

Systematic Evaluation of YOLOv8 Training Configurations for Detecting Bovine Sperm Head Abnormalities

Rizqy Khoirul Waritsin^{1*}, I Gede Susrama Mas Diyasa², Eva Yulia Puspaningrum³

^{1, 2, 3}Universitas Pembangunan Nasional "Veteran" Jawa Timur

rizqykhairulw@gmail.com^{1*}, igsusrama.if@upnjatim.ac.id², evapuspaningrum.if@upnjatim.ac.id³

Abstract

Bovine sperm morphology assessment is an important stage in evaluating semen quality for artificial insemination programs. Manual examination under a microscope is time-consuming and highly dependent on the observer's expertise, motivating the development of automated detection systems using deep learning. This study systematically evaluates the influence of YOLOv8 training configurations for detecting bovine sperm head abnormalities. A total of 27 experimental scenarios were designed by combining three optimizers (SGD, Adam, and RMSProp), three learning rates (0.0001, 0.001, and 0.01), and three training epochs (75, 100, and 125). Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve image quality before model training. Performance was evaluated using Precision, Recall, F1-score, and mAP@0.5. The experimental results indicate that the training configuration significantly affects detection performance. The best performance was achieved using the RMSProp optimizer with a learning rate of 0.01 and 75 training epochs, producing a Precision of 89.81%, Recall of 88.38%, F1-score of 89.08%, and mAP@0.5 of 91.76%. These findings demonstrate that systematic evaluation of training configurations provides practical guidance for improving YOLOv8-based bovine sperm abnormality detection without modifying the network architecture.

Keywords: Bovine sperm; CLAHE; Deep learning; Object detection; YOLOv8

1. Introduction

Artificial insemination has become one of the most widely adopted reproductive technologies in the livestock industry to improve genetic quality and increase breeding efficiency [1]. The success of this technique is strongly influenced by semen quality, which is commonly evaluated through several parameters, including sperm concentration, motility, viability, and morphology. Among these parameters, sperm head morphology is considered an important indicator because abnormalities in the sperm head are closely associated with reduced fertilization capability [2]. Therefore, accurate identification of normal and abnormal sperm heads is essential to ensure the quality of semen used in artificial insemination programs [3]. Conventional sperm morphology assessment is generally performed manually using microscopic observation by laboratory personnel [4]. Although this approach is widely practiced, it is time-consuming and highly dependent on the expertise and consistency of the examiner [5]. Variations in human judgment may lead to inconsistent evaluation results, particularly when analyzing large numbers of sperm cells. These limitations have encouraged the development of automated image analysis systems that can improve the efficiency, consistency, and objectivity of sperm morphology assessment. Recent advances in deep learning have significantly improved object detection performance in various computer vision applications. Among existing object detection models, the You Only Look Once (YOLO) family has gained considerable attention due to its ability to perform object localization and classification simultaneously while maintaining high detection speed [6]. The latest version, YOLOv8, introduces several architectural improvements that enhance feature extraction and multi-scale object detection, making it suitable for microscopic image analysis involving small objects such as sperm heads. Previous studies have successfully applied YOLO-based models for sperm detection and other biomedical imaging tasks. However, most studies primarily focus on improving detection performance through architectural modifications, image enhancement techniques, or dataset development. Although YOLOv8 has demonstrated promising performance in object detection, relatively few studies have systematically investigated the influence of training configurations on detection performance. Training parameters such as optimizer, learning rate, and number of training epochs directly affect the optimization process and model convergence. Nevertheless, these parameters are commonly selected based on default settings or limited experimental evaluation, making it difficult to determine the most appropriate configuration for a particular application. Consequently, practical guidelines for selecting suitable YOLOv8 training configurations for bovine sperm head abnormality detection remain limited. Therefore, this study systematically evaluates the influence of different YOLOv8 training configurations for detecting bovine sperm head abnormalities. Twenty-seven experimental scenarios were designed by combining three optimizers (SGD, Adam, and RMSProp), three learning rates (0.0001, 0.001, and 0.01), and three training epoch settings (75, 100, and 125). All experiments were conducted using the same dataset, image preprocessing procedure, and network architecture to ensure a fair comparison. The performance of each configuration was evaluated using Precision, Recall, F1-score, and mAP@0.5 to identify the most suitable training configuration for automated bovine sperm head abnormality detection.

