







A Novel Approach for Sperm Morphology Classification Using Generative Adversarial Network and Attention Network

Chellammal Palanichamy Thamil Selvi^{1*}, Saravanan Subramaniyan², Narmadha Ramakrishnan³,
Krishna Samy Selva Sheela⁴

¹ Department of Artificial Intelligence and Data Science, Rathinam Technical Campus, Coimbatore 641021, India

² Department of Computer Science and Engineering, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science & Technology, Tiruchirappalli Campus, Trichy 621105, India

³ Department of Mechatronics, Sathyabama Institute of Science and Technology, Chennai 60019, India

⁴ Department of Computer Science, School of Computing, VELTECH Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai 600062, India

Corresponding Author Email: drcpthamilselvi@gmail.com

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430223>

ABSTRACT

Received: 18 December 2025

Revised: 28 March 2026

Accepted: 7 April 2026

Available online: 30 April 2026

Keywords:

sperm morphology, BMAF-Net, Generative Adversarial Network, Modified Waterwheel Plant Algorithm, Gradient Boosted Random Forest

Sperm morphology classification plays a major role in fertility assessments. The accurate identification of different sperm morphology classes is needed for clinical diagnosis and treatment planning. In this work, an innovative approach is proposed for sperm classification. For data augmentation, a Modified Generative Adversarial Network (M-GAN) is proposed. This network generates high-quality sperm images using a morphology-based conditioning mechanism. To extract useful features, the Biomorphological Attention Feature Network (BMAF-Net) is introduced. BMAF-Net uses attention mechanisms and learnable Gabor filters to extract both morphological and texture features. To further enhance performance, a Modified Waterwheel Plant Algorithm (M-WWPA) is used for feature optimisation. For classification, a Gradient Boosted Random Forest (GBRF) is proposed. GBRF combines the robustness of Random Forest (RF) with the boosting mechanism of Gradient Boosting to reduce classification errors. The proposed model is validated on three different datasets of HusHem, EVISAN, and SCIAN-SpermSegGS. Also, the explainable artificial intelligence technique is used for better model interpretation. Results show that the model achieves an overall accuracy of 96%.

1. INTRODUCTION

Sperm morphology refers to the size and shape of sperm cells. The structure of sperm is very important for its function and fertilization [1]. It consists of three major parts: the Head, the Midpiece, and the Tail. The head of the sperm contains genetic material. The midpiece contains mitochondria for energy [2]. Likewise, the tail portion is used for swimming. Based on the survey, 40% of men have fertility issues. The major cause of fertility problems is increased scrotal temperature and oxidative stress. It will damage developing sperm and lead to morphological problems. In addition, chronic prostatitis, epididymitis, and sexually transmitted infections can cause inflammation. This inflammation damages sperm production and maturation processes. The proper sperm morphology analysis is used to identify these issues and supports fertility treatments and interventions.

The traditional method of sperm morphology analysis is labour-intensive and time-consuming. Also, it is dependent on the expertise of a trained analyst. However, recently, AI-based approaches have provided automated solutions to analyze sperm morphology [3]. In AI, the models can be trained on labeled datasets of sperm images to classify sperm into

categories, such as "normal" and "abnormal." These classifications are based on features like head shape, midpiece, and tail structure. There are two categories of AI models that can be used for classification. The machine learning (ML) models like support vector machine (SVM), decision tree, and CatBoost, etc. [4]. Likewise, the deep learning (DL) models of Convolutional Neural Networks (CNNs), ResNet, and U-Net architectures are used for image classification [5].

Despite significant advancements in AI models, there are several critical challenges that remain, which affect the performance of the models. Existing sperm image datasets are limited in size and may not adequately represent the full spectrum of sperm morphology. This dataset suffers from class imbalances. Most of the models lack fine-grained feature extraction. They struggled to capture complex and multiscale features inherent in sperm images.

In this work, a three-fold model is proposed to address the above challenges. The Generative Adversarial Network (GAN) model is proposed for data augmentation. Then, new models are proposed for feature extraction and classification. The proposed model is validated for multiple datasets with Explainable AI (XAI) techniques for model interpretations.

2. RELATED WORK

A hybrid system is proposed by Ilhan et al. [6] for sperm morphology analysis. This system uses clustering and group sparsity for automatic segmentation. Then, the classification is done with ensemble ML models. In validation, this system achieves 87% accuracy. Likewise, the authors proposed a joint learning model for sperm head segmentation and morphological category prediction [7]. The concept of ellipticity from a segmented image is used to determine sperm morphology. In addition, multi-angle sperm analysis is performed for accurate classification.

In their study, Lv et al. [8] proposed a DL algorithm for fully automatic sperm head segmentation. The modified U-Net architecture is proposed for sperm head segmentation. Compared to the conventional U-Net, the modified U-Net consists of dilated convolution and a custom-designed block to achieve high segmentation accuracy on both unstained and low-resolution images.

To address detection challenges, Li et al. [9] proposed an advanced sperm detection model, based on YOLOv8-STA. It involves a Space-to-Depth Conv structure for fine-grained sperm target extraction. In addition, the partial C2F modules are replaced with C2F-Triplet modules to capture dimensional interactions and attentions.

In their work, Jabbari and Bigdeli [10] proposed a new architecture for sperm image classification. This approach uses the CapsNets architecture with a GAN model to solve the data imbalance issues. Results show that CapsNets achieves balanced classification accuracy of 97.8% and maintains over 80% accuracy even with a 1:30 minority-to-majority class ratio.

In the study of Prabakaran and Saravanan [11], the authors proposed a modified Havrda-Charvat entropic technique for sperm cell segmentation. Initially, the Wiener filter is applied for preprocessing. Then, modified entropy is applied for segmentation. Results show that the model achieves an overall accuracy of 98.99% accuracy. Likewise, Yüzkat et al. [12] presented an ensemble approach for sperm cell quality classification. This approach uses a CNN model with different layer properties for morphological feature extraction. Then, the extracted features are combined using multi-level fusion for final classification. Results on datasets of SMIDS, HuSHeM, and SCIAN show that the ensemble model achieves accuracies of 91.8%, 86.2%, and 81.91%, respectively.

A three-fold model is proposed by Ilhan et al. [13] for morphology assessment. For noise removal, the multi-stage cascade preprocessing is applied. Then, region-based descriptor features are extracted. For classification, the non-linear kernel SVM is applied. Results are verified on HuSHeM and SMIDS datasets. Compared to existing models, the THREE-fold model increases classification accuracy by 10% for HuSHeM and 5% for SMIDS. In their study, to detect abnormalities in semen samples, Shahzad et al. [14] proposed a sequential DL method. The author proposed a Sequential Deep Neural Network (SDNN) for sperm analysis. The model achieved a high accuracy, with 90%, 88%, and 89.6% for the classification of irregularities in the acrosome, head, and vacuole, respectively. In the study of Marin and Chang [15], the authors analyzed the performance of different DL models for sperm head segmentation. The results show that the U-Net with transfer learning outperformed other models and achieved up to 95% overlap with hand-segmented masks.

In their study, for individual sperm segmentation, the

authors Yang et al. [16] proposed a hybrid model. It combines BlendMask with SegNet to separate the head. The system achieved 90.82% morphological accuracy and was validated on 1272 samples from multiple hospitals. Similarly, Lewandowska et al. [17] developed a modified Feature Pyramid Network (FPN) for sperm head segmentation. In addition, the attention mechanism is added in the pyramid network to concentrate on more relevant features. In a similar way, Javadi and Mirroshandel [18] proposed a DL model for sperm image processing. Also, this method uses data augmentation and a sampling technique to address data imbalance.

A transfer learning based AlexNet model is proposed by Liu et al. [19] to automatically classify sperm into four categories. This method uses AlexNet's feature extraction architecture with pre-trained parameters. The proposed method achieved an average accuracy of 96.0% and precision of 96.4% on the HuSHeM dataset.

