Deep learning-based sperm image analysis to support assessment of male reproductive health

Viet-Thang Vu*
Hanoi University of Industry, Hanoi, Vietnam

Manh-Quang Do
Phenikaa University, Hanoi, Vietnam

Trong-Hop Dang
Hanoi University of Industry, Hanoi, Vietnam

Dinh-Minh Vu
Hanoi University of Industry, Hanoi, Vietnam

Viet-Vu Vu
CMC University, Hanoi, Vietnam

Doan-Vinh Tran
University of Education, Vietnam National University, Hanoi, Vietnam

Hong-Seng Gan
Liverpool Univesity, Suzhou, China.

Abstract

Nowadays, male infertility is a worldwide issue. This problem can be caused by various factors such as low sperm count, weak sperm, anti-sperm antibodies, blocked sperm ducts, congenital infertility, and so on. In fact, if the problem can be detected early, it could be completely solved in some specific cases. This paper proposes a new, more comprehensive framework based on a deep learning model that can address two phases including detection and classification, which are the main steps in solving infertility problems. Experiments conducted on some datasets show that our proposed model has achieved high efficiency compared to other models.

Key Words: Human Sperm Analysis, Object Detection, Object Classification, Deep Learning, Computer Vision.

1   Introduction

Infertility is now a crucial problem in our life. According to the World Health Organization (WHO, 2023) statistics, it is estimated that around 17.5% of the adult population – roughly 1 in 6 worldwide – experience infertility [16]. Globally, 48.5 million couples are affected by infertility (2015). Statistic about infertility in the world shows that 9 out of 10 countries have the highest birth rates including countries in Africa and Afghanistan. Countries in Southern Europe, Eastern Europe, and East Asia have the lowest birth rates averaging 1.5 children per woman. In developing countries, one in four couples is affected by infertility. In Vietnam, according to a study by Nguyen Hoai Bac and colleagues [1], among 1,649 infertile men included in the research, the rate of unexplained male infertility accounted for 67.5% of the total study subjects and particularly up to 80.0% for cases with sperm present in the semen. Azoospermia infertility accounted for only 20.8%, but up to 80.9% of these cases found a cause. Additionally, from April 2013 to April 2019, 3,386 out of 28,963 patients visited for infertility, making up 11.7%. This is also an alarming rate for the youth today. In the works of [10, 17], 15% of couples face infertility issues, of which 30-40% of cases are attributed to male factors. Therefore, one of the early detection methods is the evaluation of sperm quality. Currently, doctors evaluate sperm quality manually through a microscope. This method has characteristics such as the ability to depend on the expertise of the evaluator, the ability to rely on the quality of the equipment, and the ability to consume a significant amount of time and manpower. According to a survey conducted in some hospitals at the lower level, the evaluation of sperm quality faces difficulties due to uneven expertise among evaluators while the demand for testing is increasing. Therefore, building a support system for doctors is essential and practical. The contribution of our paper is to propose a framework for sperm analysis to support the assessment of male reproductive health based on a deep learning model. Moreover, we also use an activation function which is introduced in our work [20] for the object detection step for our framework. Experiments conducted on

---
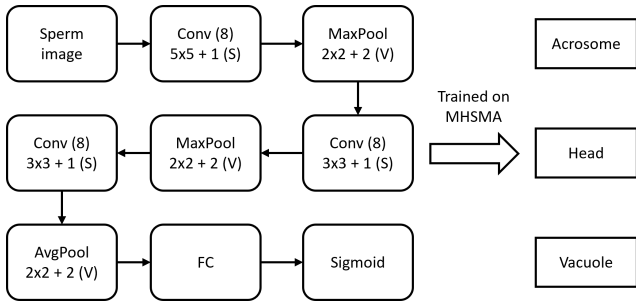*Corresponding author(s). Email: vuvietthang@haui.edu.vn.

