Diagnosis of Ovarian Cancer using a Multi-Omics inspired fusion model

Mishika Sood and Vardhaman Kalloli

Introduction

Ovarian cancer (OC) presents an alarming challenge in global health due to its insidious onset, diagnostic complexity, and high mortality rate. Often known to be the "silent killer," OC typically evades detection until it has progressed to advanced stages, with approximately 70% of cases diagnosed at stage III or IV. This delayed diagnosis contributes significantly to poor prognosis and a five-year survival rate of less than 50% for late-stage patients (Burns, 2025) (Iqbal Madakkatel et al., 2024). Despite advances in treatment modalities like frozen section analysis (Huang et al., 2018), the absence of reliable, early-stage diagnostic tools continues to undermine clinical efforts to reduce mortality.

Traditional diagnostic methods—such as pelvic examinations, imaging techniques, and serum biomarkers like CA-125 i.e. a high-molecular-weight glycoprotein antigen measured in the blood, primarily used as a tumor marker for detecting and monitoring epithelial ovarian cancer. (Gandhi & Bhatt, 2021 —have demonstrated limited accuracy, especially in early detection. These conventional approaches often yield high false-positive rates and lack the integrative power to synthesize complex, multi-modal patient data. Consequently, the medical community has turned its focus to artificial intelligence (AI) and machine learning (ML) as promising avenues to improve diagnostic precision.

This research paper investigates the application of AI-driven methodologies in diagnosing ovarian cancer, particularly emphasizing the evolution from classical machine learning models to modern, multi-modal fusion architectures (Carrillo-Pérez et al., 2022). Previous models were frequently siloed—relying exclusively on imaging, clinical, or genomic data—leading to fragmented and suboptimal diagnostic insights, as each modality captures only a portion of the complex tumor biology and fails to leverage complementary information across data types, resulting in reduced sensitivity and specificity, increased false positives/negatives, and poor generalizability (Oakden-Rayner et al., 2019). Fusion models, by contrast, combine multiple data sources, leveraging the unique strengths of each to generate a more accurate and holistic diagnostic assessment.

The central aim of this research is to explore and validate a fusion-based machine learning framework that integrates tumor marker data with clinical and hematological features to enhance the early and accurate classification of ovarian tumors. Through rigorous data analysis, model training, and evaluation, this research demonstrates the potential of ensemble learning and late fusion strategies to surpass the diagnostic performance of conventional single-modality models. By combining multi-domain information into a unified decision process, the system enables general practitioners to rapidly distinguish benign from malignant ovarian cases, improving diagnostic accuracy and patient referral decisions. Ultimately, this work underscores the transformative potential of AI in the early detection of ovarian cancer and offers a pathway toward more precise, efficient, and life-saving diagnostic practices.

Background and Significance

Ovarian cancer (OC) is recognised globally as one of the most severe threats to women's health, ranking as the fifth leading cause of cancer-related deaths among women (Sung et al., 2021). According to recent estimates, approximately 314,000 new cases of ovarian cancer are diagnosed each year, resulting in nearly 207,000 deaths globally (Sung et al., 2021) (Torre et al., 2018). Its high mortality rate is largely due to its diagnosis at advanced stages in about 70% of cases, primarily because effective screening and early detection strategies are still limited (Lheureux et al., 2019) (Armstrong et al., 2019). Despite advancements in oncology, five-year survival rates remain below 50% for late-stage diagnoses, significantly impacting patients, families, and healthcare systems worldwide

(National Cancer Institute, 2018).

The burden of ovarian cancer is compounded by economic factors, as late-stage treatments require extensive medical interventions, including surgeries, chemotherapy, and targeted therapies, significantly elevating healthcare costs (Wang et al., 2024). Moreover, Ovarian cancer incidence rises steadily with age—predominantly affecting postmenopausal women—and reaches its highest rates among those aged 65–74 years, complicating management by the frequent presence of age-related comorbidities (Jibril et al., 2024). The urgency for improved diagnostic methods is underscored not only by the clinical outcomes but also by the potential for reducing economic strain and improving patient quality of life through earlier and more accurate detection methods (Sung et al., 2021) (Torreet al., 2018).

Literature Review

Ovarian cancer is one of the most significant malignancies and remains the most lethal caner of the female reproductive system (Cai et al., 2024). Globally, it accounted for approximately 314,000 new cases and 207,000 deaths in 2020 (Nihal Abuzinadah et al., 2023). Often known as the "silent killer," early-stage ovarian cancer produces few specific symptoms, leading to the majority of patients being diagnosed at advanced stages (The Role of the Obstetrician–Gynecologist in the Early Detection of Epithelial Ovarian Cancer in Women at Average Risk, n.d.). Despite its high mortality rate, there is currently no effective population screening strategy and large trials have shown that annual pelvic exams, serum markers (like CA-125), or ultrasound screening do not reduce mortality and can lead to unnecessary interventions, according to the American College of obstetricians and gynecologists. The result is that ovarian cancer is often detected only when it has spread, underscoring the need for improved diagnostic approaches.

Diagnosis of ovarian cancer has traditionally relied on clinical evaluation, tumor markers, and imaging, followed by surgical/pathological confirmation. The initial workup usually includes a review of symptoms and medical history, a physical pelvic examination, measurement of serum cancer antigen 125 (CA-125), and imaging studies (Aleksandra Englisz et al., 2024). CA-125 is a glycoprotein tumor marker that was first identified in the early 1980s and is elevated in the blood of many ovarian cancer patients. An abnormal CA-125 level (generally >35 U/mL) raises suspicion for ovarian neoplasm; however, this marker on its own has limited utility. Only ~50% of early-stage ovarian cancers show elevated CA-125, whereas over 90% of advanced cases do. Moreover, CA-125 lacks specificity; it can be elevated in benign conditions such as endometriosis, pelvic inflammatory disease, menstruation, or even other cancers, yielding a modest specificity of around 73–78% (Dochez et al., 2019). Thus, while CA-125 is useful (especially for monitoring or in postmenopausal patients), it cannot definitively diagnose ovarian cancer by itself.

