QUBAHAN ACADEMIC JOURNAL
VOL. 5, NO. 3, September 2025
https://doi.org/10.48161/qaj.v5n3a2064

Ovarian Tumor Detection Using Machine Learning Algorithms and Local Texture Features

Nidaa Ghalib Ali [1], Farsad Zamani Boroujeni [2, 3*], Sahar Adill Kadum [1], Mehdi Akbari Kopayei [4, 5] and Mahdi Mosleh [1]

[1] Department of Computer Engineering, Isf.C., Islamic Azad University, Isfahan 81595-158, Iran;
[2] Department of Computer Engineering, SR.C., Islamic Azad University, Tehran 14515-775, Iran;
[3] Artificial Intelligence and Data Analysis Research Center, SR.C., Islamic Azad University, Tehran 14515-775, Iran;
[4] Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad 85141-43131, Iran;
[5] Big Data Research Center, Najafabad Branch, Islamic Azad University, Najafabad 85141-43131, Iran.

* **Corresponding author:** farsad.zamani@iau.ac.ir.

ABSTRACT: Ovarian cancer (OC) is considered the fifth leading cause of death among women globally, and early detection of ovarian cancer symptoms will be vital for ovarian cancer treatment. Its detection via image-based clinical diagnosis is often prone to misclassification. Therefore, this study proposes using machine learning methods to reduce such errors and deliver faster and more accurate results. Two ovarian tumor datasets consisting of ultrasound images were analyzed using nine machine learning classification algorithms. To reduce the extracted features derived from four pre-trained convolutional networks and five texture-based features to fewer than 50. A hierarchical feature selection method was proposed, combining the ReliefF filter algorithm in the first stage and ten metaheuristic algorithms in the second stage. The results showed that ensemble algorithms, such as LightGBM, achieved a high accuracy of over 95% in diagnosing various types of ovarian tumors using both 2D and 3D ultrasound images. Among the feature selection approaches, the combination of ReliefF and Quantum Approximate Neighbourhood Analysis (QANA) yielded the best performance. Experimental findings on real datasets show that the suggested method not only preserves data confidentiality but also yields excellent performance in early and precise detection of OC.

Keywords: ovarian cancer tumor, ultrasound imaging, hierarchical feature selection, deep learning features, texture features, metaheuristic algorithms.

I. INTRODUCTION

Ovarian cancer is the fifth most common cancer and one of the leading causes of cancer-related death in women worldwide, which is defined by the unchecked growth of aberrant cells in ovarian tissue [1]. Many patients with early-stage ovarian cancer show no symptoms at all. Even when symptoms do appear, they are often nonspecific and ambiguous, resembling those of other illnesses. Consequently, early detection through routine screening is highly constrained. OC is known as "the silent killer" because it is often detected at an advanced phase [2]. Research shows that early diagnosis can raise survival rates by roughly 10% to 30%, highlighting the serious need for intelligent systems capable of detecting diseases early [3].

At the moment, predictive models based on imaging have shown encouraging findings in the classification of ovarian tumors. These methods improve the precision of disease diagnosis and prognosis by utilizing case-based reasoning and avoiding the drawbacks of traditional diagnostic algorithms. Additionally, by minimizing human error, they are essential in lowering patient mortality [4-7].

In the classification of ovarian tumors, artificial neural networks (ANNs) have demonstrated significant effectiveness. Current studies have employed deep learning-based approaches to further improve classification outcomes [8,9]. In [10], a deep convolutional neural network (DCNN), inspired by the pretrained AlexNet method, was proposed for the forecasting and subtyping of OC using histopathological images. The network architecture was optimized by inserting a layer of max-pooling after every pair of convolutional layers and substituting the Rectified Linear Unit (ReLU) activation function with the exponential linear unit (ELU). This enhanced design achieved a prediction accuracy of 83.93% for both ovarian cancer detection and subtype classification. Similarly, studies [2,11] introduced methods based on the pre-trained Visual Geometry Group 16-layer network (VGG-16) and Xception CNN models, which attained ovarian cancer classification accuracies of 84.64% and 0.85, respectively. Another investigation [12] utilized ten CNN models including GoogleNet, Residual Network (ResNet), and AlexNet to evaluate their performance. After training, ensemble learning was applied to the top three models, demonstrating the best trade-off between computational cost and classification time. The highest achieved accuracy among the ten deep networks was 90.51%.

One of the critical limitations of such approaches is the insufficient consideration of textural features, which are highly informative for differentiating tumor types but are often overlooked [9, 12]. These features are generally captured in the initial convolutional layers and tend to be diminished or lost due to dimensionality reduction through successive pooling layers. Consequently, deep networks may fail to retain essential tumor-specific details, leading to suboptimal classification outcomes. Furthermore, the fine-grained classification of ovarian tumor subtypes has not been addressed by the majority of previous studies, which have concentrated on binary classification (benign vs. malignant) using Magnetic Resonance Imaging (MRI) images.

Unlike earlier approaches, this study makes use of ultrasound imaging, which has a number of benefits over MRI. Ultrasound is more widely available, more affordable, and appropriate for use in both remote and routine clinical settings. Unlike MRI, it is safe for all patients, including those with metallic implants, and enables real-time imaging for dynamic evaluation of ovarian structures. Additionally, it is ideal for routine monitoring due to its non-invasive nature, and when used appropriately, it can differentiate between benign and malignant masses. Another key distinction of this study lies in the integration of nine distinct feature categories, including five textural descriptors and four deep learning-based features, thereby facilitating accurate identification of ovarian tumor types.

Even though deep learning models have made great strides in the detection of ovarian cancer, three main issues are still not fully addressed. Firstly, the majority of previous studies has primarily focused on MRI images, which are more expensive, harder to obtain, and inappropriate for routine or remote screening. Second, previous researches have mostly ignored the fine-grained differentiation of tumor subtypes, which is essential for accurate treatment planning in favor of binary classification (benign vs. malignant). Third, during feature extraction, textural features which are highly informative for tumor characterization have often been neglected or underutilized in conventional CNN architectures. These gaps highlight the need for an integrative approach that uses both deep and textural features on ultrasound images data to enhance diagnostic accuracy, interpretability, and clinical applicability. By combining a hybrid feature-based framework with hierarchical feature selection and a variety of machine learning classifiers, this study overcomes these drawbacks.

Unlike prior studies that predominantly relied on MRI images data and binary classification frameworks, this study introduces several novel contributions. First, it is among the few attempts to employ ultrasound imaging for ovarian tumor classification, making the approach more accessible, cost-effective, and clinically practical. Second, the framework integrates nine distinct categories of features five textural descriptors and four deep learning-based representations thus bridging the gap between handcrafted and learned features, which are often treated separately in existing research. Third, the research offers a methodical way to keep the most discriminative features while cutting down on redundancy by employing a hierarchical feature selection approach that blends ReliefF with ten metaheuristic optimization algorithms. Lastly, the comparison of the suggested method with nine cutting-edge classifiers demonstrates its generalizability and resilience. When taken as a whole, these contributions create a novel approach that goes beyond current

ovarian cancer detection systems and provides a scalable framework for implementation in actual healthcare environments. The key contributions of this study are listed as follows:

- Improved diagnostic accuracy: The proposed method enhances the diagnostic accuracy for OC, surpassing the performance limitations of MRI-based methods (with an accuracy of 85%) and reducing human errors. By incorporating nine feature categories (five textural and four deep).
- Effectiveness of machine learning algorithms: The performance of nine classifiers is investigated, including Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), Attention-based Multilayer Perceptron (MLP), and Naive Bayes (NB).
- Evaluation of hierarchical feature selection: A hierarchical feature selection strategy is employed, comprising the ReliefF algorithm at the first level, followed by ten metaheuristic optimization algorithms at the second level. These include the Improved Grey Wolf Optimizer [13], QANA [14], Chinese Pangolin Algorithm [15], a hybrid Firefly Particle Swarm Optimization method [16], Whale Optimization Algorithm [17], Zebra Optimization Algorithm [18], Secretary Bird Algorithm [19], Wild Horse Optimizer [20], Hippopotamus Algorithm [21], and Marine Predators Algorithm [22].

The rest of this paper is organized as follows: Section 2 offers the materials and methods employed in the proposed study. Section 3 reviews the related literature. The suggested methodology is presented in Section 4, and the findings analysis and interpretability of AI are the main topics of Section 5. Lastly, Section 6 offers suggestions for further study, methods for improvement, and a thorough conclusion.

II. MATERIALS AND METHODOLOGY

The approach consists of efficient feature extraction (FE) and selection processes, which are followed by the application of several classification models designed to distinguish between various ovarian tumor types. Several assessment metrics, including accuracy, F1-score, precision, recall, and specificity, are carefully utilized and discussed in order to assess the predictive performance of these models. Each classifier's working mechanism is also covered in detail to give readers a comprehensive understanding of the classification procedure as a whole.

1. BASELINE CLASSIFICATION ALGORITHMS FOR COMPARATIVE ANALYSIS

A collection of sophisticated tree-based ensemble learning algorithms was used in this study to increase model generalization and classification accuracy. While XGBoost and LightGBM, as gradient boosting techniques, provide high predictive power with controls to prevent overfitting, Random Forest and Extra Trees improve prediction robustness by building diverse ensembles of decision trees. By managing missing values and handling categorical features natively, CatBoost enhances this framework even more and provides increased accuracy without requiring a lot of preprocessing [21-25].

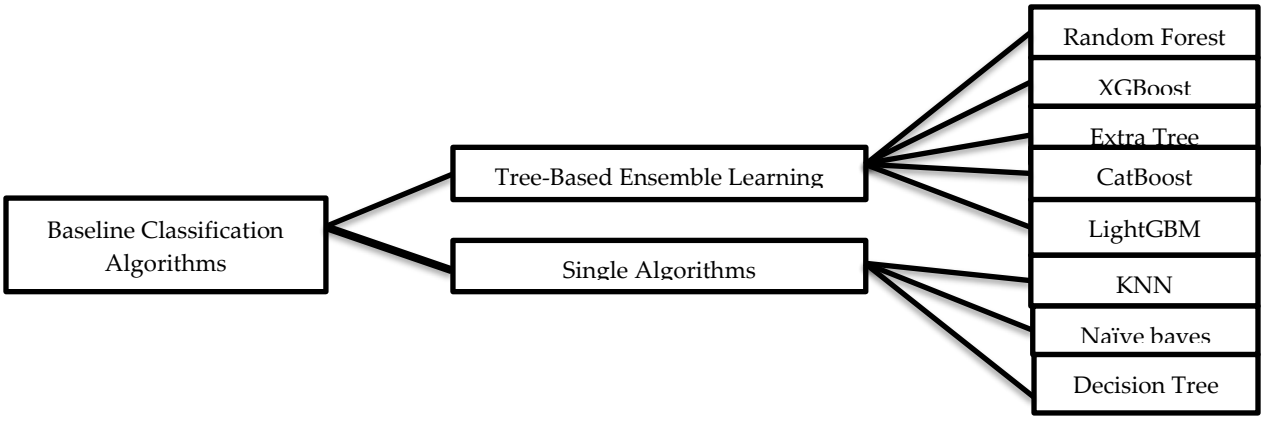


FIGURE 1. Overview of tree-based ensemble and baseline classification algorithms.

Baseline classifiers Decision Tree, K-Nearest Neighbours, and Naïve Bayes were also used for comparative analysis. Despite their simplicity, these algorithms offer competitive performance and interpretable models, making them valuable standards by which the efficacy of more intricate ensemble approaches can be assessed [26-28]. An overview of the tree-based ensemble and baseline classification algorithms used in this study is presented in Figure 1.

2. DEEP FEATURE EXTRACTION USING TRANSFER LEARNING

Transfer learning allows the reuse of pre-trained models on related tasks, reducing training time and computational cost while maintaining strong performance with limited labeled data [26]. Three effective convolutional neural network architectures were used as feature extractors in this study: ShuffleNet, MobileNet, and ResNet (ResNet-50, ResNet-101). ResNet facilitates the training of deep networks through residual connections [27], ShuffleNet provides lightweight efficiency using group convolutions with channel shuffling [28], and MobileNet achieves high accuracy in resource-constrained environments through depthwise separable convolutions [29]. The categorization of pre-trained CNN architectures employed for feature extraction is illustrated in Figure 2.

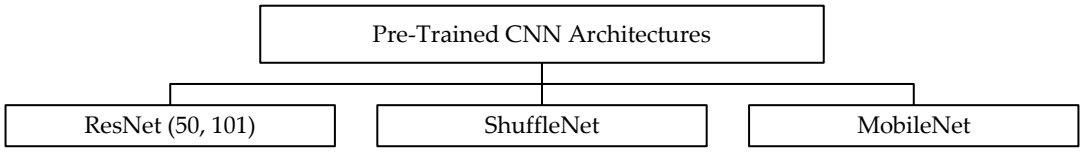


FIGURE 2. Categorization of pre-trained CNN architectures for feature extraction.

3. TEXTURE FEATURE EXTRACTION

As illustrated in FIGURE 3, five types of local texture features were extracted in the proposed model, which are detailed in the following sections.

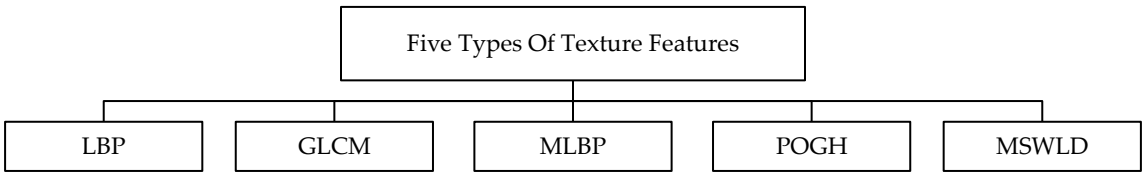


FIGURE 3. Five types of local texture features were extracted in the proposed model.

4. LOCAL BINARY PATTERN TEXTURE FEATURE

The Local Binary Pattern (LBP) is a prominent indicator in image processing and computer vision, widely used for characterizing and extracting local texture features from images [30, 31]. This descriptor is computed for each neighborhood in the image based on Equation (1):

$$LBP_{(P,R)} = \sum_{i=0}^{P-1} s(g_i - g_c)2^i, \qquad where\ s(x) = \begin{cases} 0 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{cases} \tag{1}$$

In Equation (1), P refers to the number of neighboring points surrounding the center, R represents the radius of the neighborhood, g_i corresponds to the intensity of the neighboring points, and g_c is the intensity of the central point. The function s is the sign function [30, 31].