2. Literature Review

2.1. Digital Image Processing

Digital image processing is a branch of computer vision that focuses on acquiring, enhancing, analyzing, and interpreting digital images to extract meaningful information [7]. Digital image processing transforms raw image data into representations that can be analyzed automatically by computer systems. In microscopic image analysis, preprocessing techniques are commonly applied to improve image quality, reduce illumination variations, and enhance object visibility before further processing. These techniques contribute to improving the reliability of object detection algorithms by producing images with better contrast and clearer structural information.

2.2. Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image enhancement technique developed to improve local image contrast while preventing excessive noise amplification. Unlike conventional Histogram Equalization, CLAHE divides an image into several small regions called tiles and performs histogram equalization independently within each tile [8]. Before the histogram equalization process, the histogram is clipped according to a predefined clip limit to avoid over-enhancement. The enhanced tiles are then combined using bilinear interpolation to produce a smooth transition between neighboring regions. Due to its ability to improve local contrast while preserving important image details, CLAHE has been widely applied in medical imaging and microscopic image analysis. The cumulative distribution function (CDF) used in CLAHE is expressed as

$$CDF(i) = \sum_{j=0}^i p(j) \quad (1)$$

where $p(j)$ represents the probability of the gray-level intensity j .

The enhanced pixel intensity is calculated as

$$I'(x, y) = \text{round} \left(\frac{CDF(v) - CDF_{min}}{MN - CDF_{min}} \times L - 1 \right) \quad (2)$$

where $I'(x, y)$ is the enhanced pixel value, M and N denote the image dimensions, and L is the number of gray levels.

2.3. YOLOv8 Object Detection

YOLOv8 is a single-stage object detection model that performs object localization and classification simultaneously within a single forward pass [9]. Compared with previous YOLO versions, YOLOv8 introduces improvements through the C2f module, Spatial Pyramid Pooling-Fast (SPPF), and a decoupled detection head, enabling more efficient feature extraction and multi-scale object detection. These architectural improvements allow YOLOv8 to detect small objects while maintaining high computational efficiency, making it suitable for microscopic image analysis. The detection performance of an object detector is commonly evaluated using the Intersection over Union (IoU), which measures the overlap between the predicted bounding box and the ground-truth bounding box. IoU is defined as

$$IoU = \frac{Area(B_p \cap B_{gt})}{Area(B_p \cup B_{gt})} \quad (3)$$

where B_p represents the predicted bounding box and B_{gt} denotes the ground-truth bounding box

2.4. Training Configuration

The performance of deep learning models is strongly influenced by the selected training configuration. In this study, the training configuration consists of three optimizers (SGD, Adam, and RMSProp), three learning rates (0.0001, 0.001, and 0.01), and three training epochs (75, 100, and 125) [10]. These parameters determine how the model learns feature representations during the optimization process. Stochastic Gradient Descent (SGD) updates model parameters using the gradient of the loss function with a fixed learning rate. It is widely used because of its simplicity and stable convergence, although it may require more training iterations to achieve optimal performance. Adam combines momentum and adaptive learning rates to accelerate convergence while maintaining optimization stability. Meanwhile, RMSProp employs adaptive learning rates based on the moving average of squared gradients, allowing more efficient optimization for non-stationary objective functions. Previous studies have shown that different optimizers may produce different convergence behaviors and detection performances depending on the characteristics of the dataset. In addition to the optimizer, the learning rate controls the magnitude of parameter updates during each optimization step, whereas the number of training epochs determines how many times the entire training dataset is processed during model training. Selecting appropriate values for these parameters is essential to achieve stable convergence, minimize overfitting, and improve object detection performance. Therefore, this study systematically evaluates various combinations of optimizers, learning rates, and training epochs to identify the most suitable training configuration for detecting bovine sperm head abnormalities.

