Journal of Human Centered Technology

https://www.humentech.utm.my E-ISSN: 2821-3467

Sperm Detection Based on You Only Look Once-Neural Architecture Search (YOLO-NAS)

Aleeyah Nur Sabrina Dzuren[1], Muhammad Amir As'ari[1*,2]

[1]Department of Biomedical Engineering and Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Johor, Malaysia
[2]Sport Innovation and Technology Centre (SITC), Universiti Teknologi Malaysia, Skudai, 81310, Malaysia

*Corresponding Author amir-asari@utm.my

Cite: https://doi.org/10.11113/humentech.v5n1.116 Research Article

Abstract:
Computer-Aided Sperm Analysis (CASA) systems are vital for male reproductive diagnostics. However, the CASA systems are prone with detection challenges. Therefore, this study explores You Only Look Once-Neural Architecture Search (YOLO-NAS) for precise, rapid sperm analysis, addressing its underexplored efficacy in this domain. Using the Sperm Videos and Images Analysis (SVIA) dataset (3,590 images, 125,000 objects), XML annotations were converted to YOLO TXT format. YOLO-NAS models (S, M, L) were trained for 40 epochs on Google Colab Pro using AdamW, batch size 32, and learning rates 0.002 (initial 0.0005). The performance was evaluated using mAP@0.50, Recall@0.50, Precision@0.50, and F1 Score@0.50, with an IoU threshold of 50%. The YOLO-NAS-L was most effective for sperm/impurity detection, achieving 0.8163 mAP@0.50 and 0.8203 precision@0.50. Increased model complexity enhances feature extraction, especially for small objects while higher precision may decrease recall. From the outputs, this study has successfully developed and evaluated a YOLO-NAS model for sperm detection. Future works include more epochs, diverse datasets, and offline training in sperm clustering, as well morphology and motility analyses.

Keywords: Sperm; Convolutional neural networks; Yolo-NAS; CASA

1. INTRODUCTION

Infertility is characterized, according to the World Health Organization (WHO), as a condition affecting the male or female reproductive system that hinders the achievement of pregnancy following one year or more of consistent unprotected sexual intercourse (1). This condition affects individuals and families globally, with estimates suggesting that one in six people of reproductive age has encountered infertility at some stage in their lives. Among the various contributing factors, female infertility may arise from tubal or uterine disorders, ovarian dysfunction, endocrine imbalances, or advanced maternal age. In males, infertility can result from obstructions in the reproductive tract, hormonal imbalances, testicular dysfunction, or deficiencies in sperm function and quality. This matter presents considerable psychological, social, and financial challenges for those impacted, as well as for the healthcare system (2). A study encompassing 384,419 men revealed that individuals facing infertility exhibit a heightened mortality risk in comparison to their fertile counterparts (3). Specifically, compromised male reproductive health, including low testosterone levels and abnormal sperm characteristics, was linked to a higher Charlson Comorbidity Index and a decline in overall health (3).

Human semen, a multifaceted biological fluid consisting of seminal plasma and spermatozoa, is essential for the process of fertilization. Evaluating semen quality is essential for comprehending male fertility. The guidelines established by the WHO outline various parameters for semen analysis, which encompass sperm concentration, semen volume, total sperm count, sperm morphology, sperm vitality, and motility (4). In the last five decades, a notable decrease has been recorded in both seminal fluid volume and sperm count (5), prompting a deeper exploration into the quality of male reproductive health. Manual Sperm Analysis (MSA) involves a thorough examination of both macroscopic parameters, such as color, volume, pH, and viscosity, as well as microscopic parameters, including concentration, motility, and morphology. Nonetheless, MSA is influenced by human bias, necessitating extensive training to accurately interpret nuanced features. It continues to be a time-intensive process, with the possibility of yielding inconclusive results owing to observer subjectivity (6).

In response to these challenges, Computer-Aided Sperm Analysis (CASA) systems were developed in 1985 (7). CASA employs image processing methodologies on microscopic images or video frames to evaluate sperm parameters with greater objectivity and consistency compared to MSA (8). Notwithstanding these advancements, CASA continues to encounter constraints stemming from discrepancies in detection methodologies and the necessity for rigorous parameter tuning. To enhance the reliability and scalability of sperm analysis, the application of machine learning, particularly deep

Received 24 June 2025; Accepted 05 December 2025
Available online 06 February 2026

learning, has emerged as a crucial instrument. Techniques such as Convolutional Neural Networks (CNNs), Single Shot Detectors (SSDs), and You Only Look Once (YOLO) algorithms have demonstrated significant efficacy in the domains of image recognition and analysis. Convolutional neural networks (CNNs) are extensively utilized due to their ability to extract intricate features from visual data (9). Single Shot Detectors (SSDs) employ feed-forward convolutional networks for the tasks of object localization and classification (10). Meanwhile, You Only Look Once (YOLO) approaches object detection as a unified regression problem, providing a balance of speed and precision (11). Convolutional neural networks function by applying convolution operations to input images using feature filters, such as FxF kernels, resulting in the generation of spatial feature maps (12). Region-based CNNs (R-CNNs), as an advancement within this domain, introduce candidate regions via selective search and subsequently normalize these regions prior to classification using linear support vector machines (13). Although R-CNNs demonstrate high accuracy, they necessitate significant computational resources, frequently taking several seconds to process a single image (14). Supplementary detection techniques such as SSDs utilize anchor boxes across diverse feature maps to facilitate multiscale predictions, demonstrating significant efficacy in identifying objects of differing sizes (10).