Figure 1: The proposed architecture utilizes a deep learning model for sperm classification [11].

some datasets show that our proposed model has obtained a high efficiency compared to other models. The rest of the paper is organized as follows: Section 2 presents the related work on sperm analysis to support the evaluation of male reproductive health. Section 3 describes our new framework. Then Section 4 presents the experiments that were carried out and discusses the results. Finally, Section 5 presents the conclusions and perspectives of this research work.

2   Related work

In this section, we will review some main works that are related to the topic of the paper. In [11], the authors proposed an automatic system installed on smartphones to assess the condition of sperm, DNA fragmentation, and results of hyaluronic acid binding (HBA) tests. Results from the article have shown that the smartphone-based approach performed with an accuracy of 87% in sperm classification tasks. In [4], a method based on a simple Convolutional Neural Networks deep learning model to assess the abnormal morphology of sperm through images at different parts including head, tail, and nucleus has been proposed. In the work, the author built a dataset consisting of 1,540 sperm images from 235 male infertility patients. The results showed the F0.5 score accuracy rates of 84.74%, 83.86%, and 94.65% for head, tail, and nucleus detection, respectively. The sperm morphology analysis is also an interesting direction of research because it is the basic step for other deep steps in studying infertility issues.

In [12], Principal Component Analysis (PCA) has been employed to extract features from sperm images. The k-Nearest Neighbors (KNN) method has also been utilized for diagnosing normal sperm. By applying these methods, the achieved accuracy is 87.53%. However, this work was conducted on a small dataset. A related work was studied by the author Christopher McCallum in 2019 [14]. In the research, they used a deep-learning neural network on a dataset consisting of 1000 DNA images of normal sperm to predict sperm quality. The authors also demonstrated the correlation between sperm cell images and DNA quality. Furthermore, the selection based on this deep learning approach is directly compatible with manual microscopic sperm selection and can support clinical doctors

by providing quick DNA quality predictions (under 10ms per cell) and sperm selection in the 86th percentile from a specific sample. In 2020, the author V. Valiuskaite conducted research and built a Region-Based Convolutional Neural Networks (R-CNN) deep learning model to evaluate the movement ability of sperm in videos, aiming to assess the sperm's state [19]. The model segments the sperm head, while another algorithm is used to track the center coordinates to calculate the head sperm's movement speed. The research recorded an accuracy of 91.77% (95% CI, 91.11–2.43%) in detecting sperm heads in the sample sperm video dataset VISEM (A Multimodal Video Dataset of Human Spermatozoa). The Mean Absolute Error (MAE) of sperm head viability prediction is 2.92 (95% CI, 2.46–3.37), while the Pearson correlation between actual sperm head viability and prediction is 0.969. In [23], Zhang et al. proposed a new method for analyzing sperm morphology but applied it to animals. In the method, by employing image processing techniques, parameters such as the length extension of the head, ellipticity, percentage of the acrosome, and the angle of the lens were examined. To achieve this goal, various algorithms such as K-means and the edge thinning algorithm were utilized. Therefore, based on the extracted parameters, the algorithm could determine the morphological quality of each sperm. Some studies do not focus on sperm morphology analysis but are related to the task of analyzing sperm motion. In [6], a method named SMA was introduced to detect and analyze different parts of human sperm based on some image processing techniques. First, SMA removes noises and enhances the contrast of the image. Then it detects the different parts of sperm (e.g., head, tail) and analyzes the size and shape of each part. Finally, each sperm will be classified as a normal or abnormal sample. In [15], an approach based on image processing is introduced to support the experts in fertility diagnosis. The method combines the Lambertian model based on surface reflectance with mathematical morphology, however, the method focuses on sperm cell segmentation. Other methods for image segmentation for fertility diagnosis can be cited here such as in [3, 8].