5. HARALICK TEXTURE FEATURES

The gray-level co-occurrence matrix (GLCM), a two-dimensional mapping that documents the frequency of simultaneous occurrence of pixel pairs with specific intensity values (positive or negative) at specific spatial offsets within a given image region, is the source of Haralick features. A thorough examination of the texture of the image is made possible by the co-occurrence matrix's structure, which is based on pixel pairs and their spatial relationships. Key Haralick features include energy, contrast, correlation, information measure of correlation, and entropy, whose mathematical definitions are provided in Equations (2-5) [32]:

$$Energy = \sum_{i,j} P(i,j)^2 \tag{2}$$

$$Contrast = \sum_{i,j} (i,j)^2 P(i,j) \tag{3}$$

$$Corrilation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma^2}, \qquad where \begin{cases} \sigma = \sum_{i,j} (i-\mu)^2 P(i,j) \\ \mu = \sum_{i,j} iP(i,j) \end{cases} \tag{4}$$

$$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \tag{5}$$

In Equations (2) to (5), $P(i,j)$ denotes the joint probability of the distribution of two gray stages i and j in neighboring positions [32].

6. PYRAMID OF ORIENTED GRADIENT HISTOGRAMS

The Pyramid of Oriented Gradient Histograms (POGH) is an advanced image descriptor developed based on the fundamental Histogram of Oriented Gradients (HOG) algorithm. In this approach, the image is initially partitioned into smaller, localized regions, and a histogram of gradient orientations is independently computed for each region. These histograms are then hierarchically aggregated in a pyramidal structure, where higher-level histograms are derived by combining the finer, local histograms from lower levels [33].

7. MEDIAN BINARY PATTERN AND MEDIAN TERNARY PATTERN

The Local Binary Pattern (LBP) is expanded upon by the Median Binary Pattern (MBP). This method computes the median gray-scale value, Mc, within a 3 × 3 local neighborhood that is defined surrounding each pixel. The 8-bit binary code of the central pixel is then generated through the intensity values of the neighboring pixels with the mid-value Mc. This comparison process results in the binary code MBP, computed using Equation (6) [34]:

$$MBP(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p) \times 2^p, \qquad where\ s(x) = \begin{cases} 1 & if\ x < Mc \\ 0 & otherwise \end{cases} \tag{6}$$

Here, Mc denotes the median gray-level value of the local neighboring pixels surrounding the central pixel located at (x_c, y_c), and (i_p) corresponds to the gray-level intensity of the local neighboring pixels [34].

In the Median Ternary Pattern (MTP), a 3 × 3 local neighborhood is initially defined surrounding each pixel, and the median gray-level value of the nine neighboring pixels is computed. The MTP code is then generated based on this value using Equation (7) [34]:

$$S_{MTP(v)} = \begin{cases} 1 & v > M_c + t \\ 0 & M_c - t \le v \le M_c + t \\ -1 & v < M_c - t \end{cases} \tag{7}$$

Here, v represents the gray-level intensity of a neighboring pixel, Mc denotes the local midi-value, and t is a user-defined threshold.

8. MSWLD: MULTI-SCALE WEBER LOCAL DESCRIPTOR

The Multi-Scale Weber Local Descriptor (MSWLD) quantifies the relative contrast between a pixel and its neighboring pixels by leveraging Weber's law of perceptual intensity differences. This descriptor comprises two principal components: differential excitation, which captures the local intensity variations, and orientation, which reflects the angular direction of these variations [35].

The differential excitation is computed as the ratio of responses from two convolution filters and is mapped to the range $[-\pi/2, \pi/2]$ using the arctangent function. This angle is then quantized into $T1$ discrete bins. For the orientation component, the horizontal-to-vertical gradient ratio is derived using Sobel filters. The resulting orientation angle is transformed into the range $[0,2\pi]$ and further quantized into $T2$ directional bins. Finally, the combination of these components forms a two-dimensional histogram $\{WLD(\xi i, \psi j)$, where each cell represents the frequency of a specific excitation level at a given direction. To simplify representation, this 2D histogram is linearized into a one-dimensional feature vector of length $1 \times T2$, effectively encoding the texture pattern across different scales and orientations [35].

9. FEATURE SELECTION: BASED ON METAHEURISTICS ALGORITHMS

In order to reduce dimensionality and enhance the performance of machine learning models, feature selection is essential. Current techniques can be broadly divided into three categories: filter, wrapper, and embedded. Wrapper-based approaches, especially those that use metaheuristic algorithms, have garnered a lot of attention because of their accuracy and adaptability, even though filter methods depend on statistical measures and embedded methods incorporate feature selection during model training. Intelligent search strategies are used by metaheuristics, such as swarm intelligence-based methods like Particle Swarm Optimization (PSO) and evolutionary algorithms like Genetic Algorithm (GA), to find discriminative feature subsets. Keeping a healthy balance between exploring new areas of the search space and taking advantage of promising solutions is crucial to their efficacy.

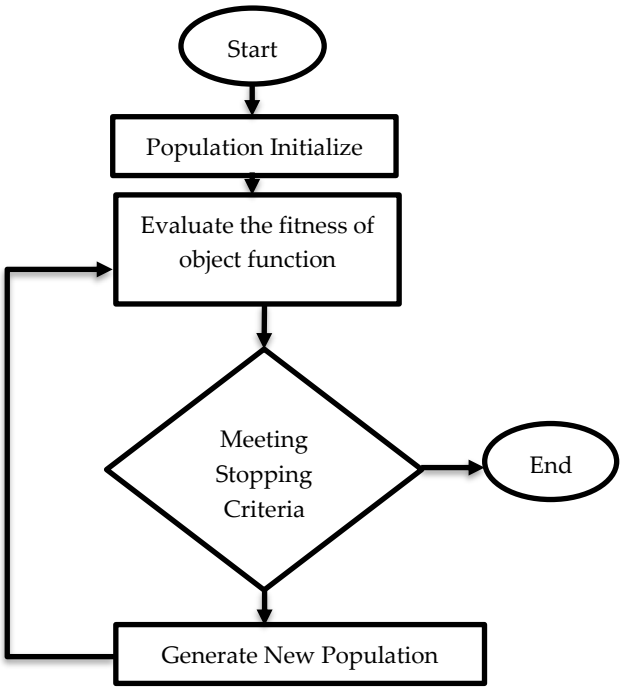


FIGURE 4. A generic diagram of the metaheuristic feature selection algorithm [39].

The general structure of metaheuristic approaches typically includes four main phases: initialization of the population, fitness evaluation, population update, and the application of termination criteria [36,37]. Several prominent algorithms utilized in this study include the Marine Predators Algorithm [38], the Whale Optimization approach [18], Wild Horse Optimizer [21], QANA Optimization method [15], Secretary Bird Algorithm [20], Firefly-PSO [17], Grey Wolf Optimizer [14], Zebra Optimization Algorithm [19], Hippopotamus Algorithm [22], and the Chinese Pangolin Algorithm [16]. A general flow diagram illustrating the metaheuristic-based feature selection process is presented in Figure 4.

III. RELATED WORKS

Although medical imaging technologies have advanced significantly, more work is still needed in terms of clinical integration and practical application. More than just technology improvements are needed to make the shift to "smart healthcare"; fundamental adjustments must be made in engineering domains as well as in response to evolving clinical standards and requirements. Handling the distinct characteristics of each imaging system while also meeting the new demands of modern healthcare, where data-driven approaches like artificial intelligence and machine learning are crucial, is therefore the primary challenge.

In this context, various studies have been conducted. For instance, in this study [40] proposed a hybrid model that merges clinical data with ultrasound imaging for the identification of polycystic ovary syndrome (PCOS). Utilizing a deep neural network to analyze ultrasound images, they achieved an accuracy of 84.81% in distinguishing between polycystic and non-polycystic ovaries. In another study [41] employed Otsu's thresholding and the Chan-Vese segmentation method to detect and segment polycystic ovaries, attaining an accuracy of 89.93% for binary classification using ultrasound data. In [42], leveraging transfer learning with deep convolutional networks such as AlexNet, Inception V3, ResNet50, and VGG16, developed a predictive system for PCOS, achieving an accuracy of 93% in binary classification. Similarly, this study [43] utilized convolutional neural networks to build a precise diagnostic model, reaching an accuracy of 83%, although this approach, like others, focused solely on binary classification.

On a different front, [44] applied a machine learning method to evaluate the efficacy of treatment strategies. Additionally, [45] investigated five approaches - CART, SVM, Naive Bayes, logistic regression, and multilayer perceptron (MLP) - to predict PCOS, with the primary aim of eliminating expensive and time-consuming clinical features while delivering a cost-effective and accurate model. Their findings showed that the MLP approach yielded optimal performance, with an accuracy of 93%. In another study, the authors [46] introduced a hybrid model for PCOS detection that involved preprocessing ultrasound images using a bilateral adaptive filter and analyzing them with a novel deep neural architecture called AResUNet. The adaptive filter effectively reduced noise in the images, while the AResUNet architecture demonstrated adaptability to both 2D and multimodal medical data. The method achieved an accuracy of 97% on the MMOTU dataset.

In [47] employed the Extreme Gradient Boosting algorithm was employed for the diagnosis of polycystic ovary syndrome (PCOS). To address the issue of data imbalance, they utilized the Synthetic Minority Over-sampling Technique (SMOTE) and achieved an accuracy of 96.03%. Similarly, this study [48] evaluated machine learning algorithms using clinical data and found that the ensemble-based Random Forest algorithm surpassed basic machine learning methods, yielding an accuracy of 93.25%. In [49] investigated the use of seven different machine learning algorithms for PCOS diagnosis was investigated, reporting that the Linear Discriminant algorithm achieved the highest accuracy, while the K-Nearest Neighbors (KNN) method demonstrated the highest sensitivity.

This study [50] developed two convolutional neural network (CNN) models for diagnosing PCOS using ovarian ultrasound images: a custom-designed model named PCONet and a fine-tuned version of InceptionV3. Their results showed that PCONet achieved superior performance with an accuracy of 98.12%, compared to 96.56% for InceptionV3. In [51] focused on ovarian cancer detection in transgenic mice using Optical Coherence Tomography (OCT) images. They employed three VGG networks, a 3D CNN, and a convolutional LSTM model. Despite the presence of noise in OCT images, the models exhibited promising

performance. The LSTM-based neural network yields the best findings, with a mean Area Under Curve (AUC) of 0.81.

In a 2022 study, the authors [52] proposed a novel architecture that integrates morphometric features with deep features extracted by a convolutional network from preprocessed images. The model achieved a validation AUC score of 0.99. The enhanced feature extraction stage enabled effective discrimination between malignant and non-malignant samples. Moreover, this study [8] developed a DCNN method for automating the analysis of ultrasound images, aiming to simplify ovarian cancer diagnosis compared to existing algorithms. Their model achieved an accuracy ranging from 83% to 87%. Lastly, this study [53] introduced a deep learning approach for classifying ovarian cancer in histopathological images using a convolutional neural network. Feature extraction was performed using the pre-trained DenseNet-201 model. Evaluations showed an accuracy of 0.91 and a recall of 0.9.

Despite these developments, the textural features of medical images have not been sufficiently considered in many studies on the diagnosis of ovarian tumors. Relying exclusively on these features often results in the loss of crucial textural information during the early stages of processing. The architecture of deep neural networks typically reduces or eliminates these features through intermediate and final layers. This limitation makes it more difficult for deep networks to accurately distinguish between different kinds of ovarian tumors, which may result in a poor or insufficient diagnosis. Therefore, in order to enhance the performance of diagnostic models and the differentiation of tumor types, it is imperative to optimize the deep learning architecture to include detailed textural and morphometric characteristics in addition to high-level and superficial features. Table 1 presents a review of related work.

Table 1. Review of related work.

Ref.	Model	Data Type	Performance
[41]	DNN	Ultrasound imaging	Acc. = 84.81%
[41]	Chan-Vese segmentation algorithm	Ultrasound imaging	Acc. = 89.93%
[42]	Transfer learning with deep AlexNet	Ultrasound images	Acc. = 93%
[43]	CNN	Ultrasound images	Acc. = 83%
[44]	Machine learning algorithms	-	-
[45]	Machine learning - CART	-	Acc. = 93%
[46]	AResUNet	Ultrasound imaging	Acc. = 97%
[47]	Extreme gradient boosting	Kaggle dataset	Acc. = 96.03%
[48]	Ensemble-based Random Forest	Clinical data	Acc. = 93.25%
[49]	Linear Discriminant algorithm	Ultrasound images	Acc. = 92.60%
[50]	PCONet	Ultrasound images	Acc. = 98.12%
[51]	LSTM	Optical Coherence Tomography (OCT) images	AUC = 0.81
[52]	CNN	immunohistochemistry images	AUC = 0.99
[8]	CNN	Ultrasound images	Acc. = 83-87%
[53]	CNN	Histopathological images	Acc. = 0.91%

Note: Acc. represents Accuracy, AUC represents Area Under Curve.

1. THEORETICAL FOUNDATION

The design of this study is also informed by established learning and behavior change theories. From the perspective of adult learning theory, the proposed system directly addresses a real-world healthcare challenge, thereby aligning with the principle that effective learning and decision-making are problem-centered and practically relevant. Moreover, constructivist perspectives emphasize that knowledge emerges from the interpretation of complex information, which in this study is facilitated by transforming raw ultrasound data into interpretable diagnostic insights. Lastly, the hierarchical feature selection strategy is supported by cognitive load theory, which helps clinicians make more accurate and effective decisions by highlighting the most discriminative patterns and reducing informational redundancy. These theoretical underpinnings support the study's methodological decisions and highlight how it can facilitate meaningful learning and clinical decision-making.

IV. PROPOSED METHOD USING HIERARCHICAL FEATURE SELECTION

This study proposes an integrated framework for the segmentation and classification of ovarian tumors in ultrasound images. To extract five distinct categories of texture-related features, precise tumor region segmentation is performed using two advanced deep learning models: UNet3++ and DeepLabV3. Following segmentation, the accurate classification of tumor types is achieved by fusing two complementary feature sets: deep features extracted from pre-trained convolutional neural networks [54, 55], and handcrafted local and texture-based features derived from the ultrasound images [56-58]. A hierarchical feature selection mechanism is then employed to refine the feature space and facilitate final tumor type recognition. The overall architecture of the proposed model is illustrated in Figure 5.

1. EXTRACTION OF LOCAL TEXTURE FEATURES FROM TUMOR REGIONS

In the initial phase of the proposed model, semantic segmentation is employed to isolate tumor area from the ultrasound images. The primary objective of this step is to prepare the images for the extraction of local and texture features specific to the tumor areas. To this end, two advanced pre-trained deep learning models, UNet++ and DeepLabV3++, are utilized. Both networks process input images resized to a fixed dimension of 256 × 256 × 3. The models are fine-tuned on the uterine ultrasound dataset used in this study, employing the Adam enhancement method with 200 training epochs and a batch size of 2, to accurately distinguish tumor regions from the surrounding background. Figure 6 illustrates the segmentation of the ovarian tumor region.