YOLO functions by processing input images through convolutional neural networks to extract a range of features, from low-level elements like edges and shapes to higher-level attributes such as colors. The features are subsequently processed through fully connected layers to generate predictions for bounding boxes and class probabilities, with non-maximum suppression employed to preserve the most reliable detections. The effectiveness of YOLO has been evidenced across various domains, such as timber defect detection (15) and precision agriculture (16). YOLO models have been utilized in various applications, including the detection of anemia through image-based systems (17), identification of breast lesions from mammography (18, 19), classification of brain tumors using MRI (20), and recognition of cervical spine structures from CT images (21). YOLO offers a framework for real-time detection by executing localization and classification concurrently across a predetermined grid (11). The evolution from YOLOv1 to the most recent YOLOv8 and YOLO-NAS models showcases a series of progressive enhancements. These include the implementation of batch normalization, the integration of residual blocks, upsampling techniques, and the transition from anchor boxes to anchor-free methodologies, all aimed at optimizing mean Average Precision (mAP), processing speed, and model size (22–27). YOLOv5 has demonstrated effective applications in defect detection within timber (15), pest monitoring in agricultural contexts (16), and classification of pressure ulcers in clinical environments (28), frequently surpassing previous methodologies in terms of both accuracy and efficiency. A recent advancement, YOLO-NAS, builds upon earlier YOLO models by integrating quantization-aware elements and Neural Architecture Search (NAS), leading to improved performance in terms of both speed and accuracy (29). For example, a study utilizing YOLO-NAS for the detection of fava bean crops reported mAP scores of 90.90% (YOLO-NASL), 94.10% (YOLO-NASM), and 92.60% (YOLO-NASS) (30).

CASA systems have progressively integrated deep learning models to improve the accuracy of sperm detection. For instance, YOLOv4 demonstrated an average precision of 90.31% in the detection of sperm objects and 68.19% for non-sperm objects, yielding a mean average precision of 79.58% throughout the training phase (31). Moreover, an enhanced YOLOv5s model for identifying small sperm targets exhibited robust performance, particularly in low-contrast and high-clutter environments, leading to a notable decrease in false negatives and an increase in detection accuracy (32). Utilizing CNN-based methodologies, specifically Faster R-CNN, has proven effective in the detection of sperm heads, attaining an accuracy rate of 91.77% when employing the Inception-V2-COCO architecture (33). The conventional dependence of CASA on manual feature extraction has been significantly enhanced by the learning capabilities of deep neural networks. Regardless of the differences in task complexity and image quality, YOLO reliably achieves high-precision outcomes within these healthcare settings. Considering its capabilities in identifying small, dense objects, YOLO models, especially YOLO-NAS, demonstrate significant potential in enhancing the accuracy of sperm detection within CASA systems.

Nonetheless, while YOLO-NAS demonstrates significant potential in precision agriculture and object detection tasks, its utilization in sperm analysis has yet to be thoroughly investigated. The absence of focused investigations assessing YOLO-NAS for sperm detection highlights a significant void in the existing body of knowledge. This study seeks to examine the efficacy of YOLO-NAS in the context of sperm analysis, aiming to develop a more precise, rapid, and standardized methodology to enhance male infertility diagnostics.

2. METHODOLOGY

2.1 Main Framework

Figure 1 illustrates the progression of the project. The project commenced with the collection of data from an online source, specifically the Sperm Video and Image Analysis (SVIA) repository on Github, which includes microscopic images of sperm and impurities. The annotation files were subsequently transformed from .xml format to .txt format in order to facilitate preparation for YOLO format. The respective images were subsequently integrated for the training and development of the YOLO-NAS models. Subsequently, various sizes of the YOLO-NAS models were developed and trained utilising Google Colab Pro. The performance of all models was evaluated.

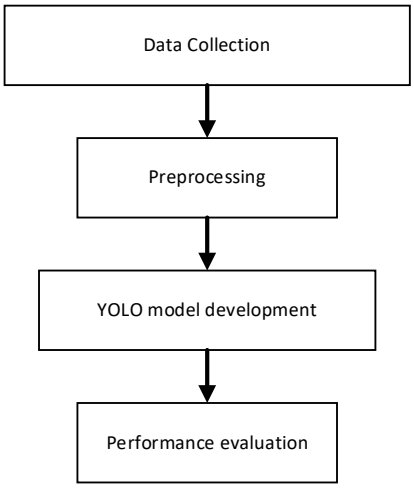

Figure 1. Project's workflow.