Imaging plays a critical role in the evaluation of an adnexal mass providing high-resolution grayscale and color-Doppler assessment to characterize lesion morphology and vascularity, with MRI (and CT when needed) reserved for further problem-solving or preoperative staging (Foti et al., 2015). Transvaginal ultrasonography (TVUS) is the first-line imaging modality for characterizing ovarian masses, given its sensitivity for detecting ovarian lesions and assessing morphology. Sonographic features such as solid areas, septations, bilateral lesions, ascites, and Doppler blood flow can raise concern for malignancy (Aleksandra Englisz et al., 2024). For instance, the International Ovarian Tumor Analysis (IOTA) group has developed ultrasound - based rules and models to distinguish benign from malignant masses. In clinical practice, a composite scoring system known as the Risk of Malignancy Index (RMI) was introduced by Jacobs et al. in 1990, which combines ultrasound findings, menopausal status, and CA-125 level into a single score (Clayton et al., 1999). An RMI above a certain cutoff (e.g. 200) is associated with a high risk of ovarian cancer and prompts referral to a gynecologic oncologist (Huwidi et al., 2022). This early "fusion" of imaging, lab, and clinical data improved diagnostic performance over any single measure and remains a widely used tool in preoperative

evaluation of ovarian masses.

Other imaging modalities like CT (Computed Tomography) and MRI (Magnetic Resonance Imaging) are often employed for further characterization of indeterminate masses and for staging once cancer is suspected. These can delineate the extent of disease (e.g. peritoneal implants, lymph node involvement) but are not routinely used for initial screening of masses due to cost and accessibility. Ultimately, definitive diagnosis of ovarian cancer is obtained through surgical exploration and histopathological examination of tissue (Wang et al., 2024). Histology not only confirms malignancy but also identifies the tumor type and grade. It is important to note that surgery itself is part of standard management; thus, diagnostic surgery (removal or biopsy of the mass) is often done with therapeutic intent. In summary, the traditional approach to ovarian cancer diagnosis involves careful clinical assessment, judicious use of tumor markers and imaging to estimate malignancy risk, and confirmatory pathology. The limitations of this approach – especially the inability to detect early disease with high sensitivity and specificity – have motivated the exploration of advanced methods, including novel biomarkers and most recently, artificial intelligence techniques.

Over the last ten years, machine learning (ML) and artificial intelligence (AI) have become promising tools to help with ovarian cancer diagnosis. AI in imaging medicine: Sometimes surpassing human performance, sophisticated algorithms can examine MRI, CT, or ultrasound images to find patterns suggestive of cancer. An early study in 1999 demonstrated that an artificial neural network could predict ovarian tumor malignancy with 95% sensitivity and 78% specificity (Clayton et al., 1999), sensitivity is the proportion of patients with disease correctly identified as positive, while specificity is the proportion of patients without disease correctly identified as negative, outperforming a traditional logistic regression model on the same data. This potential was already suggested by research conducted in the late 1990s. This proved that AI methods can enhance accuracy by combining diagnostic variables (such as age, CA -125, and Doppler ultrasound features) in non-linear ways (Clayton et al., 1999).

The advent of modern computing and larger datasets has led to deep learning methods being applied more recently. Convolutional neural networks (CNNs), in particular, have shown great promise in interpreting medical images for ovarian cancer. A 2024 systematic review and meta- analysis included 14 studies using AI for ultrasound -based ovarian cancer diagnosis (over 15,000 images) and reported an overall sensitivity of 81% and specificity of 92%, indicating that AI algorithms can achieve high diagnostic performance on par with expert sonographers. These algorithms automatically learn and identify malignancy-associated image features (tumor texture, shape, etc.) that may not be apparent to the human eye, thereby aiding radiologists in distinguishing benign from malignant ovarian masses (Mitchell et al., 2024).

Beyond imaging, AI has also been applied to other data modalities. For instance, machine learning models have been trained on patterns of clinical laboratory tests and biomarkers to detect ovarian cancer. In one recent multicenter study in China, researchers developed an ensemble AI model (nicknamed the “MCF model”) that utilized a panel of 98 routine lab test values (complete blood counts, liver function, etc.) to predict ovarian cancer (Mitchell et al., 2024). This model demonstrated robust performance across multiple validation cohorts, including an internal validation cohort of 3,007 individuals and two external validation cohorts of 5,641 and 2,344 individuals, achieving AUCs of 0.949, 0.882, and 0.884, respectively, offering a potentially cost-effective, non-invasive diagnostic adjunct – essentially finding “signal” for ovarian cancer in ordinary blood test results (Cai et al., 2024). Such approaches hint that AI could flag high-risk individuals even before imaging or assist in triaging patients for further evaluation. Similarly, AI techniques have been explored to analyze gene expression profiles and proteomics as well.

Overall, the impact of AI/ML in ovarian cancer diagnosis has been significant in research settings. Studies consistently show that AI can increase diagnostic accuracy and potentially expedite the identification of malignancies. Notably, convolutional neural network ensembles such as Ovry-Dx1

and Ovry-Dx2, transformer-based architectures like the Swin Transformer, and combined deep-learning radiomics (DLR) frameworks have each demonstrated diagnostic performance—AUCs up to 0.97, sensitivities up to 97%, and specificities up to 94%—that is statistically indistinguishable from expert sonographer interpretation of adnexal ultrasound images. In an international validation published in 2025, a transformer-based deep neural network trained on ultrasound images from thousands of patients significantly outperformed both expert and generalist examiners in detecting ovarian cancer. The AI system was able to reduce unnecessary referrals to specialists by 63% while improving diagnostic accuracy, effectively demonstrating expert-level performance in ovarian tumor classification (Christiansen et al., 2025). These advancements suggest that AI, especially in the form of deep learning algorithms, can serve as a powerful assistive tool in ovarian cancer diagnosis. However, it is also emphasized that prospective clinical validation and careful integration into workflows are needed before AI can be widely adopted in routine practice (Mitchell et al., 2024). As the field progresses, the combination of traditional expertise with AI decision support holds promise for earlier and more accurate detection of ovarian cancer.

Machine learning methods used in ovarian cancer diagnosis span from classical statistical models to cutting-edge deep learning. Classical ML models include techniques like logistic regression, decision trees, naive Bayes, and support vector machines (SVMs). These approaches typically require predefined input features – for example, a logistic regression might take CA-125 level, ultrasound score, and patient age to calculate a risk probability. Indeed, the Risk of Malignancy Index can be seen as a simple statistical model using a linear combination (product) of features identified by clinicians (Dochez et al., 2019). Such models are interpretable and were the mainstay of predictive modeling in past decades; however, they may struggle with complex, high-dimensional data and non-linear relationships.