3   The proposed framework

This section will introduce a deep learning model to the assessment of male reproductive health. The overview of our framework is presented in Fig. 2. In contrry with other methods, we focus on two main steps which are object detection and object classification. In the object detection step, we use the YOLOv6 model for detecting sperm. The purpose of sperm detection is to identify where the sperm is located and prepare for classification steps. This is an important step. In this step, we use not only common activation functions but also test our new activation function introduced in our preliminary paper [20] called SegRELU. The form of SegRELU is presented in the equation 1.
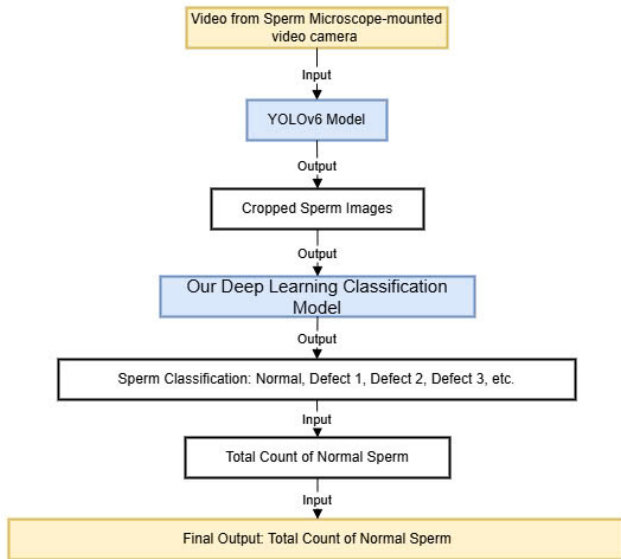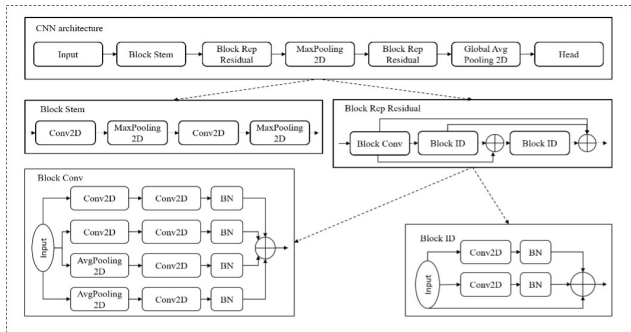
Figure 2: The overview of our framework.

Figure 3: The proposed CNN architecture.

$$f(x) = \begin{cases} \frac{x}{1+|x|} & \text{if } x \leq 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (1)$$

Following the work in [20], SegRELU has some specific properties: (1) the parameter-free configuration is designed to facilitate user implementation, (2) the lightweight architecture aims to promote reproducibility in future deep learning models, and (3) a robust weight update is applied to both the positive and negative segments of the activation functions to maintain effective feature learning throughout the training process.

In the object classification step, we propose a deep learning model with the architecture depicted in Figure 3, this classifier will classify sperms as normal or abnormal. The advantage of the model is a simple, lightweight architecture, leading to fast processing times. The model's architecture is divided into three main components: Block Stem, Block Rep Residual, and Block Head. The Block Stem is used to reduce the size of the data, the Block Rep Residual is extracts various features from images, and the Block Head is responsible for classification. The Block Rep Residual contains the Block Conv and the Block ID. In
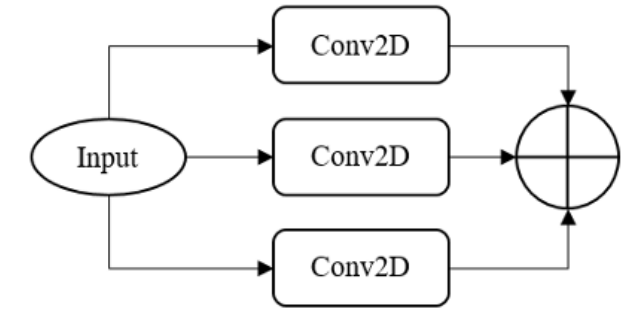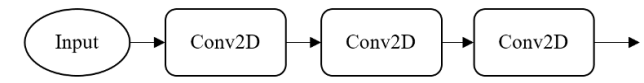