A deep fusion model based on the Swin transformer and MobileNetV3 is proposed by Mahali et al. [20] to classify sperm from impurities in the SVIA Subset-C. The Swin Transformer extracts long-range features, and MobileNetV3 focuses on local features. Also, this architecture integrates an autoencoder for automatic noise removal. In their work, Li et al. [21] proposed a developed multi-scale FPN for sperm detection. In addition, a regularisation algorithm is applied to prevent overfitting. Results show that the multi-scale FPN achieves 98.37% Mean Average Precision (mAP) on the EVISAN dataset and 85.6% mAP on the Pascal VOC 2007 dataset. Similarly, Riordon et al. [22] developed a VGG16-based model for sperm detection. Initially, the model is trained using ImageNet. Then, the model is trained using publicly available sperm head datasets of HuSHeM and SCIAN.

A two-stage model is suggested by Turkoglu et al. [23] for sperm morphology classification. This model uses two feature extractors of NFNet-F4 and vision transformers. The result shows that the two-stage model increases accuracy by 4.38% when compared to single-model baselines. Likewise, Maalej et al. [24] proposed a DL model based on the ResNet50 model. The proposed model classifies 12 morphological defects with an accuracy of 93.5%.

3. PROPOSED SYSTEM

In this work, a multi-stage model is proposed for sperm morphology classification. The overall workflow is shown in Figure 1. Initially, for data augmentation, the Modified Generative Adversarial Network (M-GAN) is proposed for realistic sperm image generation. Then, a Biomorphological Attention Feature Network (BMAF-Net) is proposed for effective feature extraction. For better feature optimisation, the Modified Waterwheel Plant Algorithm (M-WWPA) is proposed. Finally, the Gradient Boosted Random Forest (GBRF) is applied for classification. This multi-component model is proposed to increase the accuracy and reliability of automated sperm morphology analysis in medical settings.

3.1 Proposed Generative Adversarial Network for sperm image generation

In this work, a modified GAN architecture is proposed to generate images. This GAN model incorporates a morphology-based conditioning mechanism. In this GAN, the

generator produces images specific to each of the four sperm morphology classes. The generator considers a latent vector as well as a morphology class index to generate realistic sperm images conditioned on specific sperm types. It involves conditional generation, embedding layers, and attention mechanisms to improve high-fidelity generation.

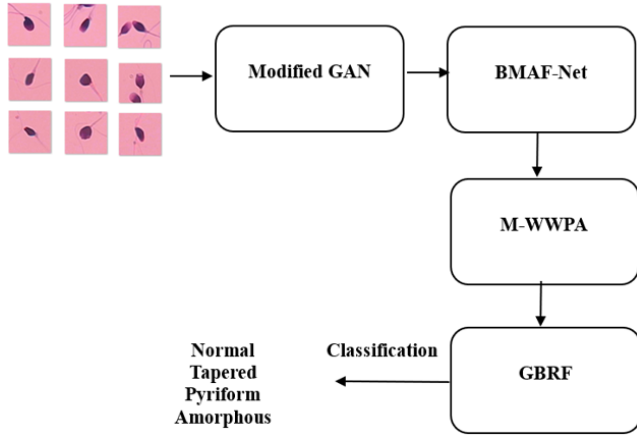


Figure 1. Workflow of proposed classification

The proposed model consists of two primary components:

1. **Generator:** This component generates sperm images conditioned on the input morphology label.
2. **Discriminator:** This component classifies images as real or fake and also determines the morphology class of the input image.

3.1.1 Generator architecture

The generator uses a latent vector z and a morphology embedding E_m as input. The morphology embedding encodes the four distinct sperm morphology types: normal, abnormal 1, abnormal 2, and defective. By conditioning the generator on this morphology label, it can generate specific sperm images corresponding to each class. The generator starts by embedding the morphology index $m \in \{0,1,2,3\}$ into a continuous space using an embedding layer. It can be expressed as a morphology vector E_m as follows:

$$E_m = \text{Embedding}(m) \text{ where } m \in \{0,1,2,3\} \quad (1)$$

The latent vector z and the morphology embedding E_m are concatenated to form an input vector for the first fully connected layer of the generator.

$$x_1 = \text{ReLU}(W_f \cdot [z, E_m] + b_f) \quad (2)$$

The feature vector x_1 is passed through a series of deconvolutional layers. It is used to progressively upscale the feature map to generate a high-resolution sperm image. In addition, attention mechanisms like non-local attention and spatial attention are applied for the model to focus on crucial regions like the sperm head and tail. The non-local attention mechanism captures long-range dependencies. The generated image captures the global structure of the sperm image.

$$A = \text{Softmax}(QK^T)V \quad (3)$$

where, Q , K , and V represent the query, key, and value matrices, respectively. The spatial attention mechanism focuses on specific regions in the image, such as the sperm

head and tail.

$$A_{\text{spatial}} = \text{Sigmoid}(\text{Conv}(x)) \quad (4)$$

where, $\text{Conv}(x)$ is a convolutional layer that generates the attention map.

3.1.2 Discriminator architecture

Here, the discriminator is responsible for two tasks: determining whether an image is real or fake and predicting the morphology class of the input image. The multiple convolutional layers are used to extract features from the input image, with a fully connected layer. The probabilities of the architecture are the Real/Fake probability and the morphology classification. For Real/Fake Classification:

$$p_{\text{real}} = \text{Sigmoid}(\text{FC}_1(x)) \quad (5)$$

For Morphology Classification:

$$p_{\text{morphology}} = \text{Softmax}(\text{FC}_2(x)) \quad (6)$$

where, FC_1 and FC_2 are fully connected layers in the discriminator.

3.1.3 Loss functions

The loss functions are applied to the generator to produce high-quality images. The adversarial loss and morphological loss are used to guide the training process. The generator tries to minimize the adversarial loss, which is calculated as the binary cross-entropy between the discriminator's prediction for real/fake:

$$\mathcal{L}_G = -\mathbb{E}[\log D_{\text{real}}(x_{\text{fake}})] + \lambda \cdot \mathcal{L}_{\text{morphology}}(x_{\text{fake}}, x_{\text{real}}) \quad (7)$$

where, $\mathcal{L}_{\text{morphology}}$ is the morphological loss. The discriminator's loss is computed as the sum of the real/fake classification loss and the morphology classification loss:

$$\mathcal{L}_D = -[\mathbb{E}[\log D_{\text{real}}(x_{\text{real}})] + \mathbb{E}[\log (1 - D_{\text{real}}(x_{\text{fake}}))]] + \mathbb{E}[\log D_{\text{morphology}}(x_{\text{real}})] \quad (8)$$

To improve the structural quality of the generated images, the edge loss and SSIM (Structural Similarity Index) loss is added. These loss functions are expressed as follows:

$$\mathcal{L}_{\text{edge}} = \| \text{Sobel}(x_{\text{fake}}) - \text{Sobel}(x_{\text{real}}) \|_2^2 \quad (9)$$

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(x_{\text{fake}}, x_{\text{real}}) \quad (10)$$

Compared to traditional conditional GANs, the M-GAN incorporates a morphology-based conditioning mechanism to generate images that are not just generic but conditioned on the specific sperm morphology types. This unique conditioning is used to generate sperm images that reflect the morphological diversity seen in real-world sperm images. Moreover, the latent vector in M-GAN is combined with a morphology class index. It is used for more controlled generation and enhancing the ability to model variations in size, shape, and scale.

In addition, attention mechanisms such as non-local

attention and spatial attention are integrated into the generator. It is used to focus on critical regions of sperm images and address the scale variations and morphological diversity. Thus, M-GAN's novelty lies in its morphology-specific conditioning and the use of advanced attention mechanisms to address specific challenges such as scale variations and morphological diversity.

3.2 Biomorphological Attention Feature Network

In this work, BMAF-Net is proposed to extract morphological and texture features from sperm images. This architecture uses attention mechanisms and learnable Gabor filters to increase the classification accuracy of the model. The architecture of BMAF-Net involves several key layers to extract useful features and produce a compact feature vector for the classification task. The architecture of BMAF-Net is shown in Figure 2.