Modern ML, on the other hand, often refers to deep learning and other advanced algorithms that automatically learn representations from data. Deep learning uses artificial neural networks with many layers (deep neural networks) to capture intricate patterns. In the context of ovarian cancer diagnosis, deep learning models (like CNNs) can take raw inputs (e.g. an ultrasound image pixel data) and learn the most predictive features implicitly, without explicit feature engineering (Wang et al., 2024). This is a key advantage over classical ML, which usually requires manual selection or crafting features (such as specific image measurements or biomarker thresholds). For example, a CNN analyzing an ultrasound image can learn to detect subtle texture differences or irregular tumor borders associated with malignancy, whereas an SVM would need those image features to be defined and fed into it by experts beforehand.

A practical illustration of the classical vs. modern ML difference was seen in the 1999 study mentioned earlier: the logistic regression model had to use the four human-chosen variables (age, CA-125, Doppler ultrasound score, etc.), but it achieved only ~82% sensitivity at ~51% specificity in distinguishing malignant vs. benign tumors. The neural network model using the same inputs was able to discern nonlinear interactions and achieved much higher accuracy (95% sensitivity at 78% specificity) (Clayton et al., 1999). Modern deep learning goes even further – models such as CNNs, recurrent neural networks (RNNs), and transformers contain sophisticated architectures that can model complex spatial and temporal patterns in the data. RNNs, for instance, are suitable for sequential data; researchers have explored RNN-based models to evaluate longitudinal changes in biomarkers (like rising CA-125 trends over time) to predict ovarian cancer earlier (Elias et al., 2017). Transformers, originally developed for language processing, have recently been applied to imaging and genomics in medicine. Their self-attention mechanism allows integration of multiple data types and long-range dependencies. In ovarian cancer diagnostics, transformer-based models have been used to analyze ultrasound images with great success, as noted in the multi-center study where they outperformed human experts. These transformer networks were able to generalize across different ultrasound machines and patient populations, highlighting the robustness of modern deep models (Christiansen et

al., 2025).

In summary, classical ML models like logistic regression and SVM provided a foundation for predictive modeling in ovarian cancer (e.g., risk indices and early computer-aided diagnosis studies). They are typically faster to train and easier to interpret but may oversimplify the data relationships. Modern ML models – including CNNs for image analysis, RNNs for sequential data, and transformers for multi-modal integration – offer superior performance by automatically learning complex features and interactions (Wang et al., 2024). The trade-off is that these models are data-hungry, computationally intensive, and often operate as "black boxes" that require techniques (like SHAP values or other explainable AI methods) to interpret their decisions. Nonetheless, the move from classical to modern ML in ovarian cancer research has markedly improved diagnostic accuracy and opened new avenues (such as multi-omics data integration) that were not feasible with earlier methods (Wang et al., 2024).

A notable challenge in ovarian cancer diagnostics is the siloed nature of data used in many predictive models. Historically, most diagnostic approaches focused on one category of information at a time for example, using only clinical features, only serum biomarkers, or only imaging. Such single-modality models are simpler to build, but they inherently ignore the complementary information that other data types might provide. In practice, a clinician evaluates a patient's condition by synthesizing multiple inputs (symptoms, physical exam, blood tests, scans); similarly, an ideal diagnostic algorithm would integrate all relevant data. Yet, until recently, few computational models attempted true multimodal integration. A 2023 review observed that despite the wealth of clinical, molecular, and imaging data available, these data "frequently remain siloed and are thus inaccessible for integrated analyses" (Sepideh Hatamikia et al., 2023). The vast majority of published studies have developed models based on a single data source (e.g., an ultrasound-based model, or a genomic signature, etc.), with only a minority of studies focusing on multi-modal AI integration (Sepideh Hatamikia et al., 2023).

For example, many early machine learning efforts in ovarian cancer dealt exclusively with serum biomarkers – producing models to distinguish benign vs. malignant masses using CA-125 alone or in combination with other blood markers like HE4, without incorporating imaging findings (Dochez et al., 2019). On the other hand, radiology researchers often built image-only models (such as CNNs on ultrasound or CT images) that do not use clinical chemistry or patient history. These "siloed" approaches can achieve decent accuracy, but they may plateau because each modality alone provides an incomplete picture. An ultrasound image might suggest malignancy, but adding a patient's menopausal status and CA-125 level could greatly improve confidence in the diagnosis – this is exactly why the RMI formula was so successful in clinical practice (Dochez et al., 2019). Siloed models miss such cross-pollination of information.

Another consequence of siloed development is inconsistency and difficulty in comparing models. One model might report excellent performance using only genetic data in a research lab, while another reports performance using only imaging; in the absence of integrated evaluation, it's unclear which approach is truly superior or how to combine them. The literature review in European Radiology Experimental noted that in ovarian cancer studies to date, about 60% of attempts at integration were just combining imaging with basic clinical data (age, etc.), and relatively few incorporated genomic or transcriptomic data together with imaging (Sepideh Hatamikia et al., 2023). This indicates that fully multi-modal models (imaging + clinical + molecular) have been rare, though they have great potential. The limitations of siloed models are increasingly recognized, and there is a growing movement toward developing fusion models that break these silos.

The concept of fusing multiple data sources for ovarian cancer diagnosis is not entirely new – as noted, the Risk of Malignancy Index in 1990 was an early example of a simple fusion model, multiplying an ultrasound score by menopausal status and CA-125 level (Dochez et al., 2019). Throughout the 1990s and 2000s, researchers continued to explore combinations of predictors to improve diagnostic

performance. By 1991, some groups had even begun experimenting with artificial neural networks to combine laboratory tests and demographic factors for cancer diagnosis (Clayton et al., 1999). In the late 2000s, the introduction of a new biomarker, HE4, led to the Risk of Ovarian Malignancy Algorithm (ROMA) – a logistic regression model that integrates CA-125 and HE4 levels along with menopausal status (Dochez et al., 2019). Moore and colleagues first described ROMA in 2009, and subsequent studies showed that ROMA could achieve an area-under-curve (AUC) of about 0.93, outperforming either CA-125 or HE4 alone in differentiating malignant from benign ovarian masses (Dochez et al., 2019). These milestones illustrate the trend from single marker to multi-marker diagnostic tools.

