

Adaptive Optimizers for Neural Network Sperm Motility Classification

I Gede Susrama Mas Diyasa¹, Vaizal Asy'ari¹, Ani Dijah Rahajoe¹, Sayyidah Humairah², Deshinta Arrova Dewi³, Muhammad Nashif Farid⁴, Faikul Umam⁵

¹ Department of Information Technology, University of Pembangunan Nasional Veteran Jawa Timur, Indonesia

² Department of Electrical and Computer Engineering, Polytechnic Faculty, University of Patras, Greece

³ Center for Data Science and Sustainable Technologies, INTI International University, Negeri Sembilan, Malaysia

⁴ Department of Data Science, University of Pembangunan Nasional Veteran Jawa Timur, Indonesia

⁵ Department of Informatics, Faculty of Engineering, Universitas Trunojoyo Madura, Indonesia

Corresponding author: I Gede Susrama Mas Diyasa (e-mail: igsusrama.if@upnjatim.ac.id), **Author(s) Email:** Vaizal Asy'ari (e-mail: vaiz.asyari@gmail.com), Ani Dijah Rahajoe (e-mail: anidijah.if@upnjatim.ac.id), Sayyidah Humairah (e-mail: humairahs@upatras.gr), Deshinta Arrova Dewi (e-mail: deshinta.ad@newinti.edu.my), Muhammad Nashif Farid (e-mail: 22083010024@student.upnjatim.ac.id), Faikul Umam (e-mail: faikul@trunojoyo.ac.id)

Abstract Infertility remains a major challenge in global reproductive healthcare, with abnormal sperm motility being one of the primary contributing factors to male infertility. Conventional sperm motility analysis relies heavily on manual microscopic observation, which is subjective, time-consuming, and prone to inter-operator variability. Although Computer-Aided Sperm Analysis (CASA) systems provide automated solutions, their high cost and limited accessibility restrict widespread clinical adoption. Therefore, there is a need for a lightweight and cost-effective computational approach that can automatically classify sperm motility abnormalities with high accuracy. This study proposes a machine learning framework based on a Multilayer Perceptron (MLP) neural network optimized using adaptive optimization algorithms to classify sperm motility trajectories. The model utilizes two kinematic features extracted from microscopic video tracking: mean velocity and trajectory linearity. A dataset consisting of 276 unique sperm trajectories was obtained from microscopic recordings and processed using the Trackpy library to generate motion trajectories. The dataset was divided using an 80:20 train-test split strategy to evaluate the generalization capability of the model. Four adaptive optimization algorithms were systematically evaluated, including Adagrad, Adadelata, FTRL, and Adamax. The neural network was trained for 50 epochs using binary cross-entropy as the loss function and evaluated using accuracy, per-class precision, recall, F1-score, specificity, balanced accuracy, and ROC-AUC. Experimental results demonstrate that the Adamax optimizer achieved the best performance with an accuracy of 96.43% and a perfect ROC-AUC of 1.00, outperforming Adagrad (92.86%), FTRL (69.64%), and Adadelata (67.86%). The results indicate that Adamax provides more stable convergence and better optimization behavior for dense kinematic feature spaces. These findings highlight the effectiveness of adaptive optimization techniques in improving neural network performance for automated sperm motility classification. The proposed approach offers a computationally efficient and accessible alternative for supporting clinical sperm analysis and reducing subjectivity in reproductive diagnostics.

Keywords Adaptive Optimizer; Neural Network; Sperm Motility; Process Innovation; Reproductive Healthcare.

1. Introduction

Infertility remains a major global reproductive health challenge, affecting approximately 17.5% of the adult population worldwide, equivalent to nearly one in six individuals during their reproductive years [1], [2]. Among the various factors contributing to male infertility, abnormalities in sperm motility represent one of the most critical determinants of fertilization success [3], [4]. Sperm motility reflects the ability of spermatozoa to move efficiently toward the oocyte, and

reduced motility is strongly associated with lower conception rates [5], [6]. Consequently, accurate evaluation of sperm motility is essential for infertility diagnosis, treatment planning, and reproductive healthcare management.

Conventional sperm motility assessment is typically performed through manual microscopic observation by trained laboratory personnel. Although widely adopted, this approach is inherently subjective, labor-intensive, and operator-dependent, leading to inconsistent

outcomes and poor reproducibility across laboratories [7], [8]. To address these limitations, Computer-Aided Sperm Analysis (CASA) systems have been developed to automate sperm detection and motility measurements using image processing and computer vision techniques [9], [10]. However, the high acquisition cost, proprietary hardware requirements, and limited accessibility of CASA systems continue to restrict their widespread implementation, particularly in resource-constrained clinical environments [11]. Therefore, there is a growing need for lightweight and cost-effective computational approaches capable of providing objective and reproducible sperm motility classification.

Recent advances in artificial intelligence and machine learning have created new opportunities for automated sperm analysis. Machine learning models are capable of learning complex patterns from biomedical data and have demonstrated promising performance in classification, detection, and diagnostic support applications [12], [13], [14]. In reproductive healthcare, various machine learning approaches have been applied to sperm morphology assessment, motility analysis, fertility prediction, and sperm abnormality detection [15]. These developments offer the potential to create more accessible and affordable alternatives to conventional CASA systems while maintaining high diagnostic accuracy.

Several studies have reported encouraging results

in sperm analysis using computational approaches. Diyasa et al. [16] utilized a Gaussian Mixture Model (GMM) combined with a matching-based algorithm for sperm trajectory tracking and abnormality detection, achieving an accuracy of 87.7%. Riordon et al. [17] employed a deep convolutional neural network (VGG16) for sperm-image classification and reported a high accuracy of 94.1%. However, their approach relied on static image segmentation and therefore could not capture the temporal dynamics of sperm movement. Meanwhile, Baksir et al. [18] applied a backpropagation neural network for fertility classification but achieved a relatively lower accuracy of 80.32%. More recent studies have also explored deep learning and transformer-based architectures for sperm detection and tracking [19], [20]. Despite these advances, many existing approaches either depend on computationally intensive models, focus primarily on static image information, or provide limited investigation into optimization strategies for trajectory-based sperm motility classification.

As summarized in Table 1, previous studies have demonstrated the effectiveness of machine learning and computer vision techniques for sperm analysis. However, several limitations remain evident. Image-based approaches primarily focus on morphological characteristics and are unable to fully represent the temporal dynamics of sperm movement. Conversely, trajectory-based studies remain relatively limited and often report lower classification performance compared

Table 1. Summary of related studies on sperm analysis, classification methods, data types, reported performance, and limitations

Study	Method	Data Type	Limitation	Reported Performance
Setiawan et al. (2020) [3]	Gaussian Mixture Model	Sperm trajectory	Limited classification capability	F-measure 0.9449 (tracking/detection)
Diyasa et al. (2024) [16]	GMM + Matching Algorithm	Sperm trajectory	Moderate classification performance	87.7% accuracy
Riordon et al. (2019) [17]	CNN	Static sperm images	No temporal motion analysis	94.1% accuracy
Baksir et al. (2020) [18]	Backpropagation Neural Network	Fertility dataset	Not applicable to trajectories	80.32% accuracy
Kosuge & Hamagami (2025) [19]	Transformer tracking	Microscopic video	High computational cost	Not reported
Hicks et al. (2019) [21]	ML based Fertility Prediction	Sperm video + demographic and semen-analysis variable	Requires heterogeneous multimodal data, limiting reproducibility	Not reported

with image-based methods. Furthermore, recent deep learning and transformer-based approaches generally require substantial computational resources, limiting their practical implementation in low-resource clinical environments. More importantly, the influence of adaptive optimization algorithms on trajectory-based sperm motility classification has not been systematically investigated.

Therefore, trajectory-based methods remain underexplored, and no study has comprehensively evaluated adaptive optimization strategies for this specific feature space. Consequently, there is insufficient evidence regarding which optimizer is most suitable for sperm kinematic characteristics defined by velocity and linearity.

To address these gaps, this study proposes a lightweight sperm motility classification framework based on trajectory-derived kinematic features and a Multilayer Perceptron (MLP) neural network. Four adaptive optimization algorithms, namely Adagrad, Adadelta, FTRL, and Adamax, are systematically evaluated within the same neural network architecture to analyze their convergence behavior and classification performance. The primary contribution of this research is the comprehensive investigation of optimizer effectiveness for trajectory-based sperm motility classification and the identification of the most suitable optimization strategy for developing low-cost, accessible, and reproducible alternatives to conventional Computer-Aided Sperm Analysis (CASA) systems

II. Basic Theory

A. Sperm Motility

Sperm motility is the ability of sperm to move progressively to reach and fertilize an egg, making it a key factor in successful fertilization. According to the WHO [22], men with sperm motility of less than 42% or only 30% with progressive motility are classified as having asthenozoospermia, which is often the main cause of infertility. Sperm motility is controlled by complex molecular and structural mechanisms. Spermatozoa consist of three main parts (Fig. 1): the head containing genetic material, the tail or flagellum for movement, and the neck connecting the head to the flagellum [23]. The flagellum is divided into the midpiece (containing mitochondria), the main piece, and the tip. Sperm motility is expressed as a percentage and measured by counting the number of motile and immotile sperm through microscopic observation. To assess motility disorders, at least 300 sperm are observed and classified. Sperm motility is divided into three categories [23], namely:

1. Progressive motility, where sperm move actively in a linear fashion,

2. Non-progressive motility, where sperm move in a non-linear pattern such as zig-zag or vibrating in place, and
3. Immotility, where sperm are dead or do not move at all.

Decreased sperm motility can prolong the time required to achieve pregnancy and necessitates further examination for male infertility issues.

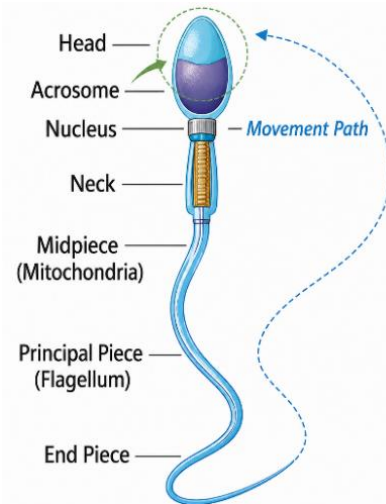


Fig. 1. Schematic representation of the structure of a human sperm cell illustrating the head, neck, midpiece, and flagellum components

B. Sperm Abnormality

Sperm motility disorders are abnormalities or reductions in the ability of spermatozoa to move properly, where spermatozoa cannot move properly or cannot move at all. Abnormal sperm motility is a sperm motility disorder that affects male infertility. According to the World Health Organization (WHO), [1] sperm motility is divided into several categories, namely fast progressive (at least 25 $\mu\text{m}/\text{second}$), slow progressive (5 to < 25 $\mu\text{m}/\text{second}$), non-progressive (less than 5 $\mu\text{m}/\text{second}$), and immotile.

