of this method, each fold is guaranteed to contain a representative proportion of instances from each class, thereby eliminating imbalanced class scenarios. During each iteration, a new fold serves as the test set and the remaining folds serve as the training set. To provide a comprehensive and reliable evaluation of the model performance, particularly when there is an imbalance in the distribution of classes, performance metrics are computed for each run, and the results are averaged over the five folds. When dealing with uneven class frequencies, stratified cross-validation prevents biases that may arise from standard cross-validation.

4.1. Confusion matrix

The confusion matrix is a matrix that has N x N columns that is used for performing measurements for ML classifications, where N is the number of target classes you wish to identify. By using this method, the number of correct and incorrect predictions is calculated and compared to determine which ML classifiers perform best when it comes to making predictions [42]. The classifiers used in this study have been evaluated and compared with the other classifiers considered in this study in this section. It is shown in Fig. 11 how each classifier performs in terms of its performance.

As a result of the confusion matrix, as shown in Fig. 5 above each classification has four outcomes that indicate how well each classifier performs independently on the positive and negative classes; for individual classifiers, two types of predictions are correct, and two types are incorrect, encompassing true positives (TPs), true negatives (TNs), false positives (FPs), and false-negatives (FNs). As a next step, each classifier is analyzed according to a set of metrics found in below section.

4.2. Performance metrics

When evaluating the performance of a ML model for ovarian cancer prediction, a variety of performance metrics are used most commonly. A predictive model can be evaluated based on these metrics, which provide insights into its effectiveness and reliability. Here are some performance metrics that can be considered for assessing the performance of an ovarian cancer prediction model:

Accuracy: A measure of accuracy can be calculated by dividing the number of correctly predicted instances by the total number of instances. As a result it provides an overall assessment of how accurately the model is able to classify cases of ovarian cancer in comparison with other models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Prediction precision refers to how many true positive predictions are made out of the total number of positive predictions. It gives an indication of the model ability to correctly identify cases of ovarian cancer in the absence of falsely classifying healthy cases as positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity): A recall calculation is a method of calculating the proportion of true positive predictions out of all actual instances of positive predictions. A test is performed to determine if the model can identify all of the ovarian cancer cases within the dataset correctly and to what extent.

$$Recall = \frac{TP}{TP + FN}$$

Specificity: Specificity measures the proportion of true negative predictions out of all actual negative instances. It evaluates the model ability to correctly classify healthy cases as negative.

$$Specificity = \frac{TN}{TN + FP}$$

F1 Score: An F1 score can be described as the harmonic mean of precision and recall in a given situation. By evaluating both precision and recall at the same time, it provides a balanced measure. The F1 score can be useful if there is an imbalance in the positive and negative cases within a dataset.

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

There are four parameters in this equation, TP, TN, FP, and FN, which represent the number of true positives, true negatives, false positives, and false negatives, respectively.

Using the above performance metrics, the performance of each classifier is evaluated. It is observed that SVM with 85 % accuracy is the highest accuracy that belongs to the base model. After combining all ensemble learning algorithms, the acquired accuracy is 89 %.