In recent years (2010s to present), the availability of large datasets and advanced ML techniques has accelerated the development of fusion models that combine heterogeneous data types. Modern ensemble learning methods have been applied – for instance, stacking or blending multiple classifiers, each trained on different features, to form a more powerful predictor. One study designed a stacked ensemble of machine learning models (mixing bagging and boosting algorithms) using 50 combined features (clinical, biochemical, etc.) and reported a diagnostic accuracy of ~97%, higher than any single model on that dataset (Nihal Abuzinadah et al., 2023). Such ensemble approaches leverage the "wisdom of models," much like an ensemble of medical opinions, to improve reliability.

Crucially, truly multi-modal AI models are now emerging. These systems incorporate clinical data (patient age, symptoms, family history), serum biomarkers (CA-125, HE4, others), imaging findings (ultrasound or CT features), and sometimes genomic/proteomic data into one predictive framework. Integrative analyses indicate that combining modalities yields superior results.

According to a 2023 review, integrative multi-omics models consistently outperform models based on any single data type for ovarian cancer classification (Sepideh Hatamikia et al., 2023). For example, a classifier that uses both ultrasound image analysis and CA-125 levels can achieve higher sensitivity and specificity than an image-only or lab-only model. In fact, multiple studies cited in that review demonstrated improved performance when imaging was paired with clinical variables (such as the IOTA models that add age and serum markers to ultrasound findings). More comprehensive fusion models that also include genomic data (e.g. BRCA mutation status or gene expression profiles) are less common but show promise in research settings (Sepideh Hatamikia et al., 2023).

Ensemble and multimodal strategies from 1991 to the present thus chart an evolution: from simple combined scores like RMI, to dual-marker algorithms like ROMA, to current AI-driven systems that can assimilate diverse data streams. One striking recent example is an AI-driven ultrasound model validated across 20 centers worldwide, which effectively "fused" knowledge from different populations and ultrasound machines; it generalized so well that it surpassed expert examiners and reduced diagnostic errors when implemented in a simulated triage setting (Christiansen et al., 2025). While that particular study focused on imaging (with the AI acting as a complex fusion of features across centers), the principles are being extended to multi-modal fusion. The integration of radiomics (quantitative imaging features) with genomics and proteomics has even given rise to the field of "radiogenomics," aiming to connect imaging phenotypes with molecular characteristics to improve diagnostic and prognostic accuracy.

In conclusion, the trajectory of ovarian cancer diagnostic models has moved from isolated, siloed approaches toward holistic fusion models that combine the strengths of various diagnostic modalities. Peer-reviewed evidence strongly supports that such integrated models provide more accurate and robust assessments (Christiansen et al., 2025). As of now, several ensemble and multimodal AI models are in development or early clinical evaluation, and they have demonstrated improved discrimination between benign and malignant ovarian masses over traditional methods. This multi-modal fusion paradigm, powered by modern AI, represents a significant advancement in our ability to diagnose ovarian cancer earlier and more reliably, which is critical to improving outcomes for this historically

deadly disease.

Data Analysis

AI has the potential to enhance OC diagnosis and management. AI techniques, including machine and deep learning, have been applied to CT, MRI, and ultrasound for cancer detection and classification (Xu et al., 2021). The fight against complex diseases like ovarian cancer increasingly relies on sophisticated data analysis techniques (Hong & Ding, 2025). Early and accurate diagnosis is crucial for improving patient outcomes but distinguishing between malignant conditions (Ovarian Cancer - OC) and benign tumors (Benign Ovarian Tumor - BOT) can be challenging. Data analysis, particularly when applied to rich datasets containing patient information and biological markers (biomarkers), offers a powerful avenue to uncover subtle patterns that might signal the presence of cancer (Piedimonte et al., 2022).

This study utilizes a dataset curated for the prediction of ovarian cancer through machine learning methodologies. The “Predict Ovarian Cancer” dataset was originally collected to support the study by Mi et al. (2020) aiming to build machine-learning models for differentiating benign ovarian tumors (BOT) from ovarian cancer (OC). The cohort comprises 349 Chinese women who underwent evaluation for adnexal masses at the Third Affiliated Hospital of Soochow University in Changzhou, Jiangsu Province, China, with all patient data de-identified in accordance with institutional ethical guidelines.

The data is publicly available on Kaggle (Predict Ovarian Cancer Dataset) and comprises five `distinct files. Each of these, serve a specific purpose in the analysis:

1. Supplementary data 1.xlsx contains the original, unprocessed data collected for the study. It represents the complete set of initial measurements.
2. Supplementary data 2.xlsx provides a comprehensive list of the biomarkers included in the study along with their abbreviations and descriptions, thus offering understanding of the clinical significance of each feature.
3. Supplementary data 3.xlsx offers a pre-processed, imputed version of the training data. The biomarker CA72-4 has been excluded from this version and missing values within the remaining features have been handled through imputation techniques, thus making it valuable for analyses.
4. Supplementary data 4.xlsx contains the raw, unprocessed training data designated for model training purposes.
5. Supplementary data 5.xlsx holds the raw, unprocessed test data, intended for evaluating the performance of the trained machine learning models on unseen data.

The target variable for prediction is the 'TYPE' column which is a binary categorical variable where a value of '1' indicates a Benign Ovarian Tumor (BOT), and a value of '0' signifies Ovarian Cancer (OC). The objective of this study is to utilize the biomarker features within the dataset to accurately classify samples into two categories.

Exploratory Data Analysis (EDA)

A thorough exploratory data analysis was conducted to gain insights into the characteristics of the dataset and the relationships between the variables. A very initial step is to drop the SUBJECT_ID column because it is not relevant for the predictive models. Also, several tumor marker columns (AFP, CA125, CA19-9) are identified as containing non-numeric data and require cleaning to convert them to a numeric format for analysis. In general, the issue of non-numeric values within tumor marker columns was addressed by defining a cleaning function because the raw data contained inconsistencies that needed to be standarized. The dataset was observed to be incomplete due to the missing values

found in the tumor markers, which were then imputed. To top it all off, the primary target variable for prediction 'TYPE' column observed a relatively balanced class distribution.

Methodology

Model Training: Two Random Forest Classifiers

We trained two separate RandomForestClassifier models (from scikit-learn) to specialize in each feature subset. Random forests were chosen for their robustness and ability to handle feature interactions without extensive parameter tuning. Both models were trained with the class_weight='balanced' option to automatically adjust for any class imbalance by giving more weight to the minority class in the loss function. Each model consisted of 200 decision trees (n_estimators=200), a number high enough to ensure stable estimates of feature importance and predictions and utilized all CPU cores in parallel (n_jobs=-1) to expedite training. A fixed random_state=42 was set for both models to make results deterministic.