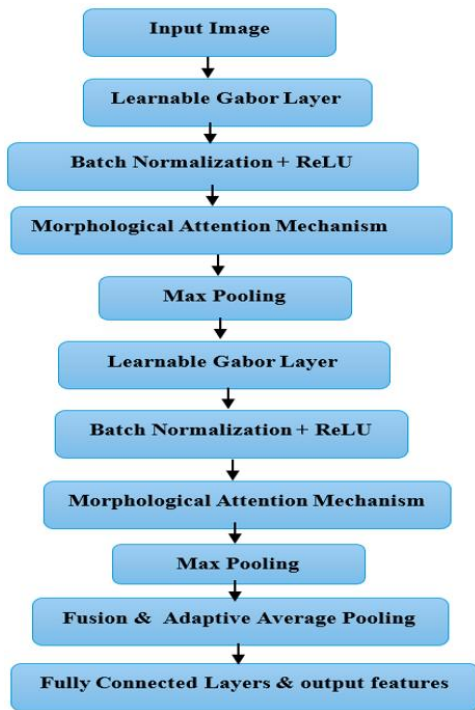


Figure 2. Architecture of Biomorphological Attention Feature Network (BMAF-Net)

3.2.1 Learnable Gabor layer

The image has dimensions $H \times W \times 3$, where H and W are the height and width of the image. Initially, learnable Gabor filters are applied to the input image to extract morphological features. The filter learns spatial frequencies and orientations that highlight structural elements in the sperm image. The output of this layer can be expressed as:

$$F_{\text{morph}} = \text{Conv}(I, W_{\text{gabor}}) + b \quad (11)$$

where, I is the input image, W_{gabor} is the learnable Gabor filter, b is the bias term, F_{morph} is the output feature map that contains the morphological features of the sperm.

The inclusion of learnable Gabor filters in BMAF-Net is motivated by the need to capture both morphological and textural features of sperm images. The learnable Gabor filters are used for the network to adaptively learn the most relevant filter parameters for sperm morphology. It includes the

specific structural patterns of sperm heads, midpieces, and tails. This flexibility enhances the network's ability to focus on both shape and texture.

3.2.2 Batch Normalization + ReLU

Next, Batch Normalisation is applied to stabilise the training process. It normalises the activations across the mini-batch. Then, the ReLU activation is used to introduce non-linearity. It can be written as:

$$F_{\text{norm}} = \text{ReLU}(\text{BatchNorm}(F_{\text{morph}})) \quad (12)$$

where, F_{morph} is the morphological feature map from the previous layer, and F_{norm} is the normalized and activated feature map.

3.2.3 Morphological Attention Mechanism

The Morphological Attention Mechanism applies channel-wise attention to focus on important morphological features like the sperm head and tail. The attention map A_{morph} is learned through a convolutional layer with a sigmoid function to produce attention weights for each channel. The attention map is given by:

$$A_{\text{morph}} = \sigma(\text{Conv}(F_{\text{norm}})) \quad (13)$$

where, σ is the sigmoid function, $\text{Conv}(F_{\text{norm}})$ is the convolution applied to the normalized feature map, A_{morph} is the attention map. Then, the feature map is recalibrated by multiplying it with the attention map:

$$F_{\text{att}} = F_{\text{norm}} \times A_{\text{morph}} \quad (14)$$

where, F_{att} is the feature map after applying attention.

3.2.4 Max pooling

Max pooling is applied to downsample the feature map and reduce its spatial dimensions. A 2×2 pooling operation is used. It is used to reduce the computational complexity and retain important spatial information. The output of the pooling operation is:

$$F_{\text{pool}} = \text{MaxPool}(F_{\text{att}}) \quad (15)$$

where, F_{pool} is the downsampled feature map after max pooling.

3.2.5 Learnable Gabor layer

In BMAF-Net, the learnable Gabor layer is applied to capture texture features from the pooled feature map. This layer learns filters that capture texture patterns. The output feature map is:

$$F_{\text{texture}} = \text{Conv}(F_{\text{pool}}, W_{\text{gabor_texture}}) + b \quad (16)$$

where, F_{pool} is the pooled feature map from the previous layer, $W_{\text{gabor_texture}}$ is the learnable texture filter, b is the bias term, F_{texture} is the texture feature map.

3.2.6 Batch Normalization + ReLU

After extracting texture features, Batch Normalization and ReLU activation are applied to the texture feature map to normalize and introduce non-linearity. The output after these

operations is:

$$F_{\text{texture_norm}} = \text{ReLU}(\text{BatchNorm}(F_{\text{texture}})) \quad (17)$$

where, F_{texture} is the texture feature map from the previous layer, and $F_{\text{texture_norm}}$ is the normalized and non-linearly activated texture feature map.

3.2.7 Morphological Attention

The second Morphological Attention Mechanism is applied to the texture feature map to focus on important texture regions in the sperm image. This attention mechanism recalibrates the texture feature map by emphasizing relevant texture patterns. The attention map for the texture features is:

$$A_{\text{texture}} = \sigma(\text{Conv}(F_{\text{texture_norm}})) \quad (18)$$

where, A_{texture} is the texture-specific attention map.

The texture feature map is recalibrated by multiplying it by the attention map:

$$F_{\text{texture_att}} = F_{\text{texture_norm}} \times A_{\text{texture}} \quad (19)$$

3.2.8 Max pooling

A second max pooling (2×2) operation is applied to the texture attention features to reduce the spatial dimensions further while retaining the most discriminative texture features. The output after max pooling is:

$$F_{\text{texture_pool}} = \text{MaxPool}(F_{\text{texture_att}}) \quad (20)$$

3.2.9 Fusion of morphological and texture features

The features from both the morphological and texture paths are fused together. This fusion is used for the network to learn a comprehensive representation of the sperm morphology. The fused feature map is:

$$F_{\text{fusion}} = F_{\text{morph}} \oplus F_{\text{texture_pool}} \quad (21)$$

where, \oplus represents concatenation. It combines the morphological and texture features.

3.2.10 Adaptive Average Pooling

Adaptive Average Pooling is applied to the fused feature map to aggregate the features into a fixed-size representation. The output after adaptive pooling is:

$$F_{\text{avg}} = \text{AdaptiveAvgPool}(F_{\text{fusion}}) \quad (22)$$

The aggregated features are passed through fully connected layers for feature recalibration. These layers are used to reduce the dimensionality of the feature vector and enable the model to focus on the most important features for classification or analysis. The output after the fully connected layers is:

$$F_{\text{recal}} = \text{FC}(F_{\text{avg}}) \quad (23)$$

The final output of the network is a feature vector that represents the sperm morphology. The final output is:

$$F_{\text{output}} = \text{FC}_2(F_{\text{recal}}) \quad (24)$$

where, F_{output} is the final output feature vector that represents

the sperm morphology.

3.3 Modified Waterwheel Plant Algorithm

The Waterwheel Plant Algorithm (WWPA) is inspired by the *Aldrovanda vesiculosa* carnivorous plant [25]. This plant hunts its prey using specialized traps. In optimization, WWPA finds an optimal solution based on exploration and stages. However, WWPA has two major drawbacks: slow Convergence and Local Optima Trapping.

3.3.1 Modified Waterwheel Plant Algorithm

To overcome these issues, in this work, a Modified Waterwheel Plant Algorithm is proposed. The strategy of differential mutation and a self-adaptive exploration-exploitation balancing is added to improve the performance of WWPA. In addition, memetic refinement is included to improve the convergence.

3.3.2 Position recognition and hunting of insects (exploration)

In this phase, the primary objective is to explore a wide area of the solution space. This phase simulates the behavior of the waterwheel plant as it searches for prey. The differential mutation introduces variability by perturbing the search agents' positions using a mutation term. The waterwheels move to new positions based on their current position and velocity to avoid local minima. The position update in the exploration phase is modelled as follows:

$$W = r_1 \cdot (P(t) + 2K) \quad (25)$$

where, W is the velocity of the search agent, r_1 is a random number in the interval $[0, 1]$, $P(t)$ is the current position of the search agent, K is an exponential factor controlling the search area. Then, the new position of the waterwheel is updated as:

$$P(t + 1) = P(t) + W \cdot (2K + r_2) + \Delta P \quad (26)$$

where, r_2 is another random variable, $\Delta P = \beta \cdot (P_r - P_s)$ is the differential mutation term, where β is a scaling factor, and P_r, P_s are randomly selected positions from the population. If the optimization does not improve for three consecutive iterations, the position is reinitialized with a Gaussian distribution to ensure continued exploration:

$$P(t + 1) = \mathcal{N}(\mu_p, \sigma) + r_1 \cdot (P(t) + 2K \cdot W) \quad (27)$$

where, $\mathcal{N}(\mu_p, \sigma)$ is a Gaussian distribution with mean μ_p and standard deviation σ .