5. Analysis of ensemble methods

By using appropriate evaluation metrics and comparing the results of the ensemble learning methods used in the model training section, the performance of the ensemble learning methods is evaluated in this section of the code. The objective of this study is to identify the most effective and efficient method for detecting ovarian cancer tumors and to identify the factors contributing to a particular ensemble learning technique performing better than others. Fig. 6 shows the visual representation of the number of features vs accuracy. This plot visually illustrates the impact of feature dimensionality on the model accuracy, aiding in identifying an optimal subset of features that strikes a balance between predictive performance and model simplicity. This analysis contributes valuable insights into the interplay between feature selection, model complexity, and accuracy, offering guidance for refining the decision-making process in cancer outcome prediction using machine learning techniques.
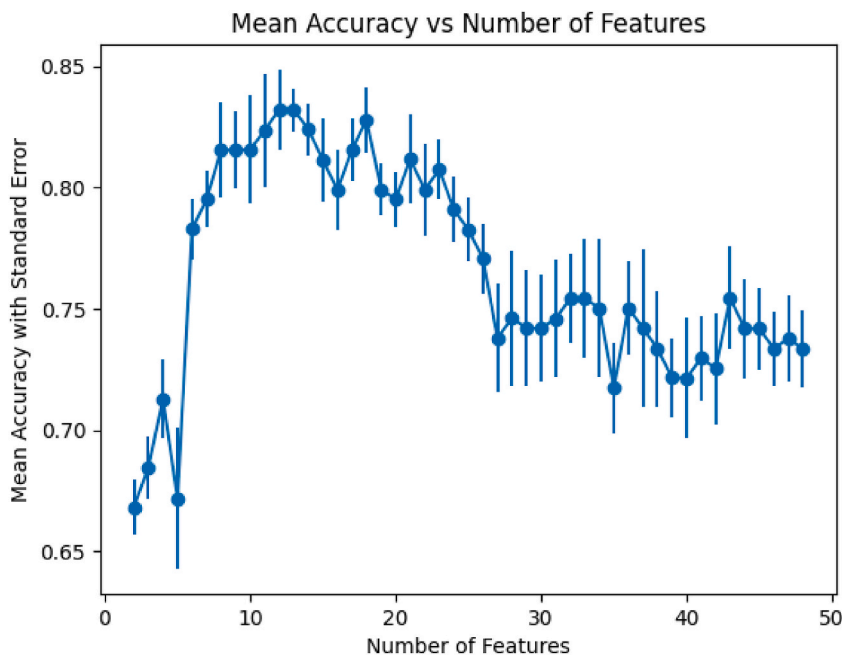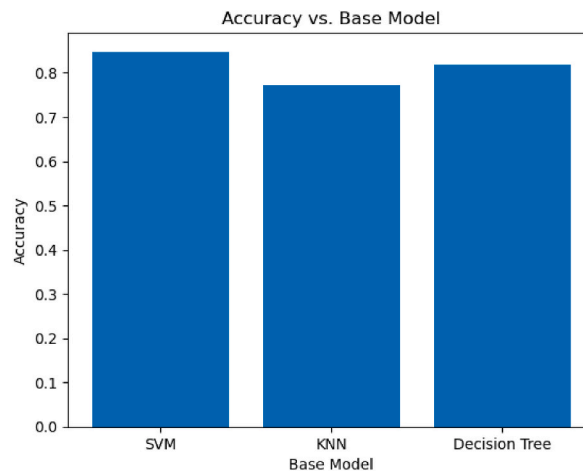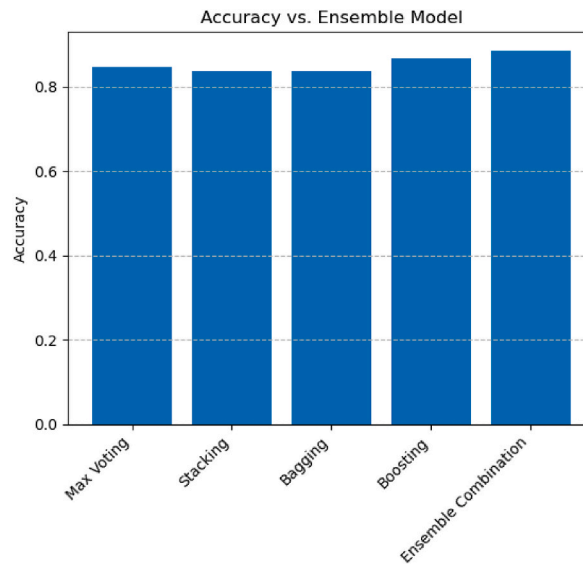

Fig. 6. Number of features vs mean Accuracy.

Fig. 7. Accuracy of base models.

Fig. 8. Accuracy of ensemble learning methods.

In terms of performance, the SVM algorithm performs the best out of the three algorithms tested with 85 % accuracy. This evaluation process used accuracy as the principal evaluation metric for comparison, which can be calculated as the proportion of correctly classified instances in a dataset. As indicated in Fig. 7 below, the accuracy of each of the base models is shown.

While the boosting method outperformed the baseline models, it was discovered that combining these methods with one another can lead to even higher performance than the boosting method because a combined ensemble method shows even better results than the baseline models. The below Fig. 8 shows the accuracy of all ensemble methods like max voting, bagging, boosting, stacking, and combinations of ensemble methods. The combination of ensemble methods provides an accuracy of 89 % is the highest accuracy.

5.1. Model interpretation

5.1.1. Explainable AI

This section discusses the techniques and methods that can be used to make ML models more intuitive to humans in order to make them more understandable [43]. Being human-centred and highly case-dependent, explainability is hard to capture by mathematical formulas. It is highly subjective, and the utility depends on the application [44]. The most active area of XAI research focuses on post-hoc explanations [36,37] that provide partial quantification of this requirement. A variety of visualization methods have also been developed to assist users [38]. It involves analyzing the model inner workings and providing insights into how it makes predictions, allowing developers and stakeholders to understand and trust the model output [45]. By incorporating XAI techniques into

the model, developers can ensure that it is not only accurate but also trustworthy and reliable.

5.1.1.1. XAI techniques used.

1. Feature Importance:
   - *eli5* permutation importance
   - SHAP permutation importance

2. Dependence Plots:
   - Partial Dependence Plot.
   - SHAP Dependence Contribution Plot

3. SHAP Values Force Plot

6. Feature importance

6.1. eli5 permutation importance

Permutation importance is a method used to calculate the importance of each feature in a ML model. It involves permuting the values of a single feature in the dataset and then observing the effect on the model accuracy. The feature that causes the largest decrease in accuracy when permuted is considered the most important feature. Table 2 represents the feature importance of the most important features.