Sperm motility is also assessed based on its linearity, which indicates how straight the sperm's swimming path is based on Eq. (1) [24] as follows:

$$LIN(\%) = (VSL/VCL) \times 100\% \quad (1)$$

Linearity (LIN) is the ratio of VSL (Straight Line Velocity) to VCL (Curve Velocity). A LIN value close to 100% indicates a straighter sperm movement trajectory, while a value below 50% indicates a more circular or irregular trajectory. Spermatozoa with a LIN of 60% are generally considered to have normal motility, while a LIN < 60% may indicate potential motility abnormalities [24].

C. Deep Learning

Deep learning is a subsection of machine learning that employs multi-layered neural networks to automatically

learn data representations, eliminating the need for manual feature extraction. Two of the most important attributes of deep learning are end-to-end learning, which is achieved by simultaneous feature learning and classification, and scalability with the introduction of larger data sets and more computational power. This method has proved to be the most effective means of managing high-volume, complex, and non-linear data sets in a myriad of fields, including but not limited to computer vision and biomedical analysis [25], [26].

There are two main parts that characterize the deep learning process: feedforward and backward propagation of errors. In the feedforward stage, the prediction of the input data depends on the processing of the data at the different levels of the network; the data is treated in weighted linear combinations and added to the biases of the activation functions and evaluated with the loss function [27]. In backpropagation, the prediction errors are sent backward to adjust the network parameters and use gradient descent, which in turn enables the model to refine the prediction accuracy. There are, however, a number of epochs in which these stages are repeated until the model reaches an optimal point with no change in the errors [28], [29], [30].

D. Neural Network

Artificial neural networks consist of layers of connected nodes that perform weighted computations, progressing through data preparation, architecture design, training, and testing [31], [32], [33]. The weight vector (w) represents the connection strengths between neurons, governing how signals propagate through the network. The first step in artificial neural network computation is to sum the input values $x_1, x_2, \dots, x_n$, each multiplied by its weight ($w_1, w_2, \dots, w_n$) as stated in the following Eq. (2) [32]:

$$net = \sum_{i=1}^n w_i x_i + b \quad (2)$$

The bias term, denoted by b , is an independent parameter that shifts the weighted sum and increases the flexibility of the neuron model. To ensure a well-calibrated response to small input signals and prevent the output from reaching extreme values, a nonlinear activation function $f(\cdot)$ is incorporated into the neuron structure [34]. Similar to the perceptron model, the final output of the neuron is expressed as Eq. (3) [34]:

$$y = f(net) \quad (3)$$

E. Multi-Layer Perceptron

A Multilayer Perceptron (MLP) is a neural network composed of an input layer, one or more hidden layers, and an output layer [35]. The input layer receives input, then the input value is given to the hidden units, as formulated in Eq. (4) [35].

$$o_j = \sigma \left(\sum_{k=1}^K w_{kj} x_k + \beta_j \right) \quad (4)$$

In the hidden layer, the inputs are processed, and the activation function (σ) for each neuron is calculated; then the results are processed to the next layer, as expressed in Eq. (5) [35]. This process is carried out in the next hidden layer until the last hidden layer, and the results are sent to the output layer.

$$v_i = \sigma \left(\sum_{j=1}^J u_{ji} o_j + \gamma_i \right) \quad (5)$$

Eq. (4) and Eq. (5) are used to calculate the output at different layers. u and w are learning parameters, β and γ represent biases, K is the number of input units, and J is the number of hidden units. Eq. (5) is simplified to Eq. (6) [35], where σ is the activation function.

$$v = \sigma(oU + Y) = \sigma((\sigma(xW + \beta))U + Y) \quad (6)$$

The determination of each layer greatly affects the performance of multilayer perceptrons, so it needs special attention in their development. Several sources state that there are no definite rules governing the formation of a good architecture in the development of multilayer perceptron methods [36].

III. Methods

This study proposes a trajectory-based sperm motility classification framework that integrates microscopic video analysis, kinematic feature extraction, and neural network optimization. The proposed methodology consists of five major stages, namely (1) video preprocessing, (2) trajectory extraction, (3) feature engineering, (4) model development, and (5) performance evaluation. The overall workflow of the proposed framework is illustrated in Fig. 2.

Fig. 2 illustrates the complete workflow of the proposed sperm motility classification framework. The process begins with microscopic video acquisition, where sperm samples are recorded using a digital microscope. The acquired videos are subsequently converted into multi-page TIFF format to preserve image quality and facilitate frame-by-frame processing. Each frame is then transformed into grayscale images to reduce computational complexity, suppress background noise, and improve contrast between sperm cells and the surrounding environment. Following preprocessing, sperm cells are detected and tracked using the Trackpy library [37]. The detected particles are associated across consecutive frames to generate continuous sperm movement trajectories. To improve trajectory reliability, drift correction and velocity smoothing procedures are applied to minimize the influence of microscope vibration and tracking

noise. The resulting trajectories are subsequently used to extract two primary kinematic features, namely mean velocity and trajectory linearity, which quantitatively describe sperm movement behavior.

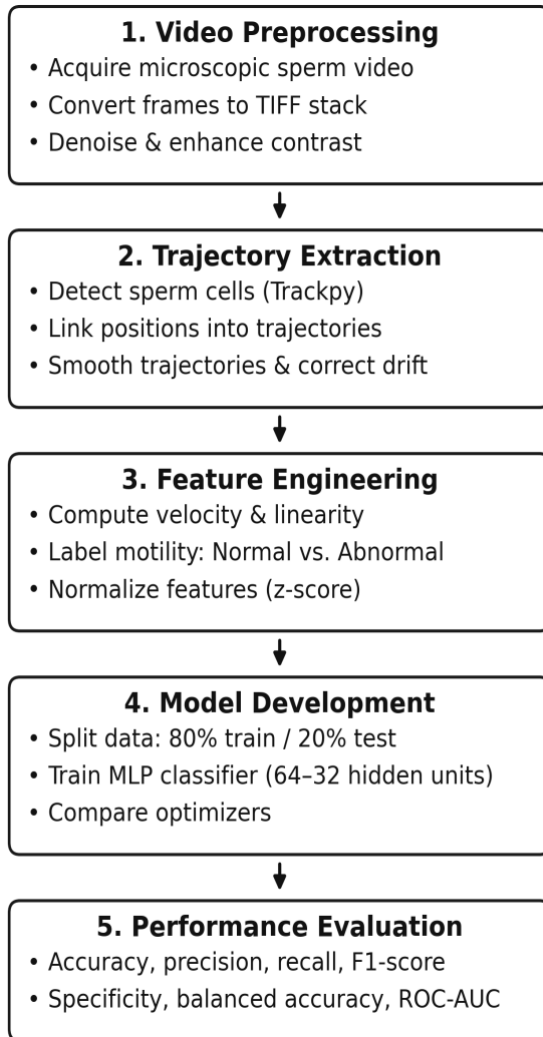


Fig. 2. Proposed research framework for sperm motility classification using trajectory-based feature extraction and neural network optimization

The extracted features are then labeled according to the World Health Organization (WHO) motility criteria. Spermatozoa are categorized as normal when the average velocity is greater than or equal to $25 \mu\text{m/s}$ and the trajectory linearity is greater than or equal to 60%. Otherwise, the spermatozoa are classified as abnormal. Prior to model training, all features are normalized using the StandardScaler technique to ensure zero mean and unit variance, thereby improving neural network convergence and training stability.

The normalized dataset is divided into training and testing subsets using an 80:20 train–test split strategy. The classification stage utilizes a Multilayer Perceptron

(MLP) neural network consisting of an input layer, two hidden layers, and a sigmoid output layer for binary classification. To investigate the impact of optimization strategies on learning performance, four adaptive optimization algorithms are systematically evaluated, namely Adagrad, Adadelata, FTRL, and Adamax. Finally, the trained models are evaluated using multiple performance metrics to determine the most effective optimizer for trajectory-based sperm motility classification. The complete step-by-step procedure of the proposed framework is summarized in [Algorithm 1](#).

Algorithm 1. Proposed Trajectory-Based Sperm Motility Classification Framework

- Input: Microscopic sperm video
Output: Motility class prediction (Normal / Abnormal)
1. Load microscopic sperm video.
 2. Convert video frames into multi-page TIFF format.
 3. Convert each TIFF frame into a grayscale image.
 4. Detect sperm particles in each frame using Trackpy.
 5. Link detected particles across consecutive frames.
 6. Generate sperm movement trajectories.
 7. Apply drift correction to reduce microscope vibration effects.
 8. Smooth velocity profiles using a 1-dimensional uniform filter.
 9. Extract mean velocity from each trajectory.
 10. Extract trajectory linearity from each trajectory.
 11. Assign class label:
 If velocity $\geq 25 \mu\text{m/s}$ AND linearity $\geq 60\%$:
 Label = Normal
 Else:
 Label = Abnormal
 12. Normalize extracted features using StandardScaler.
 13. Split dataset into training data and testing data using an 80:20 ratio.
 14. Construct an MLP model with 2 input neurons, 64 and 32 hidden neurons, and 1 sigmoid output neuron.
 15. For each optimizer in {Adagrad, Adadelata, FTRL, Adamax}:
 Train the MLP model for 50 epochs.
 Evaluate prediction results on the testing dataset.
 Compute accuracy, precision, recall, specificity, F1-score, balanced accuracy, and ROC-AUC.
 16. Compare all optimizer performances.
 17. Select the best-performing optimizer.