Figure 4: The parallel architecture

Figure 5: The sequential architecture of the AlexNet model and the VGG16 model

which, the Block Conv was designed to extract features with parallel Conv2D layers to increase the ability to extract data features through (1x1), (3x3), and (5x5) kernels with the same input data. If we use Conv2D layers in a sequential order of kernels, the ability to extract diversity will be limited because the input matrix has been changed. While the Block ID was built based on an idea partially from the ResNet network, which is also the main advantage of the ResNet network. As the model has more layers, the gradients of the later layers tend to approach zero. To avoid this, the authors of the ResNet model added X to the output of the model. This helps the model update weights more effectively during training. The model is described more detail in Figure 4.

As mentioned above, the proposed architecture has about 3.5 million parameters, which is more optimal compared to the AlexNet model (about 61 million parameters) and the VGG16 model (about 134 million parameters) and is like the MobileNet model (about 4.3 million) and the MobileNetV2 model (about 3.5 million). Moreover, we used a parallel architecture in Block ID, unlike the sequential architecture of the AlexNet and VGG16 models (Figure 5). This makes our model lighter and
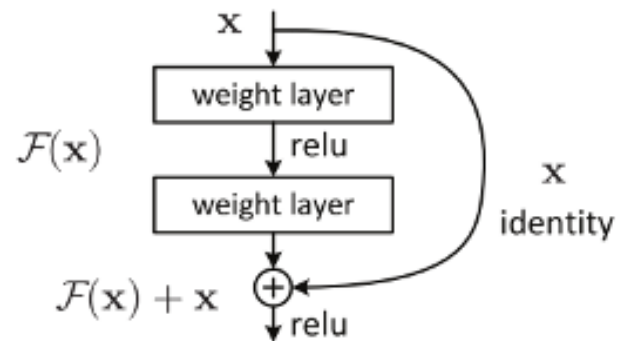

Figure 6: The Block ID architecture

faster to process than the others. Despite this, the achieved accuracy remains sufficiently good for our specific problem.

4 Experiments and results

4.1 Dataset

To evaluate our object detection model's effectiveness, we extracted 125,000 images from the SVIA dataset [5] for training, validation, and testing. The SVIA dataset is sourced from JingHua Hospital of Shenyang, and the creation of this dataset commenced in 2017. Over a span of approximately four years, more than 278,000 objects been annotated. These annotations were performed by 14 reproductive doctors and biomedical scientists and were validated by six reproductive doctors and biomedical scientists. In addition, the annotated objects encompass various sperm types, such as normal, pin, amorphous, tapered, round, or multinucleated head sperm, along with impurities like bacteria, protein lumps, and bubbles. To evaluate the effectiveness of our classification model, we conducted experiments on various datasets: HUSHEM, SCIAN [22], SMIDS [18]. The detailed datasets are described in table 1.

Table 1: Datasets for classification phase

HuShem Dataset		
Labels	HuShem Original	HuShem Original + GANs [2]
Normal	54	1054
Tapered	53	1053
Pyriform	57	1057
Amorphous	52	1052
Total	216	4216

SCIAN Dataset		
Labels	SCIAN Original	SCIAN Original + GAN [2]
Normal	100	3851
Tapered	228	3852
Pyriform	76	3852
Small	72	3852
Amorphous	656	3852
Total	1132	19259

SMIDS Dataset		
Labels	SMIDS Original	SMIDS Original + GAN [2]
Normal	1021	2021
Abnormal	1005	2005
Non-Sperm	974	1974
Total	3000	6000

4.2 Evaluation methods

To evaluate the effectiveness of machine learning models, we use mAP for object detection, and Accuracy, F1-Score, Precision, and Recall for classification models. Before delving into the details of these formulas, let's familiarize ourselves with

some related concepts such as Intersection over Union (IoU). IoU is the ratio of the intersection to the union of the predicted region and the true object region, calculated as the following equation 2.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