Shapley values, mentioned below Fig. 9, can be used to determine the relative marginal contribution of a variable towards an individual decision by a ML model. An Explainable AI (XAI) method called SHAP (SHapley Additive Explanations) assesses how individual features affect a machine learning model predictions. First, the model is evaluated against the original dataset to determine its baseline performance. As a result, each feature is evaluated by randomly permuting its value, reevaluating the model performance, and assessing how it differs from the baseline. A performance drop that is larger indicates more influential features, and a score that is negative indicates a possible reliance on other features. Feature importance scores are generated for each feature, allowing the model to identify the relative importance of each feature in influencing predictions. As a model-agnostic method, the approach provides a clear indicator of the impact of every feature on the overall performance of the model, thus improving its interpretability.These explanations do not, however, relate to validity of the model or have semantics themselves. They are only meaningful when combined with medical knowledge, where invalid contributions could also be observed by clinicians. This is also particularly relevant to personalised medicine [46]. Shapley based approaches are becoming one of the most widely used explanation techniques and have seen application in many domains [39,47], including healthcare [42].

7. Dependence plots

Dependence plots in Fig. 10 below show how a single input feature affects the predicted output of a ML model while holding all other features constant. They provide insights into non-linear relationships and help to interpret the model. XAI uses dependency plots

Table 2
Feature importance using permutation importance.
B.SHAP permutation importance

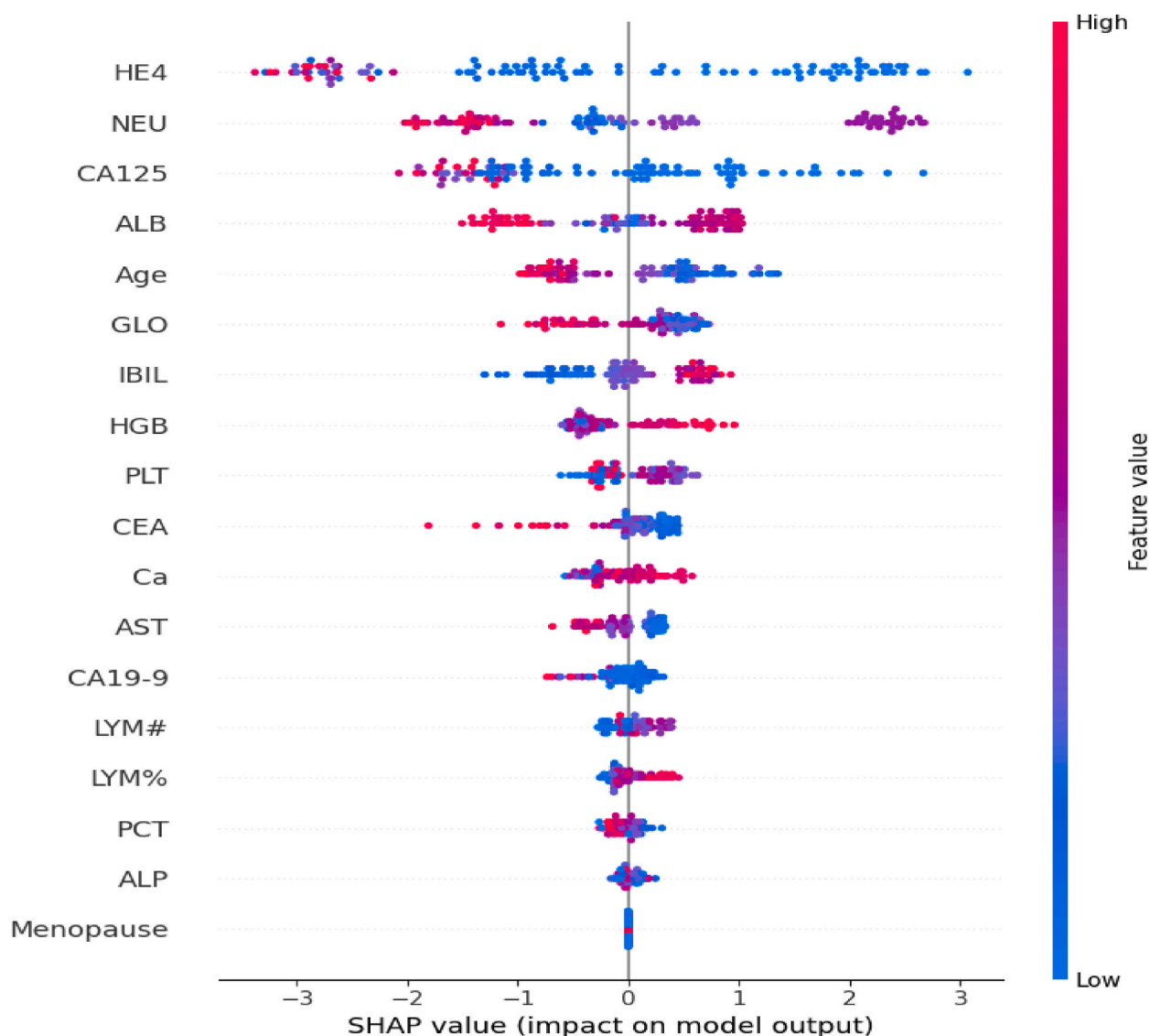| Weight	Feature
0.1695 ± 0.0349	HE4
0.0800 ± 0.0748	NEU
0.0343 ± 0.0373	Age
0.0190 ± 0.0381	CA125
0.0095 ± 0.0120	LYM%
0.0076 ± 0.0076	CA19-9
0.0076 ± 0.0143	ALB
0.0038 ± 0.0093	PCT
0.0019 ± 0.0305	HGB
0.0000 ± 0.0120	AST
0.0000 ± 0.0120	GLO
0 ± 0.0000	Menopause
−0.0019 ± 0.0076	ALP
−0.0019 ± 0.0076	LYM#
−0.0038 ± 0.0194	CEA
−0.0038 ± 0.0093	PLT
−0.0057 ± 0.0194	IBIL
−0.0076 ± 0.0076	Ca