3.3.3 Carrying the insect in a suitable tube (exploitation)

In this phase, WWPA focuses on refining the candidate solutions. It simulates the waterwheel plant's behavior of carrying prey to a feeding tube. To find the best solution, the algorithm shifts its focus broadly. The waterwheel moves closer to the best solution found so far. If there is no improvement, a mutation term is applied to avoid stagnation in local optima. The position of each waterwheel is updated to move towards the best solution found so far:

$$W = r_3 \cdot (K \cdot P_{\text{best}}(t) + r_3 \cdot P(t)) \quad (28)$$

where, $P_{\text{best}}(t)$ is the best solution found at iteration t , r_3 is a random factor in the interval $[0, 2]$. The new position of the

waterwheel is updated as follows:

$$P(t + 1) = P(t) + K \cdot W \quad (29)$$

If the solution does not improve after three iterations, a mutation term is applied to explore nearby areas:

$$P(t + 1) = (r_1 + K) \cdot \sin(FC\theta) \quad (30)$$

where, F and C are random variables in the range $[-5, 5]$, θ is the angle guiding the search direction. The exploration rate decays to focus more on exploitation. This is controlled by the equation:

$$K = \left(1 + 2 \cdot \frac{t^2}{T_{max}} + F\right) \quad (31)$$

where, t is the present iteration, T_{max} is the number of iterations.

3.3.4 Memetic refinement in exploitation

To accelerate convergence, a memetic refinement technique is incorporated. It performs simulated annealing or gradient descent to improve the best solutions found so far. It can be modelled as follows:

$$P(t + 1) = P(t) - \alpha \cdot \nabla F(P(t)) \quad (32)$$

where, α is a learning rate, $\nabla F(P(t))$ is the gradient of the objective function at the current solution. Alternatively, Simulated Annealing can be applied as follows:

$$P(t + 1) = P(t) + \alpha \cdot \delta P \quad (33)$$

where, δP is a random perturbation, α decreases over time to help fine-tune the solution. After feature extraction, M-WWPA is applied for feature selection to remove irrelevant features. The selected features are forwarded to GBRF for classification.

Here, the four major components, like M-GAN, BMAF-Net, M-WWPA, and GBRF, are essential to address specific challenges in sperm morphology classification. M-GAN handles data augmentation and class imbalance by generating realistic sperm images. The BMAF-Net uses attention mechanisms and learnable filters for detailed feature extraction. Also, M-WWPA is applied to handle the high-dimensional nature of sperm image features. M-WWPA overcomes slow convergence and local optima issues in the feature selection process.

3.4 Gradient Boosted Random Forest

For classification, the GBRF-based hybrid ML is proposed. This model combines the strengths of RF and Gradient Boosting to improve predictive performance. The main innovation in GBRF is based on the integration of RF as base learners within a Gradient Boosting model. It uses a boosting mechanism to reduce the error of the entire ensemble of trees.

In conventional gradient boosting, each tree is trained sequentially to correct the errors made by the previous trees, which uses gradient descent on the residual errors. Likewise, RF constructs multiple decision trees independently and combines their results. The proposed GBRF combines these two techniques by incorporating RF as the base learner within

the boosting framework. Here, the trees are trained iteratively with the goal of minimizing the ensemble's error. In GBRF, Random Forests is used as a base learner for gradient boosting. At each iteration, instead of training a single decision tree, an RF is trained on the negative gradient of the current ensemble. It can be mathematically expressed as follows:

$$\hat{y}^{(m)} = \hat{y}^{(m-1)} + \eta \cdot \frac{1}{T_m} \sum_{t=1}^{T_m} f_{t,m}(x) \quad (34)$$

where, $\hat{y}^{(m-1)}$ is the prediction from the previous iteration, $f_{t,m}(x)$ represents the prediction of the t -th tree from the m -th Random Forest, T_m is the number of trees in the m -th Random Forest.

Here, at each boosting iteration, instead of adding a single decision tree, an RF is added, which is trained on the residual errors from the previous iteration. The residual error at each iteration is computed as:

$$r_m = y - \hat{y}^{(m-1)} \quad (35)$$

where, y is the true label, $\hat{y}^{(m-1)}$ is the predicted value from the previous iteration. Then, the RF is trained on the negative gradient of the residuals at each step:

$$f_{t,m}(x) = \text{TrainRandomForest}(r_m) \quad (36)$$

4. RESULT AND DISCUSSION

In this work, the sperm data set of HusHem (<https://www.kaggle.com/datasets/nirmalgaud/sperm-analysis>) is used for evaluation. The data set consists of four different categories of Normal, Tapered, Pyriform, and Amorphous. For all classes, the data set consists of 60 samples with class imbalance. The sample pictures are visualised in Figure 3. For data augmentation and class balancing, the proposed GAN model is applied. The entire image is divided into training and test sets. The model training and validation curve is shown in Figure 4. The model shows strong convergence with high performance. In validation, both training and validation accuracies increase rapidly. The training and validation loss curves track each other very closely with good generalization.

Figure 5 shows the performance of the M-WWPA across 50 iterations. The green curve demonstrates an overall improvement in fitness as the number of iterations increases. It reflects the algorithm's ability to better optimize feature selection. The blue curve denotes the accuracy. The accuracy value gradually increases as the algorithm refines the selected features. This suggests that the M-WWPA is effectively enhancing the feature set to achieve higher classification performance over time.

The confusion matrix of the proposed model is given in Figure 6. The rows represent the true classes and the columns represent the predicted classes. The diagonal entries indicate the correct classifications with high values along the diagonal. It is observed that the model performs well in predicting the correct class for the majority of samples. The model correctly classified 285 instances as "Normal," 275 instances as "Tapered," 278 instances as "Pyriform," and 280 instances as "Amorphous."