The results of IoU are values in the range (0,1), and each detection will have its own IoU value. To determine whether detection is correct or incorrect, we rely on a predefined threshold (in our research, the threshold = 0.7). If the IoU is greater than or equal to the threshold, it is considered a correct detection; otherwise, it is considered a wrong detection. Based on these concepts, we calculate the True Positive (TP) values: The IoU greater than or equal to the threshold, indicating a correct detection; and False Positive (FP): The IoU less than the threshold, indicating a wrong detection; False Negative (FN): cases where the ground truth does not have a predicted bounding box. From these, we obtain metrics used in our research: Precision, which measures prediction accuracy (%) and Recall, which measure of the ability to find correct detections (%), and these metrics are calculated using the following formulas 3, 4.

From the precision and recall obtained above, we can draw the Precision-Recall (PR) curve for each individual class. It illustrates the trade-off between Precision and Recall for various Confidence Score values. The Average Precision (AP) is the area under the PR curve, which was mentioned earlier. A larger area indicates a higher level of Precision and Recall, implying a higher-quality model. After calculating the AP, we compute mAP by averaging the AP values for all classes. The higher the mAP score, the more accurately the model detects and makes correct predictions.

4.3 Results

We tested some models for object detection and the results are presented in Table 2. From the table, we can see that the YOLOv6 using the activation function SegRELU has obtained the highest performance for all indicates Precision, Recall, F1_score, and mAP. It can be explained by the fact that SegRERU has some good properties. SegRELU is based on combining the characteristics of RELU and the Softsign function. Softsign exhibits polynomial growth and features a smoother asymptote line, showcasing a heightened level of non-linearity. The non-linear nature introduced by the quadratic function is highly valued in neural network research for preserving essential features. SegRELU inherits this non-linearity and allows it to accurately delineate complex object boundaries in images. Additionally, SegRELU can generate

Table 2: Comparison of the proposed detection model with other models in the paper [21].

Model	Precision (%)	Recall (%)	F1_Score (%)	mAP (%)
SSD	92.23	65.44	76.59	41.98
RetinaNet	96.30	98.70	98.10	-
DeepSperm	88.50	96.50	93.00	94.11
YOLOv5	82.60	89.20	85.80	88.00
YOLOv6 (SiLU)	97.64	98.37	98.02	99.24
YOLOv6 (ReLU)	97.36	98.42	98.24	99.24
YOLOv6 (SegRELU)	**99.10**	**99.25**	**99.30**	**99.55**

activation during the calculation of gradients in the negative part and mitigate the issue of dead neurons during training.

The results of the classification model have been presented in Table 3, Table 4, and Table 5. From these tables, we can see that our model obtained good results compared with other methods. It can be explained by the fact that our model has been redesigned based on some properties which are presented earlier in section 3.

Table 3: Comparison of the proposed classification model with state-of-the-art classification models on the SMIDS dataset [7].

Model	Overall Accuracy (%)
DWT + SVM (Poly)	77.30
DTCWT + SVM (Poly)	80.10
DTWT + SVM (RBF)	80.30
SURF + SVM (Poly)	77.60
MSER + SVM (Poly)	80.70
KAZE + SVM (RBF)	83.80
VGG19 Aug	87.00
Mobilnet Aug	87.00
Inception Aug	87.30
Our model	**90.77**

Table 4: Comparison of the proposed classification model with state-of-the-art classification models on the SCIAN dataset [2, 9, 13].

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DL-HPM	65.9	58.7	68.9	63.2
CE-SVM	44	-	58	-
APDL	49	-	62	-
FT-VGG	49	47	62	53
MC-HSH	63	56	68	61
TL	-	-	62	-
Our model	60.02	58.34	61.32	61.86

5 Conclusions

In this paper, we have proposed an efficient framework for sperm image analysis to support the assessment of male

Table 5: Comparison of the proposed classification model with state-of-the-art classification models on the HUSHEM dataset [2, 9, 13].