Fig. 9. SHAP permutation importance.

to visualize how a particular feature is related to a model predictions. The x-axis displays the value distribution of the feature, whereas the y-axis displays the model predictions. Based on the feature value and model prediction, each data point represents an instance within the dataset. Using a trend line or fitted curve, it is possible to identify trends or non-linear patterns in the data. A variable spread of data points around the trend line illustrates the uncertainty of the model predictions. In AI decision-making, dependency plots provide an intuitive representation of how changes in particular features influence outcomes, which enhances transparency and interpretability.

Fig. 11 below represents the SHAP dependence contribution plot of age. Explainable AI provides a SHAP (SHapley Additive Explanations) dependence contribution plot which demonstrates the relationship between a chosen feature and the predictions of the model. Each feature value is represented on the x-axis and its associated SHAP value is presented on the y-axis, indicating how its value affects the model output. A color gradient represents the value of another conditioned feature for each data point, which aids in visualizing potential interactions between them. Trend lines provide an overview of the average relationship between a feature and the model output, thereby aiding in the identification of patterns. Generally, a wide range of deviations from the trend line indicate that the feature has a stable impact on predictions, while a wide range of deviations indicates that the feature has a higher variability of impact. Generally, SHAP dependence plots facilitate the interpretation of the model decision-making process by providing a comprehensive overview of how individual features contribute to the model predictions.