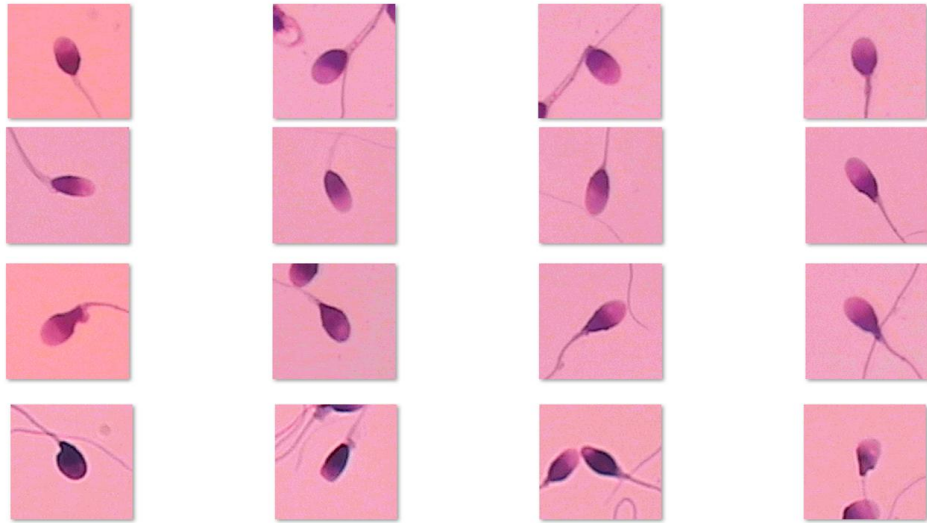


Figure 3. Sperm images

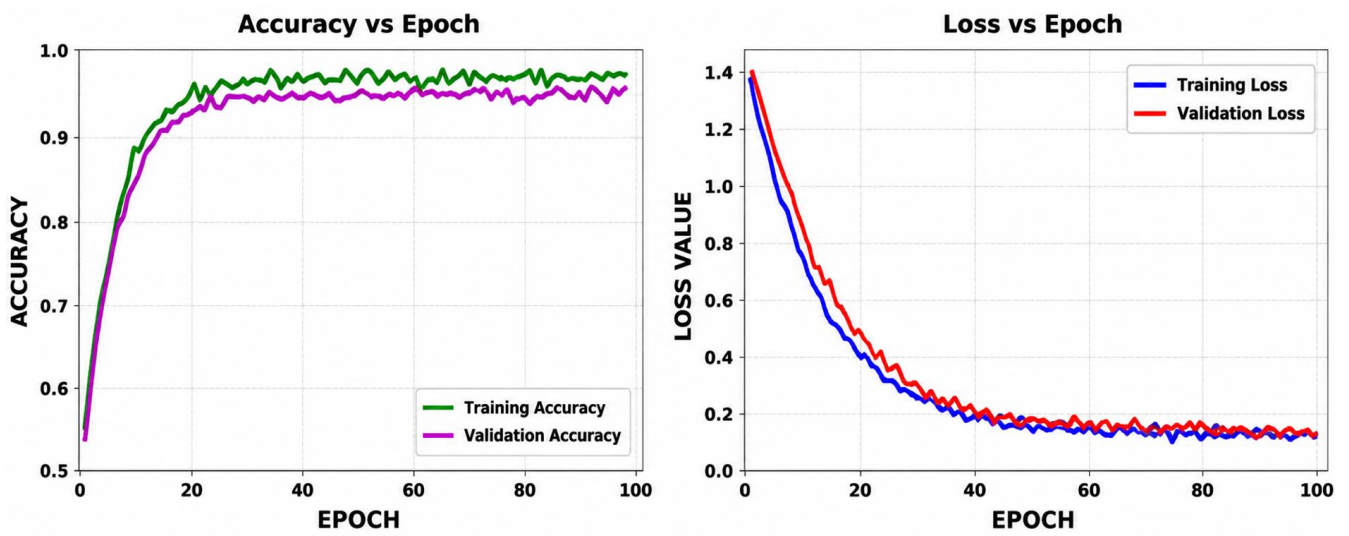


Figure 4. Model validation curves

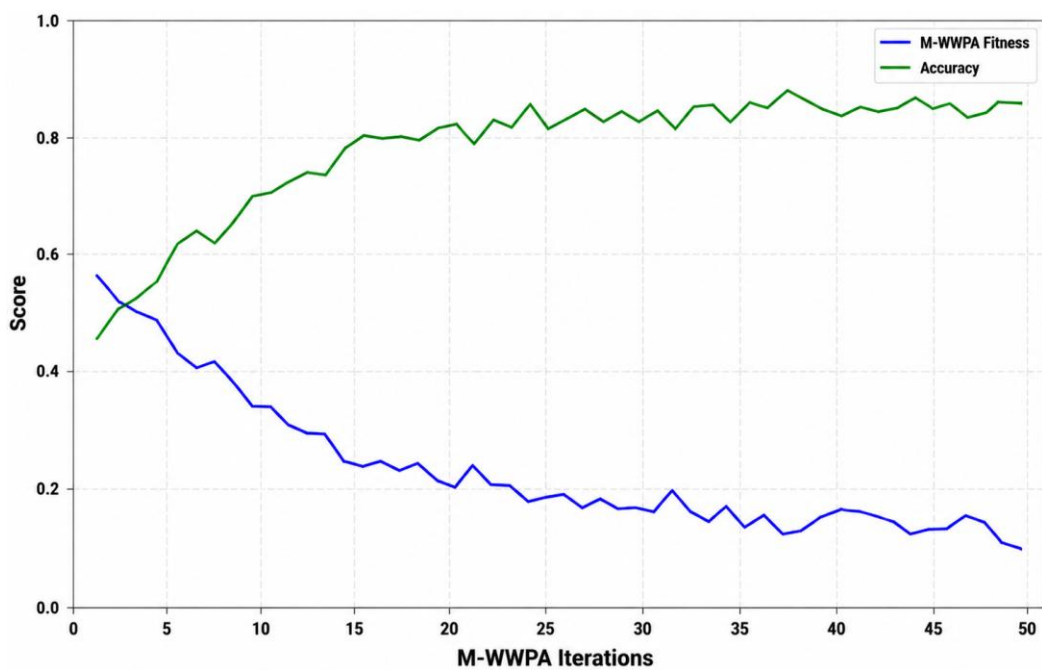


Figure 5. Modified Waterwheel Plant Algorithm fitness curves

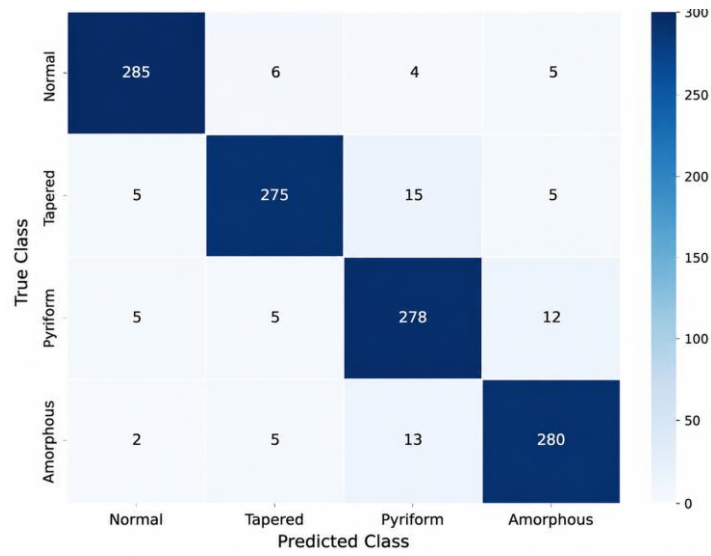


Figure 6. Confusion matrix

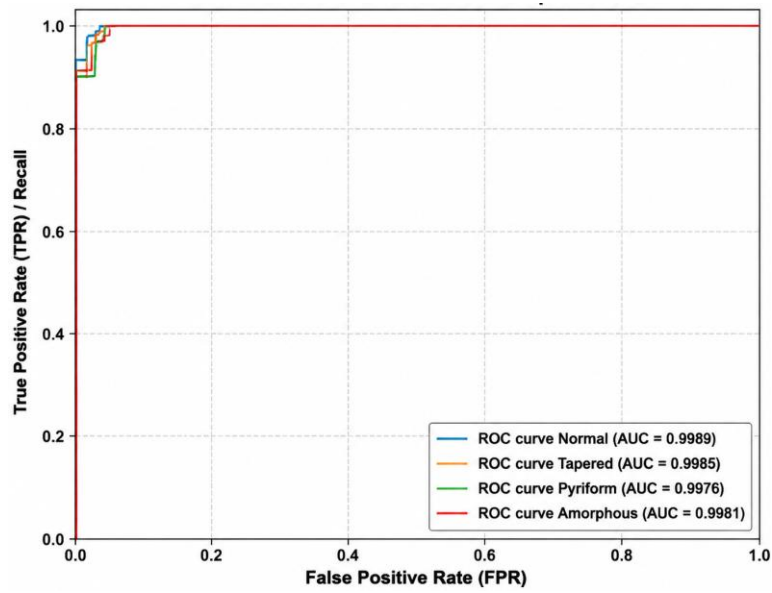


Figure 7. Receiver operating characteristic curve

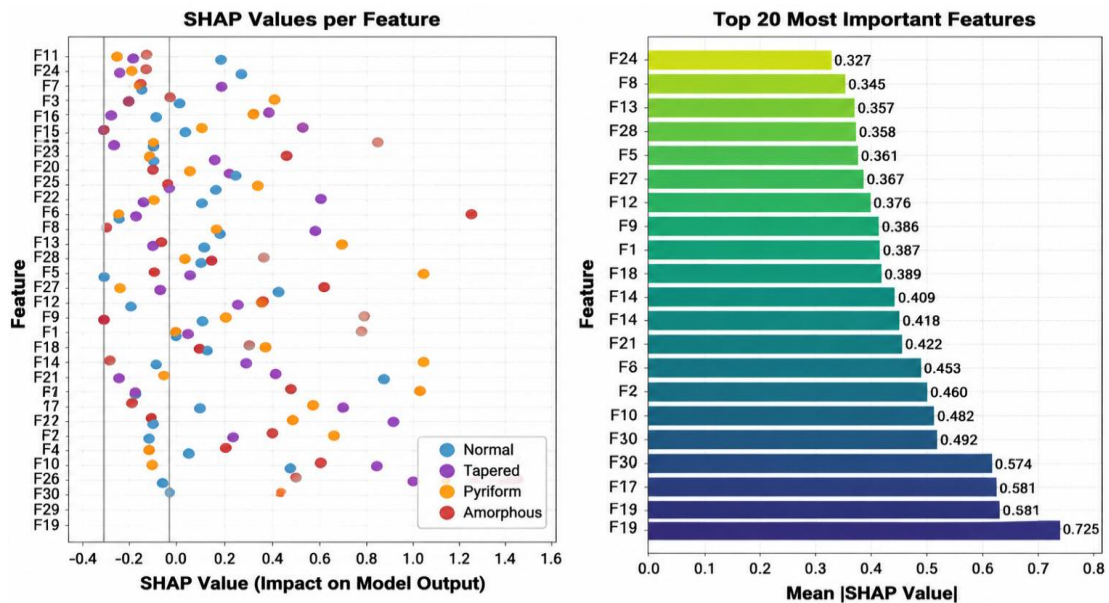


Figure 8. SHapley Additive exPlanations feature analysis

The plot in Figure 7 shows the True Positive Rate (TPR) versus the False Positive Rate (FPR) for each of the four classes. All four ROC curves are tightly clustered near the top-left corner. The AUC values are all very close to 1. This demonstrates that the classification model has excellent classification power and is highly effective at correctly.

To interpret the contribution of individual features to the model's predictions, SHapley Additive exPlanations (SHAP) values are used. The SHAP value for a feature represents its contribution to the prediction compared to the average prediction. The SHAP plot of the model is shown in Figure 8. The features exhibit varying impacts across different samples and classes. The Amorphous class frequently shows the largest positive SHAP values for features like F20, F18, and F19. It indicates that these features are strong drivers for classifying a sample as Amorphous. Similarly, the features like F17 and F19 show high SHAP for the Tapered and Amorphous classes. Overall, it is observed that F19, F17, and F29 stand out as the most critical predictors. The important feature after optimization is shown in Figure 9. The feature selection step successfully reduced the dataset to the top 30 predictors. The subsequent SHAP analysis validated the effectiveness of the selected features, which identify F19, F17, and F29 as the most significant variables. This feature contributes to the model for accurate classification across multiple categories. To strengthen the model's interpretability, the clinical expert's knowledge is added to validate the identified features. For example, the features like F19 could be linked to sperm head morphology, and F17 could relate to sperm tail length. By comparing these findings with clinical practices, the model's predictions can be better understood and aligned with biological factors.