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DL-HPM	96.50	96.80	96.60	96.50
CE-SVM	78.50	80.50	78.50	78.90
APDL	92.20	93.50	92.30	92.90
FT-VGG	94.00	94.70	94.10	94.10
MC-HSH	95.70	96.10	95.50	95.50
TL	96.00	96.40	96.10	96.00
Our model	98.22	98.22	98.22	98.23

reproductive health. The framework is based on deep learning techniques some new ideas will be integrated with the developing deep learning model making it more efficient. Moreover, we also use an activation function which was introduced in our previous work for the object detection step. Some limitations of our paper can be noted here: the results are not a significant improvement; the data set needs update. The results conducted on some real datasets show the effectiveness of our model. In the future, we will continue to extend our research by using other data sets, developing new techniques, and applying our framework in real applications.

6 Acknowledgments

References

[1] Nguyen Hoai Bac and Pham Minh Quan. Study on causes of male infertility. *Journal of Science, Hanoi Medical University*, 1(125):119–128, 2020.

[2] Kamran Balayev, Nihad Guluzade, Sercan Aygün, and Hamza O.ilhan. The implementation of dcgan in the data augmentation for the sperm morphology datasets. *Avrupa Bilim ve Teknoloji Dergisi*, (26):307–314, 2021.

[3] Karima Boumaza, Abdelhamid Loukil, and Kaouthar Aarizou. Automatic human sperm concentrartion in microscopic videos. *Medical Technologies Journal*, 2019.

[4] Violeta Chang, Jose M. Saavedra, Victor Castañeda, Luis Sarabia, Nancy Hitschfeld-Kahler, and Steffen Härtel. Gold-standard and improved framework for sperm head segmentation. *Computer methods and programs in biomedicine*, 117 2:225–37, 2014.

[5] Ao Chen, Chen Li, Shuojia Zou, Md Mamunur Rahaman, Yudong Yao, Haoyuan Chen, Hechen Yang, Peng Zhao, Weiming Hu, Wanli Liu, and Marcin Grzegorzek. Svia dataset: A new dataset of microscopic videos and images for computer-aided sperm analysis. *Biocybernetics and Biomedical Engineering*, 42(1):204–214, 2022.

[6] Fatemeh Ghasemian, Seyed Abolghasem Mirroshandel, Sara Monji-Azad, Mahnaz Azarnia, and Ziba Zahiri. An efficient method for automatic morphological abnormality detection from human sperm images. *Computer Methods and Programs in Biomedicine*, 122(3):409–420, 2015.

[7] Hamza O Ilhan, I Onur Sigirci, Gorkem Serbes, and Nizamettin Aydin. A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. *Med. Biol. Eng. Comput.*, 58(5):1047–1068, May 2020.

[8] Hamza Osman Ilhan and Nizamettin Aydin. A novel data acquisition and analyzing approach to spermiogram tests. *Biomedical Signal Processing and Control*, 41:129–139, March 2018. Publisher Copyright: © 2017 Elsevier Ltd.

[9] Imran Iqbal, Ghulam Mustafa, and Jinwen Ma. Deep learning-based morphological classification of human sperm heads. *Diagnostics (Basel)*, 10(5):325, May 2020.

[10] Aldo Isidori, Maurizio Latini, and Francesco Romanelli. Treatment of male infertility. *Contraception*, 72(4):314–318, 2005.

[11] Soroush Javadi and Seyed Abolghasem Mirroshandel. A novel deep learning method for automatic assessment of human sperm images. *Computers in biology and medicine*, 109:182–194, 2019.

[12] Jiaqian Li, Kuo-Kun Tseng, Haiting Dong, Yifan Li, Ming Zhao, and Mingyue Ding. Human sperm health diagnosis with principal component analysis and k-nearest neighbor algorithm. *2014 International Conference on Medical Biometrics*, pages 108–113, 2014.

[13] Rui Liu, Mingmei Wang, Min Wang, Jianqin Yin, Yixuan Yuan, and Jun Liu. Automatic microscopy analysis with transfer learning for classification of human sperm. *Applied Sciences*, 2021.