Table 2. Summary of the dataset used for sperm motility classification

Dataset Component	Description
Data source	Microscopic video recordings of human sperm samples
Total trajectories	276 sperm movement trajectories
Tracking method	Trackpy library used to detect and track sperm movement
Feature extraction	Mean velocity ($\mu\text{m/s}$) and trajectory linearity (%)
Feature dimension	2 numerical features
Normal motility samples	39 trajectories
Abnormal motility samples	237 trajectories
Classification criteria	Velocity $\geq 25 \mu\text{m/s}$ and linearity $\geq 60\%$
Data normalization	StandardScaler
Train–test split	80% training (198 abnormal, 22 normal) and 20% testing (39 abnormal, 17 normal)
Frame rate	30 frames per second (fps)
Video resolution	1152 × 912 pixels
Recording duration	30 seconds

A. Data Acquisition and Pre-Processing

All data acquisition procedures strictly complied with standard ethical and privacy protocols and received ethical clearance from the Health Research Ethics Committee (Approval No: EA/3985/KEPK-Poltekkes_Sby/V/2025). The dataset used in this study consists of 276 unique sperm trajectories extracted from microscopic videos. The videos were first converted to opaque multi-page TIFFs to allow for lossless quality frame capture [38]. Each frame was processed into a grayscale image to reduce background noise, alleviate lighting issues (chiaroscuro), and enhance the contrast between the sperm cells and the background [39]. This pre-processing is the critical first step in ensuring the visual data is ready for trajectory tracking. Furthermore, the extracted numerical data (velocity and linearity) were normalized using a standard scaler (StandardScaler) to ensure zero mean and unit variance, which effectively accelerates the neural network's convergence. The characteristics of the dataset used in this study are summarized in Table 2.

B. Sperm Motility Detection

Once the grayscale frames were prepared, the Trackpy algorithm was utilized to associate detected sperm cells across sequential frames, encapsulating individual movement trajectories. To improve accuracy, drift corrections were applied to counteract shifts from

microscope vibrations. Feature extraction was performed mathematically: the temporal velocity data was first smoothed using a 1-dimensional uniform filter (size = 20) to reduce tracking noise. Subsequently, path linearity was derived as the ratio of the straight-line displacement to the total curvilinear distance traveled by the sperm.

These two parameters (mean velocity and linearity) served as the primary features. Spermatozoa were categorized using conjunction logic (AND), i.e., sperm is considered normal if it meets the criteria of speed $\geq 25 \mu\text{m/s}$ and linearity $\geq 60\%$. Sperm that did not meet these specific kinematic thresholds were labeled as 'abnormal'. These extracted spatial-temporal features were then structured into a 2-dimensional input vector for the neural network. The dataset is markedly class-imbalanced, comprising 39 normal (14.1%) and 237 abnormal (85.9%) trajectories. This imbalance is accounted for by reporting balanced accuracy and ROC-AUC alongside overall accuracy. The 80:20 split yields a training set of 220 samples (198 abnormal, 22 normal) and a test set of 56 samples (39 abnormal, 17 normal). The velocity threshold of $25 \mu\text{m/s}$ and the linearity threshold of 60% are applied as the authors' operational classification rule, based on WHO motility criteria [1], [22].

C. Mathematical Formulation

To improve the transparency and reproducibility of the proposed framework, the mathematical formulation of trajectory feature extraction, feature normalization, neural network computation, loss minimization, and adaptive optimization algorithms is presented in this subsection.

1. Trajectory-Based Feature Extraction

The proposed framework utilizes sperm trajectories generated by the Trackpy tracking algorithm. Let (x_i, y_i) denote the sperm position at frame i . The displacement between two consecutive positions is calculated using the Euclidean distance, as shown in Eq. (7) [24], [37]:

$$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (7)$$

where d_i is the displacement distance between two consecutive frames, x_i and y_i are the sperm coordinates at frame i , and x_{i-1} and y_{i-1} are the sperm coordinates at the previous frame.

The curvilinear velocity (VCL), which represents the actual velocity of sperm movement along its complete trajectory, is calculated as shown in Eq. (8) [24]:

$$VCL = \frac{\sum_{i=2}^n d_i}{T} \quad (8)$$

where VCL denotes the curvilinear velocity ($\mu\text{m/s}$), d_i denotes the displacement segment, n denotes the total number of trajectory points, and T denotes the observation time (s).

The straight-line velocity (VSL) is calculated to represent the shortest distance travelled between the

initial and final sperm positions, as defined in Eq. (9) [24]:

$$VSL = \frac{\sqrt{(x_n - x_1)^2 + (y_n - y_1)^2}}{T} \quad (9)$$

where VSL denotes the straight-line velocity ($\mu\text{m/s}$), (x_1, y_1) represents the initial sperm coordinates, and (x_n, y_n) represents the final sperm coordinates.

The average path velocity (VAP) is computed from the smoothed trajectory path, as formulated in Eq. (10) [24]:

$$VAP = \frac{\sum_{i=2}^n p_i}{T} \quad (10)$$

where VAP denotes the average path velocity ($\mu\text{m/s}$) and p_i denotes the displacement segment along the average path.

2. Feature Normalization

Since the extracted kinematic features have different numerical scales, feature normalization is applied using the StandardScaler method to ensure that all features contribute equally during model training, as expressed in Eq. (11) [40]:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (11)$$

where z_i denotes the normalized feature value, x_i denotes the original feature value, μ denotes the feature mean, and σ denotes the feature standard deviation.

3. Objective Function and Gradient

The full objective function minimized during training is the mean Binary Cross-Entropy loss computed across all N training samples, as defined in Eq. (12) [25]:

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N L_i \quad (12)$$

where $J(\theta)$ denotes the objective function, θ denotes the trainable parameters, and N denotes the number of training samples.

The gradient used by the optimization algorithms is defined as shown in Eq. (13) [25]:

$$g_t = \nabla_{\theta} J(\theta_t) \quad (13)$$

where g_t denotes the gradient at iteration t and ∇_{θ} denotes the gradient operator. This gradient is subsequently used by each adaptive optimization algorithm to perform parameter updates, as described below.

4. Adaptive Optimization Algorithms

To investigate the influence of optimization strategies on classification performance, four adaptive optimization algorithms are evaluated under identical experimental settings. For the Adagrad optimizer, the parameter update rule is defined as shown in Eq. (14) [41]:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} g_t \quad (14)$$

where θ_t denotes the parameter vector at iteration t , η denotes the learning rate, G_t denotes the accumulated squared gradient, and ϵ denotes a numerical stability constant.

For the Adadelta optimizer, the parameter update step is calculated as shown in Eq. (15) [42]:

$$\Delta\theta_t = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g_t \quad (15)$$

where $\Delta\theta_t$ denotes the parameter update, $RMS[\Delta\theta]_{t-1}$ denotes the root mean square of previous updates, and $RMS[g]_t$ denotes the root mean square of the current gradient.

For the FTRL optimizer, the parameter update rule is expressed as shown in Eq. (16) [43]:

$$\theta_{t+1} = \begin{cases} 0, & |z_t| \leq \lambda_1 \\ -\frac{z_t - \text{sgn}(z_t)\lambda_1}{\frac{\beta + \sqrt{n_t}}{\alpha} + \lambda_2}, & |z_t| > \lambda_1 \end{cases} \quad (16)$$

where z_t denotes the accumulated gradient state, λ_1 and λ_2 denote the L1 and L2 regularization parameters, respectively, α denotes the learning-rate parameter, and β denotes the smoothing coefficient.

For the Adamax optimizer, the parameter update rule is defined as shown in Eq. (17) [44]:

$$\theta_{t+1} = \theta_t - \frac{\eta}{u_t} \hat{m}_t \quad (17)$$

where u_t denotes the exponentially weighted infinity norm estimate, $\hat{m}_t$ denotes the bias-corrected first moment estimate, and η denotes the learning rate.

The mathematical formulations presented above describe the computational processes employed in the proposed framework, ranging from trajectory-based feature extraction to adaptive parameter optimization.

D. Classification Model Training

The structured dataset was used to train a classification model based on a Multilayer Perceptron (MLP) architecture using the TensorFlow/Keras library. The MLP was specifically chosen for its capability to learn complex, non-linear boundaries between the extracted numerical parameters [45].

The network was constructed with an input layer receiving the 2 extracted features, followed by two hidden layers comprising 64 and 32 neurons, respectively. The Rectified Linear Unit (ReLU) activation function was applied to both hidden layers to effectively mitigate the vanishing gradient problem during training. Finally, the output layer consisted of a single neuron utilizing a Sigmoid activation function to facilitate binary classification (Normal vs. Abnormal). Each neuron computes a weighted sum of its inputs and bias as described in Eq. (2). The hidden layer

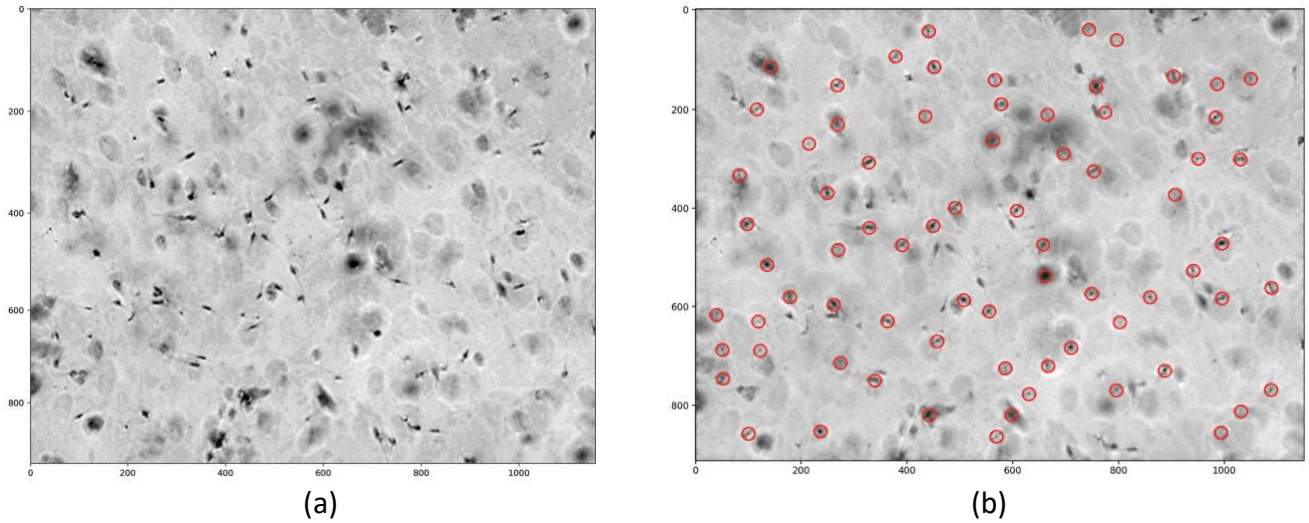


Fig. 3. Sperm preprocessing and detection: (a) grayscale-converted sperm frame, and (b) sperm detection in the initial frame.

output is obtained using the Rectified Linear Unit activation function given in Eq. (18) [46]:

$$h_j = \max(0, Net_j) \quad (18)$$

For the output layer, the predicted probability is calculated using the sigmoid function given in Eq. (19) [25]:

$$\hat{y} = \frac{1}{1 + e^{-z}} \quad (19)$$

where $\hat{y}$ represents the predicted probability of abnormal sperm motility. To train the network, the parameters are optimized by minimizing the binary cross-entropy loss function defined in Eq. (20) [25]:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (20)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability. During the optimization process, the network parameters are updated iteratively using adaptive optimization algorithms.