2.2 Data Collection and Preprocessing

The dataset was sourced from Github (https://github.com/Demozsj/Detection-Sperm/tree/main/MIaMIA-SpermVideo%20data%20set). For the object detection task in this study, Subset A was employed, comprising a total of 3,590 images and encompassing 125,000 object locations. The dataset was initially gathered using a WLJY-9000 CASA, equipped with a 20× objective and 20× electronic glasses, in .mp4 format, capturing at a frame rate of 30 frames per second. A total of 3590 images and their corresponding annotation files were derived from 19 microscopic videos, serving as object detection datasets. The images are formatted in .png with dimensions of 698 × 528 × 3, while the annotation files are formatted in .xml.

The dataset was partitioned into three subsets according to the following distribution: 70% allocated for training, 15% designated for validation, and 15% reserved for testing. The division was designed to improve the model's performance by ensuring an adequate amount of data for training, along with dependable sets for validation and testing, in accordance with recommendations from prior research that employed this dataset for sperm detection in YOLOv4 (34). The dataset consisted of 2537 images designated for training, 550 images allocated for validation, and 552 images reserved for testing purposes. All images along with their respective annotations were systematically organized in Google Drive. The annotation files were originally formatted in .xml, rendering them incompatible with the YOLO (You Only Look Once) framework. YOLO necessitates annotation files in the .txt format because of its distinct structure for labelling objects. Although .xml files, frequently utilized by PASCAL VOC datasets, offer comprehensive metadata such as image dimensions, bounding box coordinates, and class labels in a structured format, YOLO requires a more straightforward, line-oriented text format. In a .txt file, each line delineates an object, comprising the class label succeeded by the normalized coordinates and dimensions of the bounding box, which includes the centre x, centre y, width, and height, all proportionately adjusted to the image dimensions. This format enhances the efficiency of YOLO's parsing mechanism, emphasizing speed and simplicity within its detection pipeline.

2.3 YOLO Model Development

2.3.1 Model Architecture of YOLO-NAS

Figure 2 illustrates the architecture of YOLO-NAS. The backbone comprised a neural network designed to extract features from the images input into the model. The approach encompassed multiple layers of convolutional neural networks (CNNs) designed to extract various levels of information, ranging from edges to more complex patterns. Within the cervical region, it comprises neural networks that exhibit a higher degree of refinement and processing. The YOLO-NAS models employed Deci's Neural Architecture Search (NAS) technology, AutoNAC, which is utilized to ascertain the appropriate block type, optimal sizes, stage structures, and the number of channels within each stage. In YOLO-NAS, the neck incorporates supplementary layers designed to improve the representation of its features. This component of the model aims to enhance the network's capability to comprehend and identify objects. Finally, at the head, detection takes place as it utilizes the refined features from the neck to generate predictions. This process involves predicting bounding boxes to identify objects and class probabilities based on the data from each grid cell within the image. The integration of NAS is a fundamental aspect of the design process for the model's backbone, neck, and head. Finally, the architecture of the model will incorporate Quantization-Aware REpVGG (QA-RepVGG) blocks to facilitate Post-Training Quantization (PTQ). Utilizing quantization aware QSP and QCL modules that incorporate QA-RepVGG blocks facilitates 8-bit quantisation and reparameterization, thereby minimizing accuracy loss during post-training quantization (35).