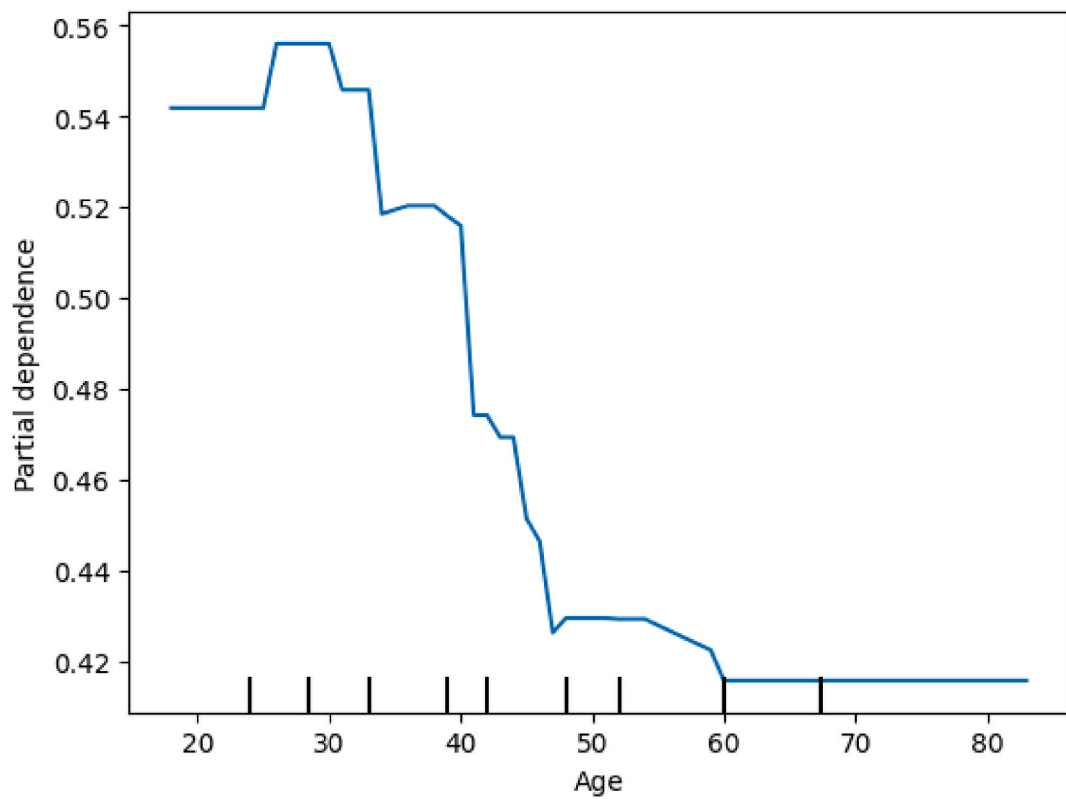
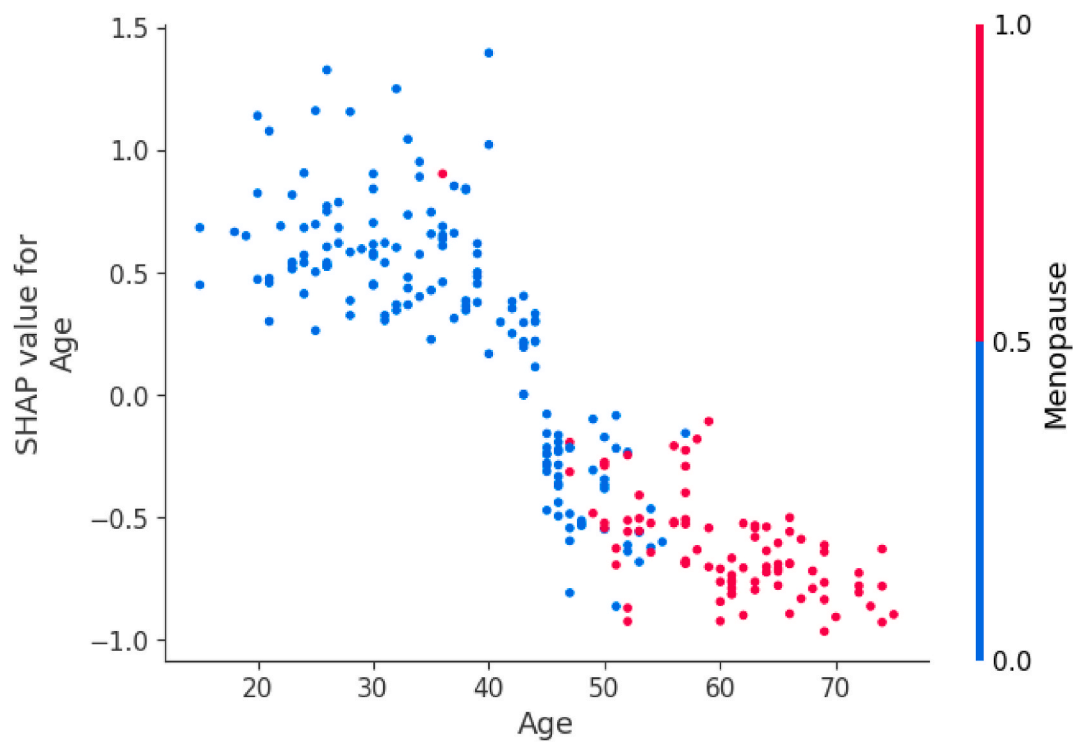**Fig. 10.** Dependence plot of age.

Fig. 11. SHAP dependence contribution plot.

7.2. SHAP values force plot

As part of Explainable AI (XAI), the SHAP (SHapley Additive Explanations) Values Force Plot serves as a visual aid for interpreting machine learning predictions. The model begins with a base value that represents the average prediction, which is then pushed away from by each feature. Each arrow in the plot represents a contribution made by a particular feature, whose length and direction indicate its magnitude and direction of impact. The features which contribute to the prediction being higher appear above and those which contribute to the prediction being lower appear below. There is a numerical value associated with each feature arrow that quantifies its impact on the prediction. Based on the base value and individual contributions to the instance, the final prediction is calculated. It provides a clear and concise picture of how each feature influences a particular prediction, thus increasing the transparency and interpretability of machine learning models.

Fig. 12 above illustrates the force plot for a random observation for the model we developed. According to the results of the correlation, it appears that the CA-125 is the factor that most significantly influences the prediction. This is followed by IBIL, HGB, IBIL, Ca, CA 19-9, PLT, LYM%, and PCT, in that order. In addition to these predictive characteristics, the model will score well (to the right), which will result in a score of 1 for the final prediction. The model predictive accuracy will be reduced to a greater or lesser extent depending on ALB, AST, and Age if they have less impact (to the left).