E. Model Validation and Performance Testing

To robustly evaluate the model's generalization capability on unseen data, the dataset of 276 trajectories was partitioned using a fixed 80:20 train-test split strategy (random_state = 42). This allocated approximately 220 samples for training and 56 samples for the validation and prediction testing phase.

The models were compiled using binary cross-entropy as the loss function and trained for exactly 50 epochs with a batch size of 32. Early stopping was not employed; instead, the convergence criteria were visually and quantitatively confirmed by observing the stabilization of the training accuracy and loss curves throughout the entire 50-epoch timeframe. Performance was subsequently evaluated using a comprehensive set of metrics including overall

accuracy, per-class precision, per-class recall, per-class F1-score, specificity, and balanced accuracy. Balanced accuracy, defined as the arithmetic mean of sensitivity and specificity, was specifically included to provide an imbalance-corrected measure of classification performance. Utilizing these metrics ensures an objective evaluation of the model's predictive capabilities across the established spatial-temporal thresholds [47], [48].

F. Comparative Analysis

In the final stage, as outlined in the research methodology framework (Fig. 2), a comparative analysis was conducted to evaluate the performance differences among the four optimization algorithms (Adagrad, Adadelta, FTRL, and Adamax). The comparative analysis was carried out by extracting and evaluating the prediction results on the unseen test data. The convergence behavior of each optimizer was quantitatively analyzed by observing the stabilization of the training accuracy and loss curves over the 50-epoch timeframe. Furthermore, confusion matrices were generated to provide a detailed comparative breakdown of true and false predictions across the normal and abnormal classes. This comprehensive performance analysis ensures an equitable comparison to determine the most optimal and robust algorithm for classifying sperm motility abnormalities.

IV. Results

This section presents the outcomes of trajectory tracking, the classification of sperm motility, and a comparative performance analysis of the Multilayer Perceptron (MLP) model utilizing four different optimization algorithms.

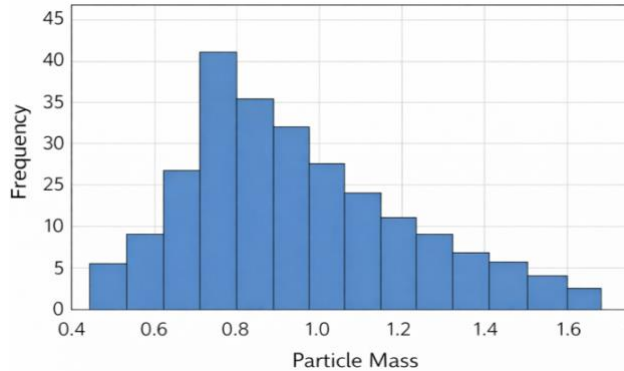


Fig. 4. Histogram of particle mass distribution

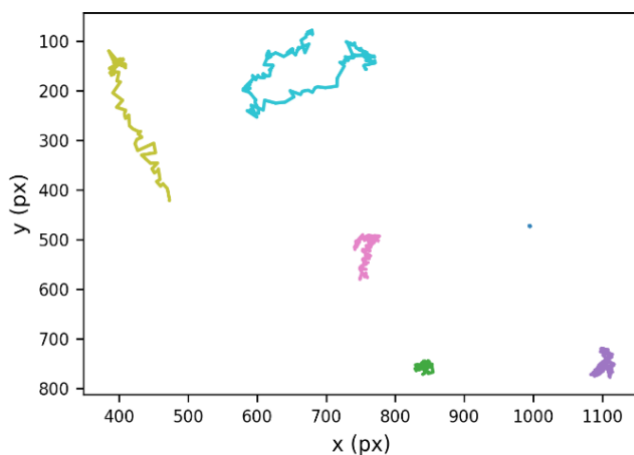
A. Trajectory Tracking

The initial conversion of the video frames into a grayscale format successfully simplified the image complexity while enhancing the contrast between the sperm cells and the background. Fig. 3(a) illustrates a processed sperm frame, demonstrating the clarity achieved. Subsequently, the Trackpy detection algorithm successfully identified moving sperm events. Displayed in Fig. 3(b), the detection results highlight sperm heads and tails traced to the initial frame. The particle mass distribution presented in Fig. 4 illustrates the statistical distribution of detected particle masses obtained during the preprocessing stage. The histogram indicates that the majority of detected particles fall within the expected mass range of sperm cells, confirming that the particle detection process successfully isolates biologically relevant sperm candidates. This distribution also demonstrates that potential background noise and non-cellular artifacts have been effectively filtered, ensuring that only valid sperm cell trajectories are used for subsequent motion

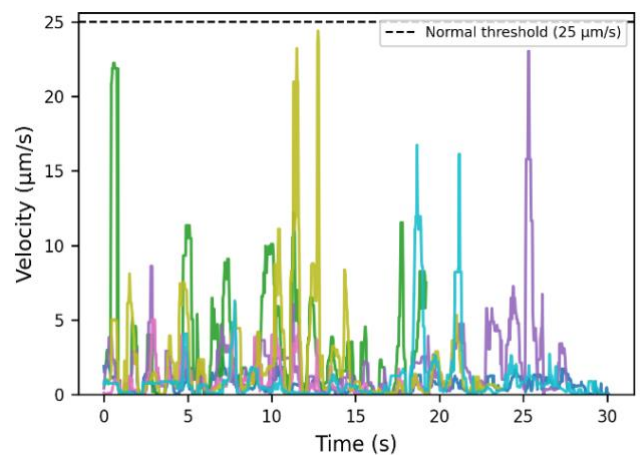
tracking and velocity analysis. The trajectory tracking algorithm then links detected sperm cells across consecutive frames to construct continuous motion paths. For clarity, Fig. 5(a) shows a representative subset of tracked sperm trajectories, each in a distinct color, enabling clear observation of individual movement patterns across the microscopic field. Finally, the spatial displacement of each sperm cell over time is used to compute the dynamic velocity profiles. The velocity profiles of the same representative sperm are presented in Fig. 5(b), where each line shows the velocity variation of an individual sperm across time, with the dashed line marking the 25 $\mu\text{m/s}$ normal-motility threshold. These velocity patterns form the fundamental quantitative features used for the subsequent sperm motility classification process.

B. Classification of Sperm Motility

Sperm are then classified as normal or abnormal depending on their average velocity and trajectory linearity. For sperm to be classified as normal, they are required to have average velocity and linearity that are both above threshold. Conversely, abnormal sperm have velocity and linearity that are below their respective thresholds. This classification is done by withholding a linearity value and comparing a sperm particle's velocity to the threshold to be classified as normal or abnormal. The data are then split into two parts, training and test, respectively, in an 80:20 ratio. A classification model is built using average speed and linearity characteristics via a Multilayer Perceptron (MLP). Once the model is constructed, training is performed using the training data and model performance is evaluated on the test data. MLP training has to be done to the point that the model is able to classify sperm correctly, thus monitoring accuracy and loss during training to avoid underfitting or overfitting. Various optimization algorithms have been used in



(a)



(b)

Fig. 5. Sperm kinematic analysis for representative sperm cells: (a) movement trajectories, and (b) the corresponding velocity profiles over time relative to the 25 $\mu\text{m/s}$ normal-motility threshold.

training this particular Multilayer Perceptron (MLP) model to classify sperm motility in four different approaches, and the performance of individual optimizers has been compared using accuracy and loss trends observed in a 50-epoch training timeframe.

Using the extracted mean velocity and trajectory linearity, the MLP model was trained to classify the

abnormal (69.6%) and 17 normal (30.4%) trajectories, meaning that high overall accuracy does not necessarily indicate strong model performance. For instance, a baseline model that always predicts the majority ("abnormal") class yields an accuracy of 69.6%, which closely matches the FTRL optimizer's result. Evaluating the models using balanced accuracy

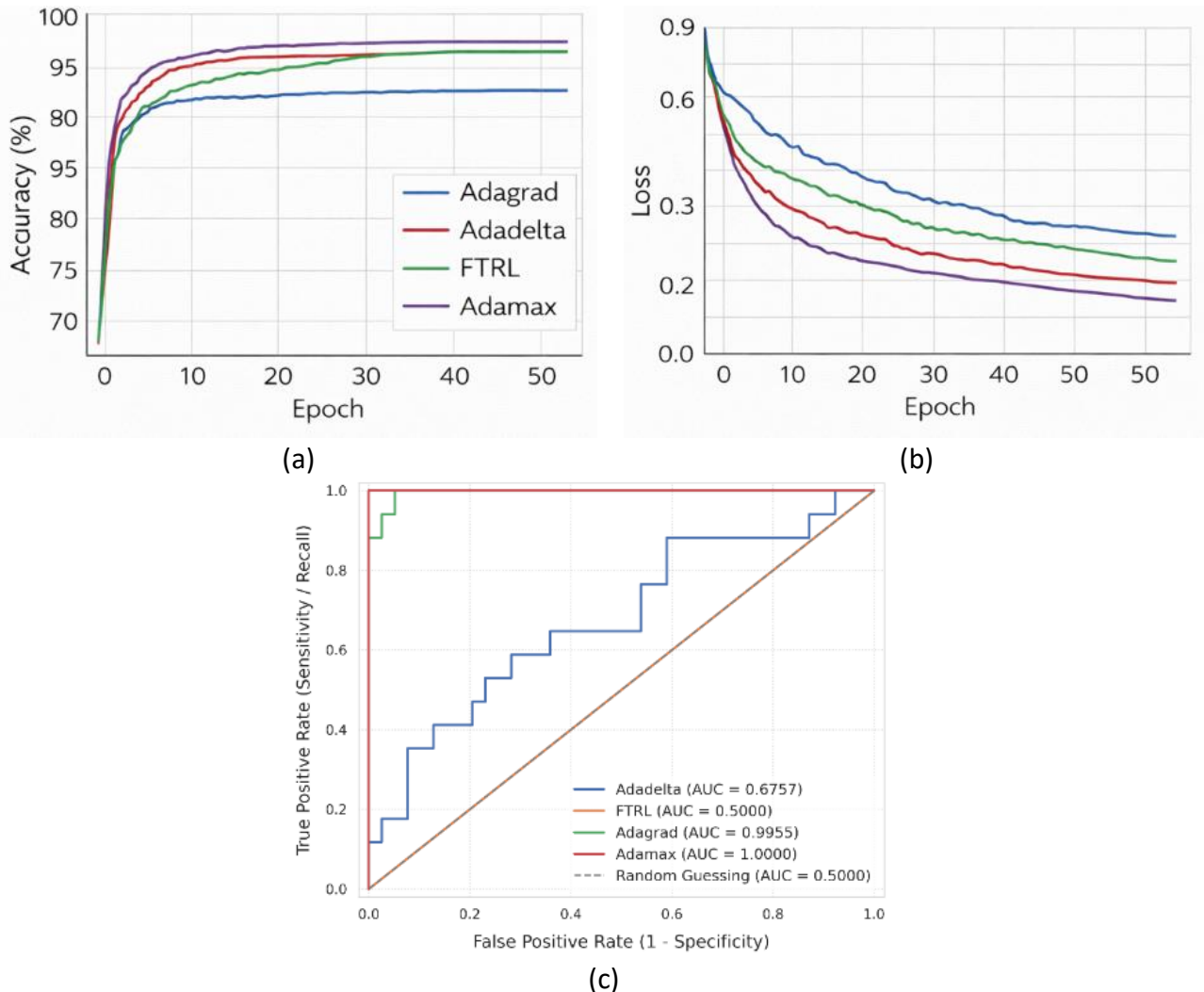


Fig. 6. Comparative performance of the MLP model using different optimizers: (a) training accuracy curves, (b) training loss curves, and (c) ROC curves with corresponding AUC values