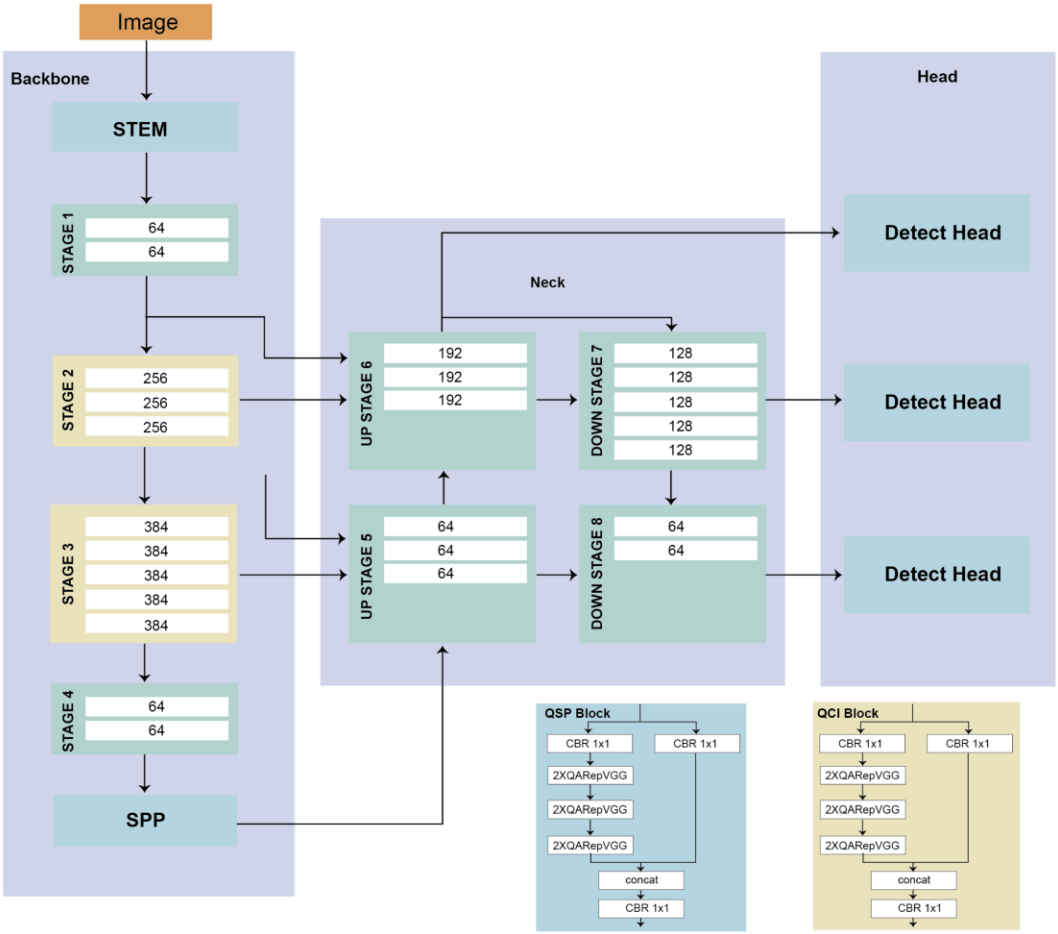

Figure 2. YOLO-NAS architecture.

2.3.2 Model Development and Configuration

The YOLO-NAS model employs a modified iteration of RepVGG backbone architecture. The architecture utilizes quantization-aware Rep-VGG blocks as the foundational elements of the network. A series of quantization-aware RepVGG blocks is integrated to create the quantization-aware 'QSP' and 'QSL' blocks that constitute the framework of the YOLO-NAS architecture. The implementation of these quantization-aware blocks guarantees the model's proficiency with Post-Training Quantization (PTQ), facilitating a reduction in accuracy loss during the conversion to integer models. Before initiating the model training, the configurations including learning rate, batch size, epochs, and optimizers were established based on findings from earlier studies utilizing the YOLO family for object detection tasks.

The training was carried out on Google Colab Pro, which offered a restricted number of computing units. In light of these limitations, each variant of the YOLO-NAS model was subjected to a total of 40 epochs independently. This project employed the AdamW optimizer, recognized for its enhanced characteristics relative to the conventional Adam optimizer. Its application is particularly advantageous in object detection tasks, as it contributes to stabilizing the training process and enhancing convergence. The batch size was established at 32, and the learning rate was determined to be 0.002, while the initial learning rate was configured to 0.0005. The selection of these hyperparameters was informed by established methodologies to promote efficient and stable training (36), particularly within the constraints of the Google Colab Pro environment. It was observed that YOLO-NAS-S contained 19.0 million parameters, YOLO-NAS-M had 51.1 million parameters, and YOLO-NAS-L included 66.9 million parameters. This configuration of parameters enables the model to effectively capture more intricate patterns, such as sperm, within the context of this project (37). All the parameters and YOLO-NAS variants used are summarized in Table 1.

This study explores the application of various YOLO-NAS variants, including YOLO-NAS-S, YOLO-NAS-M, and YOLO-NAS-L, to determine which model demonstrates the most effective capability in detecting sperm and other impurities within the dataset. In a manner akin to (38), the application of these three models provides a degree of flexibility and adjustability, allowing for customization or optimization based on the specific type of sperm analysis intended. This provides an opportunity for health professionals to further refine this model according to their preferred levels of precision.

In this study, the YOLO-NAS model employs an automated procedure through the NAS algorithm, which systematically designs the architecture of neural networks to optimize performance for the specific task at hand. The incorporation of an architectural framework aims to minimize the necessity for human intervention while utilizing currently available resources effectively. NAS architecture can typically be divided into three primary components. The initial segment focusses on the

exploration of the current search space. Secondly, the search process encompasses the selection of the appropriate controller type and the assessment of the currently capable candidates (39), culminating in the methodology for evaluating the performance of the selected candidate. "Search strategy" denotes the methodical approach employed in NAS to ascertain the most appropriate architecture within the specified search space. The categorization of NAS algorithms is based on their respective search methods, which include random search, reinforcement learning, evolutionary algorithms, sequential model-based optimization, and gradient optimization (40, 41).

Table 1. Summary of training parameter.

Hyperparameter	**Value**
Warmup Mode	Linear Epoch Step
Learning Warmup Epoch	10
Learning Mode	Cosine
Optimizer	AdamW
Epochs	40
Batch Size	32
Initial Learning Rate	0.0005
Learning Rate	0.002
Parameters of YOLO-NAS model variants	
YOLO-NAS-S	19.0 params (M)
YOLO-NAS-M	51.1 params (M)
YOLO-NAS-L	66.9 params (M)

2.4 Performance Evaluation

At this stage, the performance of the developed YOLO-NAS-S, YOLO-NAS-M, and YOLO-NAS-L models was assessed in detecting sperm and impurities from microscopic images. Four parameters were selected to characterize the model's performance: Mean Average Precision (mAP), Recall, Precision, and F1 score.