Model 1: Tumor Marker Classifier

The first model was dedicated to tumor marker features only. Six tumor marker measurements (including AFP, CA125, CA19-9, CA72-4, CEA, and HE4) were used as predictors. After MinMax scaling these six features, we trained a RandomForestClassifier with the parameters noted above (class_weight='balanced', n_estimators=200, n_jobs=-1, random_state=42). The balanced class weights help ensure that the model gives due attention to the ovarian cancer class even if benign cases slightly outnumber malignancies in training. No explicit depth limit was set for this model's trees, allowing the ensemble to fully explore the small feature set for any relevant splits. Model 1 effectively acts as a "biomarker expert," focusing solely on patterns in the tumor marker panel.

Model 2: Clinical & Hematological Classifier

The second model was trained on the clinical and hematological features (a larger set of 43 variables, such as blood cell counts, metabolic panel results, patient age, and menopausal status). This feature group provides a different perspective, capturing general health and blood chemistry data. We applied MinMaxScaler to these features separately, then trained a RandomForestClassifier with n_estimators=200 and class_weight='balanced' plus some constraints to prevent overfitting given the higher dimensionality: a maximum tree depth of 10 (max_depth=10) and minimum requirements of 5 samples to split a node and 3 samples per leaf (min_samples_split=5, min_samples_leaf=3). These hyperparameters regularize the tree growth, forcing each decision tree to consider more data points per split and limiting complexity. This way, Model 2 can capture the most significant clinical/hematological patterns without modeling spurious noise. As with Model 1, a random state of 42 ensured consistent results across runs. Model 2 serves as a "clinical expert," learning from the broader health indicators.

Both Random Forest models were trained on their respective training subsets (after scaling). During training, the algorithms internally bootstrap sampled the data and built an ensemble of decision trees that vote on the outcome. The balanced weighting was especially useful to make the models sensitive to the ovarian cancer class (the positive class of interest), addressing the critical goal of not missing cancer diagnosis. We monitored performance metrics on the training data (and could use cross-validation if needed) to verify that each model was fitting appropriately without obvious overfitting (the built-in randomness and our parameter choices provide resistance to overfitting).

Model Persistence

After training, we saved each model along with its scaler to enable easy reuse for inference. Using Python's joblib library, Model 1 and its MinMaxScaler were serialized to disk (e.g., rf_model_tumor_markers.joblib and scaler_tumor_markers.joblib), as were Model 2 and its scaler (rf_model_clin_hem.joblib and scaler_clin_hem.joblib). Persisting on the models ensures that the

exact trained parameters and scaling transformations can be loaded later for making predictions on new data, maintaining consistency with the training environment. This practice is important for deployment of the model in a clinical decision support setting – it prevents the need to retrain and guarantees that the inference uses the same scaling fitted on the training data (to avoid data leakage or scaling mismatches).

Ensemble Inference via Late Fusion

For the prediction phase, we implemented a simple late fusion ensemble to combine the strengths of the two models. Each model produces a probability estimate for the positive class (i.e., the probability that a given patient sample is Ovarian Cancer). We loaded the saved scalers and models, then processed the test set in parallel through each pipeline. The tumor marker test features were scaled with the tumor marker MinMaxScaler, and fed into Model 1 to obtain a probability P1(OC)=Model1(Scaled Tumor Marker Features) for ovarian cancer. Separately, the clinical/hematological test features were scaled with the second scaler and passed into Model 2, yielding a probability P2(OC)=Model2(Scaled Clinical/Hematological Features). These probabilities correspond to the model's confidence that the case is malignant (ovarian cancer).

We then combined the two probabilities by a simple unweighted average:

P(avg) (OC) = [P1(OC) + P2(OC)]/2

This averaging constitutes a late fusion at the decision level, meaning we integrate the modalities only after each has made its own prediction. Finally, to produce a binary classification, we applied a threshold of 0.5 to the average probability. If $Pavg(OC) \geq 0.5$, the ensemble predicts Ovarian Cancer; if $Pavg(OC) < 0.5$, it predicts a benign tumor. This 0.5 cutoff effectively means that if the models together are at least equally confident in cancer as not, we label it as cancer (a standard approach for balanced classes). The entire inference process was performed on the held-out test set to evaluate how well the methodology generalizes to unseen patients.

Justification for Late Fusion Approach

We deliberately used a late fusion ensemble (probability averaging) to combine the two learned models because it leverages complementary information in the tumor markers versus clinical data. In multi-omics and multi-modal settings, late fusion allows separate models to specialize on different data types and then merge their outputs, capitalizing on each model's "expertise" (Stahlschmidt et al., 2022). This approach maintains modality-specific interpretability – we can inspect how the tumor marker model behaves independently from the clinical model – while still achieving a unified prediction. By keeping the modalities separate until the final step, we avoided forcing a single model to learn a potentially very complex relationship spanning all features at once. Instead, each Random Forest could focus on capturing patterns in its feature domain (e.g., one on biomarker profiles, another on clinical blood panels). The literature supports this divide-and-conquer strategy: late fusion ensembles have been shown to yield robust performance improvements in biomedical classification by combining diverse feature sources. For example, Carrillo-Pérez et al. (2022) demonstrated that an ensemble which fused five different omics and imaging modalities for cancer diagnosis achieved significantly higher accuracy than any single-modality model (Carrillo-Pérez et al., 2022). In our context, the tumor marker model and the clinical data model offer different "views" of the patient – one biochemical and one physiological – and fusing them is expected to improve diagnostic accuracy by averaging out their individual errors and reinforcing signals that are consistently detected by both. This simple averaging fusion assumes each model's contribution is equally important, a reasonable choice when we do not have prior knowledge favoring one modality over the other. Research in multi-omics cancer detection has found that such unweighted late fusion can effectively harness complementary strengths of each data type (Carrillo-Pérez et al., 2022). In summary, late fusion provided a straightforward and justifiable way to integrate multi-omic features in our ovarian cancer classifier, aligning with current

best practices for combining heterogeneous data in machine learning models.

Results and Findings

Performance Metrics

In this study, we trained and assessed Random Forest models using distinct feature groups: tumor markers, clinical and hematological features, and a combination of all features; keeping our approach multi-omic inspired. Thus, the predictive capabilities of different sets of biomarkers for ovarian cancer classification were evaluated.