In summary, this study demonstrated a workflow from data collection to result evaluation through the use of classification and XAI. During the process, we discussed, explained, evaluated, and justified key decisions in the pipeline that helped make the result more explainable. To evaluate the results, we gave examples of some of the ways SHAP results can be interpreted and utilized in combination with domain knowledge. This pipeline was effective, general, and applicable to any similar tabular set of biomarkers.

8. Conclusion

The threat of ovarian cancer to the world is significant, emphasizing the need for timely treatment. The early detection of ovarian cancer is crucial to mitigating the severity of the disease, and machine learning (ML) models have emerged as critical tools for predicting the outcome of the disease. An ensemble machine learning classification method is employed in this study in order to advance early detection and prediction. As outlined in the proposed methodology, there are five steps: loading the Ovarian Cancer dataset, data preparation, feature selection, implementation of the classifier, and analysis of the SHAPley score. An impressive 89 % accuracy rate was achieved by the ensemble learning classifiers. The implications of this approach extend to both the computer and health sciences. In addition to providing clear guidelines regarding the adoption of machine learning methods in medical research, it addresses issues related to accountability. Collaboration between medical experts and computer scientists is instrumental in exploring novel disease biomarkers through explainable models. Enhanced accuracy will be achieved through the use of deep learning and image processing in future research. For the advancement of ovarian cancer treatment, it is imperative to integrate these techniques, optimize explainability methodologies such as SHAPley scores, and engage in interdisciplinary collaboration. Research efforts and wider evaluations using diverse datasets may lead to significant advancements in the treatment of ovarian cancer.

Data availability

Supplementary data to this article can be found online at Mendeley Data, V11, https://doi.org/10.17632/th7fztbrv9.11.

Code availability

All of the methods are implemented in Python. The original data and source code are available on GitHub page: https://github.
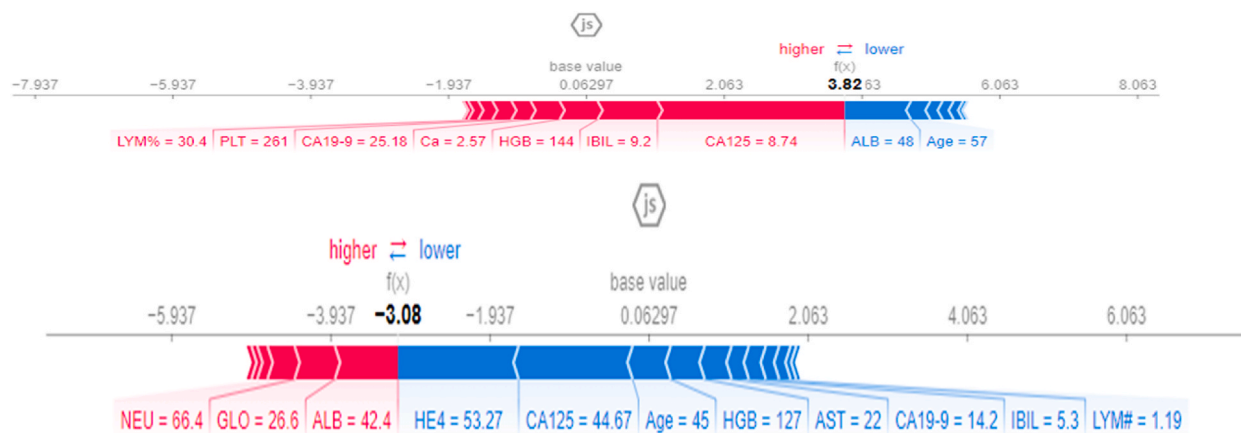

Fig. 12. SHAP Force plot.

com/sheelajeba/XAI.

Declarations

Human and animal ethics Not Applicable.
Ethics approval and consent to participate Not Applicable.

CRediT authorship contribution statement

Sheela Lavanya J M: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Subbulakshmi P:** Writing – review & editing, Validation, Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ratana Lim, Martha Lappas, Clyde Riley, Niels Borregaard, Holger J. Moller, Nuzhat Ahmed, Gregory E. Rice, Investigation of human cationic antimicrobial protein-18 (hCAP-18), lactoferrin and CD163 as potential biomarkers for ovarian cancer, J. Ovarian Res. (2013).

[2] M. Mathur, V. Jindal, G. Wadhwa, Detecting malignancy of ovarian tumour using convolutional neural network: a review, in: 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, 2020, November, pp. 351–356.

[3] Y. Zhang, S. Wang, K. Xia, Y. Jiang, P. Qian, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion, Inf. Fusion 66 (2021) 170–183.