Following semantic segmentation and the isolation of tumor regions from the ultrasound images, each segmented tumor patch was resized to dimensions of 300 × 300 pixels. Subsequently, six categories of local texture features were extracted from each patch as detailed below:

- Local Binary Pattern (LBP): For LBP extraction, the radius (R) was fixed at 1 and the number of neighbors (P) was set to 8. Each image produced a 256-dimensional feature vector with this setup [30, 31].
- Haralick Texture Features: Four statistical texture descriptors were included in this set: homogeneity, contrast, energy, and correlation. A Gray-Level Co-occurrence Matrix (GLCM) of size 8 × 8 was calculated after the image's grayscale intensity was quantized into 8 levels (Num Levels = 8). With a directional angle of 0°, the offset values for the row and column were set to 0 and 1, respectively. A 4-dimensional feature vector was produced as a result of this step [32].
- Pyramid Histogram of Oriented Gradients (PHOG): This descriptor had the pyramid level set to 2, the number of bins set to 8, and the gradient orientation angle set to 360°. The dimensionality of the resultant feature vector was 168 [33].
- Median Binary Pattern (MBP): This descriptor creates a 256-dimensional vector by capturing median-based patterns from textured areas [34].
- Triple Median Pattern (TMP): As an extension of MBP, this descriptor generated a feature vector of length 512 for each image [34].
- Weber Local Descriptor (WLD): For WLD extraction, the number of quantization levels for gradient orientations was set to 8, the differential excitation to 4, and the number of WLD scales to 1. The final output of this descriptor was a 32-dimensional feature vector [35].

By concatenating the feature vectors from all six descriptors, a comprehensive texture representation of each image was obtained, resulting in a final feature vector of length 1228, formulated as follows:

$$Final\ Feature\ Vector\ Length = 256(LBP) + 4(Haralick) + 168(PHOG) + 256(MBP) + 512(TMP) + 32(WLD) = 1228$$

2. DEEP FEATURE EXTRACTION FROM ENTIRE ULTRASOUND IMAGES

To extract deep representations from the full ultrasound images, four pre-trained convolutional neural networks - ResNet101, ResNet50, Shuffle Net, and MobileNetV2 - were employed. Since these architectures require input images of size 224 × 224 × 3, the original ultrasound images were resized accordingly.

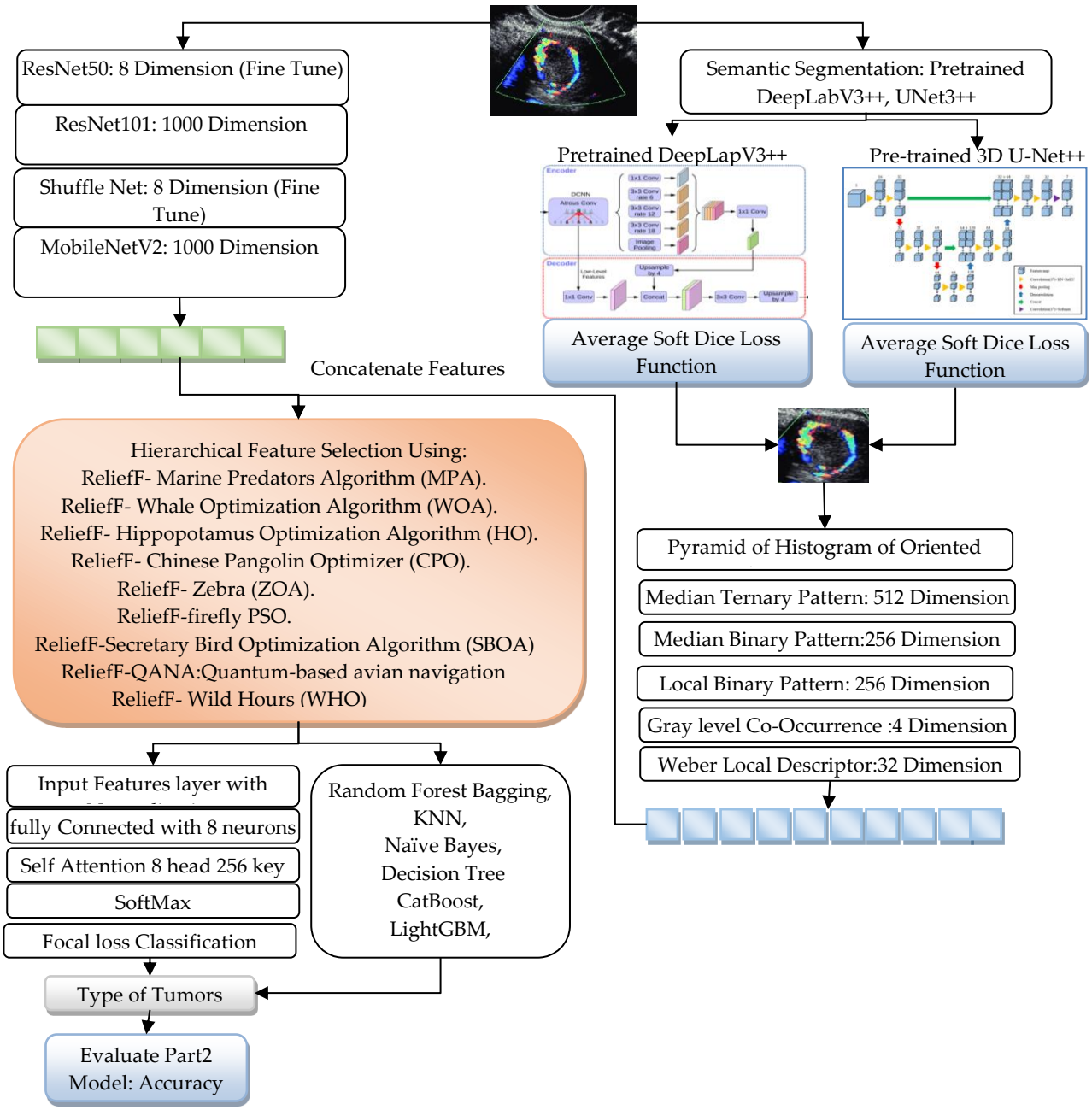


FIGURE 5. The overall framework of the proposed ovarian cancer tumor detection methodology.

Among these, ResNet50 and Shuffle Net were specifically fine-tuned using ultrasound images from the two datasets utilized in this study. Fine-tuning was conducted by freezing the weights of the early layers and retraining only the final layers to adapt the networks to the target domain. The training process was performed using the Adam improver, a fixed learning rate, 30 epochs, and a mini-batch size of 128. Feature extraction was carried out by retrieving the output of the final fully connected layer of each network. By concatenating the feature vectors extracted from all networks, a final deep feature vector of length 2016 was obtained.

FIGURE 6. Segmentation of the ovarian tumor region.

3. HIERARCHICAL FEATURE SELECTION

In medical image analysis, feature extraction is essential. Several feature extraction methods are frequently used to produce high-dimensional feature vectors in order to obtain thorough diagnostic insight. However, such increased dimensionality significantly intensifies the computational complexity and runtime of subsequent classification models. To deal with this limitation, an efficient feature selection strategy becomes essential to preserve model performance while minimizing computational burden.

Accordingly, a hierarchical feature selection method is proposed, combining the ReliefF algorithm with ten distinct metaheuristic optimization algorithms. These include: Marine Predators Algorithm (MPA), Improved Grey Wolf Optimizer (IGWO), Whale Optimization Algorithm (WOA), Wild Horse Optimizer (WHO), Secretary Bird Optimization Algorithm (SBOA), Hybrid Firefly-Particle Swarm Optimization (FiPSO), Quantum-based (QANA), Zebra Optimization Algorithm (ZOA), Hippopotamus Optimization Algorithm (HO), and Chinese Pangolin Optimizer (CPO). The overall structure of the proposed hierarchical selection framework is illustrated in Figure 7. In this approach, deep features and texture-based features are first concatenated, yielding a comprehensive feature vector of length 3244, which serves as the input to the proposed selection pipeline.

In the first stage of the hierarchical method, the full feature matrix is processed by the ReliefF algorithm, which was configured with five nearest neighbors. ReliefF estimates the importance of each feature by comparing its values among neighboring instances within the same class and across different classes. This method effectively identifies features that are highly discriminative in distinguishing between classes. Subsequently, the top 50 features selected by ReliefF are provided as inputs to the fitness function for evaluation within the metaheuristic optimization phase. In this phase, the initial population is randomly generated, with each individual represented as a binary vector of 50 features (columns) corresponding to the features retained from the ReliefF output. Consequently, the initial population matrix has a 50×50 size, with rows representing potential solutions (search agents) and columns representing particular features.

Each metaheuristic algorithm has a maximum of 500 iterations, and the initial population values are uniformly sampled from the interval [-100, 100]. Each algorithm looks for the ideal subset of features that minimizes redundancy and maximizes classification performance through iterative refinement. The fitness (cost) function at this stage is defined as the classification error obtained by employing the Decision Tree (DT) algorithm, computed according to Equation (8). Given the non-parametric nature of decision trees, their use as an evaluation criterion within metaheuristic optimization algorithms proves highly appropriate. This is because the performance of DTs is independent of input parameter tuning, allowing for a more reliable and robust assessment of the quality of candidate feature subsets.

$$Cost = 1 - \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

During each iteration of the metaheuristic algorithms, after the population members are updated and moved toward the global optimum, the components of each individual with values greater than zero are identified. The corresponding indices are then extracted from the matrix of selected features, which was obtained from the output of the ReliefF algorithm. These indices determine the feature subset evaluated in that iteration.

Subsequently, the feature matrix corresponding to the training set is provided to the Decision Tree (DT) approach to perform the learning process. The model then utilizes the feature matrix of the test set to predict the tumor type. By comparing the predicted labels with the ground truth of the test set, the classification accuracy for each individual in the population is computed. This accuracy is substituted into Equation (9) and recorded as the cost (fitness) value associated with that population member. This process is iteratively performed for all individuals across 500 iterations.

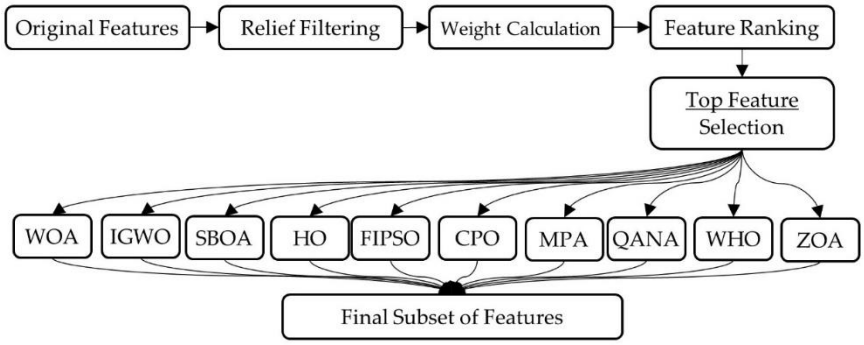


FIGURE 7. Schematic diagram of the proposed hierarchical feature selection process.

At the conclusion of the optimization process, the individual yielding the minimum cost value (i.e., the highest classification accuracy and lowest classification error) is identified as the optimal solution. This individual represents the indices of the most informative and discriminative features. Let x denote the best individual in the population, and xi its i −th component. The features corresponding to elements with values greater than zero are selected using Equation (9). The convergence curves of several metaheuristic algorithms employed in this framework are illustrated in Figures 8-9.

$$Best\ member\ vector = index[x' > 0], \qquad [x_i'] = \begin{cases} 1 & if\ x_i > 0 \\ 0 & otherwise \end{cases} \quad (9)$$

The output of the metaheuristic algorithms typically comprises over 20 selected features, often approximating N/2, where N represents the total number of features generated by the ReliefF algorithm.

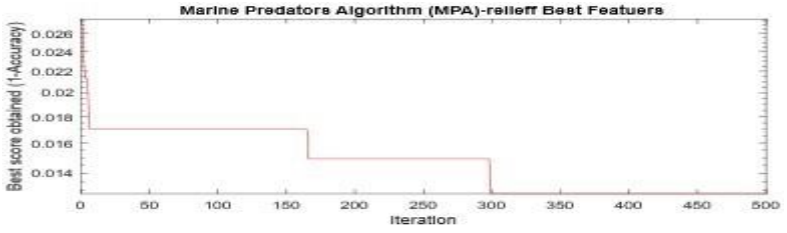


FIGURE 8. Convergence curve of the Marine predator algorithm.

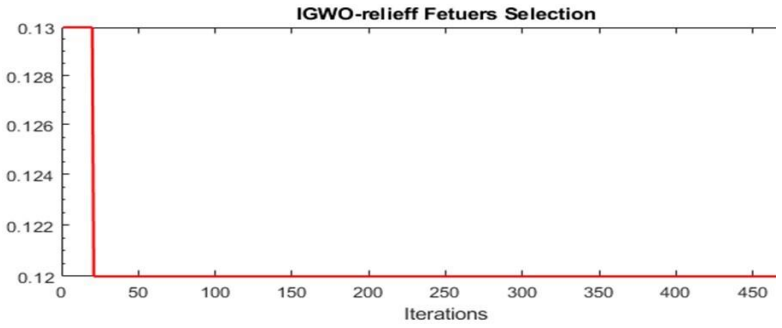


FIGURE 9. Convergence curve of the improved grey wolf optimizer.

4. TUMOR DETECTION AND CLASSIFICATION

- The best feature subset was selected, and nine different classifiers were used for tumor detection and classification: K-Nearest Neighbours (KNN), Decision Tree (DT), Naïve Bayes, Random Forest (RF), XGBoost, Extra Trees, LightGBM, CatBoost, and Multilayer Perceptron with Self-Attention (MLP-SA). The KNN algorithm applied Euclidean distance and standardised Euclidean distance to the two-dimensional and three-dimensional ultrasound datasets, respectively, and set the number of neighbours to 1.
- A bagging technique was used for the Random Forest classifier, and 700 learning cycles were set.
- Without any additional adjustments, the Naïve Bayes and Decision Tree classifiers were implemented with their default parameter settings.
- The XGBoost model was set up with 200 trees, a learning rate of 0.05, and a maximum tree depth of 7. An L2 regularization parameter (λ) of 1 was used, and random sampling rates for both data and features were set at 0.8 to reduce overfitting.
- There were 100 decision trees in the Extra Trees algorithm, with a maximum depth of 25. Two samples were needed to split a node, and one sample was needed per leaf. The square root of the total number of features was utilized to determine the number of features were taken into account at each split.
- There were 500 trees in the LightGBM model, and the learning rate was set at 0.03. To balance accuracy and model complexity, the number of leaves was 64 and the maximum tree depth was 10. The minimum number of samples required to create a child node was fixed at 30. To reduce overfitting, 80% random sampling of both data and features was applied at each tree. Regularization parameters for L1 and L2 were both set to 1.
- A learning rate of 0.1 and 500 training iterations were used in the CatBoost algorithm. To strike a balance between approach complexity and accuracy, the decision tree depth was set to 8. An early stopping mechanism was employed, halting training after 50 consecutive rounds without improvement to prevent overfitting.
- The defined Multilayer Perceptron (MLP) network in this model consists of five layers. In the input layer, Z-score normalization was applied, calculated according to Equation (10). The Z-score function for a random variable X with mean μ and standard deviation σ is defined as follows:
-

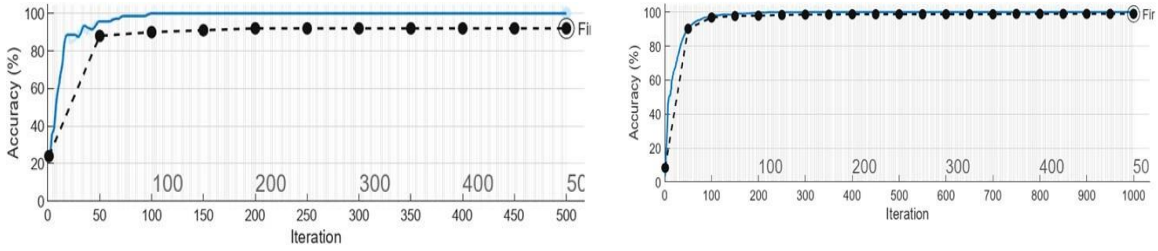
$$Z = \frac{X - \mu}{\sigma} \tag{10}$$

The second layer is a fully connected layer consisting of 8 neurons. The third layer implements a self-attention mechanism, where the number of heads is set to 8 and the key dimension is defined as 256. The self-attention mechanism computes attention scores between different positions within the input sequence.