Mean Average Precision (mAP) serves as a critical performance metric in the evaluation of object detection tasks (11). The average precision for each class will be calculated, and the mean of these values will be computed to derive the mean Average Precision (mAP). A high mean Average Precision (mAP) is indicative of superior performance, implying that the model is capable of attaining high precision across various object categories. Equation (1) represents the formula for calculating mAP, where AP_i denotes the Average Precision for the ith class, and N signifies the total number of classes under evaluation.

Recall serves as an additional metric that assesses the efficacy of our object detection models in the identification of sperm and impurities. In this project, recall quantifies the ratio of true positive instances accurately identified by the model. A high recall score suggests that the model successfully identified a significant number of relevant positive instances, even if it also included some irrelevant ones. Recall is articulated in Equation (2), where True Positives represent instances in which the identified objects belong to the positive class, whether categorized as Sperm or Impurity. Conversely, False Negatives refer to instances where the predictions indicate a negative class, despite the original classification being positive.

Similarly, it is important to note that precision was employed as a metric to evaluate the performance of the YOLO-NAS model variants in this study. The concept of precision in this context quantifies the ratio of true positive predictions generated by the model relative to the overall count of positive predictions made. A model exhibiting a high precision score signifies that it has successfully made accurate positive predictions while maintaining a minimal occurrence of false positives. This implies that the likelihood of the model's prediction aligning with its positive class is expected to be highly accurate. Recall is articulated in Equation (3), where False Positives in this equation delineate the instances that were misclassified as belonging to the positive class, despite originally being part of the negative class.

An F-1 score, as presented in Equation (4), serves as a performance metric that integrates both precision and recall into a single value, offering insights into the equilibrium between accuracy and identification capability. The F1 score typically ranges from 0 to 1, where a score of 1 indicates that the model achieves perfect precision and recall, while a score of 0 signifies that the model is evidently ineffective.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{1}$$

$$\mathrm{Re}\,call = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

$$\Pr ecision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{3}$$

$$F1\ Score = \frac{2 \times \Pr ecision \times \mathrm{Re} call}{\Pr ecision + \mathrm{Re} call} \tag{4}$$

3. RESULTS AND DISCUSSION

3.1 Evaluation of YOLO-NAS performance

This study will evaluate the performance of YOLO-NAS models through the application of several metrics, including mean Average Precision (mAP@0.5), Recall@0.5, Precision@0.5, and F1 Score@0.5. The "@0.5" notation signifies an Intersection over Union (IoU) threshold of 0.5, which means that a predicted bounding box is considered a true positive if it overlaps with the ground truth bounding box by a minimum of 50%. The mAP@0.5 metric serves to quantify the average precision across all classes, thereby offering a thorough assessment of the model's accuracy in object detection. Recall@0.5 measures the ratio of true positives accurately recognised by the model, whereas Precision@0.5 examines the ratio of predicted positives that are indeed correct. The F1 Score@0.5, representing the harmonic mean of precision and recall, provides a comprehensive evaluation of the model's performance. The application of these metrics, utilising an IoU threshold of 0.5, guarantees a thorough and consistent assessment of the model's object detection performance. As indicated in Table 2, the mAP@0.50 values are as follows: YOLO-NAS-S at 0.5713, YOLO-NAS-M at 0.7880, and YOLO-NAS-L at 0.8163. This calculation is derived from the previously established mAP formula presented in Equation (1). The recall values indicate that the model's capacity to predict positive instances is 0.6553 for YOLO-NAS-S, 0.6746 for YOLO-NAS-M, and 0.5547 for YOLO-NAS-L. It is noteworthy that YOLO-NAS-L demonstrates the highest precision at 0.8203, succeeded by YOLO-NAS-M at 0.6303, and YOLO-NAS-S at 0.5173. This might be due to the significant complex architecture in YOLO-NAS-L capable of preserving and extracting fine-grained details of tiny objects that represent the sperm and impurity.

Table 2. Summary of YOLO-NAS model variants performances.

Model Variant	Precision@0.50	Recall@0.50	Average Precision of 'S'	Average Precision of 'Impurity'	mAP@0.50	F1@0.50 Score
YOLO-NAS-S	0.5178	0.6553	0.5906	0.5520	0.5713	0.5785
YOLO-NAS-M	0.6303	0.5746	0.8000	0.7760	0.7880	0.6012
YOLO-NAS-L	0.8203	0.5547	0.8300	0.8030	0.8163	0.6618

3.2 Sperm and Impurity Detection

Figures 3 to 5 illustrate the variants of YOLO-NAS models that have been evaluated on the same image. It is evident that the YOLO-NAS-S model fails to detect any impurities in the image, while both the YOLO-NAS-M and YOLO-NAS-L models successfully identify impurities present in the image. Upon further examination of the images produced by YOLO-NAS-M and YOLO-NAS-L, it is evident that the predictions made by YOLO-NAS-L exhibit a higher confidence interval in comparison to those of YOLO-NAS-M. This indicates that the precision performance is distinctly evident in this comparison. Despite the detection of impurities, YOLO-NAS-L exhibits a marginally elevated confidence interval in comparison to YOLO-NAS-M.