sperm samples into normal and abnormal categories. The model's performance was evaluated over a 50-epoch timeframe. The training accuracy and loss trends for all four optimizers are depicted in Fig. 6(a) and (b), respectively.

The experimental results demonstrate that the choice of optimizer significantly influences the classification performance of the proposed neural network model. The Adamax optimizer achieved the highest classification accuracy of 96.43%, outperforming Adagrad (92.86%), FTRL (69.64%), and Adadelata (67.86%). The test set consists of 39

reveals that FTRL (0.50) and Adadelata (0.59) perform near random chance, whereas Adagrad (0.95) and Adamax (0.94) effectively classify both categories despite the data imbalance. These classification differences are closely tied to the underlying training dynamics of each optimizer. Analysis of the training curves in Fig. 6(a) indicates that Adamax converged faster than the other optimizers. The model reached stable accuracy within approximately 15 epochs, while Adagrad required nearly 30 epochs to reach convergence. In contrast, Adadelata and FTRL failed to achieve stable learning behavior and produced

significantly lower classification accuracy. The corresponding training accuracy and loss curves for each optimizer are illustrated in Fig. 6(a) and (b), respectively.

The overall threshold-independent class separation capacity of the evaluated models is further illustrated by the Receiver Operating Characteristic (ROC) curves in Fig. 6(c). Both the Adamax and Adagrad optimizers demonstrated exceptional discriminative capabilities, achieving an Area Under the Curve (AUC) of 1.00. Conversely, the FTRL curve strictly tracked the

diagonal reference line (AUC = 0.50), confirming its lack of predictive discrimination under majority-class bias, while Adadelta showed moderate sensitivity with an intermediate AUC value of 0.68. It should be noted that ROC-AUC is a threshold-independent measure of class separability, whereas balanced accuracy depends on the selected classification threshold. Consequently, Adamax and Adagrad can share an identical ROC-AUC of 1.00 while exhibiting slightly different balanced-accuracy values (0.94 and 0.95, respectively).

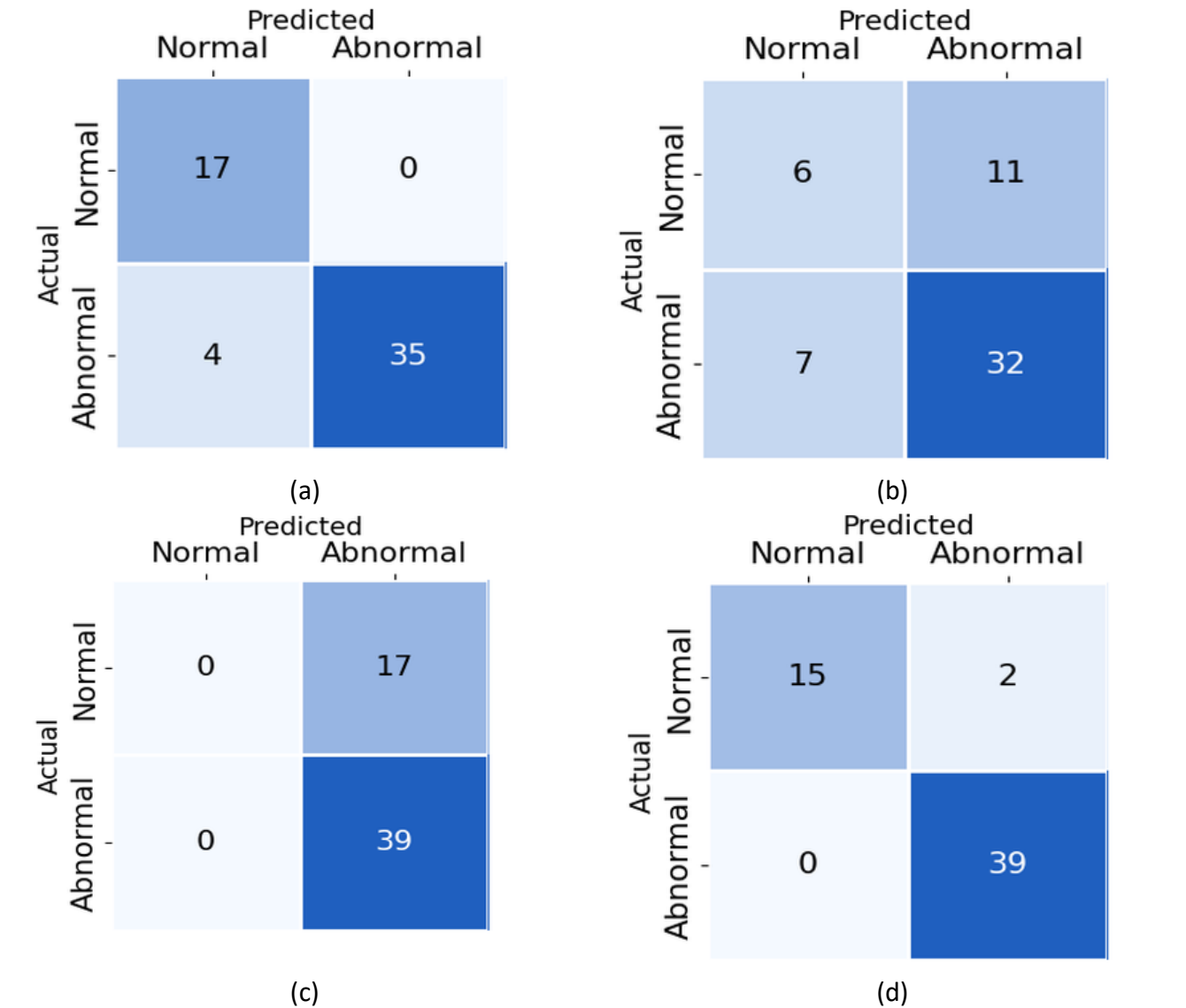


Fig. 7. Comparative confusion matrices of the MLP classifier using different optimizers:(a) Adagrad, (b) Adadelta, (c) FTRL, and (d) Adamax.

C. Comparative Analysis

To rigorously evaluate the classification performance, confusion matrices were generated based on the test dataset predictions (Fig. 7). The model optimized with Adagrad (Fig. 7(a)) successfully identified all 17 normal

samples (recall = 1.00) but misclassified 4 of the 39 abnormal samples as normal (yielding 4 false negatives for the abnormal class), resulting in an overall accuracy of 92.86%. Per-class metrics for Adagrad yielded a precision of 1.00, recall of 0.90, and

an F1-score of 0.95 for the abnormal class, alongside a precision of 0.81, recall of 1.00, and an F1-score of 0.89 for the normal class. The corresponding specificity was 0.90, and the balanced accuracy was 0.95. In contrast, Adadelta (Fig. 7(b)) produced 11 false negatives and 7 false positives. This optimizer yielded a precision of 0.74, recall of 0.82, and an F1-score of 0.78 for the abnormal class, but demonstrated substantially weaker performance on the normal class with a precision of 0.46, recall of 0.35, and an F1-score of 0.40. Its specificity was 0.82, and its balanced accuracy was 0.59, indicating limited discriminative capability on the minority class.

The FTRL model (Fig. 7(c)) exhibited extreme class bias, predicting every trajectory as abnormal (39 TN, 17 FN, 0 TP, 0 FP). Consequently, while recall for the majority abnormal class reached 1.00, the minority normal class yielded a precision, recall, and F1-score of 0.00. The absence of false positives resulted in a specificity of 1.00. However, the balanced accuracy was exactly 0.50. This performance is equivalent to random guessing, confirming that FTRL's 69.64% overall accuracy purely reflects the underlying class prevalence of the test set rather than authentic discriminative learning. The Adamax optimizer (Fig. 7(d)) delivered the most outstanding results. It produced only 2 false negatives (normal samples incorrectly classified as abnormal) while successfully identifying all 39 abnormal samples, yielding an overall accuracy of 96.43%. Furthermore, detailed metric analysis reveals that Adamax achieved a superior F1-Score of 0.97 for the abnormal class and 0.94 for the normal class, alongside perfect precision (1.00) in identifying normal sperm and perfect recall (1.00) for the abnormal class (0 false positives). Specificity was 1.00 and balanced accuracy was 0.94, confirming reliable performance across both classes under imbalanced conditions. The quantitative performance comparison of the evaluated optimizers is summarized in Table 3. As shown in the table, the Adamax optimizer achieved the highest classification accuracy of 96.43%, followed by Adagrad with 92.86%, while FTRL and Adadelta obtained lower accuracies of 69.64% and 67.86%, respectively.