- The fourth and fifth layers are responsible for classifying the input data and computing the error. In this perceptron network, the focal loss function was employed instead of the conventional cross-entropy loss. The focal loss function is defined by Equation (11), where γ denotes a positive focusing parameter:
-

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(P_t) \quad where\ \alpha_t = \begin{cases} \alpha & if \quad y=1 \\ 1-\alpha & otherwise \end{cases} \tag{11}$$

The focal loss function combines the cross-entropy loss with a modulating factor $(1-p_t)^{\gamma}$, which enhances classification accuracy [59-61]. The deep multilayer perceptron network was trained utilizing the Adam optimizer with 500 epochs and a minibatch size set accordingly. Figure 10 illustrates the learning curves of the network on the 3D and 2D datasets, respectively.



(a) (b)

FIGURE 10. Learning curve of the attention-based MLP on (a) 3D, (b) 2D ultrasound images.

V. PERFORMANCE EVALUATION

This section examines the proposed method for ultrasound imaging. It also provides detailed descriptions of the datasets, evaluation metrics, and comparisons with other state-of-the-art approaches.

1. SIMULATION SETUP

The experiments were conducted using MATLAB R2024 on a system configured to obtain simulation results. The hardware platform consists of an Intel(R) Core (TM) i9-11900K processor operating at 3.5 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 3080 GPU, all running on the Microsoft Windows 10 operating system.

2. IMAGE ACQUISITION

To assess the proposed approach, the multimodal ovarian ultrasound image dataset MMOTU, publicly available on GitHub [62], was utilized. This dataset was originally introduced by [63]. The MMOTU dataset comprises two distinct subsets acquired through different imaging modalities: the OTU_2d subset containing 1,469 two-dimensional ultrasound images, and the OTU_3d subset consisting of 170 Contrast-Enhanced Ultrasound (CEUS) or three-dimensional images. Both subsets include pixel-level semantic annotations and tumor classification labels provided by [63].

The data were collected from Shijitan Hospital, affiliated with Capital Medical University, encompassing a total of 1,639 ovarian ultrasound images from 294 patients. The distribution of images in the OTU_2d subset by class is as follows: mucinous cystadenoma (104), high-grade serous carcinoma (56), theca cell tumor (88), simple cyst (66), normal ovary (267), endometrioma (336), serous cystadenoma (219), and teratoma (336). The model's performance may be impacted by this noticeably unbalanced class distribution, especially since classes like endometrioma and teratoma have significantly more samples than classes like high-grade serous carcinoma or simple cyst.

Similar to this, the OTU_3d subset shows an uneven distribution of images by class, including the following: simple cyst (18), normal ovary (5), mucinous cystadenoma (6), endometrioma (35), serous cystadenoma (40), teratoma (38), theca cell tumor (25), and high-grade serous carcinoma (9). The model's

ability to accurately identify some classes may be hampered by the small number of samples in those classes, particularly high-grade serous carcinoma, mucinous cystadenoma, and normal ovary. Table 2 represents the dataset partitioning.

Table 2. The data partitioning of the MMOTU image dataset.

Data type	Categories	Training data		Testing data	
		Samples	Patients	Samples	Patients
OTU_2d	8	1000	171	469	76
OTU_CEUS	8	70	20	100	27

For this study, an ovarian tumor ultrasound dataset retrieved from GitHub was used, since it was benchmarked in previous studies for the classification of medical images using machine learning, making it accessible and reproducible. The dataset consists of 1639 images of both benign and malignant tumors, which captures the diversity needed for evaluating the proposed methodology. While this study's ovarian tumor dataset sample size is small compared to large-scale clinical datasets, in the context of the clinical internally benchmarked studies, it was in line with other ovarian cancer classification studies and was enough to robustly measure the system's performance. Having an open dataset increases the reproducibility and comparability of the teamwork results across diverse research groups. Regardless of this, the results have limited their relevance for routine clinical practice, and further research using datasets from numerous institutions that better capture real-world patient demographics is necessary to validate the results.

For this research, no observation checklists or other self-created tools were used. Instead, the main source of data was a publicly available ultrasound dataset. The accuracy of this dataset is justified by its use in peer-reviewed research, where it was used to benchmark the performance of machine learning algorithms in medical imaging evaluation. The dataset includes images with pre-labels assigned by clinical specialists, which enhances the precision of the diagnosis and reduces labeling bias. Reliability was ensured by the imaging protocols used during data collection and the ground truth verifiable model with expert adjudication. In addition, the fact that the dataset was publicly available increases the transparency of the research and the possibility of reproducibility which strengthens the reliability and external validity of the study.

3. EVALUATION METRICS

To assess the performance of the proposed approach, the metrics of accuracy, recall, precision, specificity, and F1-score were employed. These metrics were calculated based on Equations (12-16).

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{13}$$

$$F_{Measure} = \frac{2 * precision * Recall}{precision + Recall} \tag{14}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$Specificity = \frac{TN}{FP + TN} \tag{16}$$

4. EVALUATION OF VARIOUS FEATURE SELECTION METHODS ON OTU_CEUS DATA

A key component introduced in the proposed model is the hierarchical feature selection framework. In this section, ten different filter-based methods combined with the ReliefF feature selection algorithm were employed. To identify the most effective approach, comprehensive evaluations were conducted on both datasets. The comparative results of this experiment, performed on the 3D ultrasound data, are presented in Table 3.

Table 3. Evaluation of different feature selection and classification methods on OTU_CEUS Data.

Feature selction methods	Algorithm	Accuracy	Recall	Precision	F1_score	Specificity	Area Under Curve (AUC)
	KNN	0.910	0.931	0.931	0.928	0.985	0.941
	Decision Tree	0.870	0.777	0.869	0.879	0.978	0.941
	Naive Bayes	0.710	0.627	0.711	0.704	0.949	0.977
	Random Forest(Bagging)	0.880	0.822	0.830	0.815	0.980	0.979
MPA	Sel-Attention MLP	0.900	0.923	0.923	0.922	0.982	0.996
	XGBoost	0.830	0.745	0.715	0.832	0.972	0.984
	ExtraTree	0.680	0.615	0.582	0.673	0.944	0.909
	LightGBM	0.900	0.833	0.933	0.854	0.982	0.977
	CatBoost	0.840	0.749	0.727	0.849	0.972	0.961
	KNN	0.930	0.916	0.951	0.928	0.988	0.912
	Decision Tree	0.880	0.795	0.917	0.826	0.979	0.911
	Naive Bayes	0.710	0.547	0.779	0.708	0.949	0.931
	Random Forest(Bagging)	0.850	0.756	0.886	0.778	0.973	0.983
IGWO	Sel-Attention MLP	0.930	0.919	0.920	0.912	0.988	0.991
	XGBoost	0.880	0.773	0.912	0.806	0.979	0.970
	ExtraTree	0.680	0.611	0.564	0.655	0.944	0.912
	LightGBM	0.890	0.758	0.867	0.874	0.981	0.965
	CatBoost	0.900	0.800	0.883	0.904	0.983	0.960
	KNN	0.910	0.921	0.948	0.924	0.984	0.952
	Decision Tree	0.890	0.908	0.872	0.863	0.982	0.934
	Naive Bayes	0.740	0.647	0.782	0.834	0.954	0.916
	Random Forest(Bagging)	0.910	0.913	0.913	0.896	0.985	0.977
WHO	Sel-Attention MLP	0.890	0.906	0.936	0.909	0.981	0.967
	XGBoost	0.870	0.886	0.840	0.830	0.979	0.985
	ExtraTree	0.690	0.660	0.638	0.619	0.946	0.919
	LightGBM	0.920	0.884	0.909	0.886	0.986	0.993
	CatBoost	0.830	0.778	0.744	0.749	0.972	0.982
	KNN	0.940	0.954	0.950	0.951	0.990	0.935
	Decision Tree	0.950	0.843	0.954	0.966	0.992	0.971
	Naive Bayes	0.850	0.759	0.901	0.885	0.972	0.997
	Random Forest (Bagging)	0.930	0.860	0.888	0.870	0.988	0.984
QANA	Sel-Attention MLP	0.970	0.977	0.974	0.974	0.995	0.993
	XGBoost	0.970	0.979	0.977	0.977	0.995	0.997
	ExtraTree	0.690	0.660	0.706	0.631	0.946	0.905
	LightGBM	0.950	0.880	0.898	0.886	0.992	0.989
	CatBoost	0.890	0.819	0.793	0.801	0.982	0.986
	KNN	0.830	0.769	0.853	0.785	0.970	0.876
SBOA	Decision Tree	0.770	0.682	0.775	0.762	0.961	0.839
	Naive Bayes	0.700	0.627	0.751	0.807	0.946	0.995

Feature selction methods	Algorithm	Accuracy	Recall	Precision	F1_score	Specificity	Area Under Curve (AUC)
	Random Forest(Bagging)	0.760	0.765	0.766	0.737	0.958	0.989
	Sel-Attention MLP	0.810	0.810	0.884	0.824	0.966	0.994
	XGBoost	0.770	0.741	0.804	0.712	0.960	0.976
	ExtraTree	0.650	0.591	0.620	0.625	0.938	0.905
	LightGBM	0.770	0.741	0.854	0.755	0.959	0.977
	CatBoost	0.680	0.638	0.718	0.687	0.943	0.896
	KNN	0.750	0.719	0.713	0.696	0.958	0.810
	Decision Tree	0.700	0.554	0.644	0.642	0.948	0.778
	Naive Bayes	0.660	0.568	0.615	0.601	0.941	0.960
	Random Forest(Bagging)	0.750	0.687	0.816	0.697	0.956	0.978
FIPSO	Sel-Attention MLP	0.670	0.639	0.608	0.610	0.944	0.793
	XGBoost	0.780	0.757	0.735	0.721	0.963	0.952
	ExtraTree	0.620	0.608	0.651	0.580	0.933	0.868
	LightGBM	0.840	0.785	0.817	0.789	0.972	0.974
	CatBoost	0.830	0.778	0.864	0.779	0.971	0.924
	KNN	0.850	0.888	0.895	0.887	0.974	0.870
	Decision Tree	0.910	0.929	0.900	0.911	0.985	0.929
	Naive Bayes	0.630	0.585	0.774	0.690	0.934	1.000
	Random Forest(Bagging)	0.830	0.787	0.846	0.778	0.971	0.991
Zebra	Sel-Attention MLP	0.850	0.893	0.887	0.888	0.974	0.987
	XGBoost	0.870	0.848	0.831	0.827	0.978	0.991
	ExtraTree	0.640	0.624	0.613	0.590	0.937	0.898
	LightGBM	0.850	0.784	0.796	0.774	0.974	0.987
	CatBoost	0.860	0.797	0.885	0.816	0.976	0.976
	KNN	0.880	0.827	0.904	0.838	0.978	0.906
	Decision Tree	0.910	0.812	0.922	0.931	0.985	0.981
	Naive Bayes	0.730	0.629	0.814	0.731	0.952	0.988
	Random Forest(Bagging)	0.900	0.837	0.932	0.857	0.982	0.999
WOA	Sel-Attention MLP	0.900	0.835	0.871	0.847	0.982	0.997
	XGBoost	0.940	0.873	0.961	0.892	0.989	0.991
	ExtraTree	0.670	0.644	0.620	0.604	0.942	0.904
	LightGBM	0.890	0.827	0.927	0.848	0.980	0.982
	CatBoost	0.930	0.862	0.921	0.866	0.988	0.986
	KNN	0.920	0.849	0.950	0.873	0.985	0.894
	Decision Tree	0.900	0.804	0.938	0.933	0.981	0.861
	Naive Bayes	0.730	0.648	0.802	0.840	0.952	0.926
	Random Forest(Bagging)	0.840	0.780	0.889	0.784	0.971	0.978
HO	Sel-Attention MLP	0.910	0.847	0.873	0.855	0.984	0.991
	XGBoost	0.830	0.777	0.819	0.779	0.969	0.986
	ExtraTree	0.690	0.623	0.580	0.671	0.946	0.910
	LightGBM	0.910	0.875	0.863	0.855	0.985	0.992
	CatBoost	0.870	0.810	0.807	0.804	0.978	0.972
	KNN	0.960	0.972	0.969	0.970	0.993	0.971
	Decision Tree	0.950	0.958	0.966	0.960	0.991	0.971
CPO	Naive Bayes	0.800	0.702	0.854	0.800	0.964	1.000
	Random Forest(Bagging)	0.930	0.936	0.955	0.938	0.988	1.000
	Sel-Attention MLP	0.920	0.936	0.948	0.938	0.986	0.999

Feature selction methods	Algorithm	Accuracy	Recall	Precision	F1_score	Specificity	Area Under Curve (AUC)
	XGBoost	0.920	0.892	0.951	0.910	0.986	0.995
	ExtraTree	0.670	0.604	0.558	0.646	0.942	0.911
	LightGBM	0.930	0.860	0.948	0.878	0.988	0.996
	CatBoost	0.840	0.703	0.795	0.715	0.972	0.977

The results presented in Table 3 indicate that the ReliefF-QANA method a feature selection algorithm combining ReliefF with the QANA optimization algorithm - achieved the highest or near-highest accuracy across most models. For instance, the accuracy of the KNN model with this method reached 0.94, while the Sel-Attention MLP model attained 0.97. Additionally, when combined with XGBoost, an accuracy of 0.97 was achieved, representing the highest accuracy among all tested combinations. These findings demonstrate the strong capability of QANA in selecting effective features, as it significantly enhanced the performance of various models at different levels. This improvement likely stems from QANA's efficient search of the feature space and its ability to eliminate redundant and less informative features, thereby enabling models to train on richer, noise-free information. The QANA algorithm leverages both long-term and short-term memory to provide meaningful knowledge during partial analysis of the search space. And employs a quantum crossover operator to generate subsequent generations of search agents–three factors that distinguish it from other methods. It is also noteworthy that QANA incorporates a population division mechanism, which further contributes to achieving superior results.