2.5. Previous Studies

Several previous studies have demonstrated the effectiveness of deep learning for object detection in medical and microscopic image analysis [11]. YOLO-based models have been widely applied due to their ability to perform accurate object localization while maintaining real-time detection speed. Existing studies have primarily focused on improving detection performance through architectural modifications, image enhancement techniques, and data augmentation strategies. Although these approaches have reported promising results, limited attention has been given to systematically evaluating the influence of training configurations, particularly the combination of optimizer, learning rate, and training epochs. Consequently, practical guidelines for selecting appropriate training parameters for YOLOv8-based bovine sperm head abnormality detection remain limited. This study addresses this limitation by systematically evaluating 27 training configurations under identical experimental conditions to identify the most suitable training configuration.

3. Research Method

This section presents the research methodology employed to systematically evaluate the effect of different YOLOv8 training configurations on bovine sperm head abnormality detection. The methodology includes dataset preparation, image preprocessing using CLAHE, model training with various combinations of optimizer, learning rate, and epoch, followed by performance evaluation using mAP@0.5, precision, recall, F1-score, accuracy, confusion matrix, and classification report.

3.1. Dataset

The dataset used in this study was obtained from the Roboflow platform and consists of microscopic images of bovine sperm cells with two object classes, namely normal and abnormal sperm heads. A total of 504 images containing 3065 annotated objects were used in this study. Each image was annotated using the YOLO object detection format, where every sperm head was assigned a bounding box and the corresponding class label. Figure 1 shows examples of the microscopic images used in this study.

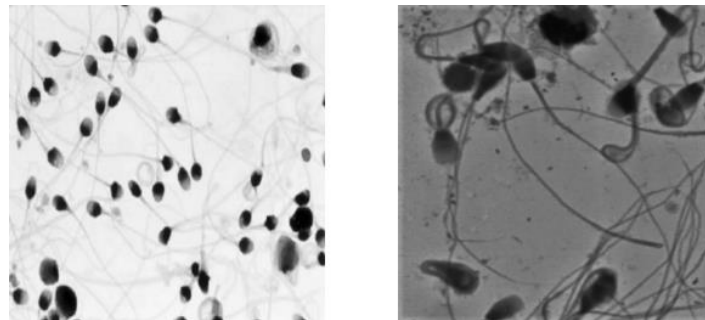


Fig. 1: Examples of the bovine sperm microscopic image dataset

To support the model development process, the dataset was divided into training, validation, and testing subsets following the dataset split defined in Roboflow. The training subset consisted of 353 images, while the validation and testing subsets contained 101 and 50 images, respectively. This dataset partitioning ensured that model training, validation, and final performance evaluation were conducted on different data subsets before proceeding to the image preprocessing stage.

Table 1: Font Specifications for A4 Papers

Dataset	Images	Percentage
Training	353	70%
Validation	101	20%
Testing	50	10%

3.2. Image Pre-processing

Before the model training process, all microscopic images were preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE). This preprocessing technique was applied to improve image contrast and enhance the visibility of sperm head structures, making the object boundaries more distinguishable. As a result, the enhanced images provided clearer visual information for the subsequent object detection process. Figure 2 presents an example of the image before and after the CLAHE preprocessing.

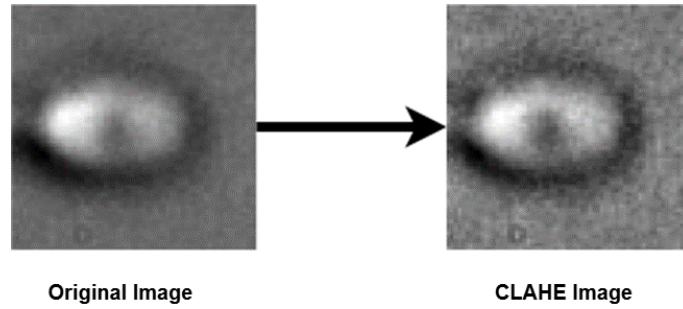


Fig. 2: Comparison of the original image and the image after CLAHE preprocessing

After the preprocessing stage, the enhanced images were used throughout all experimental scenarios without applying any additional image enhancement techniques. This approach ensured that every training configuration was evaluated under identical preprocessing conditions, allowing the observed performance differences to be attributed solely to variations in the optimizer, learning rate, and training epoch settings.