V. Discussion

A. Interpretation of Optimizer Behavior

The experimental results indicate that the choice of optimization algorithm significantly influences the learning behavior and classification performance of the proposed Multilayer Perceptron (MLP) model. As presented in Table 3, the Adamax optimizer achieved the highest classification accuracy of 96.43%, outperforming Adagrad (92.86%), FTRL (69.64%), and Adadelta (67.86%). These results demonstrate that Adamax provides the most stable optimization

behavior for the trajectory-based sperm motility dataset used in this study. A critical consideration in interpreting these results is the class imbalance present in the dataset. The test set consisted of 39 abnormal samples (69.6%) and 17 normal samples (30.4%). Under this distribution, overall accuracy alone does not provide a complete assessment of model performance. FTRL achieved an accuracy of 69.64% by assigning all samples to the abnormal class, resulting in a balanced accuracy of 0.50 and an ROC-AUC of 0.50, which signifies no diagnostic discrimination. Similarly, Adadelta's weak separating power is verified by its low ROC-AUC score of 0.68. A similar limitation was observed for Adadelta, which achieved 67.86% accuracy but a balanced accuracy of only 0.59 due to low recall for the normal class (0.35). In contrast, Adagrad and Adamax achieved balanced accuracies of 0.95 and 0.94, respectively, indicating strong discrimination between normal and abnormal classes. Adagrad attained a recall of 1.00 for normal samples, with no false negatives in the test set. These results indicate that Adagrad and Adamax were less affected by the class imbalance and maintained strong performance across both classes. The superior performance of Adamax can be attributed to its parameter update mechanism based on the infinity norm (L_∞), which allows the optimizer to maintain stable parameter updates even when gradient magnitudes vary during training. This characteristic makes Adamax particularly suitable for handling the dense numerical feature space formed by the kinematic parameters used in this study, namely velocity and linearity. Consequently, the optimizer achieves smoother convergence compared to other adaptive optimization methods evaluated in this research.

In comparison, the Adagrad optimizer achieved an accuracy of 92.86%, which is relatively competitive but slightly lower than Adamax. This behavior can be explained by Adagrad's cumulative gradient mechanism, where the learning rate continuously decreases as the squared gradients accumulate over time. Although this mechanism stabilizes early learning, it may limit the model's ability to further refine its weights during later training epochs. Meanwhile, Adadelta and FTRL produced considerably lower accuracies of 67.86% and 69.64%, respectively. Their corresponding balanced accuracies of 0.59 and 0.50 confirm that neither optimizer achieves meaningful discrimination between the two classes within this feature space. These results suggest that both optimizers are less suitable for the dense and low-dimensional feature space used in trajectory-based sperm motility classification. In particular, the FTRL optimizer is generally designed for large-scale sparse datasets with categorical features, which may explain

its poor performance when applied to the continuous numerical features used in this study. Overall, the findings indicate that adaptive optimization algorithms capable of maintaining stable learning rates, such as Adamax, are more effective for neural network-based classification of sperm motility trajectories.

B. Comparison with Existing Studies

To position the proposed framework within the broader body of research, Table 4 compares this study with seven representative works in automated sperm analysis and neural-network optimizer evaluation. The comparison considers the underlying method, the type of data analyzed, and the reported performance. Two observations emerge. First, among trajectory and video-based sperm studies, the proposed MLP with the

Adamax optimizer attains a competitive accuracy of 96.43% while relying on only two interpretable kinematic features (mean velocity and linearity), in contrast to computationally heavier convolutional pipelines. Image-based convolutional models such as Riordon et al. [17] report a comparable accuracy (94.1%) but operate on static images and therefore cannot capture the temporal motion characteristics that are central to motility assessment. Second, the optimizer-comparison literature [49], [50] consistently reports that adaptive optimizers outperform non-adaptive ones, which agrees with the present finding that Adamax and Adagrad clearly surpass Adadelta and FTRL on this dense, low-dimensional feature space.

Table 4. Comparison of our work with related sperm-analysis and optimizer-evaluation studies

Study	Method	Data Type	Reported Performance	Task Type
Setiawan et al. (2020) [3]	Mixture of Gaussian (MoG) V2	Microscopic video of human spermatozoa	F-measure of 0.9449 for moving sperm tracking/detection	Detection & tracking
Diyasa et al. (2024) [16]	Gaussian Mixture Model + matching-based algorithm	Microscopic video of human spermatozoa	87.7% accuracy for motility abnormality determination	Detection & abnormality determination
Dewan et al. (2018) [39]	Deep CNN + tracking/association algorithm	Microscopic video of live human semen	Pearson correlation of 0.86 (concentration) and 0.818 (motility) with SQA-V Gold, 7.44% mean tracking miss rate	Tracking & motility prediction
Hicks et al. (2019) [21]	Classical ML (regression) & 2D CNNs	VISEM dataset (Semen video + participant data)	Automated motility prediction with deep learning methods yielding consistent, generalized results	Motility prediction
Riordon et al. (2019) [17]	CNN (VGG16 transfer learning)	Static human sperm images	94.1% accuracy (static images, no temporal motion captured)	Classification
Baksir et al. (2020) [18]	Backpropagation neural network	Fertility dataset	80.32% accuracy	Classification
Elbert et al. (2025) [50]	CNN with optimizer comparison (SGD, Adam, RMSprop, Adamax, etc.)	Real-time face-recognition images	Adam achieved the best accuracy among the compared optimizers	Optimizer evaluation
This Study	MLP + adaptive optimizers (Adagrad, Adadelta, FTRL, Adamax)	Trajectory-derived kinematic features	96.43% accuracy, ROC-AUC 1.00 (Adamax), 276 trajectories, fixed 80:20 split	Classification

Note: The reported-performance values correspond to different tasks (detection/tracking, classification, motility prediction, and optimizer evaluation, as indicated in the Task Type column) and are therefore not directly comparable. For example, an F-measure for sperm detection is not equivalent to a classification accuracy.

A direct comparison with the detection and tracking studies in Table 4 shows a methodological similarity and a clear difference in scope. Setiawan et al. [3] and Diyasa et al. [16] analyzed the same type of data as the present work, namely microscopic videos of human spermatozoa. Setiawan et al. reported an F-measure of 0.9449 for moving sperm detection, whereas Diyasa et al. combined a Gaussian Mixture Model with a

matching-based algorithm and reported 87.7% accuracy for motility abnormality determination. The similarity lies in the use of trajectory information derived from frame-by-frame motion analysis. The difference is that both studies derive the decision from the tracking stage itself rather than from a trained classifier, and neither examines the effect of the optimization algorithm. The present study extends that pipeline by

treating the tracked trajectory as the input of a supervised classifier, so the detection and abnormality results reported in those works cannot be equated with the 96.43% classification accuracy obtained here. Dewan et al. [39] also remained at the measurement level. They combined a deep convolutional network with an association algorithm and validated the output against the SQA-V Gold analyzer, obtaining a Pearson correlation of 0.86 for concentration and 0.818 for motility with a 7.44% mean tracking miss rate. Their approach agrees with ours in the use of temporal information, but it requires a convolutional backbone and a reference analyzer, while the proposed model uses only two scalar kinematic descriptors and no external instrument.

Among the classification and prediction studies, the differences come mainly from the input representation. Riordon et al. [17] obtained 94.1% accuracy with a VGG16 transfer learning model, which is close to the 96.43% obtained here. The important distinction is that their model operates on static sperm images and therefore encodes morphology rather than motion, so it cannot express velocity or path linearity, which are the defining quantities of motility. Baksir et al. [18] reported 80.32% accuracy using a backpropagation neural network, the same family of architecture used in this study. The gap of roughly 16 percentage points therefore comes from the feature space, not the learning algorithm, because their model was trained on a tabular fertility questionnaire dataset instead of measured sperm kinematics. Hicks et al. [21] applied classical regression and 2D convolutional networks to the VISEM dataset and reported consistent and generalizable motility prediction. Their work is similar to ours in targeting motility, but it requires video together with participant and semen analysis variables, which limits reproducibility in routine laboratories, while the proposed framework depends on a single video modality.

The optimizer evaluation study of Elbert et al. [50] provides the closest methodological parallel to the design used here, because it also compares SGD, Adam, RMSprop and Adamax within one fixed architecture. Both studies agree that adaptive optimizers outperform non-adaptive ones, and Fatimah [49] reached the same conclusion for medical image processing. The two studies differ in which adaptive optimizer prevails. Adam performed best for their real-time face recognition CNN, while Adamax was superior in the present setting. This difference is consistent with the nature of the feature space, because the infinity norm update of Adamax suits a dense, two-dimensional, and low-noise kinematic representation better than the high-dimensional sparse gradients produced by convolutional image models. The comparison with the seven studies above therefore indicates that the contribution of this study is not a

higher accuracy value alone, but the combination of trajectory-level input, an interpretable two-feature representation, and an explicit choice of optimizer.

C. Limitations of the Study

Several limitations should be acknowledged when interpreting these results. First, the dataset is relatively small and class-imbalanced, comprising 276 trajectories with a test set of 39 abnormal (69.6%) and 17 normal (30.4%) samples; this imbalance inflates raw accuracy and is the reason balanced accuracy and ROC-AUC are reported alongside it. Second, model evaluation used a single fixed 80:20 train-test split rather than k-fold cross-validation, so the reported metrics may not fully reflect variance across different data partitions. Third, classification relies on only two kinematic features (mean velocity and linearity), which, while interpretable and lightweight, omit morphological and additional dynamic descriptors (e.g., VCL, VAP, ALH, BCF) that could improve discrimination. Fourth, the trajectories were acquired under a single set of microscopes and recording conditions; performance under varying magnifications, frame rates, illumination, and sample-preparation protocols remains to be validated. Finally, FTRL and Adadelta were evaluated with default hyperparameters, so their weak performance reflects suitability for this dense, low-dimensional feature space rather than an exhaustively tuned optimum.

D. Practical Implications

The findings carry direct practical implications for reproductive healthcare. Because the proposed model is lightweight and depends only on two trajectory-derived features, it can be deployed as a real-time pre-screening module on existing laboratory microscopes or low-cost Computer-Aided Sperm Analysis (CASA) systems without specialized GPU hardware. Acting as an automated second opinion, it can rapidly flag abnormal motility patterns, reduce the manual verification workload for embryologists, and mitigate the inter-observer subjectivity that limits manual semen analysis. The optimizer comparison also offers methodological guidance for similar small-scale, low-dimensional biomedical classification tasks: adaptive optimizers that maintain stable learning rates, particularly Adamax, are preferable to optimizers designed for large-scale sparse data such as FTRL. More broadly, demonstrating clinically useful accuracy from a simple, interpretable feature set supports the development of affordable, explainable fertility-screening tools for resource-limited settings.

VI. Conclusion

This research successfully developed a fully automated system for the identification of abnormally motile sperm using an artificial neural network (ANN) model trained with various optimization algorithms. The Adamax algorithm achieved the highest accuracy

(96.43%), followed by Adagrad (92.86%), while the other algorithms (Adadelta and FTRL) yielded significantly lower accuracies (67.86% and 69.64%, respectively). These results demonstrate that the choice of optimization algorithm heavily impacts the overall competency and diagnostic capability of the model in differentiating sperm trajectories.