The performance of each model was evaluated on a sacrosanct test set using several key metrics including accuracy, balanced accuracy, precision, recall, F1-score and confusion matrix.

Feature Set	Accuracy	Balanced Accuracy	Precision (OC)	Recall (OC)	F-1 Score (OC)	Precision (BOT)	Recall (BOT)	F-1 Score (BOT)
Tumor	0.90	0.9003	0.91	0.89	0.90	0.89	0.91	0.90
Markers								
Clinical and Hermatological	0.90	0.8995	0.89	0.92	0.90	0.91	0.88	0.90
Combined Features	0.9429	0.9420	0.92	0.97	0.95	0.97	0.91	0.94

Table 1: Performance Metrics of Random Forest Models

The performance metrics presented in Table 1 reveal that the Random Forest model trained on the combined set of tumor markers, clinical, and hematological features achieved the highest overall performance. This model demonstrated the best accuracy (0.9429) and balanced accuracy (0.9420) indicating superior classification across both Ovarian Cancer (OC) and Benign Ovarian Tumor (BOT) cases. Notably, the combined model also exhibited the highest precision (0.92) and recall (0.97) for the Ovarian Cancer class, resulting in a strong F1-score of 0.95 for this critical diagnostic category.

Confusion Matrix

The confusion matrices visualizes and summarizes the performance of the model's classification, highlighting the counts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) for each class (Ovarian Cancer - OC, Benign Ovarian Tumor - BOT).

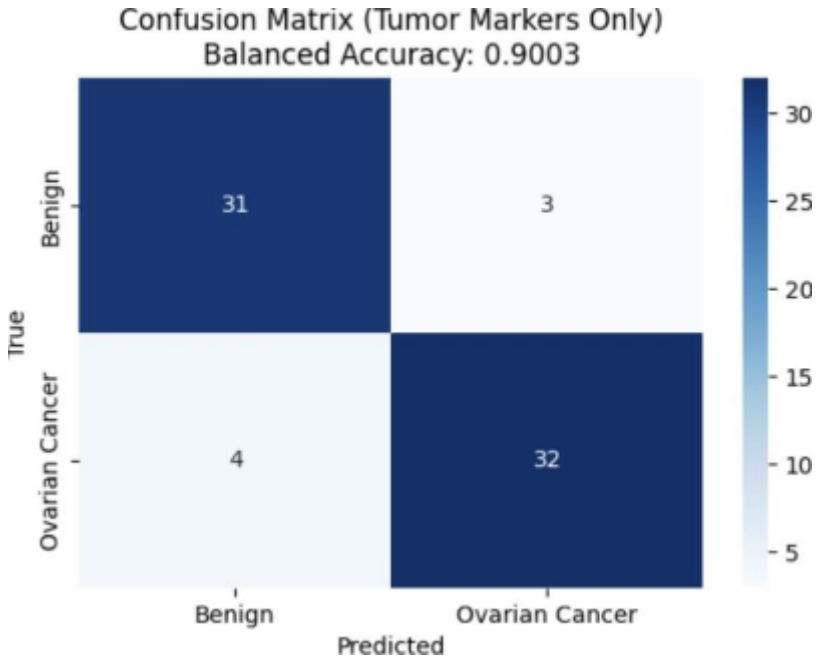

Figure 1: Confusion Matrix for Random Forest Model (Tumor Markers Only)

This matrix shows the number of actual Ovarian Cancer cases that were correctly identified as such (True Positives) are 32, the number of actual Benign Ovarian Tumor cases correctly identified (True Negatives) are 31, the number of Benign cases incorrectly classified as Ovarian Cancer (False Positives) are 3, and the number of Ovarian Cancer cases incorrectly classified as Benign (False Negatives) are 4.

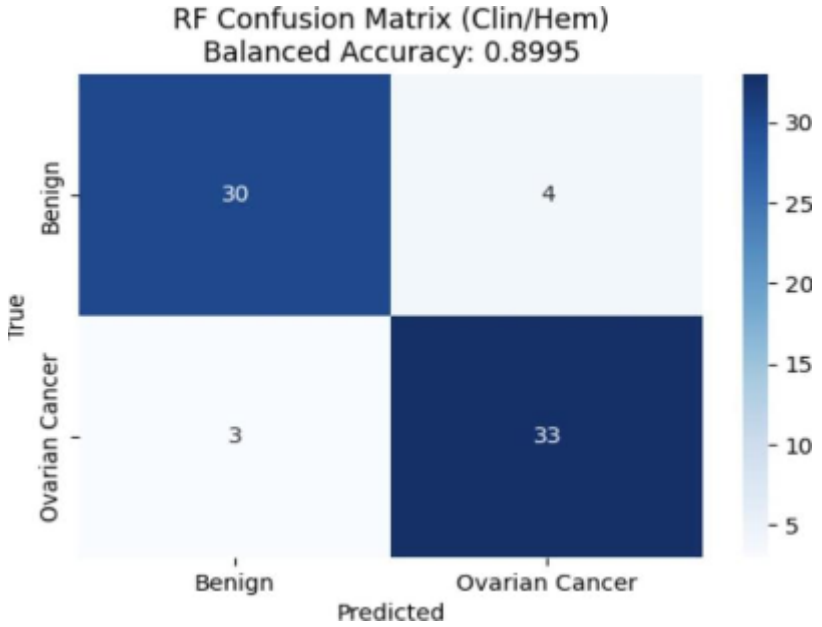

Figure 2: Confusion Matrix for Random Forest Model (Clinical & Hematological Features Only)

This matrix for the model trained used only clinical and hematological features. Assessment of the model's ability to distinguish between the two classes based on this specific set of biomarkers gives a balanced accuracy of 89.95%.