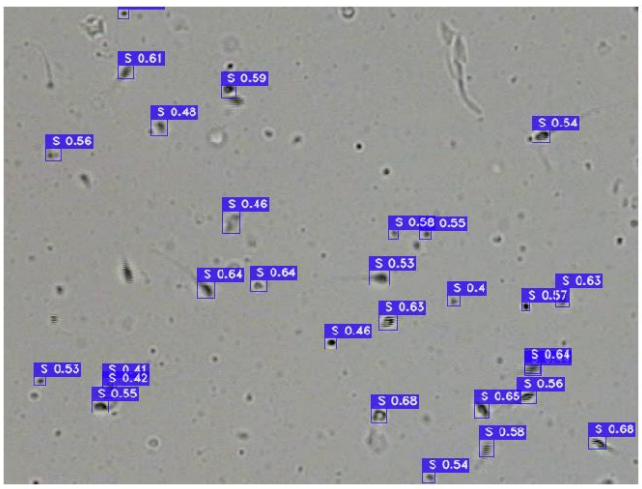

Figure 3. Output image with prediction bounding boxes for YOLO-NAS-S.

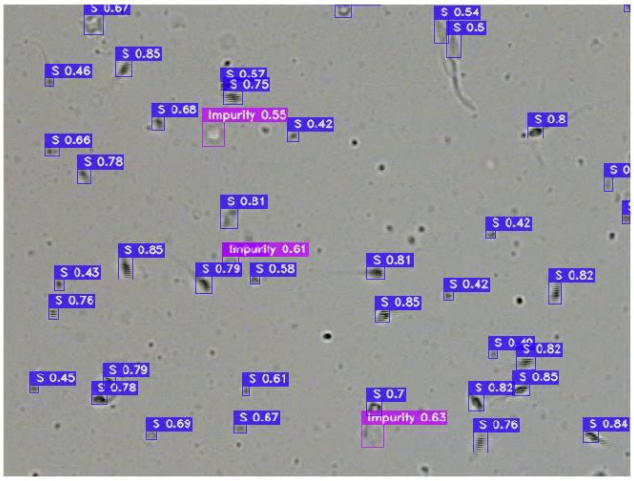

Figure 4. Output image with prediction bounding boxes for YOLO-NAS-M.

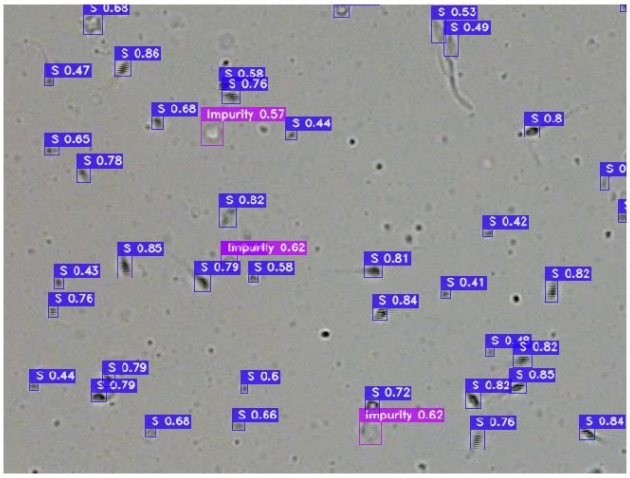

Figure 5. Output image with prediction bounding boxes for YOLO-NAS-L.

4. CONCLUSION

In conclusion of this project, both objectives were successfully met: the development of a YOLO-NAS model for the detection of sperm and impurities, as well as the evaluation of its performance in detecting these elements. The analysis revealed that the most effective YOLO-NAS model for detecting sperm and other impurities in microscopic images was YOLO-NAS-L, achieving a noteworthy mAP@0.50 of 0.8163 and a precision@0.50 of 0.8203. The findings indicate that as the model's complexity escalates with size, it facilitates greater computational complexity, thereby enhancing its reliability in extracting additional features from images, particularly in the context of object detection tasks involving small objects. With an increase in model precision, there is a corresponding decrease in recall, as a higher confidence threshold is established for predictions to minimise the occurrence of false positives.

The performance of this model can be further refined and improved through specific adjustments to this project, including: 1) Training the model with increased epochs and extended training duration to enhance its capacity for recognising and extracting features from images; 2) Utilising a dataset that encompasses various sizes, resolutions, background lighting, and contrasts to facilitate the extraction of additional features by the model; 3) Conducting training on an offline platform to prevent interruptions during the training process or model development; 4) Pursuing additional research and development on YOLO-NAS to enable the clustering of healthy and unhealthy sperm using an alternative dataset; and 5) Advancing YOLO-NAS to analyse sperm morphology and motility through video tracking or image segmentation techniques.