Despite these promising results, this study has several limitations. The dataset is currently constrained to 276 unique sperm trajectories, and the model evaluation utilized a fixed 80:20 train-test split rather than k-fold cross-validation, which may limit the absolute generalizability of the findings. Furthermore, the classification relies solely on two kinematic features (mean velocity and linearity). Future work should aim to expand the dataset across diverse sperm samples captured under varying clinical conditions, and incorporate additional morphological or dynamic features. Researchers could also explore more advanced heuristic optimization algorithms to provide even greater predictive competency. The immediate sustainable value of this work lies in its practical clinical application. This lightweight MLP model can be seamlessly integrated as a real-time, pre-screening software module within existing laboratory microscopes or affordable Computer-Aided Sperm Analysis (CASA) setups. Acting as an automated "second opinion," it can rapidly flag abnormal trajectories, thereby reducing the manual verification workload for embryologists, minimizing inter-operator subjectivity, and providing appropriate, timely clinical assistance in fertility diagnostics.

Acknowledgment

The authors would like to express their sincere gratitude to the Institute for Research and Community Service (LPPM) of Universitas Pembangunan Nasional Veteran Jawa Timur. The support and commitment of the institution in promoting research and academic development have played a significant role in the successful completion of this work.

Funding

This research was funded by the Institute for Research and Community Service (LPPM) of Universitas Pembangunan Nasional Veteran Jawa Timur as part of an internal research funding scheme with decree 129/SPK/C.C4/PPK.DHK/IV/2026 and grant number 30/UN63.8/LT-Kontrak/IV/2026

Data Availability

The dataset and software code that support the findings of this study are available in a Google Drive repository at the following URL: <https://s.id/yIF5J>. Access to any sensitive or personally identifiable information is restricted in accordance with the ethical approval and participant confidentiality requirements.

Code Availability

The source code used for preprocessing, feature extraction, model training, and evaluation is available in the repository stated in the Data Availability section.

Author Contribution

I Gede Susrama Mas Diyasa contributed to conceptualization, methodology, validation, formal analysis, investigation, resources, writing—original draft, visualization, supervision, and project administration. Vaizal Asy'ari contributed to conceptualization, methodology, software, validation, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, and project administration. Ani Dijah Rahajoe contributed to conceptualization, methodology, validation, investigation, data curation, writing—original draft, and supervision. Deshinta Arrova Dewi contributed to conceptualization, formal analysis, investigation, writing—original draft, and funding acquisition. Sayyidah Humairah contributed to software, validation, formal analysis, and investigation. Muhammad Nashif Farid contributed to resources, data curation, writing—original draft, and writing—review and editing. Faikul Umam contributed to methodology, validation, formal analysis, and data curation. All authors have read and agreed to the published version of the manuscript.

Declarations

Ethical Approval

This study has received ethical clearance in the form of an ethical exemption from the Health Research Ethics Committee (KEPK) of Poltekkes Kemenkes Surabaya, Indonesia, under approval number EA/3985/KEPK-Poltekkes_Sby/V/2025. The research protocol was reviewed and declared ethically appropriate in accordance with the seven ethical standards of WHO (2011), referring to the CIOMS Guidelines (2016), including social value, scientific validity, equitable assessment of benefits and risks, avoidance of exploitation, confidentiality and privacy, and informed consent. The ethical approval is valid from November 03, 2025, to November 03, 2026. All research procedures were conducted in compliance with the approved ethical guidelines, and data confidentiality and privacy were maintained throughout the study.

Consent for Publication

Written informed consent for publication was obtained from all participants involved in this study.

Informed Consent

Written informed consent was obtained from all participants prior to sample collection and data use.

Competing Interests

The authors declare that they have no competing interests.

References

- [1]. WHO, WHO Manual for the Laboratory Examination and Processing of Human Semen, 6th Edition. Geneva, Switzerland: World Health Organization, 2021.
- [2]. P. Choudhary, P. Dogra, and K. Sharma, "Infertility and lifestyle factors: how habits shape reproductive health," *Middle East Fertil. Soc. J.*, vol. 30, Art. no. 14, pp. 1–9, 2025, doi: 10.1186/s43043-025-00228-7.
- [3]. A. Setiawan, I. G. S. Mas Diyasa, M. Hatta and E. Y. Puspaningrum, "Mixture gaussian V2 based microscopic movement detection of human spermatozoa," *International Journal of Advances in Intelligent Informatics* Vol. 6, No. 2, pp. 210-222, July 2020, doi: 10.26555/ijain.v6i2.507.
- [4]. V. Chang et al., "Gold-standard for computer-assisted morphological sperm analysis," *Comput. Biol. Med.*, vol. 83, pp. 143–150, Apr. 2017, doi: 10.1016/j.combiomed.2017.03.004.
- [5]. D. S. Guzick et al., "Sperm Morphology, Motility, and Concentration in Fertile and Infertile Men," *N. Engl. J. Med.*, vol. 345, no. 19, pp. 1388–1393, Nov. 2001, doi: 10.1056/NEJMoa003005.
- [6]. T. G. Cooper et al., "World Health Organization reference values for human semen characteristics," *Hum. Reprod. Update*, vol. 16, no. 3, pp. 231–245, May 2010, doi: 10.1093/humupd/dmq020.
- [7]. F. Hamzah, Z. A. Mulud, M. M. Napes, S. A. Mubarak, and R. Shafie, "Psychological Distress and Quality of Life Among Infertility Couples Undergoing Infertility Treatment in Malaysia," *J. Keperawatan Indones.*, vol. 28, no. 3, pp. 224–235, Nov. 2025, doi: 10.7454/jki.v28i3.1271.
- [8]. L. Bueno-Sánchez, T. Alhambra-Borrás, A. Gallego-Valadés, and J. Garcés-Ferrer, "Psychosocial Impact of Infertility Diagnosis and Conformity to Gender Norms on the Quality of Life of Infertile Spanish Couples," *Int. J. Environ. Res. Public. Health*, vol. 21, no. 2, p. 158, Jan. 2024, doi: 10.3390/ijerph21020158.
- [9]. M. Rubessa et al., "High-throughput sperm assay using label-free microscopy: morphometric comparison between different sperm structures of boar and stallion spermatozoa," *Anim. Reprod. Sci.*, vol. 219, p. 106509, Aug. 2020, doi: 10.1016/j.anireprosci.2020.106509.
- [10]. J. Yaniz, C. Alquézar-Baeta, and J. Yagüe-Martínez, "Expanding the Limits of Computer-Assisted Sperm Analysis through the Development of Open Software," *Biology*, vol. 9, no. 8, p. 207, August 2020, doi: 10.3390/biology9080207.
- [11]. M. K. Panner Selvam, A. K. Moharana, S. Baskaran, R. Finelli, M. C. Hudnall, and S. C. Sikka, "Current updates on involvement of artificial intelligence and machine learning in semen analysis," *Medicina*, vol. 60, no. 2, p. 279, Feb. 2024, doi: 10.3390/medicina60020279.
- [12]. J. Lammers, S. Chtourou, A. Reignier, S. Loubersac, P. Barrière, and T. Fréour, "Comparison of two automated sperm analyzers using 2 different detection methods versus manual semen assessment," *J. Gynecol. Obstet. Hum. Reprod.*, vol. 50, no. 8, p. 102084, Oct. 2021, doi: 10.1016/j.jogoh.2021.102084.
- [13]. R. Finelli, K. Leisegang, S. Tumallapalli, R. Henkel, and A. Agarwal, "The validity and reliability of computer-aided semen analyzers in performing semen analysis: a systematic review," *Transl. Androl. Urol.*, vol. 10, no. 7, pp. 3069–3079, Jul. 2021, doi: 10.21037/tau-21-276.
- [14]. E. Molina and J. Parraga-Alava, "Artificial Neural Networks for Classification Tasks: A Systematic Literature Review," *Enfoque UTE*, vol. 15, no. 4, pp. 1–10, Oct. 2024, doi: 10.29019/enfoqueute.1058.
- [15]. K. Qaderi, F. Sharifipour, M. Dabir, R. Shams, and A. Behmanesh, "Artificial intelligence (AI) approaches to male infertility in IVF: a mapping.review," *Eur. J. Med. Res.*, vol. 30, no. 1, p. 246, Apr. 2025, doi: 10.1186/s40001-025-02479-6.
- [16]. I. G. S. Mas Diyasa, W. S. J. Saputra, A. A. N. Gunawan, D. Herawati, S. Munir, and S. Humairah, "Abnormality Determination of Spermatozoa Motility Using Gaussian Mixture Model and Matching-based Algorithm," *J. Robot. Control JRC*, vol. 5, no. 1, pp. 103–116, Jan. 2024, doi: 10.18196/jrc.v5i1.20686.
- [17]. J. Riordon, C. McCallum, and D. Sinton, "Deep learning for the classification of human sperm," *Comput. Biol. Med.*, vol. 111, p. 103342, Aug. 2019, doi: 10.1016/j.combiomed.2019.103342.
- [18]. A. H. Baksir, A. Fuad, F. Tempola, and R. Rosihan, "Fertility Prediction Using Artificial Neural Networks Using the Backpropagation Method," *JIKO J. Inform. Dan Komput.*, vol. 3, no. 2, pp. 107–112, Aug. 2020, doi: 10.33387/jiko.v3i2.1922.
- [19]. S. Kosuge and T. Hamagami, "Sperm Detection and Tracking Model Using HDE Transformer with Spatio-Temporal Deformable Attention for Sperm Analysis Automation," *IEEE Access*, vol. 13, pp. 51978–51985, 2025, doi: 10.1109/ACCESS.2025.3552792.