The model's performance across different classes is given in Table 1. For the Normal class, the model shows strong

performance with an accuracy of 95%. The precision is high at 96%. For the Tapered class, the model shows an accuracy of 91%. For the Pyriform class, the model achieves a precision of 89%. Finally, the Amorphous class shows the best results in terms of precision, recall, and F1-score, all at 93%.

Table 1. Results of the model for different classes

Class	Accuracy	Precision	Recall	F1-Score
Normal	0.95	0.96	0.95	0.95
Tapered	0.91	0.91	0.92	0.91
Pyriform	0.91	0.89	0.93	0.91
Amorphous	0.93	0.93	0.93	0.93

To validate the performance further, the model is validated on EVISAN and SCIAN-SpermSegGS. The results are given in Table 2. For the EVISAN dataset, the model achieves accuracy, precision, recall, and F1-Score of 94%, 93%, 91% and 92%, respectively. For SCIAN-SpermSegGS, the model achieves accuracy, precision, recall, and F1-Score of 96%, 95%, 94%, and 94%, respectively.

Table 2. Performance of the model for other datasets

Dataset	Accuracy	Precision	Recall	F1-Score
EVISAN	0.94	0.93	0.91	0.92
SCIAN-SpermSegGS	0.96	0.95	0.94	0.94

The ablation examination of the model is given in Table 3. From the full model, the removal of any of the components results in a drop in accuracy. The BMAF-Net Only variant achieves the lowest overall performance and highlights the importance of combining generative techniques and optimization strategies for improved results.

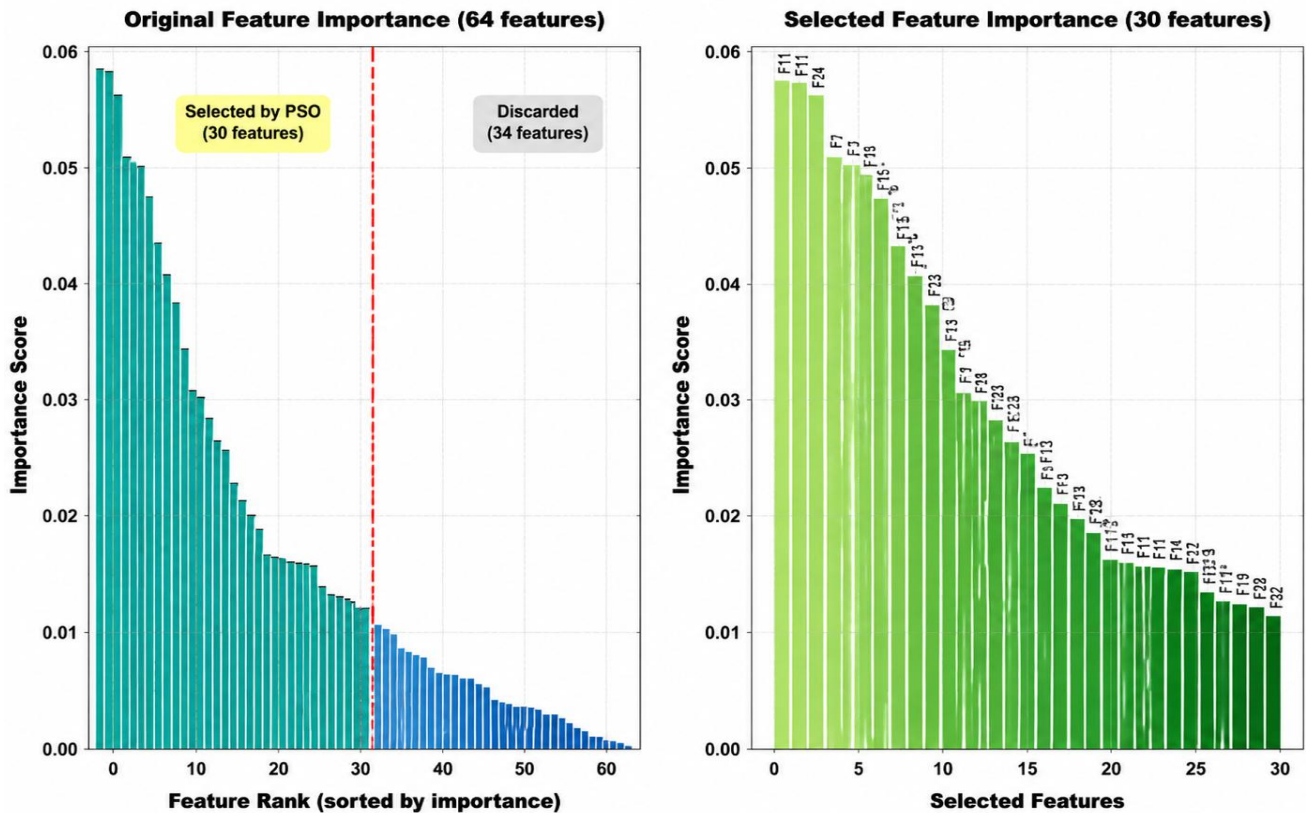


Figure 9. Feature analysis after optimization

Table 3. Ablation study for model components

Model Variant	Accuracy	Precision	Recall	F1-Score
Full Model (M-GAN + BMAF-Net + M-WWPA + GBRF)	96.5%	96%	95%	95%
Without BMAF-Net (No Attention Feature Network)	93%	94%	92%	93%
Without M-WWPA (No Feature Optimization)	94.2%	94.5%	92.5%	93.4%
Without GBRF (No Gradient Boosting in RF)	95%	94%	93%	93.5%
BMAF-Net Only (No Generative Network or Optimization)	91.5%	90%	89%	89.7%
Modified Waterwheel Plant Algorithm Only	92.8%	91%	91%	91.3%