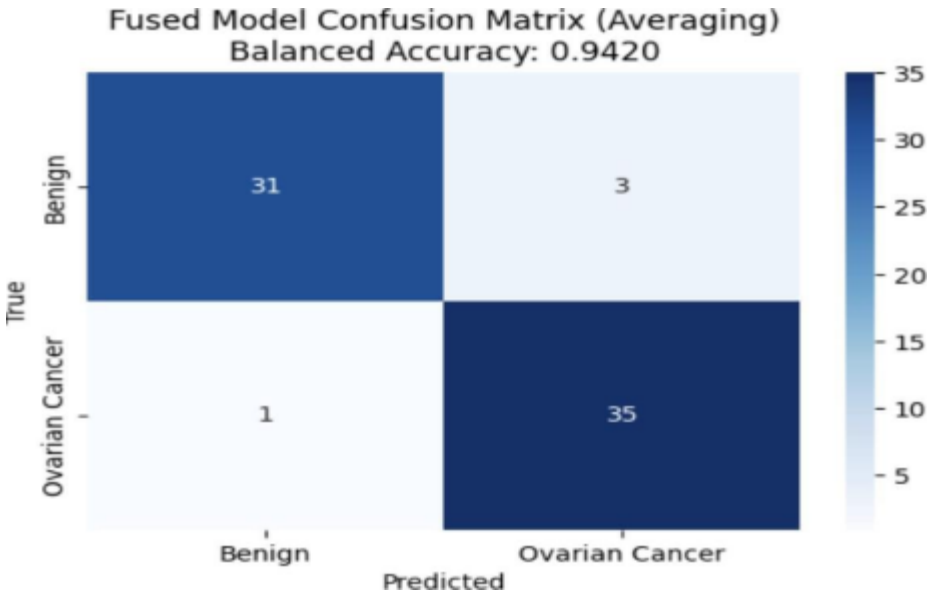

Figure 3: Confusion Matrix for Random Forest Model (Combined Features)

This matrix for the Random Forest model trained on the combined set of tumor markers, clinical, and hematological features, the fusion model as per the methodology. By comparing this matrix with those in Figures 1 and 2, we can observe the impact of incorporating a broader range of biomarkers on the model's classification accuracy and error patterns. By examining both the overall performance metrics and the detailed breakdown of classifications in the confusion matrices, the effectiveness of different biomarker combinations in predicting ovarian cancer can be realized and that can further pave the way for identification of potential areas for improvement.

Conclusion

In this study, we showed how a multi-omics-inspired machine learning approach can greatly improve the diagnosis of ovarian cancer (OC). We outperformed single-modality models in diagnostics by training different Random Forest classifiers on different domains of tumor marker data and clinical/hematological features, then using late fusion through probability averaging. The fusion model's notable gains in accuracy, precision, recall, and F1-score highlight the importance of combining various data types.

Our results provide compelling evidence that ensemble approaches are both feasible and effective in addressing the drawbacks of conventional OC diagnostic techniques, which frequently depend on discrete biomarkers or clinical indicators. The late fusion strategy used here proved beneficial, leveraging the complementary strengths of the distinct feature sets without compromising interpretability.

This study opens up a number of future research directions. To further improve model accuracy and diagnostic comprehensiveness, future research may investigate integrating additional data modalities like genomic, proteomic, or radiomic features. Additional insights and advancements could be obtained by assessing different fusion methods, such as weighted averaging or more intricate machine learning ensembles. Furthermore, evaluating the generalizability and clinical utility of the suggested strategy would require prospective clinical validation studies involving bigger, multi-center patient cohorts.

Finally, enabling earlier and more precise detection, the incorporation of such advanced multi-modal models into clinical practice holds promise for revolutionizing the diagnosis of ovarian cancer. This development has the potential to improve patient outcomes and quality of life by drastically lowering mortality and healthcare expenses related to late-stage treatments. Continued interdisciplinary research

combining oncology expertise, computational biology, and machine learning innovation remains critical to realizing the full potential of AI-driven diagnostics in ovarian cancer and beyond.

References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660
2. Torre, L. A., Trabert, B., DeSantis, C. E., Miller, K. D., Samimi, G., Runowicz, C. D., ... & Siegel, R. L. (2018). Ovarian cancer statistics, 2018. CA: A Cancer Journal for Clinicians, 68(4), 284–296. https://doi.org/10.3322/caac.21456
3. Lheureux, S., Gourley, C., Vergote, I., & Oza, A. M. (2019). Epithelial ovarian cancer. The Lancet, 393(10177), 1240–1253. https://doi.org/10.1016/S0140-6736(18)32552-25.
4. National Cancer Institute. (2022). SEER Cancer Stat Facts: Ovarian Cancer. Retrieved from https://seer.cancer.gov/statfacts/html/ovary.htmlSEER
5. Armstrong, D. K., Alvarez, R. D., Bakkum-Gamez, J. N., Barroilhet, L., Behbakht, K., Berchuck, A., ... & National Comprehensive Cancer Network. (2019). NCCN Guidelines Insights: Ovarian Cancer, Version 1.2019. Journal of the National Comprehensive Cancer Network, 17(8), 896–909. https://doi.org/10.6004/jnccn.2019.0039
6. Cai, G., Huang, F., Gao, Y., Li, X., Chi, J., Xie, J., Zhou, L., Feng, Y., Huang, H., Deng, T., Zhou, Y., Zhang, C., Luo, X., Xie, X., Gao, Q., Zhen, X., & Liu, J. (2024). Artificial intelligence-based models enabling accurate diagnosis of ovarian cancer using laboratory tests in China: a multicentre, retrospective cohort study. The Lancet Digital Health. https://doi.org/10.1016/s2589-7500(23)00245-5
7. Nihal Abuzinadah, Sarath Kumar Posa, Aisha Ahmed Alarfaj, Ebtisam Abdullah Alabdulqader, Umer, M., Kim, T.-H., Shtwai Alsubai, & Ashraf, I. (2023). Improved Prediction of Ovarian Cancer Using Ensemble Classifier and Shaply Explainable AI. Cancers, 15(24), 5793–5793. https://doi.org/10.3390/cancers15245793
8. The Role of the Obstetrician–Gynecologist in the Early Detection of Epithelial Ovarian Cancer in Women at Average Risk. (n.d.). Www.acog.org. https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2017/09/the- role-of-the-obstetriciangynecologist-in-the-early-detection-of-epithelial-ovarian-cancer- in-women-at-average-risk
9. Huwidi, A., Abobrege, A., Assidi, M., Buhmeida, A., & Ermiah, E. (2022). Diagnostic value of risk of malignancy index in the clinical evaluation of ovarian mass. Molecular and Clinical Oncology, 17(1). https://doi.org/10.3892/mco.2022.2551
10. Aleksandra Englisz, Smycz-Kubańska, M., & Aleksandra Mielczarek-Palacz. (2024). Sensitivity and Specificity of Selected Biomarkers and Their Combinations in the Diagnosis of Ovarian Cancer (Review). Diagnostics, 14(9), 949–949. https://doi.org/10.3390/diagnostics14090949
11. Dochez, V., Caillon, H., Vaucel, E., Dimet, J., Winer, N., & Ducarme, G. (2019). Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. Journal of Ovarian Research, 12(1). https://doi.org/10.1186/s13048- 019-0503-7
12. Clayton, R. D., Snowden, S., Weston, M. J., Mogensen, O., J. Eastaugh, & Lane, G. (1999). Neural networks in the diagnosis of malignant ovarian tumours. BJOG an International Journal of Obstetrics & Gynaecology, 106(10), 1078–1082. https://doi.org/10.1111/j.1471-