AUTHORSHIP CONTRIBUTION STATEMENT

Aleeyah Nur Sabrina Dzuren: formal analysis, investigation, writing – original draft. Muhammad Amir As'ari: supervision, validation, writing – review & editing.

DATA AVAILABILITY

The data used in this are available on https://github.com/Demozsj/Detection-Sperm .

DECLARATION OF COMPETING INTEREST

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

DECLARATION OF GENERATIVE AI

The author(s) confirms that Gemini 2.5 Pro was used solely to assist with the refinement and presentation of the written text.

ACKNOWLEDGMENT

The authors would like to express their appreciation to Universiti Teknologi Malaysia (UTM) for endowing this research and the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UTM/02/1) for funding this research.

REFERENCES

(1) World Health Organization. Infertility [Internet]. 2023 [cited 2025 Dec 5]. Available from: https://www.who.int/news-room/fact-sheets/detail/infertility.
(2) Bak CW, Seok HH, Song SH, Kim ES, Her YS, Yoon TK. Hormonal imbalances and psychological scars left behind in infertile men. J Androl. 2012;33(2):181–189. https://doi.org/10.2164/jandrol.110.012351.
(3) Glazer CH, Eisenberg ML, Tøttenborg SS, Giwercman A, Flachs EM, Bräuner EV, Vassard D, Pinborg A, Schmidt L, Bonde JP. Male factor infertility and risk of death: a nationwide record-linkage study. Hum Reprod. 2019;34(11):2266–2273. https://doi.org/10.1093/humrep/dez189.
(4) World Health Organization. WHO laboratory manual for the examination and processing of human semen [Internet]. 2021 [cited 2025 Dec 5]. Available from: https://www.who.int/publications/i/item/9789240030787.
(5) Carlsen AGE, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. BMJ. 1992;305:609. https://doi.org/10.1136/bmj.305.6854.609.
(6) Agarwal A, Selvam MKP, Ambar RF. Validation of LensHooke® X1 PRO and computer-assisted semen analyzer compared with laboratory-based manual semen analysis. World J Mens Health. 2021;39(3):496. https://doi.org/10.5534/wjmh.200185.
(7) Mirroshandel SA, Ghasemian F. Automated morphology detection from human sperm images. In: Palermo GD, Sills ES, editors. Intracytoplasmic sperm injection: Indications, techniques and applications. Spinger Cham: Berlin; 2018. p. 99–122. https://doi.org/10.1007/978-3-319-70497-5.
(8) Jati G, Gunawan AA, Lestari SW, Jatmiko W, Hilman M. Multi-sperm tracking using Hungarian Kalman Filter on low frame rate video. 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE. 2016;530–535. https://doi.org/10.1109/ICACSIS.2016.7872796.
(9) Bhattacharjee B, Rahim J. Predicting semen motility using three-dimensional convolutional neural networks. arXiv:2101.02888. 2021. https://arxiv.org/abs/2101.02888.
(10) Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. SSD: Single shot multibox detector. Computer Vision – ECCV 2016: 14th European Conference on Computer Vision. 2016;Part I 14:21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
(11) Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;779–788. https://doi.org/10.1109/CVPR.2016.91.
(12) Chauhan R, Ghanshala KK, Joshi R. Convolutional Neural Network (CNN) for image detection and recognition. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). IEEE. 2018;278–282. https://doi.org/10.1109/ICSCCC.2018.8703316.
(13) Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014;580–587. https://doi.org/10.1109/CVPR.2014.81.
(14) Chahal KS, Dey K. A survey of modern object detection literature using deep learning. arXiv:1808.07256. 2018. https://arxiv.org/abs/1808.07256.
(15) Fang Y, Guo X, Chen K, Zhou Z, Ye Q. Accurate and automated detection of surface knots on sawn timbers using YOLO-V5 model. Bioresources. 2021;16(3):5390. https://doi.org/10.15376/biores.16.3.5390-5406.