In contrast, the ReliefF-FIPSO and ReliefF-SBOA methods generally yielded inferior results. Specifically, ReliefF-FIPSO recorded the lowest accuracies in most models, notably achieving only 0.66 in the Naive Bayes classifier, 0.70 in the decision tree, and 0.67 in the MLP model. Similarly, ReliefF-SBOA exhibited poor performance in models such as KNN with an accuracy of 0.81, random forest with 0.76, and CatBoost with 0.68. This subpar performance can be attributed to the inefficiency of the FIPSO and SBOA algorithms in selecting an optimal subset of features from complex 3D ultrasound images characterized by significant class overlap. These algorithms suffered from premature convergence or became trapped in local optima during the optimization process, resulting in the selection of less informative or redundant features and consequently reduced classifier accuracy.

Conversely, the ReliefF-CPO method demonstrated highly favorable outcomes, particularly in models such as KNN with 0.96 accuracy, decision tree with 0.95, and XGBoost with 0.92, highlighting the algorithm's capability to maintain a balanced exploration and exploitation of the feature space. By leveraging chaotic dynamics and adaptive search mechanisms, this algorithm effectively identifies impactful features. This performance disparity underscores the critical importance of selecting an appropriate metaheuristic algorithm in conjunction with filter-based methods like ReliefF to enhance machine learning model efficacy. Table 4 presents the findings of an ANOVA test on the accuracy values of nine machine learning methods across ten different feature selection groups, with a significance level (alpha, α) set at 0.05.

In addition to the numerical improvements reported, there are additional critical insights from the findings. First, the consideration of both textural and deep features appears to be an important contributor for the enhancement of classification accuracy. The textural features, although not considered in previous studies using CNNs, added value by providing important tumor-specific details that were lost in the deeper layers of the neural networks. Second, the feature selection technique of hierarchical feature selection—specifically ReliefF in combination with sophisticated metaheuristic was highly effective in eliminating redundant features and enhancing the most relevant features in the ultrasound images. Of these, QANA was the most distinctive, because of its balance of exploration and exploitation, population diversity and its ability to avoid premature convergence led to richer feature subsets that were provided to the classifiers.

In the same manner, the comparison across classifiers showed that no single algorithm excelled under all conditions, and instead, peak performance was achieved by the combination of robust feature subsets with specific classifiers KNN, XGBoost, or Self-Attention MLP. This indicates that the effectiveness of a diagnostic

system is determined not solely by the choice of model, but by the interplay between the features and the model's capacity to learn through the features provided.

Taken together, these results demonstrate that the superior performance observed in this study stems from three intertwined factors: (1) the inclusion of texture-aware descriptors, (2) the adoption of a two-level feature selection pipeline, and (3) the careful pairing of optimized feature sets with classifiers well-suited to ultrasound image characteristics. These insights provide a foundation for developing more reliable CAD (computer-aided diagnosis) systems for ovarian cancer detection in real-world clinical settings.

Table 4. Findings of the ANOVA test on the accuracy of feature selection methods in the OTU_CEUS Dataset.

ANOVA: Single Factor (Accuracy)				
Summary				
Groups	**Count**	**Sum**	**Average**	**Variance**
ReliefF-MPA	9	7.52	0.835556	0.007128
ReliefF-IGWO	9	7.65	0.85	0.0084
ReliefF-WHO	9	7.65	0.85	0.006725
ReliefF-QANA	9	8.14	0.904444	0.007978
ReliefF-SBOA	9	6.74	0.748889	0.003586
ReliefF-FIPSO	9	6.6	0.733333	0.00585
ReliefF-Zebra	9	7.29	0.81	0.010325
ReliefF-WOA	9	7.75	0.861111	0.008911
ReliefF-HO	9	7.6	0.844444	0.006903
ReliefF-CPO	9	7.92	0.88	0.00895

ANOVA						
Source of Variation	**Sum of Squares (SS)**	**Degrees of Freedom (DF)**	**Mean Square (MS)**	**F-Ratio**	**Probability Value P-value**	**F crit**
Between Groups	0.237071	9	0.026341	3.523649	0.001016	1.999115
Within Groups	0.598044	80	0.007476			
Total	0.835116	89				

The results of the ANOVA test conducted on the accuracy of 10 hierarchical hybrid feature selection methods demonstrate statistically significant differences in their performance. Based on the statistical analyses, the ReliefF-QANA method, with a mean accuracy of 0.904 and a variance of 0.007978, was identified as the best-performing approach. This superiority is primarily attributed to the inherent nature of the QANA algorithm, which leverages quantum mechanisms to avoid entrapment in local optima. Additionally, this algorithm utilizes both long-term and short-term memory to provide meaningful

knowledge during partial search space analysis, alongside a quantum crossover operator to generate the next generation of search agents. Such features confer considerable advantages, especially when processing complex, high-dimensional ultrasound imaging data.

Conversely, the ReliefF-FIPSO method exhibited the weakest performance, with a mean accuracy of 0.733. This deficiency can be ascribed to the intrinsic challenges faced by particle swarm optimization-based algorithms in navigating high-dimensional search spaces, where premature convergence often occurs, resulting in the selection of suboptimal feature subsets. The ANOVA results, with an F-value of 3.52 exceeding the critical F-value of 1.99, clearly indicate significant differences among the methods. This finding is further confirmed by a very low p-value (0.001016), affirming that the observed differences are not random but are due to the genuine impact of the feature selection algorithms.

Variance analysis among the methods also provides valuable insights. For instance, ReliefF-SBOA demonstrated the most stable performance across various classifiers, with a variance of 0.003586. Such stability is particularly important for clinical applications requiring high reliability. On the other hand, methods like QANA and CPO exhibited excellent adaptability with a wide range of classifiers, including KNN, XGBoost, and self-attention neural networks, owing to their high flexibility.

In summary, the selection of an appropriate feature selection approach in ultrasound image-based medical diagnostic systems has a direct effect on the final accuracy of the system. Novel approaches inspired by quantum physics concepts and natural behaviors tend to outperform others due to their ability to effectively manage the intrinsic complexity of medical data. Table 5 presents the findings of the ANOVA test comparing the classification models' accuracy. The significance level (alpha or α) was set at 0.05.

Table 5. ANOVA test findings on the accuracy of various machine learning methods using the OTU_CEUS dataset.

Anova: Single Factor (Accuracy)						
Summary						
Groups	**Count**	**Sum**	**Average**	**Variance**		
KNN	10	8.88	0.888	0.003951		
Decision Tree	10	8.73	0.873	0.006246		
Naive Bayes	10	7.26	0.726	0.003982		
Random Forest(Bagging)	10	8.58	0.858	0.004196		
Sel-Attention MLP	10	8.75	0.875	0.007072		
XGBoost	10	8.66	0.866	0.004293		
ExtraTree	10	6.68	0.668	0.000573		
LightGBM	10	8.85	0.885	0.002761		
CatBoost	10	8.47	0.847	0.004534		
ANOVA						
Source of Variation	**Sum of Squares (SS)**	**Degrees of Freedom (DF)**	**Mean Squares (MS)**	**F-ratio**	**P-value**	**F crit**
Between Groups	0.496636	8	0.062079	14.85593	3.69E-13	2.054882
Within Groups	0.33848	81	0.004179			
Total	0.835116	89				

The findings of the ANOVA analysis, conducted to compare the accuracy of nine different classification models, indicate statistically significant differences in their performance ($F(8, 81) = 14.86$, p-value = 3.69E-13). This finding confirms that the selection of classification model has a significant impact on predictive accuracy.

With a mean accuracy of 0.888 and a variance of 0.003951, the K-Nearest Neighbors (KNN) model and LightGBM model performed the best among the models that were evaluated, according to descriptive statistics. These findings imply that within this particular dataset, neighborhood-based models like KNN and gradient boosting optimizers like LightGBM show better generalizability.

With a mean accuracy of 0.668, the ExtraTree model, on the other hand, performed the worst. This is because its structure is comparatively simpler than that of other approaches. According to variance comparisons, ExtraTree produced the most consistent results (variance of 0.000573), while the Sel-Attention MLP model showed the most instability (variance of 0.007072).

The observed differences between models are statistically significant and not the product of chance, as this analysis shows. Figures 11 to 14 present the ROC curves and confusion matrices for several of the aforementioned models evaluated on the OTU_CEUS ultrasound dataset.

The findings in Figure 17 indicate that the ReliefF-QANA method significantly outperforms others across most models, particularly achieving an accuracy of 0.97 with the Sel-Attention MLP and XGBoost. More complex models like LightGBM and Random Forest have higher average accuracies (ranging from 0.89 to 0.93) than simpler models like Naive Bayes (0.71 to 0.80) and ExtraTree (0.62 to 0.69). These trends suggest that the best results are obtained when combining advanced feature selection algorithms like QANA with deep learning or ensemble-based models, even though simpler models are usually more affected by the limitations of the feature selection algorithms used.

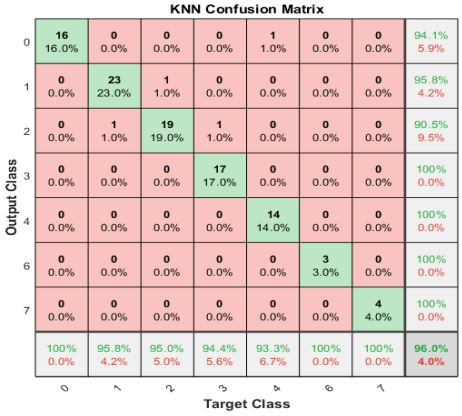


FIGURE 11. Confusion matrix of the CPO-KNN model.

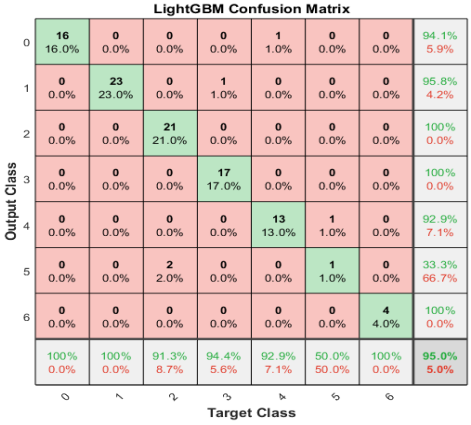


FIGURE 12. Confusion matrix of the QANA-LightGBM model.

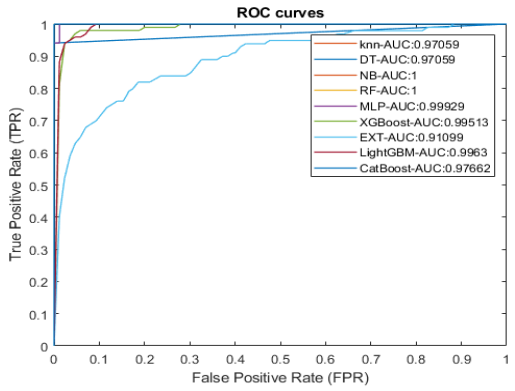


FIGURE 13. ROC curves of methods with CPO feature selection.

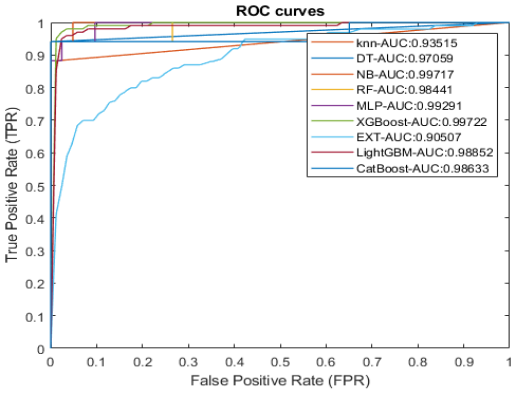


FIGURE 14. ROC curves of methods with QANA feature selection.

5. COMPARISON OF METHOD RESULTS ON OTU_CEUS DATA

Table 6 presents a comparison between the top-performing methods from the present study and the methods reported in [63].

Table 6. Comparison of method performance on OTU_CEUS ultrasound image dataset.

	Model	Accuracy	Recall	Precision	F1score	Specificity	AUC
Best Proposed Methods in Each FS	QANA-XGBoost	0.970	0.979	0.977	0.977	0.995	0.997
	CPO-KNN	0.960	0.972	0.969	0.970	0.993	0.971
	WOA-XGBoost	0.940	0.873	0.961	0.892	0.989	0.991
	IGWO-Sel-Attention MLP	0.930	0.919	0.920	0.912	0.988	0.991
	HO-KNN	0.920	0.849	0.950	0.873	0.985	0.894
	WHO-Random Forest(Bagging)	0.910	0.913	0.913	0.896	0.985	0.977
	Zebra-Decision Tree	0.910	0.929	0.900	0.911	0.985	0.929
	MAP-Sel-Attention MLP	0.900	0.923	0.923	0.922	0.982	0.996
	SBOA-Sel-Attention MLP	0.810	0.810	0.884	0.824	0.966	0.994
	FIPSO-KNN	0.750	0.719	0.713	0.696	0.958	0.810
[63]	VGG-16	0.740	0.647	0.782	0.834	0.954	0.916
	ResNet-34	0.830	0.769	0.853	0.785	0.970	0.876
	ResNet-50	0.825	0.741	0.721	0.832	0.962	0.982
	DenseNet-121	0.770	0.741	0.804	0.712	0.960	0.976
	MobileNetV2	0.780	0.757	0.735	0.721	0.963	0.952
	EfficientNet-b0	0.900	0.923	0.923	0.922	0.982	0.996
	EfficientNet-b1	0.730	0.629	0.814	0.731	0.952	0.988
	EfficientNetV2-S	0.830	0.787	0.846	0.778	0.971	0.991
	EfficientNetV2-M	0.85	0.89	0.89	0.89	0.97	0.87

In this study, the proposed methods demonstrated superior performance compared to the deep learning approaches presented by with an accuracy of 97%, recall of 97.9%, and precision of 97.7%. The QANA-XGBoost approach notably outperformed the others. This superiority can be ascribed to the best possible combination of efficient feature selection and metaheuristic optimization algorithms, which reduced dimensionality and improved model efficiency. With accuracies of 74% and 83%, respectively, Zhao's methods, such as VGG-16 and ResNet-34, fared worse because of the constraints imposed by the small dataset size. In order to avoid overfitting or failing to recognize intricate patterns, deep neural networks usually need a large amount of data to be trained.