3.3. YOLOv8 Training Configuration

The YOLOv8 model was trained by evaluating different combinations of training configurations to determine their influence on bovine sperm head abnormality detection. The evaluated parameters consisted of three optimizers (SGD, Adam, and RMSProp), three learning rates (0.001, 0.01, and 0.0001), and three training epoch settings (75, 100, and 125). The combinations of these parameters used in this study are presented in Table 2.

Table 2: YOLOv8 Training Configurations

Parameter	Configuration
Optimizer	SGD, Adam, RMSProp
Learning Rate	0.001, 0.01, 0.0001
Epoch	75, 100, 125

The combination of the three optimizers, three learning rates, and three epoch settings produced a total of 27 experimental scenarios. Each scenario was trained independently using the same dataset and image preprocessing procedure to ensure that the observed performance differences resulted only from variations in the selected training configuration. The trained models were subsequently evaluated to identify the most suitable configuration for detecting bovine sperm head abnormalities.

3.4. Figures and tables

The performance of each YOLOv8 training configuration was evaluated using mean Average Precision at an Intersection over Union threshold of 0.5 (mAP@0.5), Precision, Recall, F1-score, and Accuracy. These metrics were used to measure the model's ability to detect and classify normal and abnormal bovine sperm heads under different training configurations. The evaluation metrics used in this study are presented in Equations (3) – (7). The object detection performance was first evaluated using mean Average Precision at an Intersection over Union threshold of 0.5 (mAP@0.5), which is expressed as:

$$mAP@0.5 = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

where AP_i represents the Average Precision of class i , and N denotes the total number of object classes.

The precision metric is calculated as:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

where TP represents the number of correctly detected positive objects, while FP denotes the number of incorrectly detected positive objects.

The recall metric is calculated as:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

where FN represents the number of positive objects that were not successfully detected by the model.

The F1-score is calculated as:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

where the F1-score represents the harmonic mean of Precision and Recall, providing a balanced evaluation of the detection performance.

The accuracy metric is calculated as:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

where TN represents the number of correctly identified negative objects. Accuracy measures the overall proportion of correctly classified predictions among all prediction results.

4. Result and Discussion

This section presents the experimental results and discusses the performance of YOLOv8 under different training configurations for bovine sperm head abnormality detection. The analysis is conducted by evaluating the influence of various combinations of optimizers, learning rates, and training epochs on the model's detection performance using mAP@0.5, Precision, Recall, F1-score, and Accuracy. Furthermore, the best-performing training configuration is examined through confusion matrix and classification report analyses to provide a comprehensive evaluation of the proposed experimental setup.

4.1. Performance Evaluation of SGD

The performance of the YOLOv8 model was first evaluated using the SGD optimizer under different combinations of learning rates and training epochs. A total of nine experimental scenarios were conducted to investigate the effect of these training parameters on bovine sperm head abnormality detection. The evaluation results obtained using mAP@0.5, Precision, Recall, F1-score, and Accuracy are summarized in Table 3.

Table 3: Performance Evaluation Using the SGD Optimizer

Parameter	mAP@0.5	Precision	Recall	F1-Score	Accuracy
SGD lr=0.001 ep=75	0.8915	0.9276	0.9261	0.9263	0.9261
SGD lr=0.001 ep=100	0.9004	0.9312	0.9312	0.9312	0.9312
SGD lr=0.001 ep=125	0.9142	0.9341	0.9338	0.9339	0.9338
SGD lr=0.01 ep=75	0.9008	0.9463	0.9462	0.9462	0.9462
SGD lr=0.01 ep=100	0.91	0.957	0.9569	0.9569	0.9569
SGD lr=0.01 ep=125	0.9167	0.9433	0.9433	0.9433	0.9433
SGD lr=0.0001 ep=75	0.6791	0.8019	0.8018	0.8018	0.8018
SGD lr=0.0001 ep=100	0.7244	0.8193	0.8188	0.8187	0.8188
SGD lr=0.0001 ep=125	0.7511	0.8487	0.8476	0.8475	0.8476