[4] Y. Zhang, H. Ishibuchi, S. Wang, Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules, IEEE Trans. Fuzzy Syst. 26 (3) (2017) 1535–1549.

[5] A. Panigrahi, A. Pati, B. Sahu, M.N. Das, D.S.K. Nayak, G. Sahoo, S. Kant, En-MinWhale: an ensemble approach based on MRMR and Whale optimization for Cancer diagnosis, IEEE Access (2023).

[6] A. Almohimeed, R.M. Saad, S. Mostafa, N. El-Rashidy, S. Farag, A. Gaballah, M. Abd Elaziz, S. El-Sappagh, H. Saleh, Explainable artificial intelligence of multi-level stacking ensemble for detection of alzheimer's disease based on particle swarm optimization and the sub-scores of cognitive biomarkers, IEEE Access (2023).

[7] Y. Zhang, G. Wang, T. Zhou, X. Huang, S. Lam, J. Sheng, K.S. Choi, J. Cai, W. Ding, Takagi-Sugeno-Kang fuzzy system fusion: a survey at hierarchical, wide and stacked levels, Inf. Fusion 101 (2024) 101977.

[8] M. Lu, Z. Fan, B. Xu, L. Chen, X. Zheng, J. Li, T. Znati, Q. Mi, J. Jiang, Using machine learning to predict ovarian cancer, Int. J. Med. Inf. 141 (2020) 104195.

[9] E.L. Barber, R. Garg, C. Persenaire, M. Simon, Natural language processing with machine learning to predict outcomes after ovarian cancer surgery, Gynecol. Oncol. 160 (1) (2021) 182–186.

[10] E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, K. Koseki, Y. Iida, M. Saito, H. Komazaki, J.S. Shapiro, C. Goto, Y. Akiyama, Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers, Clin. Cancer Res. 25 (10) (2019) 3006–3015.

[11] E.S. Paik, J.W. Lee, J.Y. Park, J.H. Kim, M. Kim, T.J. Kim, S.W. Seo, Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods, Journal of gynecologic oncology 30 (4) (2019).

[12] K. Chen, H. Xu, Y. Lei, P. Lio, Y. Li, H. Guo, M. Ali Moni, Integration and interplay of machine learning and bioinformatics approach to identify genetic interaction related to ovarian cancer chemoresistance, Briefings Bioinf. 22 (6) (2021) bbab100.

[13] A. Laios, A. Gryparis, D. DeJong, R. Hutson, G. Theophilou, C. Leach, Predicting complete cytoreduction for advanced ovarian cancer patients using nearest-neighbor models, J. Ovarian Res. 13 (2020) 1–8.

[14] S.Y. Liu, R.H. Zhu, Z.T. Wang, W. Tan, L. Zhang, Y.Q. Wang, F.F. Dai, M.Q. Yuan, Y.J. Zheng, D.Y. Yang, F.Y. Wang, Landscape of immune microenvironment in epithelial ovarian cancer and establishing risk model by machine learning, Journal of oncology 2021 (2021).

[15] R. Chen, Machine learning for ovarian cancer: lasso regression-based predictive model of early mortality in patients with stage I and stage II ovarian cancer, medRxiv (2020) 2020–2025.

[16] Z. Yue, C. Sun, F. Chen, Y. Zhang, W. Xu, S. Shabbir, L. Zou, W. Lu, W. Wang, Z. Xie, L. Zhou, Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis, Biomed. Opt Express 12 (5) (2021) 2559–2574.

[17] V.V.P. Wibowo, Z. Rustam, S. Hartini, F. Maulidina, I. Wirasati, W. Sadewo, Ovarian cancer classification using K-nearest neighbor and support vector machine, J. Phys. Conf. 1821 (1) (2021) 012007.

[18] J. Ma, J. Yang, Y. Jin, S. Cheng, S. Huang, N. Zhang, Y. Wang, Artificial intelligence based on blood biomarkers including CTCs predicts outcomes in epithelial ovarian cancer: a prospective study, OncoTargets Ther. (2021) 3267–3280.

[19] S.B. Cho, H.H. Won, Machine learning in DNA microarray analysis for cancer classification, Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 2003 19 (2003, January) 189–198.

[20] R. Xu, X. Cai, D.C. Wunsch, Gene expression data for DLBCL cancer survival prediction with a combination of machine learning technologies, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2006, January, pp. 894–897.

[21] M. Pirooznia, J.Y. Yang, M.Q. Yang, Y. Deng, A comparative study of different machine learning methods on microarray gene expression data, BMC Genom. 9 (2008) 1–13.

[22] B. Sahu, D. Mishra, A novel feature selection algorithm using particle swarm optimization for cancer microarray data, Procedia Eng. 38 (2012) 27–31.