Table 4. Performance comparison of the model

Model	Accuracy
Proposed Model	96.5%
Hybrid system	87%
YOLOv8-STA for sperm detection	92%
CapsNets with Generative Adversarial Network for sperm classification	93%
Ensemble approach with Convolutional Neural Network	91.8%
Three-fold model for sperm morphology	89%
Sequential Deep Neural Network	90%
U-Net with transfer learning	94%
BlendMask + SegNet hybrid model	90.82%
Modified Feature Pyramid Network with attention	91%
AlexNet model for sperm classification	94.5%
Swin Transformer + MobileNetV3 model	90%
Multi-scale Feature Pyramid Network	95%
VGG16-based sperm detection	87%
Two-stage model with NFNet-F4 + Vision Transformers	94%
ResNet50 model	93.5%

Table 5. Comparison of the Modified Waterwheel Plant Algorithm with other optimizers

Optimization Algorithm	Accuracy	Iterations to Convergence
Modified Waterwheel Plant Algorithm	96.5%	50
Particle Swarm Optimization	92.0%	100
Genetic Algorithm	90.5%	150
Simulated Annealing	89.0%	200

The proposed model is compared with other models in terms of accuracy. All the models are trained based on the standardized setup with 100 epochs. The batch size is assigned to 32. The optimiser is the Adam optimizer with identical data augmentation techniques. The results are given in Table 4. The same test ratio with the parameter setting used for other models. The average accuracy was calculated for existing models. The Proposed Model shows the maximum accuracy at 96.5%. It proves its superior ability to correctly identify sperm morphology compared to other approaches. The hybrid system shows a lower accuracy of 87%.

The comparison of M-WWPA with other well known optimization algorithms like Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Simulated Annealing

(SA) is given in Table 5.

The M-WWPA algorithm achieved the highest accuracy of 96.5% with the fastest convergence. It denotes that it is the most efficient in terms of both accuracy and speed. The GA and SA require more iterations to achieve less optimal results.

To validate the efficiency of hybrid GBRF, the performance are compared with sole GB & RF. The results are given in Table 6. The GBRF achieves the highest accuracy and outperforms both RF and GB. This suggests that combining Random Forest's robustness with Gradient Boosting's bias correction results in superior generalization performance. RF exhibits the lowest accuracy due to its tendency to underfit and fail to capture complex nonlinear patterns in the data. GB shows an improvement over RF but still lags behind GBRF.

The computational analysis of the model is given in Table 7. The Proposed Model achieves 96.5% accuracy with a training time of 24 hours and 50 ms/image inference time on an NVIDIA RTX 3090. Despite its longer training time, it outperforms existing models in terms of classification accuracy and shows the effectiveness of the proposed architecture. The Proposed Model reaches the best performance in terms of accuracy and still maintains competitive inference time.

Table 6. Comparison of Gradient Boosted Random Forest (GBRF), Random Forest (RF), and Gradient Boosting Machine (GBM) models

Model	Accuracy	Precision	Recall	F1-Score	Training Time (Seconds)
RF	91.5%	92.1%	90.3%	91.1%	45
GB	94.3%	93.5%	92.7%	93.1%	60
GBRF	96.5%	96.0%	95.0%	95.5%	55

Table 7. Computational analysis of the proposed model

Model	Training Time (hrs)	Inference Time (ms/image)
Proposed Model (M-GAN + BMAF-Net + M-WWPA + GBRF)	21	50
Hybrid System	10	20
YOLOv8-STA for Sperm Detection	12	35
CapsNets with Generative Adversarial Network for Sperm Classification	18	40
Ensemble Approach with Convolutional Neural Network	8	25
Three-fold Model for Sperm Morphology	16	60
Sequential DNN	4	15
U-Net with Transfer Learning	14	45
BlendMask + SegNet Hybrid Model	10	30
Modified FPN with Attention	15	50
AlexNet Model for Sperm Classification	6	10
Swin Transformer + MobileNetV3 Model	20	55
Multi-scale FPN	22	65
VGG16-based Sperm Detection	8	20
Two-stage Model with NFNet-F4 + Vision Transformers	25	60
ResNet50 Model	12	40

The performance of the model is validated for different noisy environment conditions as given in Table 8. Under optimal conditions, the model reaches 96.5% accuracy without any adjustments needed. For Low-Resolution Images, a 4.5% drop in accuracy is observed due to image quality loss. It can be mitigated by super-resolution techniques, which increase the accuracy back to 94%. In Real-World Distortions, the model accuracy is reduced to 6.5%. Likewise, the combined noise and magnification issues cause a significant drop of 7.5%.

The cross-validation is used to assess a model's performance more reliably by testing it on multiple subsets of the data. The five-fold cross-validation results for the HusHem Dataset are given in Table 9. The slight variations in results across different folds indicate that the model is stable and adaptable. It supports the claim of its reliability despite the small dataset size.

The statistical test is used to assess the significance of the performance differences between the proposed model and the existing models. The p-value is used to determine whether the observed difference in performance is due to chance or reflects a true difference in model effectiveness. The Statistical Significance results are given in Table 10. The model demonstrates a statistically significant improvement over other models with a p-value of 0.001. It suggests a very high

likelihood that the improvement is not due to chance. The ablation study of BMAF-Net's sub-modules is given in Table 11. It is observed that both learnable Gabor filters and attention mechanisms contribute significantly to the model's success.

Table 8. Performance of the model under various conditions

Condition	Accuracy (No Adjustments)	Post-Adjustment Accuracy
Ideal Condition (High-Resolution, Standard Staining, No Noise)	96.5%	96.5%
Low-Resolution Images	92%	94%
High Noise Levels	93%	94.5%
Variation in Staining Methods	93%	94.2%
Low Magnification/Zoom Level	91.5%	93.5%
Real-World Image Distortions	90%	92%
Unbalanced Class Distribution	94%	95.5%
Noise + Low Magnification	88%	91.8%
Real-Time Inference (High Latency)	96.5%	95%

Table 9. Cross-validation strategy for HusHem dataset

Cross-Validation Strategy	Fold	Accuracy	Precision	Recall	F1-Score
1-Fold (Standard Split)	1st Fold	96.5%	96%	95%	95%
2-Fold (2-Cross Validation)	2nd Fold	96.3%	95.5%	94%	94.8%
3-Fold (3-Cross Validation)	3rd Fold	96.2%	95%	94.5%	94.6%
5-Fold (5-Cross Validation)	5th Fold	95.8%	94.8%	93.5%	94.1%

Table 10. Statistical significance testing for model performance

Model	Accuracy	Precision	Recall	F1-Score	Standard Deviation (Accuracy)	Confidence Interval (Accuracy)	p-Value (vs. Existing Models)
Proposed Model	96.5%	96%	95%	95%	±1.2%	[95.1%, 97.9%]	0.001
Hybrid System	87%	87%	85%	86%	±2.1%	[84.9%, 89.1%]	0.03
YOLOv8-STA for Sperm Detection	92%	91%	90%	90.5%	±1.5%	[90.1%, 93.9%]	0.05
CapsNets with Generative Adversarial Network for Sperm Classification	93%	92%	91%	91.5%	±1.8%	[91.2%, 94.8%]	0.02
Ensemble Approach with Convolutional Neural Network	91.8%	92%	90%	91%	±1.9%	[90.2%, 93.4%]	0.04
Three-Fold Model for Sperm Morphology	89%	88%	86%	87%	±2.0%	[86%, 92%]	0.06

Table 11. Ablation study of Biomorphological Attention Feature Network (BMAF-Net) sub-modules

Sub-Module	Accuracy	Precision	Recall	F1-Score	Contribution Analysis
BMAF-Net (Full)	96.5%	96%	95%	95%	Full model with Gabor filters + Attention mechanisms.
Without Learnable Gabor Filters	94.3%	93%	91%	92%	Reduction in feature quality without Gabor filters.
Without Attention Mechanisms	95.1%	94%	93%	94%	An attention mechanism is used to focus on key features. Removal leads to a slight performance drop.
Only Learnable Gabor Filters	94.8%	94%	92%	93%	The Gabor filters alone improve texture feature extraction. But, it loses context from attention mechanisms.
Only Attention Mechanisms	95.6%	94.5%	93%	93.7%	Attention alone provides some improvement. But, its morphological feature extraction is still suboptimal without Gabor filters.