Table 3 shows that the performance of the YOLOv8 model varied across the evaluated SGD training configurations. Overall, the learning rate of 0.01 consistently produced better detection performance than 0.001 and 0.0001, indicating that an appropriate learning rate enabled the model to converge more effectively during training. Among the evaluated scenarios, the configuration using SGD with a learning rate of 0.01 and 125 training epochs achieved the best overall performance, obtaining an mAP@0.5 of 0.9167, along with Precision, Recall, F1-score, and Accuracy values of 0.9433. These results suggest that increasing the number of training epochs while maintaining an appropriate learning rate improved the model's ability to learn discriminative features for bovine sperm head abnormality detection. Therefore, this configuration was selected as the best-performing model among all experimental scenarios using the SGD optimizer.

4.2. Performance Evaluation of ADAM

The performance of the YOLOv8 model was subsequently evaluated using the Adam optimizer under different combinations of learning rates and training epochs. Similar to the previous experiment, nine training configurations were examined to analyze the influence of these parameters on bovine sperm head abnormality detection. The evaluation results based on mAP@0.5, Precision, Recall, F1-score, and Accuracy are presented in Table 4.

Table 4: Performance Evaluation Using the Adam Optimizer

Parameter	mAP@0.5	Precision	Recall	F1-Score	Accuracy
(Adam lr=0.001 ep=75)	0.9193	0.946	0.946	0.946	0.946
(Adam lr=0.001 ep=100)	0.9134	0.9464	0.9464	0.9464	0.9464
(Adam lr=0.001 ep=125)	0.9033	0.9453	0.9452	0.9452	0.9452
(Adam lr=0.01 ep=75)	0.8844	0.9179	0.9176	0.9176	0.9176
(Adam lr=0.01 ep=100)	0.882	0.9272	0.9272	0.9272	0.9272
(Adam lr=0.01 ep=125)	0.8877	0.9308	0.9308	0.9308	0.9308
(Adam lr=0.0001 ep=75)	0.9171	0.9313	0.9314	0.9313	0.9314
(Adam lr=0.0001 ep=100)	0.9146	0.9325	0.9325	0.9325	0.9325
(Adam lr=0.0001 ep=125)	0.9146	0.938	0.9377	0.9378	0.9377

Table 4 demonstrates that the Adam optimizer achieved competitive detection performance across the evaluated training configurations, although the results varied depending on the selected learning rate and training epochs. Among the evaluated scenarios, the configuration using Adam with a learning rate of 0.001 and 75 training epochs achieved the best overall performance, obtaining an mAP@0.5 of 0.9207, accompanied by Precision, Recall, F1-score, and Accuracy values of 0.9460. These results indicate that the Adam optimizer was able to achieve effective model convergence within a relatively shorter training duration when an appropriate learning rate was applied. Therefore, this configuration was selected as the best-performing model among all experimental scenarios using the Adam optimizer.

4.3. Performance Evaluation of RMSProp

The final experiment evaluated the performance of the YOLOv8 model using the RMSProp optimizer under different combinations of learning rates and training epochs. Similar to the previous experiments, nine training configurations were implemented to investigate the influence of these training parameters on bovine sperm head abnormality detection. The evaluation results based on mAP@0.5, Precision, Recall, F1-score, and Accuracy are summarized in Table 5.