The use of intelligent hierarchical feature selection algorithms made up of CPO (Penguin Population Optimization) and QANA (Quantum-inspired Adaptive Network Algorithm) is a major component underpinning the benefits of the suggested methods. These algorithms help the model focus on crucial discriminative features by eliminating irrelevant or noisy features. Furthermore, there is a reasonable trade-off between maintaining computational efficiency and extracting complex features when metaheuristic algorithms are combined with traditional machine learning methods like XGBoost and KNN. In contrast, Zhao's methods rely primarily on pre-built deep neural network architectures, which are less adaptable to small-scale datasets.

Interestingly, even among Zhao's approaches, EfficientNet-b0 achieved a competitive performance with 90% accuracy and an AUC of 99.6%, indicating that efficient architectures like Efficient Net can partly mitigate the challenges posed by limited data. However, it still underperformed compared to QANA-XGBoost in terms of accuracy and other key metrics. In particular, QANA-XGBoost performed about 12% better in accuracy and 9% better in recall than Zhao's top model, EfficientNetV2-M (85% accuracy). This notable distinction emphasizes how hybrid approaches based on metaheuristic optimization and machine learning can provide better results than pure deep learning models in situations with limited data.

The proposed framework for deep learning tumor diagnosis is effective due to its combination of handcrafted textural features and deep representations. The hierarchical feature selection mechanism, combining ReliefF with advanced metaheuristics like QANA and CPO, filters out noisy variables while retaining discriminative descriptors. This process improves classification accuracy and reduces computational overhead, making it more practical for clinical deployment. QANA is particularly impactful due to its ability to balance exploration and exploitation, maintain population diversity, and prevent premature convergence. The framework's most effective components are the hybridization of textural and deep features, the two-level feature selection strategy, and the integration of metaheuristic-driven feature optimization with flexible machine learning classifiers.

6. EVALUATION OF DIFFERENT FEATURE SELECTION METHODS ON THE OTU_2D DATASET

In this section, ten metaheuristic optimization (MO) algorithms were applied to select the most pertinent subset of features from extracted ultrasound image features. To select the most effective approach, comprehensive evaluations were carried out on the OTU_2d dataset as presented in Table 7.

Table 7. Evaluation of various feature selection and classification methods on the OTU_2d dataset.

Feature Selection	Algorithms	Accuracy	Recall	Precision	F1_score	Specificity	AUC
MPA	KNN	0.981	0.972	0.984	0.978	0.997	0.991
	Decision Tree	0.998	0.993	0.999	0.996	1.000	1.000
	Naive Bayes	0.968	0.967	0.957	0.961	0.995	1.000
	Random Forest(Bagging)	0.994	0.988	0.994	0.990	0.999	1.000
	Sel-Attention MLP	0.987	0.984	0.988	0.986	0.998	1.000
	XGBoost	0.996	0.990	0.997	0.993	0.999	0.999
	ExtraTree	0.759	0.571	0.820	0.595	0.963	0.978
	LightGBM	0.994	0.988	0.996	0.992	0.999	1.000
	CatBoost	0.996	0.990	0.997	0.993	0.999	1.000
IGWO	KNN	0.979	0.975	0.970	0.972	0.997	0.993
	Decision Tree	0.998	0.996	0.998	0.997	1.000	1.000
	Naive Bayes	0.953	0.911	0.908	0.909	0.993	1.000
	Random Forest(Bagging)	0.996	0.998	0.997	0.997	0.999	1.000
	Sel-Attention MLP	0.991	0.987	0.987	0.987	0.999	1.000
	XGBoost	0.996	0.998	0.997	0.997	0.999	1.000
	ExtraTree	0.746	0.565	0.822	0.593	0.961	0.971
	LightGBM	0.994	0.990	0.993	0.991	0.999	1.000
	CatBoost	0.994	0.988	0.996	0.992	0.999	1.000
WHO	KNN	0.991	0.991	0.995	0.993	0.999	0.997
	Decision Tree	0.987	0.970	0.987	0.978	0.998	0.993
	Naive Bayes	0.979	0.976	0.987	0.981	0.997	0.996
	Random Forest(Bagging)	0.994	0.986	0.996	0.991	0.999	0.999
	Sel-Attention MLP	0.987	0.980	0.992	0.986	0.998	1.000
	XGBoost	0.996	0.990	0.998	0.994	0.999	1.000
	ExtraTree	0.744	0.549	0.830	0.571	0.960	0.977
	LightGBM	0.996	0.990	0.997	0.994	0.999	1.000
	CatBoost	0.987	0.970	0.988	0.978	0.998	0.999
QANA	KNN	0.981	0.981	0.986	0.983	0.997	0.990
	Decision Tree	0.996	0.992	0.997	0.994	0.999	1.000

Feature Selection	Algorithms	Accuracy	Recall	Precision	F1_score	Specificity	AUC
SBOA	Naive Bayes	0.983	0.984	0.981	0.983	0.997	1.000
	Random Forest(Bagging)	0.996	0.992	0.997	0.994	0.999	1.000
	Sel-Attention MLP	0.991	0.990	0.987	0.988	0.999	1.000
	XGBoost	0.996	0.992	0.997	0.994	0.999	1.000
	ExtraTree	0.776	0.626	0.848	0.670	0.965	0.974
	LightGBM	0.996	0.992	0.997	0.994	0.999	1.000
	CatBoost	0.991	0.985	0.982	0.982	0.999	1.000
	KNN	0.985	0.982	0.987	0.984	0.998	0.994
	Decision Tree	0.996	0.992	0.997	0.995	0.999	1.000
FI-PSO	Naive Bayes	0.981	0.981	0.975	0.977	0.997	1.000
	Random Forest(Bagging)	0.998	0.993	0.999	0.996	1.000	1.000
	Sel-Attention MLP	0.994	0.986	0.993	0.989	0.999	1.000
	XGBoost	0.996	0.992	0.997	0.994	0.999	1.000
	ExtraTree	0.776	0.626	0.848	0.670	0.965	0.974
	LightGBM	0.996	0.992	0.997	0.994	0.999	1.000
	CatBoost	0.991	0.985	0.982	0.982	0.999	1.000
	KNN	0.981	0.974	0.964	0.968	0.997	0.991
	Decision Tree	0.998	0.993	0.999	0.996	1.000	1.000
Zebra	Naive Bayes	0.996	0.992	0.997	0.994	0.999	1.000
	Random Forest(Bagging)	0.996	0.992	0.997	0.994	0.999	1.000
	Sel-Attention MLP	0.987	0.973	0.979	0.976	0.998	1.000
	XGBoost	0.996	0.992	0.997	0.994	0.999	1.000
	ExtraTree	0.753	0.577	0.839	0.609	0.962	0.976
	LightGBM	0.998	0.993	0.999	0.996	1.000	1.000
	CatBoost	0.998	0.993	0.999	0.996	1.000	1.000
	KNN	0.983	0.983	0.984	0.983	0.997	0.990
	Decision Tree	0.989	0.986	0.991	0.988	0.998	0.994
WOA	Naive Bayes	0.981	0.978	0.978	0.978	0.997	1.000
	Random Forest(Bagging)	0.994	0.993	0.996	0.994	0.999	1.000
	Sel-Attention MLP	0.985	0.983	0.978	0.980	0.998	1.000
	XGBoost	0.991	0.995	0.993	0.994	0.999	1.000
	ExtraTree	0.731	0.550	0.814	0.575	0.958	0.971
	LightGBM	0.994	0.993	0.996	0.994	0.999	1.000
	CatBoost	0.981	0.976	0.973	0.974	0.997	1.000
	KNN	0.974	0.977	0.980	0.978	0.996	0.975
	Decision Tree	0.989	0.994	0.994	0.994	0.998	0.998
	Naive Bayes	0.959	0.971	0.971	0.971	0.994	0.985
	Random Forest(Bagging)	0.985	0.986	0.989	0.987	0.998	1.000
	Sel-Attention MLP	0.989	0.993	0.986	0.990	0.998	1.000
	XGBoost	0.985	0.985	0.989	0.987	0.998	1.000
	ExtraTree	0.780	0.589	0.842	0.618	0.966	0.980
	LightGBM	0.994	0.991	0.994	0.992	0.999	1.000
	CatBoost	0.987	0.984	0.977	0.981	0.998	1.000

Feature Selection	Algorithms	Accuracy	Recall	Precision	F1_score	Specificity	AUC
HO	KNN	0.981	0.981	0.978	0.979	0.997	0.986
	Decision Tree	0.994	0.994	0.993	0.994	0.999	0.991
	Naive Bayes	0.957	0.955	0.963	0.958	0.994	0.992
	Random Forest(Bagging)	0.994	0.997	0.991	0.994	0.999	0.996
	Sel-Attention MLP	0.987	0.989	0.987	0.988	0.998	0.999
	XGBoost	0.994	0.997	0.996	0.996	0.999	0.999
	ExtraTree	0.759	0.568	0.828	0.589	0.963	0.982
	LightGBM	0.994	0.993	0.993	0.993	0.999	0.999
	CatBoost	0.991	0.992	0.991	0.992	0.999	0.999
CPO	KNN	0.983	0.980	0.983	0.982	0.997	0.994
	Decision Tree	0.998	0.993	0.998	0.996	1.000	1.000
	Naive Bayes	0.964	0.965	0.954	0.959	0.995	1.000
	Random Forest(Bagging)	0.996	0.998	0.997	0.997	0.999	1.000
	Sel-Attention MLP	0.985	0.982	0.986	0.984	0.998	1.000
	XGBoost	0.994	0.990	0.989	0.989	0.999	1.000
	ExtraTree	0.746	0.565	0.822	0.593	0.961	0.971
	LightGBM	0.994	0.990	0.993	0.991	0.999	1.000
	CatBoost	0.994	0.988	0.996	0.992	0.999	1.000

A variety of feature selection techniques, such as MPA, IGWO, WHO, QANA, SBOA, FI-PSO, Zebra, WOA, HO, and CPO, were used to extract the most relevant features from the ultrasound images, and these were combined with a number of classification algorithms, including KNN, Decision Tree, Naïve Bayes, Random Forest (Bagging), Sel-Attention MLP, XGBoost, ExtraTree, LightGBM, and CatBoost. The evaluation of various feature selection and classification methods on the OUT_2d dataset shows that both the classifier and feature selection technique have a significant impact on performance.

The results show that ensemble and tree-based approaches consistently outperformed simpler classifiers, such as Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost were able to capture complex, nonlinear relationships among the extracted features with accuracy values between 0.996 and 0.998 and AUCs of 1.0, as well as precision and recall values that generally hovered around 0.99, with Random Forest and the MPA feature selection method producing precision and recall of 0.994 and 0.988, respectively.

However, other approaches, like ExtraTree, consistently performed worse, with accuracy ranging from 0.74 to 0.78, probably because of overfitting tendencies; Naïve Bayes, although sometimes achieving acceptable accuracy, demonstrated variability across feature sets, highlighting its practical limitations; and KNN and Sel-Attention MLP performed moderately, with their accuracy and F1 scores improving when paired with more successful feature selection methods, such as IGWO and QANA. Taken together, these results suggest that the most effective components of the framework were (1) the hybridization of textural and deep features, (2) the two-level feature selection strategy, and (3) the integration of metaheuristic-driven feature optimization with flexible machine learning classifiers. These elements worked synergistically to enhance performance, providing an alternative pathway to high-accuracy diagnosis in scenarios where large annotated datasets for deep learning are not readily available.

The outcomes of testing different classification models for the diagnosis of ovarian tumors utilizing two-dimensional ultrasound images show that ensemble and tree-based learning algorithms perform noticeably better than alternative strategies. With accuracy ranging from 0.996 to 0.998 and an area under the curve (AUC) of 1, decision tree algorithms Random Forest, XGBoost, LightGBM, and CatBoost performed the best among them. The ability of these methods to capture nonlinear relationships among ultrasound image features is the main reason for their superiority. The best-performing models' precision and recall metrics continuously hovered around 0.99, suggesting the ideal ratio of accurately detecting positive cases to

reducing false positives. Precision and recall values of 0.994 and 0.988, respectively, were obtained by combining the Random Forest model with the MPA feature selection method. This indicates that the features extracted from the ultrasound images were of high quality.

All things considered, these results highlight how feature selection and classifier selection have a significant influence on model performance. One particularly successful method for creating AI systems for the diagnosis of ovarian tumors from 2D ultrasound images is to combine ensemble learning with decision tree-based classifiers.

Tables 8 and 9 report the results of ANOVA statistical tests from two different perspectives, with a significance level (alpha, α) set at 0.05. The findings of the one-way ANOVA analysis indicate no statistically significant difference between the mean accuracies of the different groups (F = 0.015649, p-value = 1). This finding suggests that all feature selection methods applied to 2D ultrasound images exhibit comparable performance in selecting effective features for ovarian tumor classification. However, in terms of mean accuracy, the ReliefF-SBOA method performed slightly better than the others, with a mean accuracy of 0.968017 and a variance of 0.005210.

Table 8. Results of ANOVA statistical test on the accuracy of feature selection methods in OTU_2d Data.

ANOVA: Single Factor (Accuracy)				
Summary				
Groups	**Count**	**Sum**	**Average**	**Variance**
ReliefF-MPA	9	8.671642	0.963516	0.005969
ReliefF-IGWO	9	8.646055	0.960673	0.006664
ReliefF-WHO	9	8.660981	0.962331	0.006724
ReliefF-QANA	9	8.705757	0.967306	0.005173
ReliefF-SBOA	9	8.712154	0.968017	0.00521
ReliefF-FIPSO	9	8.701493	0.966833	0.006484
ReliefF-Zebra	9	8.628998	0.958778	0.0073
ReliefF-WOA	9	8.643923	0.960436	0.004665
ReliefF-HO	9	8.65032	0.961147	0.005881
ReliefF-CPO	9	8.652452	0.961384	0.006616

ANOVA

Source of Variation	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Squares (MS)**	**F-Ratio**	**P-value**	**F crit**
Between Groups	0.000855	9	9.5E-05	0.015649	1	1.999115
Within Groups	0.485486	80	0.006069			
Total	0.486341	89				

Table 9. ANOVA test results on the accuracy of machine learning methods in the OTU_2d Dataset.