[23] C. Gunavathi, K. Premalatha, A comparative analysis of swarm intelligence techniques for feature selection in cancer classification, Sci. World J. 2014 (2014).

[24] P. Yasodha, N.R. Ananthanarayanan, Analysing big data to build knowledge based system for early detection of ovarian cancer, Indian J. Sci. Technol. 8 (14) (2015) 1.

[25] H. Asri, H. Mousannif, H. Al Moatassime, T. Noel, Using machine learning algorithms for breast cancer risk prediction and diagnosis, Proc. Comput. Sci. 83 (2016) 1064–1069.

[26] M.N. Elgedawy, Prediction of breast cancer using random forest, support vector machines and naïve Bayes, International Journal of Engineering and Computer Science 6 (1) (2017) 19884–19889.

[27] M.W. Huang, C.W. Chen, W.C. Lin, S.W. Ke, C.F. Tsai, SVM and SVM ensembles in breast cancer prediction, PLoS One 12 (1) (2017) e0161501.

[28] T. Glasmachers, C. Igel, A unified view on multi-class support vector classification, J. Mach. Learn. Res. 17 (45) (2016) 1–32.

[29] M. Amer, M. Goldstein, S. Abdennadher, Enhancing one-class support vector machines for unsupervised anomaly detection, in: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, 2013, August, pp. 8–15.

[30] R. He, X. Li, G. Chen, G. Chen, Y. Liu, Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries, Expert Syst. Appl. 150 (2020) 113244.

[31] M. Akazawa, K. Hashimoto, Artificial intelligence in ovarian cancer diagnosis, Anticancer Res. 40 (8) (2020) 4795–4800.

[32] T. Muinao, H.P.D. Boruah, M. Pal, Multi-biomarker panel signature as the key to diagnosis of ovarian cancer, Heliyon 5 (12) (2019) 12345.

[33] Nitish Biswas, Khandaker Mohammad Mohi Uddin, Sarreha Tasmin Rikta, Samrat Kumar Dey, A comparative analysis of machinelearning classifiers for stroke prediction: a predictive analytics approach, Healthcare Analytics (2022).

[34] S. Sharmin, T. Ahammad, M.A. Talukder, P. Ghose, A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection, IEEE Access (2023).

[35] H. Salem, M.Y. Shams, O.M. Elzeki, M. Abd Elfattah, J. F. Al-Amri, S. Elnazer, Fine-tuning fuzzy KNN classifier based on uncertainty membership for the medical diagnosis of diabetes, Appl. Sci. 12 (3) (2022) 950.

[36] M. Ghassemi, L. Oakden-Rayner, A.L. Beam, The false hope of current approaches to explainable artificial intelligence in health care, The Lancet Digital Health 3 (11) (2021) e745–e750.

[37] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond, Inf. Fusion 77 (2022) 29–52.

[38] M. Pocevičiūtė, G. Eilertsen, C. Lundström, Survey of XAI in digital pathology, Artificial intelligence and machine learning for digital pathology: state-of-the-art and future challenges (2020) 56–88.

[39] R. Rodríguez-Pérez, J. Bajorath, Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions, J. Comput. Aided Mol. Des. 34 (2020) 1013–1026.

[40] J.S. Lavanya, P. Subbulakshmi, Machine learning techniques for the prediction of non-communicable diseases, in: 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), 2023, January, pp. 1–8.

[41] Z.A. Hadi, N.C. Dom, Development of machine learning modelling and dengue risk mapping: a concept framework, IOP Conf. Ser. Earth Environ. Sci. 1217 (1) (2023) 012038.

[42] L. Ibrahim, M. Mesinovic, K.W. Yang, M.A. Eid, Explainable prediction of acute myocardial infarction using machine learning and shapley values, IEEE Access 8 (2020) 210410–210417.

[43] Khandaker Mohammad Mohi Uddin, Nitish Biswas, Sarreha Tasmin Rikta, Samrat Kumar Dey, Atika Qazi, A self-driven interactive mobile application utilizing explainable machine learning for breast cancer diagnosis, Engineering Reports (2023).

[44] M. Ribera, À. Lapedriza García, Can we do better explanations? A proposal of user-centered explainable AI. CEUR Workshop Proceedings, 2019.

[45] W. Huang, H. Suominen, T. Liu, G. Rice, C. Salomon, A.S. Barnard, Explainable discovery of disease biomarkers: the case of ovarian cancer to illustrate the best practice in machine learning and Shapley analysis, J. Biomed. Inf. 141 (2023) 104365.

[46] G.D. Kitsios, D.M. Kent, Personalised medicine: not just in our genes, BMJ 344 (2012).

[47] Y. Nohara, K. Matsumoto, H. Soejima, N. Nakashima, Explanation of machine learning models using shapley additive explanation and application for real data in hospital, Comput. Methods Progr. Biomed. 214 (2022) 106584.