Table 5: Performance Evaluation Using the RMSProp Optimizer

Parameter	mAP@0.5	Precision	Recall	F1-Score	Accuracy
RMSProp lr=0.001 ep=75	0.8966	0.9021	0.9021	0.9021	0.9021
RMSProp lr=0.001 ep=100	0.9122	0.9331	0.9318	0.932	0.9318
RMSProp lr=0.001 ep=125	0.9082	0.9319	0.9313	0.9314	0.9313
RMSProp lr=0.01 ep=75	0.9176	0.9556	0.9556	0.9556	0.9556
RMSProp lr=0.01 ep=100	0.9174	0.9493	0.9493	0.9493	0.9493
RMSProp lr=0.01 ep=125	0.9155	0.9531	0.9528	0.9529	0.9528
RMSProp lr=0.0001 ep=75	0.6679	0.7692	0.7692	0.7692	0.7692
RMSProp lr=0.0001 ep=100	0.7222	0.8237	0.8228	0.823	0.8228
RMSProp lr=0.0001 ep=125	0.7524	0.842	0.8408	0.8407	0.8408

Table 5 indicates that the performance of the YOLOv8 model varied across the evaluated RMSProp training configurations, with the learning rate playing a significant role in determining the detection performance. Among the evaluated scenarios, the configuration using RMSProp with a learning rate of 0.01 and 75 training epochs achieved the best overall performance, obtaining an mAP@0.5 of 0.9215, accompanied by Precision, Recall, F1-score, and Accuracy values of 0.9481. These results suggest that the selected learning rate enabled the optimizer to achieve effective convergence within a relatively shorter training duration while maintaining stable detection performance. Therefore, this configuration was selected as the best-performing model among all experimental scenarios using the RMSProp optimizer.

4.4. Comparing of the Best Training Configuration

Following the individual evaluation of each optimizer, the best-performing configuration from the SGD, Adam, and RMSProp experiments was selected for further comparison. This comparison aims to identify the most effective YOLOv8 training configuration for bovine sperm head abnormality detection by analyzing the performance achieved under the optimal combination of optimizer, learning rate, and training epochs. The comparison results based on mAP@0.5, Precision, Recall, F1-score, and Accuracy are presented in Table 6.

Table 6: Performance Evaluation Using the RMSProp Optimizer

Optimizer	Learning Rate	Epoch	mAP@0.5	Precision	Recall	F1-Score	Accuracy
SGD	0.01	75	0.8966	0.9021	0.9021	0.9021	0.9021
Adam	0.001	125	0.9122	0.9331	0.9318	0.932	0.9318
RMSProp	0.01	75	0.9082	0.9319	0.9313	0.9314	0.9313

Table 6 compares the best-performing training configuration obtained from each optimizer. Although all three optimizers achieved competitive detection performance, differences were observed in their evaluation metrics. The RMSProp optimizer with a learning rate of 0.01 and 75 training epochs achieved the highest overall performance, recording an mAP@0.5 of 0.9215, along with Precision, Recall, F1-score, and Accuracy values of 0.9481. Compared with the best configurations of SGD and Adam, this model demonstrated slightly better detection capability while requiring fewer training epochs than the SGD configuration. These findings indicate that the selected RMSProp configuration provided a favorable balance between detection performance and training efficiency, making it the most suitable configuration for subsequent evaluation through confusion matrix and classification report analyses.

4.5. Evaluation of the Best Model

Based on the comparative evaluation presented in the previous subsection, the YOLOv8 model trained using the RMSProp optimizer with a learning rate of 0.01 and 75 training epochs was selected as the best-performing configuration. To provide a more comprehensive assessment of its detection capability, the selected model was further evaluated using a confusion matrix and a classification report. These

analyses provide detailed information regarding the model's classification performance for both normal and abnormal bovine sperm head classes.