- [20]. O. Ogunmolu, X. Gu, S. Jiang, and N. Gans, "Nonlinear Systems Identification Using Deep Dynamic Neural Networks," 2016, arXiv. doi: 10.48550/ARXIV.1610.01439.
- [21]. S. A. Hicks et al., "Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction," *Sci. Rep.*, vol. 9, no. 1, p. 16770, Nov. 2019, doi: 10.1038/s41598-019-53217-y.
- [22]. WHO, "Infertility Prevalence Estimates, 1990–2021. Geneva, Switzerland: World Health Organization, 2023.
- [23]. R. Dcunha et al., "Current Insights and Latest Updates in Sperm Motility and Associated Applications in Assisted Reproduction," *Reprod. Sci.*, vol. 29, no. 1, pp. 7–25, Jan. 2022, doi: 10.1007/s43032-020-00408-y.
- [24]. M. Kraemer, C. Fillion, B. Martin-Pont, and J. Auger, "Factors influencing human sperm kinematic measurements by the Celltrak computer-assisted sperm analysis system," *Hum. Reprod.*, vol. 13, no. 3, pp. 611–619, Mar. 1998, doi: 10.1093/humrep/13.3.611.
- [25]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [26]. O. Shobayo and R. Saatchi, "Developments in Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation," *Diagnostics*, vol. 15, no. 9, Art. no. 1072, 2025, doi: 10.3390/diagnostics15091072.
- [27]. A. R. Pasaribu, S. Sawaluddin, S. Sitorus, and M. Syahputra, "Artificial Neural Networks Predict Covid-19 Cases in Indonesia Using the Backpropagation Method," *Leibniz J. Mat.*, vol. 4, no. 1, pp. 34–46, Jan. 2024, doi: 10.59632/leibniz.v4i1.388.
- [28]. B. Shi, "An Exploration of the Early Warning System for College Students' Academic Performance Based on BP Neural Network Driven by Multidimensional Data," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 20, no. 1, pp. 1–19, 2025, doi: 10.4018/IJWLTT.384801.
- [29]. F. Y. H. Ahmed, M. Al-Bahri, M. Zakarya, N. Khan, B. K. Joseph, and A. Abdullah, "Improving SpikeProp's Training Efficiency in Spiking Neural Networks for Large Language Models Through Innovative Weight Initialization," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 1, Art. no. 286, 2025, doi: 10.1007/s44196-025-00961-x.
- [30]. E. M. S. Rochman et al., "Classification of hypertension disease using Artificial Neural Network (ANN) backpropagation method case study in mitigating health risk: UPT Modopuro Mojokerto Health Center," *BIO Web Conf.*, vol. 146, p. 01083, 2024, doi: 10.1051/bioconf/202414601083.
- [31]. H. Qiu et al., "Comparative analysis of Kalman Filters, Gaussian Sum Filters, and Artificial Neural Networks for state estimation in energy management," *Energy Rep.*, vol. 13, pp. 4417–4440, Jun. 2025, doi: 10.1016/j.egy.2025.03.054.
- [32]. R. Qamar and B. A. Zardari, "Artificial Neural Networks: An Overview," *Mesopotamian Journal of Computer Science*, vol. 2023, pp. 124–133, 2023, doi: 10.58496/MJCSC/2023/015.
- [33]. C. Liu and B. Ye, "Artificial neural network and the prospect of AGI: an argument from architecture," *Discov. Artif. Intell.*, vol. 5, Art. no. 299, 2025, doi: 10.1007/s44163-025-00561-w.
- [34]. M. Guerra-Marín et al., "A convolutional neural network method to obtain the dynamic model parameters of a three-linear-axis Cartesian robot," *Neural Comput. Appl.*, vol. 37, pp. 15439–15467, 2025, doi: 10.1007/s00521-025-11297-0.
- [35]. H. Taud and J. F. Mas, "Multilayer Perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*, Cham: Springer, 2018, pp. 451–455, doi: 10.1007/978-3-319-60801-3_27.
- [36]. K. Danil, "Pengenalan Jenis Kelamin dalam Lingkungan Multiaksen Menggunakan Metode Multi-Layer Perceptron (MLP) dan Gated Recurrent Unit (GRU): Gender Recognition in a Multiaccent Environment Using Multi-Layer Perceptron (MLP) and Gated Recurrent Unit (GRU) Methods," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 803–811, May 2024, doi: 10.57152/malcom.v4i3.1323.
- [37]. D. Allan, T. Caswell, N. Keim, and C. van der Wel, *trackpy: Trackpy v0.3.0.* (Nov. 12, 2015). Zenodo. doi: 10.5281/ZENODO.34028.
- [38]. S. van der Walt, J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014, doi: 10.7717/peerj.453.
- [39]. K. Dewan, T. R. Dastidar, and M. Ahmad, "Estimation of Sperm Concentration and Total Motility from Microscopic Videos of Human Semen Samples," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA: IEEE, Jun. 2018, pp. 2380–2387. doi: 10.1109/CVPRW.2018.00307.
- [40]. F. Pedregosa et al., "Scikit-learn: Machine

- learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011.
- [41]. J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," J. Mach. Learn. Res., vol. 12, pp. 2121–2159, Jul. 2011.
- [42]. M. D. Zeiler, "ADADELTA: An adaptive learning rate method," Dec. 22, 2012, arXiv: arXiv:1212.5701. doi: 10.48550/arXiv.1212.5701.
- [43]. H. B. McMahan et al., "Ad click prediction: a view from the trenches," in Proc. 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), Chicago, IL, USA, Aug. 2013, pp. 1222–1230. doi: 10.1145/2487575.2488200.
- [44]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 2015. doi: 10.48550/arXiv.1412.6980.
- [45]. W. Zhao et al., "A Survey of Semen Quality Evaluation in Microscopic Videos Using Computer Assisted Sperm Analysis," Feb. 17, 2022, arXiv: arXiv:2202.07820. doi: 10.48550/arXiv.2202.07820.
- [46]. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in Proc. 27th Int. Conf. Machine Learning (ICML), Haifa, Israel, Jun. 2010, pp. 807–814.
- [47]. M. J. Somers and J. C. Casal, "Using Artificial Neural Networks to Model Nonlinearity: The Case of the Job Satisfaction–Job Performance Relationship," Organ. Res. Methods, vol. 12, no. 3, pp. 403–417, Jul. 2009, doi: 10.1177/1094428107309326.
- [48]. S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," Afr. J. Biomed. Res., pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [49]. S. Fatimah, "Neural Network Optimization Optimization For Medical Image Processing," J. Komput. Indones., vol. 2, no. 1, pp. 33–40, Jun. 2023, doi: 10.37676/jki.v2i1.566.
- [50]. Elbert, M. Wulandari and J. Fat, "Optimizer Comparison In Convolutional Neural Network For Real Time Face Recognition," Jurnal EMACS (Engineering, Mathematics and Computer Science) Vol.7 No.1 January 2025: 15-24, doi: 10.21512/emacsjournal.v7i1.12058.



I Gede Susrama Mas Diyasa is a lecturer and researcher at the Faculty of Computer Science, University of Pembangunan Nasional "Veteran" Jawa Timur (UPNVJT), where he also serves as Vice Dean I. His expertise includes image processing, computer vision, and biomedical applications. He earned his bachelor's degree from Institut Teknologi Adhi Tama Surabaya (ITATS) and both his master's and doctoral degrees in Information Technology from Institut Teknologi Sepuluh Nopember (ITS), Surabaya. His research focuses on intelligent systems, particularly in biomedical applications, precision agriculture, and smart village development. Known for his interactive and practice-oriented teaching style, he is also a frequent speaker at national and international conferences. Contact: igsusrama.if@upnjatim.ac.id.



Vaizal Asy'ari is a graduate student of the Department of Information Technology, University of Pembangunan Nasional "Veteran" Jawa Timur (UPNVJT) Surabaya, Indonesia. He holds a Bachelor of Computer from Universitas Nahdlatul Ulama Surabaya (UNUSA) and a Bachelor of Economics in Accounting from Sekolah Tinggi Ilmu Ekonomi Mahardhika (STIE Mahardhika) Surabaya, providing him with a unique multidisciplinary perspective. His research interests include Intelligent Systems, Data Mining, and Machine Learning. His current research focuses on the application of machine learning and computer vision techniques for biomedical data analysis. His work involves developing machine learning models to enhance the accuracy and efficiency of medical diagnostic processes. He can be contacted at vaiz.asyari@gmail.com



Ani Dijah Rahajoe is a lecturer in the Master of Information Technology study program at the Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia. She is a graduate of the master's and doctoral programs at Universitas Gadjah Mada, Yogyakarta, Indonesia. Her research specialty is in artificial intelligence, including intelligent systems. Her research also discusses feature selection for high data dimensions in her intelligent system algorithm. One of her research projects is modifying harmony search combined with genetic algorithms to

Author Biography

improve accuracy. In addition to being a lecturer and researcher, she has also served as the general chair of the IEEE International Seminar for two years and as a TPC chair. She can be contacted at anidijah.if@upnjatim.ac.id



Sayyidah Humairah is a graduate of the undergraduate Informatics program, with a research background in biomedical informatics, particularly in breast cancer detection. During her bachelor's studies, her research focused on

applying computational and intelligent methods to support medical diagnosis, demonstrating strong interest in interdisciplinary research between informatics and health sciences. She is currently pursuing advanced studies in the biomedical field at the University of Patras, Department of Electrical and Computer Engineering, Polytechnic Faculty, Greece. Her current academic focus lies in biomedical engineering, where she continues to develop research competencies at the intersection of computing, engineering, and medical applications. She can be contacted at humairahs@upatras.gr



Deshinta Arrova Dewi is a distinguished academic, researcher, and innovator specializing in artificial intelligence (AI) and big data analytics. She earned her doctorate from Universiti Kebangsaan Malaysia (UKM),

focusing on applying AI to enhance efficiency and innovation across education, economics, and healthcare sectors. Deshinta has an impressive portfolio of award-winning AI solutions, including predictive machine learning algorithms and adaptive learning models that have transformed technology-based education. As an educator, he inspires students through interactive, practice-oriented teaching and frequently serves as a keynote speaker at international AI conferences, sharing insights on leveraging AI to address digital-era challenges. She can be contacted at deshinta.ad@newinti.edu.my



Muhammad Nashif Farid is an undergraduate student in the Data Science program at the Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia. He is currently conducting research in biomedical data science domain, focusing on sperm analysis using

data-driven and computational approaches. His research involves data processing, analysis, and interpretation to support objective assessment in biomedical applications, reflecting his interest in applying data science methods to health-related problems. Through this research, he is developing competencies in data analytics, modeling, and interdisciplinary problem solving at the intersection of data science and biomedicine. He can be contacted at nashifrd@gmail.com.

Faikul Umam is a lecturer at the Faculty of Engineering, Universitas Trunojoyo Madura, Bangkalan, Indonesia, where he



currently serves as the Dean of the Faculty. He completed his undergraduate studies in Informatics and earned his Master's degree in Engineering from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. His research interests

focus on engineering and applied technology areas, including robotics, control systems, automation, embedded systems, and intelligent engineering applications. In his academic career, he has been actively involved in research, teaching, and institutional development, particularly in strengthening the quality of engineering education and research collaboration. In addition to his role as a lecturer and researcher, he holds strategic leadership responsibilities in managing academic programs, research activities, and partnerships within the Faculty of Engineering. He can be contacted via his institutional email at faikul@trunojoyo.ac.id.