ANOVA: Single Factor (Accuracy)				
Summary				
Groups	**Count**	**Sum**	**Average**	**Variance**
KNN	10	9.818763	0.981876	1.94E-05
Decision Tree	10	9.942431	0.994243	1.72E-05
Naive Bayes	10	9.720682	0.972068	0.000188
Random Forest(Bagging)	10	9.940299	0.99403	1.19E-05

Sel-Attention MLP	10	9.884861	0.988486	8.28E-06
XGBoost	10	9.938166	0.993817	1.16E-05
ExtraTree	10	7.571429	0.757143	0.000262
LightGBM	10	9.946695	0.99467	2.27E-06
CatBoost	10	9.910448	0.991045	2.4E-05

ANOVA

Source of Variation	Sum of Square (SS)	Degrees of Freedom (DF)	Mean Squares (MS)	F-Ratio	P-value	F crit
Between Groups	0.481434	8	0.060179	993.3325	1.79E-77	2.054882
Within Groups	0.004907	81	6.06E-05			
Total	0.486341	89				

These results imply that the combination of texture and deep features in ultrasound images can achieve high diagnostic accuracy for ovarian tumors regardless of the specific feature selection method used. This outcome likely stems from the richness of the extracted features, which leads different feature selection algorithms to yield similar results. Furthermore, the low variance values across all groups (ranging from 0.004665 to 0.007300) indicate consistent and stable performance of the methods in selecting effective features. The results of the above analysis indicate highly significant statistical differences among the various groups (F = 993.3325, p-value = 1.79E-77). This result unequivocally shows that, in the two-dimensional ultrasound dataset used in this study, the accuracy of ovarian tumor diagnosis is significantly impacted by the machine learning algorithm selection.

The LightGBM approach is the best-performing algorithm, according to the results obtained, with a mean accuracy of 0.99467 and an exceptionally low variance (2.27E-06). As demonstrated by its low variance, this approach not only performs the most consistently out of all the methods that were evaluated, but it also attains the highest average accuracy. Outstanding performance was also shown by other tree-based and ensemble learning algorithms, including Random Forest (0.99403), XGBoost (0.993817), and Decision Tree (0.994243). These results confirm the superiority of this family of algorithms in medical diagnostic applications, particularly in the focused domain of this research - ovarian tumor type classification. Comparison of variance values across groups reveals that more advanced methods, such as LightGBM, Random Forest, and XGBoost not only yield higher accuracy but also produce more stable results, with variances ranging between 1.16E-05 and 2.27E-06.

7. COMPARISON OF METHOD RESULTS ON OTU_2D DATA

In the final set of experiments, Table 10 presents a comparison of the results of each proposed model in this study with those reported in previous works, such as [46, 63, 64, 65].

The comparative results of different methods on the 2D ultrasound dataset used in this study demonstrate that hybrid approaches based on optimization and machine learning significantly outperform models built on pretrained architectures. In particular, hybrid algorithms like MPA Decision Tree and IGWO–Decision Tree obtained flawless AUC and specificity scores, demonstrating the efficacy of this strategy. However, despite their intricate and deep architectures, pretrained models such as VGG-16 and ResNet-34 were unable to achieve comparable performance levels. There are a number of important reasons for this deficiency. First of all, these models frequently experience overfitting when used on smaller datasets because they are usually built to function on large-scale datasets. Second, the lack of targeted optimization mechanisms for selecting relevant features has reduced their efficiency. For instance, although MobileNetV2 is lightweight, it underperforms due to the loss of critical features in its compression layers. Among the reviewed methods, the AResUNet model was introduced by [46]. achieved the highest performance in ovarian tumor classification. However, the best-performing proposed methods in this study outperformed AResUNet by approximately 2% in accuracy, 1% in precision, and 1% in recall. A closer inspection of the results indicates

that the success of the proposed methods can be attributed to two key factors: the integration of advanced optimization algorithms with machine learning models, and the application of ensemble-based hybrid approaches. Algorithms such as MPA and IGWO enhance feature selection by reducing dimensionality and improving the efficiency of the base classifiers. In Figures 19 to 22, the confusion matrices and ROC curves of the proposed ensemble learning methods are presented.

Table 10. Comparison of method results on the OTU_2d ultrasound image dataset.

	Model	Accuracy	Recall	Precision	F1score	Specificity	AUC
Best Proposed Methods in Each FS	MPA-Decision Tree	0.998	0.993	0.999	0.996	1.000	1.000
	IGWO- Decision Tree	0.998	0.996	0.998	0.997	1.000	1.000
	WHO-XGBoost	0.996	0.990	0.998	0.994	0.999	1.000
	QANA-XGBoost	0.996	0.992	0.997	0.994	0.999	1.000
	SBOA-Random Forest (Bagging)	0.998	0.993	0.999	0.996	1.000	1.000
	FIPSO-LightGBM	0.998	0.993	0.999	0.996	1.000	1.000
	ZOA-LightGBM	0.994	0.993	0.996	0.994	0.999	1.000
	WOA-LightGBM	0.994	0.991	0.994	0.992	0.999	1.000
	HO-Decision Tree	0.994	0.994	0.993	0.994	0.999	0.991
	CPO-Decision Tree	0.998	0.993	0.998	0.996	1.000	1.000
[63]	VGG-16	0.779	0.726	0.848	0.782	0.965	0.974
	ResNet-34	0.889	0.834	0.851	0.842	0.981	0.979
	ResNet-50	0.951	0.952	0.951	0.952	0.959	0.993
	DenseNet-121	0.899	0.894	0.893	0.893	0.965	0.981
	MobileNetV2	0.876	0.889	0.871	0.880	0.943	0.974
	EfficientNet-b0	0.859	0.848	0.859	0.853465	0.911	0.974
[46]	AResUNet	0.976	0.985	0.986	0.985	-	-
[64]	Sq-SSD-Lite Modified	0.924	0.784	0.852	-	-	-
[65]	LinkNet+DenseNet	0.965	-	-	-	-	-

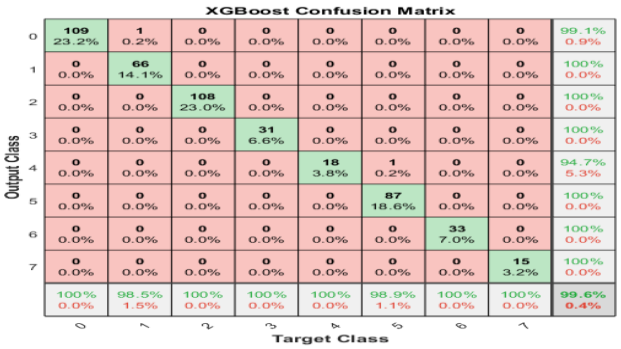


FIGURE 15. Confusion matrix of QANA-XGBoost on the OUT_2d dataset.

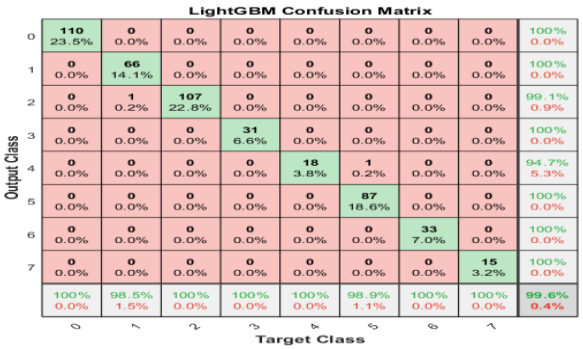


FIGURE 16. Confusion matrix of QANA-LGBM on the OUT_2d dataset.

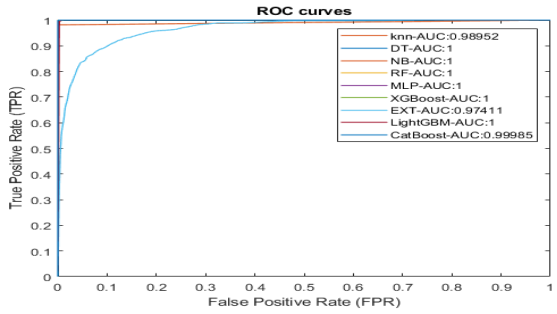


FIGURE 17. Confusion matrix of ReliefF-QANA on the OUT_2d dataset.

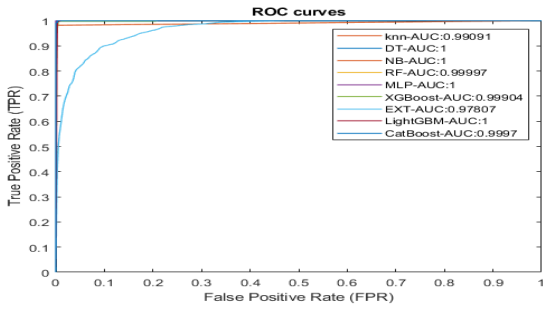


FIGURE 18. Confusion matrix of ReliefF-MPA on the OUT_2d dataset.

8. MODEL INTERPRETABILITY WITH AI EXPLAIN

Figures 19 and 20 present the interpretability of the models developed in this study using the SHAP method. Based on the mean absolute SHAP values, Figures 17 and 18 show how important each predictor variable is overall in the decision tree model. The results of this analysis indicate that variable x4 in 2D ultrasound images has the highest contribution to model prediction, with a mean absolute Shapley value of approximately 0.5. According to this research, x4 is the most important characteristic in ovarian tumor differentiation. With respective values of roughly 0.4 and 0.35, the next most significant variables are x1 and x2, suggesting that these three attributes collectively form the core of the model's decision-making process. The significant difference between the top variables and the others, especially x14, which has the lowest importance, indicates that the model is highly dependent on a specific subset of features. This observation holds importance from both clinical and model optimization perspectives. Clinically, the highly important variables can serve as potential key features for targeted diagnosis. From the standpoint of model optimization, it suggests that the model can be simplified by removing less influential variables like x14 without a significant drop in accuracy. Figures 27 and 28 also present the feature importance plots for the Random Forest and KNN models.

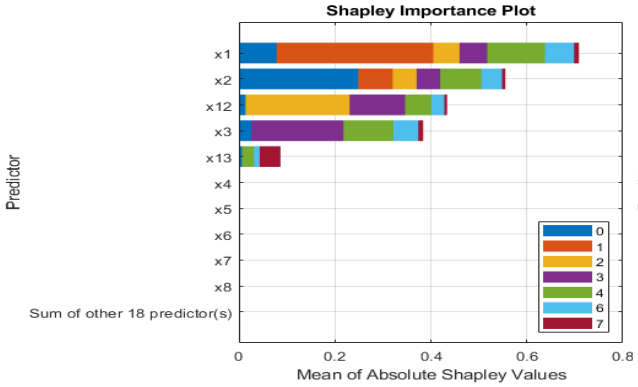


FIGURE 19. Feature importance plot in the performance of the decision tree on 3D ultrasound images.

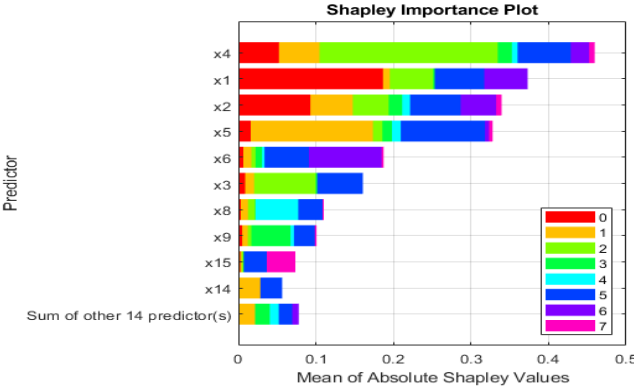


FIGURE 20. Feature importance plot in the performance of the decision tree on 2D ultrasound images.

9. DISCUSSION

Comparing the evaluation of the proposed model to recent studies from 2022–2025, significant findings were found. In the 2D dataset (OUT_2D), ensemble-based traditional machine learning methods like MPA-Decision Tree, IGWO-Decision Tree, SBOA-Random Forest, and FIPSO-LightGBM performed exceptionally

well, with accuracies of 0.998 and AUC of 1.000, outperforming several deep learning architectures like VGG-16 (0.779), ResNet-50 (0.951), and AResUNet (0.976) [46, 63, 64, 65]. Similarly, in the CEUS dataset (OUT_CEUS), the proposed methods, particularly QANA-XGBoost and CPO-KNN, achieved high accuracies (0.970 and 0.960) and F1-scores (0.977 and 0.970), outperforming recent CNN models like EfficientNet-b0 (0.900) and ResNet-50 (0.825) [63; 64]. Although deep learning is becoming more and more popular in medical image analysis, these results show that intelligently combining feature selection with ensemble classifiers is still very effective for ovarian tumor detection. The comparison with recent publications shows that the suggested approach not only achieves competitive or superior performance but also maintains robustness across various ultrasound modalities. Overall, the results emphasize the significance of hybrid strategies that integrate advanced feature selection and classification techniques, highlighting their potential to improve diagnostic accuracy in real-world clinical applications. Beyond conventional performance metrics, the use of ANOVA provided interpretive depth by quantifying whether the differences in accuracy and other metrics across feature selection–classifier combinations were statistically meaningful. This enhances the robustness of the reported findings.

Given these promising results, it is crucial to provide practical guidance for implementing the proposed framework in real-world clinical settings. The proposed framework can be implemented in other clinical settings to facilitate replication as follows. First, high-quality ultrasound images should be collected and preprocessed, ensuring consistency and proper handling of missing values. Second, both textural and deep features should be extracted to preserve discriminative information while minimizing noise, followed by hierarchical feature selection using techniques such as ReliefF combined with QANA or CPO. Third, robust classifiers, including XGBoost and KNN, should be trained with optimized hyperparameters. Model performance should then be evaluated using standard metrics, including Accuracy, Recall, Precision, F1-score, and AUC, complemented by statistical analyses such as ANOVA. Finally, the trained model should be integrated with existing clinical imaging systems, continuously monitored for performance, and deployed in compliance with ethical and privacy standards.

10. LIMITATIONS

The use of publicly available ultrasound datasets, which might not precisely reflect the ranges of ovarian tumors in clinical practice, is one of the study's limitations. Another is that the findings might not apply to other imaging environments or hospitals. Ultrasound image quality and preprocessing methods may affect the model's performance, and observer bias may make practical application challenging. Prospective clinical validation and multi-center datasets should be used in future research.

VI. CONCLUSION

This study developed and evaluated hybrid models based on optimization and machine learning for the classification of ovarian tumor types using both 2D and 3D ultrasound images. Unlike previous studies, the approach employed ten categories of features, including six texture-based and four deep features. To optimize feature selection and reduce model complexity, a hierarchical feature selection method was introduced, consisting of the ReliefF filter algorithm at the first stage and ten metaheuristic algorithms at the second level. Tumor classification was performed using nine machine learning and neural network methods The findings demonstrated that intelligent feature selection combined with ensemble classification algorithms significantly enhances diagnostic accuracy. Notably, ReliefF-QANA with KNN and LightGBM produced better results on 3D data, while the LightGBM algorithm performed best on 2D data. Additionally, pre-trained models were unable to compete with the suggested approaches, highlighting the importance of selecting the right algorithm and optimizing feature selection. In such situations where there is a shortage of training data, future research could investigate combining the suggested model with generative networks like GANs to produce high-quality synthetic samples. This would improve the accuracy of the local method and eventually the network as a whole.