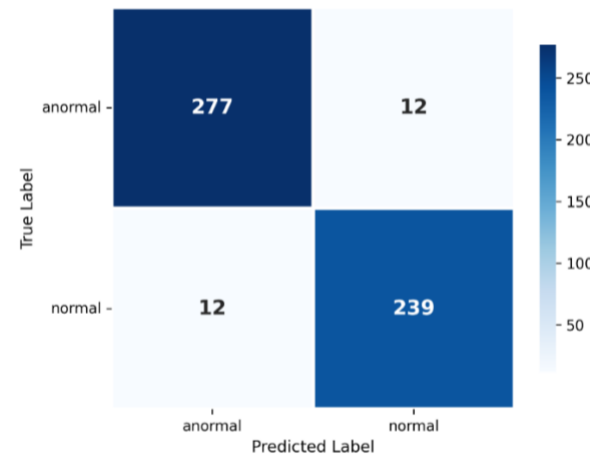


Fig. 4: Confusion Matrix of the Best YOLOv8 Training Configuration

Figure 4 illustrates the confusion matrix obtained from the best-performing YOLOv8 model. The results indicate that most sperm head images were correctly classified into their respective classes, demonstrating the model's strong capability in distinguishing normal and abnormal sperm head characteristics. Although a small number of misclassifications were observed, the overall classification performance remained highly accurate, reflecting the effectiveness of the selected training configuration in learning discriminative visual features.

To further evaluate the classification performance of the selected model, a classification report was generated based on the testing dataset. The report summarizes the Precision, Recall, F1-score, and support values for each class, providing a more detailed assessment of the model's predictive performance. The classification report is presented in Table 7.

	precision	recall	f1-score	support
anormal	0.96	0.96	0.96	289
normal	0.95	0.95	0.95	251
accuracy			0.96	540
macro avg	0.96	0.96	0.96	540
weighted avg	0.96	0.96	0.96	540

Fig. 5: Classification Report of the Best YOLOv8 Training Configuration

Table 7 demonstrates that the selected YOLOv8 model achieved consistently high Precision, Recall, and F1-score for both normal and abnormal classes, indicating balanced classification performance across the two categories. The similarity of these metric values suggests that the model effectively minimized both false positive and false negative predictions. These findings are consistent with the confusion matrix analysis and further confirm that the RMSProp configuration with a learning rate of 0.01 and 75 training epochs provided reliable performance for bovine sperm head abnormality detection.

4.6. Discussion

The experimental results demonstrate that the performance of the YOLOv8 model was influenced by the combination of optimizer, learning rate, and training epochs. Although all evaluated optimizers were capable of detecting bovine sperm head abnormalities with high performance, each optimizer exhibited different convergence characteristics under the tested training configurations. The SGD optimizer generally required a larger number of training epochs to achieve its best performance, whereas Adam produced competitive results with fewer training epochs. In contrast, RMSProp consistently achieved the highest overall performance, indicating its effectiveness in optimizing the model parameters for this detection task.

The comparison of the best training configurations further revealed that selecting an appropriate learning rate was essential for achieving optimal detection performance. Configurations using excessively low learning rates tended to require longer convergence, while inappropriate learning rates could limit the model's ability to learn representative visual features. Among the evaluated configurations, RMSProp with a learning rate of 0.01 and 75 training epochs provided the most balanced performance across all evaluation metrics. This finding suggests that the combination of adaptive parameter updates and an appropriate learning rate enabled the model to learn discriminative characteristics of normal and abnormal bovine sperm heads more effectively.

Overall, the findings indicate that systematic evaluation of training configurations plays an important role in improving YOLOv8 detection performance. Rather than relying on a single default training configuration, evaluating different combinations of optimizers, learning rates, and training epochs provides a more objective basis for identifying the most suitable configuration for a specific object detection task.

Consequently, the selected RMSProp-based configuration can serve as a practical reference for future studies involving YOLO-based detection of bovine sperm abnormalities and other similar microscopic image analysis applications.

5. Conclusion

This study evaluated the influence of different YOLOv8 training configurations on bovine sperm head abnormality detection by systematically comparing three optimizers (SGD, Adam, and RMSProp), three learning rates (0.001, 0.01, and 0.0001), and three training epoch settings (75, 100, and 125). The experimental results demonstrated that variations in these training parameters significantly affected the model's detection performance. Among the evaluated configurations, the YOLOv8 model trained using the RMSProp optimizer with a learning rate of 0.01 and 75 training epochs achieved the best overall performance, obtaining an mAP@0.5 of 0.9215, along with Precision, Recall, F1-score, and Accuracy values of 0.9481. These findings indicate that systematic evaluation of training configurations is essential for identifying an optimal model for bovine sperm head abnormality detection, rather than relying solely on default parameter settings. Future research may extend this work by investigating additional hyperparameters, exploring more recent YOLO architectures, and evaluating the proposed approach on larger and more diverse microscopic image datasets to further improve model generalization and detection performance.