(16) Mamdouh N, Khattab A. YOLO-based deep learning framework for olive fruit fly detection and counting. IEEE Access. 2021;9:84252–84262. https://doi.org/10.1109/ACCESS.2021.3088075.
(17) Rivero-Palacio M, Alfonso-Morales W, Caicedo-Bravo E. Anemia detection using a full embedded mobile application with YOLO algorithm. IEEE Colombian Conference on Applications of Computational Intelligence. 2021; 3–17. https://doi.org/10.1007/978-3-030-91308-3_1.
(18) Al-Masni MA, Al-Antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, Valarezo E, Choi M-T, Han S-M, Kim T-S. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Programs Biomed. 2018;157:85–94. https://doi.org/10.1016/j.cmpb.2018.01.017.
(19) Baccouche A, Garcia-Zapirain B, Olea CC, Elmaghraby AS. Breast lesions detection and classification via YOLO-based fusion models. Comput Mater Continua. 2021;69(1). https://doi.org/10.32604/CMC.2021.018461.
(20) Montalbo FJP. A Computer-aided diagnosis of brain tumors using a fine-tuned YOLO-based model with transfer learning. KSII Trans Internet Inf Syst. 2020;14(12):4816–4834. https://doi.org/10.3837/tiis.2020.12.011.
(21) Yaseen M, Ali M, Ali S, Hussain A, Athar A, Kim H-C. Deep learning based cervical spine bones detection: A case study using YOLO. 2024 26th International Conference on Advanced Communications Technology (ICACT). IEEE. 2024;1–5. https://doi.org/10.23919/ICACT60172.2024.10472013.
(22) Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;7263–7271. https://doi.org/10.1109/CVPR.2017.690.
(23) Wu D, Lv S, Jiang M, Song H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Comput Electron Agric. 2020;178:105742. https://doi.org/10.1016/j.compag.2020.105742.
(24) Li C, Li L, Jiang H, Weng K, Geng Y, Li L, Ke Z, Li Q, Cheng M, Nie W, Li Y, Zhang B, Liang Y, Zhou L, Xu X, Chu X, Wei X, Wei X. YOLOv6: A single-stage object detection framework for industrial applications. arXiv:2209.02976. 2022. https://arxiv.org/abs/2209.02976.
(25) Li C, Li L, Geng Y, Jiang H, Cheng M, Zhang B, Ke Z, Xu X, Chu X. YOLOv6 v3.0: A full-scale reloading. arXiv:2301.05586. 2023. https://arxiv.org/abs/2301.05586.
(26) Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023;7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
(27) What is YOLOv8? [Internet]. [cited 2025 Dec 5]. Available from: https://roboflow.com/model/yolov8.
(28) Aldughayfiq B, Ashfaq F, Jhanjhi N, Humayun M. YOLO-based deep learning model for pressure ulcer detection and classification. Healthcare. 2023;11(9):1222. https://doi.org/10.3390/healthcare11091222.
(29) Roboflow. What is YOLO-NAS? [Internet]. [cited 2025 Dec 5]. Available from: https://roboflow.com/model/yolo-nas.
(30) Slimani H, El Mhamdi J, Jilbab A. Advancing disease identification in fava bean crops: A novel deep learning solution integrating YOLO-NAS for precise rust. J Intell Fuzzy Syst. 2024;46(2):1–15. https://doi.org/10.3233/JIFS-236154.
(31) Aristoteles A, Syarif A, Sutyarso S, Lumbanraja FR. Identification of human sperm based on morphology using the you only look once version 4 algorithm. Int J Adv Comput Sci Appl. 2022;13(7):424–431. https://doi.org/10.14569/IJACSA.2022.0130752.
(32) Zhu R, Cui Y, Huang J, Hou E, Zhao J, Zhou Z, Li H. YOLOv5s-SA: Light-weighted and improved YOLOv5s for sperm detection. Diagnostics. 2023;13(6):1100. https://doi.org/10.3390/diagnostics13061100.
(33) Valiuškaitė V, Raudonis V, Maskeliūnas R, Damaševičius R, Krilavičius T. Deep learning based evaluation of spermatozoid motility for artificial insemination. Sensors. 2020;21(1):72. https://doi.org/10.3390/s21010072.
(34) Chen A, Li C, Zou S, Rahaman MM, Yao Y, Chen H, Yang H, Zhao P, Hu W, Liu W, Grzegorzek M. SVIA Dataset: A new dataset of microscopic videos and images for computer-aided sperm analysis. Biocybern Biomed Eng. 2022;42(1):204–214. https://doi.org/10.1016/j.bbe.2021.12.010.
(35) Deci AI Team. YOLO-NAS (You Only Look Once- Neural Architecture Search) [Internet]. 2023 [cited 2025 Dec 5]. Available from: https://deci.ai/model-zoo/yolo-nas/.
(36) Thakur A, Kumar R, Bhadoria RS. YOLO-NAS based deep learning approach for breast lesion identification. 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). IEEE. 2024;2:1–4. https://doi.org/10.1109/IATMSI60426.2024.10502989.
(37) Casas E, Ramos L, Bendek E, Rivas-Echeverría F. Assessing the effectiveness of YOLO architectures for smoke and wildfire detection. IEEE Access. 2023. https://doi.org/10.1109/ACCESS.2023.3312217.
(38) Shukla S, Raghuvanshi C, Sharan HO. Firearm detection using YOLO-NAS L/M/S in advanced surveillance system. Res Militaris. 2023;13(1):4248–4262.
(39) Chen RC, Dewi C, Zhuang YC, Chen JK. Contrast limited adaptive histogram equalization for recognizing road marking at night based on YOLO models. IEEE Access. 2023. https://doi.org/10.1109/ACCESS.2023.3309410.

(40) Zhang H, Yao Q, Kwok JT, Bai X. Searching a high-performance feature extractor for text recognition network. IEEE Trans Pattern Anal Mach Intell. 2022;45(5):6231–6246. https://doi.org/10.1109/TPAMI.2022.3205748.
(41) Dai G, Fan J, Dewi C. ITF-WPI: Image and text based cross-modal feature fusion model for wolfberry pest recognition. Comput Electron Agric. 2023;212:108129. https://doi.org/10.1016/j.compag.2023.108129.