Funding Statement

This research received no external funding.

Author Contributions

Nidaa Ghalib Ali; Conceptualization, Software, Investigation, Writing - Original Draft. Farsad Zamani Boroujeni; Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision. Sahar Adill Kadum; Methodology, Validation, Writing - Review & Editing, Supervision. Mehdi Akbari Kopayei; Software, Formal analysis, Investigation, Data Curation. Mahdi Mosleh; Validation, Formal analysis, Data Curation.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are available upon reasonable request.

Acknowledgments

The authors would like to thank Islamic Azad University for support in the present work, as well as the editor and reviewers for their contributions to the preparation of the article for publication.

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.
2. Saida, T., Mori, K., Hoshiai, S., Sakai, M., Urushibara, A., Ishiguro, T., Minami, M., Satoh, T., & Nakajima, T. (**2022**). Diagnosing Ovarian Cancer on MRI: A Preliminary Study Comparing Deep Learning and Radiologist Assessments. *Cancers*, *14*(4), 1-12.
3. Bjerkan, J., Valderaune, V., & Olsen, R. M. (**2021**). Patient safety through nursing documentation: Barriers identified by healthcare professionals and students. *Frontiers in Computer Science*, *3*, 1-11.
4. Chen, K., Xu, H., Lei, Y., Lio, P., Li, Y., Guo, H., & Ali Moni, M. (**2021**). Integration and interplay of machine learning and bioinformatics approach to identify genetic interaction related to ovarian cancer chemoresistance. *Briefings in bioinformatics*, *22*(6), 1-11.
5. Ghoniem, R. M., Algarni, A. D., Refky, B., & Ewees, A. A. (**2021**). Multi-modal evolutionary deep learning model for ovarian cancer diagnosis. *Symmetry*, *13*(4), 1-25.
6. Yue, Z., Sun, C., Chen, F., Zhang, Y., Xu, W., Shabbir, S., Zou, L., Lu, W., Wang, W., & Xie, Z. (**2021**). Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis. *Biomedical optics express*, *12*(5), 2559-2574.
7. Zhang, Z., & Han, Y. (**2020**). Detection of ovarian tumors in obstetric ultrasound imaging using logistic regression classifier with an advanced machine learning approach. *IEEE Access*, *8*, 44999-45008.
8. Gao, Y., Zeng, S., Xu, X., Li, H., Yao, S., Song, K., Li, X., Chen, L., Tang, J., & Xing, H. (**2022**). Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. *The Lancet Digital Health*, *4*(3), 179-187.
9. Wang, R., Cai, Y., Lee, I. K., Hu, R., Purkayastha, S., Pan, I., Yi, T., Tran, T. M. L., Lu, S., & Liu, T. (**2021**). Evaluation of a convolutional neural network for ovarian tumor differentiation based on magnetic resonance imaging. *European radiology*, *31*, 4960-4971.
10. Kasture, K. R., Shah, D. D., & Matte, P. N. (**2021**). A new deep learning method for automatic ovarian cancer prediction & subtype classification. *Turkish Journal of Computer and Mathematics Education*, *12*(12), 1233-1242.
11. Kasture, K. R., Sayankar, B. B., & Matte, P. N. (**2021**). Multi-class Classification of Ovarian Cancer from Histopathological Images using Deep Learning-VGG-16. 2021 2nd Global Conference for Advancement in Technology (GCAT),
12. Hsu, S.-T., Su, Y.-J., Hung, C.-H., Chen, M.-J., Lu, C.-H., & Kuo, C.-E. (**2022**). Automatic ovarian tumors recognition system based on ensemble convolutional neural network with ultrasound imaging. *BMC Medical Informatics and Decision Making*, *22*(298), 1-12.
13. Wang, Y., Zhang, H., Wang, T., Yao, L., Zhang, G., Liu, X., Yang, G., & Yuan, L. (**2023**). Deep learning for the ovarian lesion localization and discrimination between borderline and malignant ovarian tumors based on routine MR imaging. *Scientific Reports*, *13*(2770), 1-10.
14. Li, Y., Lin, X., & Liu, J. (**2021**). An Improved Gray Wolf Optimization Algorithm to Solve Engineering Problems. *Sustainability*, *13*(6), 1-32.

15. Zamani, H., Nadimi-Shahraki, M. H., & Gandomi, A. H. (**2021**). QANA: Quantum-based avian navigation optimizer algorithm. *Engineering Applications of Artificial Intelligence*, *104*, 104314.
16. Guo, Z., Liu, G., & Jiang, F. (**2025**). Chinese Pangolin Optimizer: A novel bio-inspired metaheuristic for solving optimization problems. *The Journal of Supercomputing*, *81*, 1-32.
17. Aydilek, I. B. (**2018**). A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Applied Soft Computing*, *66*, 232-249.
18. Mirjalili, S., & Lewis, A. (**2016**). The whale optimization algorithm. *Advances in engineering software*, *95*, 51-67.
19. Trojovská, E., Dehghani, M., & Trojovský, P. (**2022**). Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm. *IEEE Access*, *10*, 49445-49473.
20. Fu, Y., Liu, D., Chen, J., & He, L. (**2024**). Secretary bird optimization algorithm: a new metaheuristic for solving global optimization problems. *Artificial Intelligence Review*, *57*(123), 1-102.
21. Naruei, I., & Keynia, F. (**2022**). Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems. *Engineering with computers*, *38*, 3025-3056.
22. Amiri, M. H., Mehrabi Hashjin, N., Montazeri, M., Mirjalili, S., & Khodadadi, N. (**2024**). Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm. *Scientific Reports*, *14*(5032), 1-50.
23. Faramarzi, A., Heidarinejad, M., Mirjalili, S., & Gandomi, A. H. (**2020**). Marine Predators Algorithm: A nature-inspired metaheuristic. *Expert systems with applications*, *152*, 1-28.
24. Chaudhary, A., Kolhe, S., & Kamal, R. (**2016**). An improved random forest classifier for multi-class classification. *Information Processing in Agriculture*, *3*(4), 215-222.
25. Breima, L. (**2010**). Random forests. *Machine Learning*, *45*, 5-32.
26. Al Amien, J., Ab Ghani, H., Saleh, N. I. M., Ismanto, E., & Gunawan, R. (**2023**). Intrusion detection system for imbalance ratio class using weighted XGBoost classifier. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *21*(5), 1102-1112.
27. Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (**2019**). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural water management*, *225*, 1-15.
28. Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (**2020**). Comparison of the CatBoost classifier with other machine learning methods. *International Journal of Advanced Computer Science and Applications*, *11*(11), 738-748.
29. Ying, L. (**2015**). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130-135.
30. Zebari, D. A., Abrahim, A. R., Ibrahim, D. A., Othman, G. M., & Ahmed, F. Y. (**2021**). Analysis of dense descriptors in 3D face recognition. In *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)* (pp. 171-176). IEEE.
31. Song, Y., Sa, J., Luo, Y., & Zhang, Z. (**2024**). A comprehensively improved local binary pattern framework for texture classification. *Multimedia Tools and Applications*, *84*(20), 1-26.
32. Zebari, D. A., Sulaiman, D. M., Sadiq, S. S., Zebari, N. A., & Salih, M. S. (**2022**). Automated Detection of Covid-19 from X-ray Using SVM. In *2022 4th International Conference on Advanced Science and Engineering (ICOASE)* (pp. 130-135). IEEE.
33. Gour, N., & Khanna, P. (**2020**). Automated glaucoma detection using GIST and pyramid histogram of oriented gradients (PHOG) descriptors. *Pattern Recognition Letters*, *137*, 3-11.
34. Swapna, S., & Garapati, Y. (**2025**). Median Binary Pattern with Statistical model for feature extraction and Mobile DB-Net based brain tumor detection using MRI image. *Expert systems with applications*, *277*, 127-201.
35. Banerjee, A., Das, N., & Santosh, K. (**2022**). Weber local descriptor for image analysis and recognition: a survey. *The Visual Computer*, *38*(1), 321-343.
36. Dhal, P., & Azad, C. (**2022**). A comprehensive survey on feature selection in the various fields of machine learning. *Applied intelligence*, *52*(4), 4543-4581.
37. Islam, M. R., Lima, A. A., Das, S. C., Mridha, M. F., Prodeep, A. R., & Watanobe, Y. (**2022**). A comprehensive survey on the process, methods, evaluation, and challenges of feature selection. *IEEE Access*, *10*, 99595-99632.
38. Ewees, A. A., Ismail, F. H., Ghoniem, R. M., & Gaheen, M. A. (**2022**). Enhanced marine predators algorithm for solving global optimization and feature selection problems. *Mathematics*, *10*(21), 1-21.
39. Chong, H. Y., Yap, H. J., Tan, S. C., Yap, K. S., & Wong, S. Y. (**2021**). Advances of metaheuristic algorithms in training neural networks for industrial applications. *Soft Computing*, *25*, 11209-11233.
40. Alamoudi, A., Khan, I. U., Aslam, N., Alqahtani, N., Alsaif, H. S., Al Dandan, O., Al Gadeeb, M., & Al Bahrani, R. (**2023**). A deep learning fusion approach to diagnosis the polycystic ovary syndrome (PCOS). *Applied Computational Intelligence and Soft Computing*, *2023*, 1-15.

41. Nazarudin, A. A., Zulkarnain, N., Mokri, S. S., Zaki, W. M. D. W., Hussain, A., Ahmad, M. F., & Nordin, I. N. A. M. (**2023**). Performance analysis of a novel hybrid segmentation method for polycystic ovarian syndrome monitoring. *Diagnostics*, *13*(4), 1-18.
42. Chitra, P., Srilatha, K., Sumathi, M., Jayasudha, F. V., Bernatin, T., & Jagadeesh, M. (2023, March). Classification of ultrasound PCOS image using deep learning based hybrid models. In 2023 Second International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1389–1394). IEEE.
43. Panicker, P. H., Shah, K., & Karamchandani, S. **(2023)**. CNN based image descriptor for polycystic ovarian morphology from transvaginal ultrasound. In 2023 International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 148–152). IEEE.
44. Panicker, P. H., Shah, K., & Karamchandani, S. **(2023)**. CNN based image descriptor for polycystic ovarian morphology from transvaginal ultrasound. In 2023 International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 148–152). IEEE.
45. Baweja, A. K., & Kanchana, M. **(2023)**. Prediction of polycystic ovarian syndrome using machine learning techniques. In *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021* (pp. 53–63). Springer Nature Singapore.
46. Bedi, P., Goyal, S., Rajawat, A. S., & Kumar, M. (**2024**). An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting polycystic ovary syndrome. *Decision Analytics Journal*, *10*, 1-10.
47. Inan, M. S. K., Ulfath, R. E., Alam, F. I., Bappee, F. K., & Hasan, R. **(2021)**. Improved sampling and feature selection to support extreme gradient boosting for PCOS diagnosis. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 1046–1050). IEEE.
48. Tiwari, S., Kane, L., Koundal, D., Jain, A., Alhudhaif, A., Polat, K., Zaguia, A., Alenezi, F., & Althubiti, S. A. (**2022**). SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning. *Expert systems with applications*, *203*, 117-132.
49. Hdaib, D., Almajali, N., Alquran, H., Mustafa, W. A., Al-Azzawi, W., & Alkhayyat, A. (2022, May). Detection of polycystic ovary syndrome (PCOS) using machine learning algorithms. In 2022 5th International Conference on Engineering Technology and Its Applications (IICETA) (pp. 532–536). IEEE.
50. Hosain, A. S., Mehedi, M. H. K., & Kabir, I. E. **(2022)**. PcoNet: A convolutional neural network architecture to detect polycystic ovary syndrome (PCOS) from ovarian ultrasound images. In 2022 International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1–6). IEEE.
51. Schwartz, D., Sawyer, T. W., Thurston, N., Barton, J., & Ditzler, G. (**2022**). Ovarian cancer detection using optical coherence tomography and convolutional neural networks. *Neural computing and applications*, *34*, 8977-8987.
52. Sengupta, D., Ali, S. N., Bhattacharya, A., Mustafi, J., Mukhopadhyay, A., & Sengupta, K. (**2022**). A deep hybrid learning pipeline for accurate diagnosis of ovarian cancer based on nuclear morphology. *PloS one*, *19*(12), 1-20.
53. Wadhwa, G., Jayanthi, N., & Mathur, M. (**2021**). A deep convolutional neural network approach for detecting malignancy of ovarian cancer using densenet model. *Annals of the Romanian Society for Cell Biology*, *25*(2), 4449-4458.
54. Hasan, M. S. **(2017)**. An application of pre-trained CNN for image classification. In 2017 20th International Conference of Computer and Information Technology (ICCIT) (pp. 1–6). IEEE.
55. Simon, M., Rodner, E., & Denzler, J. (**2016**). ImageNet pre-trained models with batch normalization. *arXiv*, 1-4.
56. Ali, H., Sharif, M., Yasmin, M., & Rehmani, M. H. (**2022**). A shallow extraction of texture features for classification of abnormal video endoscopy frames. *Biomedical Signal Processing and Control*, *77*, 1-17.
57. Faust, O., Acharya, U. R., Meiburger, K. M., Molinari, F., Koh, J. E., Yeong, C. H., Kongmebhol, P., & Ng, K. H. (**2018**). Comparative assessment of texture features for the identification of cancer in ultrasound images: a review. *Biocybernetics and Biomedical Engineering*, *38*(2), 275-296.
58. Liu, L., Lao, S., Fieguth, P. W., Guo, Y., Wang, X., & Pietikäinen, M. (**2016**). Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing*, *25*(3), 1368-1381.
59. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. **(2017)**. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980–2988). IEEE.
60. Tian, X., Wu, D., Wang, R., & Cao, X. **(2018)**. Focal text: An accurate text detection with focal loss. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 2984–2988). IEEE.
61. Nemoto, K., Hamaguchi, R., Imaizumi, T., & Hikosaka, S. (**2018**). Classification of rare building change using cnn with multi-class focal loss. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain.
62. GitHub. (**2025**). *cv516Buaa/MMOTU_DS2Net*. https://github.com/cv516Buaa/MMOTU_DS2Net

63. Zhao, Q., Lyu, S., Bai, W., Cai, L., Liu, B., Cheng, G., Wu, M., Sang, X., Yang, M., & Chen, L. (**2022**). MMOTU: A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation. *arXiv*, 1-18.
64. Le, V.-H., & Pham, T.-L. (**2024**). Ovarian Tumors Detection and Classification on Ultrasound Images Using One-stage Convolutional Neural Networks. *Journal of Robotics and Control (JRC)*, *5*(2), 561-581.
65. Rani, N. D., & Arrama, M. B. (**2025**). Novel Deep Learning-Based Framework for Multi-class Detection of Ovarian Tumors Using Ultrasound Imaging. *Indian Journal of Gynecologic Oncology*, *23*(3), 1-19.