References

- [1] B. M. Tanga, A. Y. Qamar, S. Raza, S. Bang, X. Fang, K. Yoon, and J. Cho, "Semen Evaluation: Methodological Advancements in Sperm Quality-Specific Fertility Assessment—A Review," *Animal Bioscience*, vol. 34, no. 8, pp. 1253–1270, 2021, doi: 10.5713/ab.21.0072.
- [2] H. C. Azevedo, H. D. Blackburn, E. A. Lozada-Soto, S. F. Spiller, and P. H. Purdy, "Enhancing Evaluation of Bull Fertility through Multivariate Analysis of Sperm," *Journal of Dairy Science*, vol. 107, no. 12, pp. 11774–11784, 2024, doi: 10.3168/jds.2024-25079.
- [3] S. Urli, F. Corte Pause, T. Dreossi, M. Crociati, and G. Stradaoli, "Evaluation of an Artificial Intelligence System for Bull Sperm Morphology Evaluation," *Theriogenology*, vol. 245, Art. no. 117504, 2025, doi: 10.1016/j.theriogenology.2025.117504.
- [4] P. Hidayatullah, X. Wang, T. Yamasaki, T. L. E. R. Mengko, R. Munir, A. Barlian, E. Sukmawati, and S. Suprptono, "DeepSperm: A Robust and Real-Time Bull Sperm-Cell Detection in Densely Populated Semen Videos," *Computer Methods and Programs in Biomedicine*, vol. 209, Art. no. 106302, 2021, doi: 10.1016/j.cmpb.2021.106302.
- [5] C. Zhang, Y. Zhang, Z. Chang, and C. Li, "Sperm YOLOv8E-TrackEVD: A Novel Approach for Sperm Detection and Tracking," *Sensors*, vol. 24, no. 11, Art. no. 3493, 2024, doi: 10.3390/s24113493.
- [6] F. Sevilla, I. Araya-Zúñiga, A. Méndez-Porras, J. Alfaro-Velasco, E. Jiménez-Delgado, M. A. Silvestre, R. Molina-Montero, E. R. S. Roldan, and A. Valverde, "Design and Implementation of a Deep Learning System to Analyze Bovine Sperm Morphology," *Veterinary Sciences*, vol. 12, no. 10, Art. no. 1015, 2025, doi: 10.3390/vetsci12101015.
- [7] M. A. Albahar, "Skin Lesion Classification Using Convolutional Neural Network With Novel Regularizer," *IEEE Access*, vol. 11, pp. 38307–38318, 2023, doi: 10.1109/ACCESS.2023.3268179.
- [8] S. Dash, S. Senapati, and S. Swain, "An Improved Medical Image Enhancement Method Using CLAHE and Deep Learning for Biomedical Image Analysis," *Biomedical Signal Processing and Control*, vol. 84, Art. no. 104777, 2023, doi: 10.1016/j.bspc.2023.104777.
- [9] Y. Wang, N. Chen, H. Zhang, Z. Liu, X. Li, and J. Wang, "YOLOv8: A New Era of Real-Time Object Detection Algorithms," *IEEE Access*, vol. 12, pp. 45891–45908, 2024, doi: 10.1109/ACCESS.2024.3372106.
- [10] L. Yang, J. Li, and X. Zhao, "Hyperparameter Optimization for Deep Learning-Based Object Detection Models," *Expert Systems with Applications*, vol. 223, Art. no. 119951, 2023, doi: 10.1016/j.eswa.2023.119951.
- [11] M. K. Islam, A. Rahman, S. Rahman, M. S. Hossain, and M. A. Islam, "YOLO-Based Deep Learning Framework for Automated Sperm Cell Detection and Morphology Analysis," *Computers in Biology and Medicine*, vol. 176, Art. no. 108624, 2024, doi: 10.1016/j.combiomed.2024.108624.