5. CONCLUSION

In this work, an integrated system is proposed for sperm morphology classification. A M-GAN model is proposed for image generation. To extract features, BMAF-Net is

developed. For effective feature optimization, the M-WWPA is proposed with GBRF-based classification. This framework not only addresses key challenges in automated sperm morphology classification but also offers a highly interpretable and clinically applicable solution. The proposed model

achieves an overall accuracy of 96.5% for validation sets. Overall, the proposed system has the potential to revolutionise automated sperm analysis with improved efficiency.

REFERENCES

- [1] Danis, R.B., Samplaski, M.K. (2019). Sperm morphology: History, challenges, and impact on natural and assisted fertility. *Current Urology Reports*, 20(8): 43. <https://doi.org/10.1007/s11934-019-0911-7>
- [2] Chang, V., Garcia, A., Hitschfeld, N., Härtel, S. (2017). Gold-standard for computer-assisted morphological sperm analysis. *Computers in Biology and Medicine*, 83: 143-150. <https://doi.org/10.1016/j.compbiomed.2017.03.004>
- [3] Liang, B., Wang, M. (2025). Deep learning-based approach for sperm morphology analysis. *BMC Urology*, 25(1): 261. <https://doi.org/10.1186/s12894-025-01946-w>
- [4] Kaveh, S., Ghaffari, A., Sohrabei, S. (2025). Investigating artificial intelligence in predicting and evaluating sperm and embryo quality in the in vitro fertilization (IVF): A systematic review. *Discover Artificial Intelligence*, 5(1): 183. <https://doi.org/10.1007/s44163-025-00420-8>
- [5] Dobrovolny, M., Benes, J., Langer, J., Krejcar, O., Selamat, A. (2023). Study on sperm-cell detection using yolov5 architecture with labaled dataset. *Genes*, 14(2): 451. <https://doi.org/10.3390/genes14020451>
- [6] Ilhan, H.O., Sigirci, I.O., Serbes, G., Aydin, N. (2020). A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. *Medical & Biological Engineering & Computing*, 58(5): 1047-1068. <https://doi.org/10.1007/s11517-019-02101-y>
- [7] Xu, Y., Chen, Y., Zhang, B., Yan, Y., Liao, H., Liu, R. (2026). Deep learning-based morphological analysis of human sperm. *Medical & Biological Engineering & Computing*, 64(1): 49-59. <https://doi.org/10.1007/s11517-025-03418-7>
- [8] Lv, Q., Yuan, X., Qian, J., Li, X., Zhang, H., Zhan, S. (2022). An improved U-Net for human sperm head segmentation. *Neural Processing Letters*, 54(1): 537-557. <https://doi.org/10.1007/s11063-021-10643-2>
- [9] Li, C., Xia, W., Li, A., Gao, L., Zhang, C., Zhi, E., Li, Z. (2025). An efficient advanced YOLOv8 framework for sperm motility detection. *Journal of Assisted Reproduction and Genetics*, 42(9): 3095-3108. <https://doi.org/10.1007/s10815-025-03589-0>
- [10] Jabbari, H., Bigdeli, N. (2023). New conditional generative adversarial capsule network for imbalanced classification of human sperm head images. *Neural Computing and Applications*, 35(27): 19919-19934. <https://doi.org/10.1007/s00521-023-08742-3>
- [11] Prabakaran, L., Saravanan, N. (2025). A three stage framework for abnormality detection in sperm cell images using CNN. *Biomedical Signal Processing and Control*, 99: 106827. <https://doi.org/10.1016/j.bspc.2024.106827>
- [12] Yüzkat, M., Ilhan, H.O., Aydin, N. (2021). Multi-model CNN fusion for sperm morphology analysis. *Computers in Biology and Medicine*, 137: 104790. <https://doi.org/10.1016/j.compbiomed.2021.104790>
- [13] Ilhan, H.O., Serbes, G., Aydin, N. (2020). Automated sperm morphology analysis approach using a directional masking technique. *Computers in Biology and Medicine*, 122: 103845. <https://doi.org/10.1016/j.compbiomed.2020.103845>
- [14] Shahzad, S., Ilyas, M., Lali, M.I.U., Rauf, H.T., Kadry, S., Nasr, E.A. (2023). Sperm abnormality detection using sequential deep neural network. *Mathematics*, 11(3): 515. <https://doi.org/10.3390/math11030515>
- [15] Marin, R., Chang, V. (2021). Impact of transfer learning for human sperm segmentation using deep learning. *Computers in Biology and Medicine*, 136: 104687. <https://doi.org/10.1016/j.compbiomed.2021.104687>
- [16] Yang, H., Ma, M., Chen, X., Chen, G., et al. (2024). Multidimensional morphological analysis of live sperm based on multiple-target tracking. *Computational and Structural Biotechnology Journal*, 24: 176-184. <https://doi.org/10.1016/j.csbj.2024.02.025>
- [17] Lewandowska, E., Weśierski, D., Mazur-Milecka, M., Liss, J., Jezierska, A. (2023). Ensembling noisy segmentation masks of blurred sperm images. *Computers in Biology and Medicine*, 166: 107520. <https://doi.org/10.1016/j.compbiomed.2023.107520>
- [18] Javadi, S., Mirroshandel, S.A. (2019). A novel deep learning method for automatic assessment of human sperm images. *Computers in Biology and Medicine*, 109: 182-194. <https://doi.org/10.1016/j.compbiomed.2019.04.030>
- [19] Liu, R., Wang, M., Wang, M., Yin, J., Yuan, Y., Liu, J. (2021). Automatic microscopy analysis with transfer learning for classification of human sperm. *Applied Sciences*, 11(12): 5369. <https://doi.org/10.3390/app11125369>
- [20] Mahali, M.I., Leu, J.S., Darmawan, J.T., Avian, C., et al. (2023). A dual architecture fusion and AutoEncoder for automatic morphological classification of human sperm. *Sensors*, 23(14): 6613. <https://doi.org/10.3390/s23146613>
- [21] Li, C., Xia, W., Han, H., Li, A., et al. (2024). A novel approach for one-stage sperm detection using advanced multi-scale feature pyramid networks. *Biomedical Signal Processing and Control*, 93: 106152. <https://doi.org/10.1016/j.bspc.2024.106152>
- [22] Riordon, J., McCallum, C., Sinton, D. (2019). Deep learning for the classification of human sperm. *Computers in Biology and Medicine*, 111: 103342. <https://doi.org/10.1016/j.compbiomed.2019.103342>
- [23] Turkoglu, A.K., Serbes, G., Uzun, H., Aktas, A., Yigit, M.H., Ilhan, H.O. (2025). Category-aware two-stage divide-and-ensemble framework for sperm morphology classification. *Diagnostics*, 15(17): 2234. <https://doi.org/10.3390/diagnostics15172234>
- [24] Maalej, R., Abdelkefi, O., Daoud, S. (2025). Advancements in automated sperm morphology analysis: A deep learning approach with comprehensive classification and model evaluation. *Multimedia Tools and Applications*, 84(23): 27345-27378. <https://doi.org/10.1007/s11042-024-20188-w>
- [25] Abdelhamid, A.A., Towfek, S.K., Khodadadi, N., Alhussan, A.A., Khafaga, D.S., Eid, M.M., Ibrahim, A. (2023). Waterwheel plant algorithm: A novel metaheuristic optimization method. *Processes*, 11(5): 1502. <https://doi.org/10.3390/pr11051502>