0528.1999.tb08117.x

13. Mitchell, S., Nikolopoulos, M., El-Zarka, A., Al-Karawi, D., Al-Zaidi, S., Ghai, A., Gaughran, J. E., & Sayasneh, A. (2024). Artificial Intelligence in Ultrasound Diagnoses of Ovarian Cancer: A Systematic Review and Meta-Analysis. Cancers, 16(2), 422. https://doi.org/10.3390/cancers16020422
14. Christiansen, F., Konuk, E., Ganeshan, A. R., Welch, R., Palés Huix, J., Czekierdowski, A., Leone, F. P. G., Haak, L. A., Fruscio, R., Gaurilcikas, A., Franchi, D., Fischerova, D., Mor, E., Savelli, L., Pascual, M. À., Kudla, M. J., Guerriero, S., Buonomo, F., Liuba, K., & Montik, N. (2025). International multicenter validation of AI-driven ultrasound detection of ovarian cancer. Nature Medicine. https://doi.org/10.1038/s41591-024-03329- 4
15. Elias, K. M., Fendler, W., Konrad Stawiski, Fiascone, S., Vitonis, A. F., Berkowitz, R. S., Gyorgy Frendl, Konstantinopoulos, P. A., Crum, C. P., Kędzierska, M., Cramer, D. W., & Chowdhury, D. (2017). Diagnostic potential for a serum miRNA neural network for detection of ovarian cancer. ELife, 6. https://doi.org/10.7554/elife.28932
16. Sepideh Hatamikia, Nougaret, S., Panico, C., Giacomo Avesani, Nero, C., Luca Boldrini, Sala, E., & Woitek, R. (2023). Ovarian cancer beyond imaging: integration of AI and multiomics biomarkers. European Radiology Experimental, 7(1). https://doi.org/10.1186/s41747-023-00364-7
17. Carrillo-Perez, F., Morales, J. C., Castillo-Secilla, D., Gevaert, O., Rojas, I., & Herrera, L. J. (2022). Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis. Journal of Personalized Medicine, 12(4), 601. https://doi.org/10.3390/jpm12040601
18. Xu, C., Xiaofei, H., Zhe, W., & Jingjing, L. (2021). Advances of Artificial Intelligence Application in Medical Imaging of Ovarian Cancers. Chinese Medical Sciences Journal, 36(3), 196–203. https://doi.org/10.1016/S1001-9294(21)00057-2
19. Hong, M.-K., & Ding, D.-C. (2025). Early Diagnosis of Ovarian Cancer: A Comprehensive Review of the Advances, Challenges, and Future Directions. Diagnostics, 15(4), 406. https://doi.org/10.3390/diagnostics15040406
20. Piedimonte, S., Rosa, G., Gerstl, B., Sopocado, M., Coronel, A., Llenno, S., & Vicus, D. (2022). Evaluating the use of machine learning use in ovarian cancer: A systematic review. Journal of Clinical Oncology, 40(16_suppl), e17570–e17570. https://doi.org/10.1200/jco.2022.40.16_suppl.e17570
21. Wang, Y., Lin, W., Zhuang, X., Wang, X., He, Y., Li, L., & Lyu, G. (2024). Advances in artificial intelligence for the diagnosis and treatment of ovarian cancer (Review). Oncology Reports, 51(3). https://doi.org/10.3892/or.2024.8705
22. Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics, 23(2). https://doi.org/10.1093/bib/bbab569
23. Burns, S. (2025, February 17). Gestational diabetes in rural Australia - screening, treatment, challenges. Rural Doctors Foundation. https://ruraldoctorsfoundation.org.au/ovarian-cancer-2/
24. Iqbal Madakkatel, Lumsden, A. L., Mulugeta, A., Mäenpää, J., & Hyppönen, E. (2024). Large-scale analysis to identify risk factors for ovarian cancer. International Journal of Gynecological Cancer, Suppl 10. https://doi.org/10.1136/ijgc-2024-005424
25. Huang Z, Li L, Li C, Ngaujah S, Yao S, Chu R, Xie L, Yang X, Zhang X, Liu P, Jiang J, Zhang

Y, Cui B, Song K, Kong B. Diagnostic accuracy of frozen section analysis of borderline ovarian tumors: a meta-analysis with emphasis on misdiagnosis factors. J Cancer 2018; 9(16):2817-2824. doi:10.7150/jca.25883. https://www.jcancer.org/v09p2817.htm

26. Gandhi, T., & Bhatt, H. (2021). Cancer Antigen 125. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK562245/
27. Oakden-Rayner, L., Dunnmon, J., Carneiro, G., & Ré, C. (2019). Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. ArXiv:1909.12475 [Cs, Stat]. https://arxiv.org/abs/1909.12475
28. Foti, P. V., Attinà, G., Spadola, S., Caltabiano, R., Farina, R., Palmucci, S., Zarbo, G., Zarbo, R., D'Arrigo, M., Milone, P., & Ettorre, G. C. (2015). MR imaging of ovarian masses: classification and differential diagnosis. Insights into Imaging, 7(1), 21–41. https://doi.org/10.1007/s13244-015-0455-4
29. Clayton, R. D., Snowden, S., Weston, M. J., Mogensen, O., J. Eastaugh, & Lane, G. (1999). Neural networks in the diagnosis of malignant ovarian tumours. BJOG an International Journal of Obstetrics & Gynaecology, 106(10), 1078–1082. https://doi.org/10.1111/j.1471-0528.1999.tb08117.x
30. Jibril, M. M., Baffah, M. D., & Mohammed, M. B. (2024). Investigating the Prevalence and Risk Determinants of Ovarian Cancer in Women from Bauchi State North-East, Nigeria. MedRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2024.10.03.24314789