[14] Christopher McCallum, Jason Riordon, Yihe Wang, Tian Kong, Jae Bem You, Scott Sanner, Alexander Lagunov, Thomas G. Hannam, Keith A. Jarvi, and David Sinton. Deep learning-based selection of human sperm with high dna integrity. *Communications Biology*, 2, 2019.

[15] Rosario Medina-Rodríguez, Luis Guzmán-Masías, Hugo Alatrista-Salas, and Cesar Beltrán-Castañón. Sperm cells segmentation in micrographic images through lambertian reflectance model. In George Azzopardi and Nicolai Petkov, editors, *Computer Analysis of Images and Patterns*, pages 664–674, Cham, 2015. Springer International Publishing.

[16] Purity Njagi, Wim Groot, Jelena Arsenijevic, Silke Dyer, Gitau Mburu, and James Kiarie. Financial costs of assisted reproductive technology for patients in low- and middle-income countries: a systematic review. *Human Reproduction Open*, 2023(2):hoad007, 03 2023.

[17] Katrien Stouffs, Herman Tournaye, Josiane Van Der Elst, Ingeborg Liebaers, and Willy Lissens. Is there a role for the nuclear export factor 2 gene in male infertility ? *Fertility and Sterility*, 90(November):1787–1791, November 2008.

[18] Omer Lutfu Tortumlu and Hamza Osman Ilhan. The analysis of mobile platform based cnn networks in the classification of sperm morphology. In *2020 Medical Technologies Congress (TIPTEKNO)*, pages 1–4, 2020.

[19] Viktorija Valiuškaitė, Vidas Raudonis, Rytis Maskeliūnas, Robertas Damaševičius, and Tomas Krilavičius. Deep learning based evaluation of spermatozoid motility for artificial insemination. *Sensors (Basel)*, 21(1):72, December 2020.

[20] Viet-Thang Vu, Thanh Quyen Bui Thi, Hong-Seng Gan, Viet-Vu Vu, Do Manh Quang, Vu Thanh Duc, and Dinh-Lam Pham. Activation functions for deep learning: an application for rare attack detection in wireless local area network (wlan). In *2023 25th International Conference on Advanced Communication Technology (ICACT)*, pages 59–64, 2023.

[21] Mecit Yuzkat, Hamza Osman Ilhan, and Nizamettin Aydin. Detection of sperm cells by single-stage and two-stage deep object detectors. *Biomedical Signal Processing and Control*, 83:104630, 2023.

[22] Yejia Zhang, Jingjing Zhang, Xiaomin Zha, Yiru Zhou, Yunxia Cao, and Danny Z. Chen. Improving human sperm head morphology classification with unsupervised anatomical feature distillation, 2022.

[23] Yu Zhang. Animal sperm morphology analysis system based on computer vision. *2017 Eighth International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 338–341, 2017.

Authors

Dr. Viet-Thang Vu is a lecturer and the head of Information System division at Hanoi University of Industry, Vietnam. His research interests include cyber-security, image processing, computer vision, machine learning, and data mining.

Msc. Manh-Quang Do Master's degree in information systems. Researcher at the Faculty of Interdisciplinary Digital Technology, Phenikaa University. The field of research is computer vision

Dr. Trong-Hop Dang received the B.SC and M.SC degrees from Hanoi University of Science and Technology, Vietnam, in 2001 and 2007, respectively. He also received the Ph.D. I earned a degree in Information Technology at Le Quy Don

Technical University, Vietnam, in 2019. He currently works at the Faculty of Information Technology, Hanoi University of Industry, Vietnam. His current topic interests clustering, fuzzy sets, supervised and unsupervised learning, deep learning, and computer vision.

Dr. Dinh-Minh Vu completed his bachelor's degree in computer science from Vietnam National University, Hanoi, his master's degree in information technology from Thai Nguyen University, and his Ph.D. from the Academy of Military Science and Technology, Vietnam, in 1999, 2004, and 2019, respectively. He is currently working at the Faculty of Information Technology, Hanoi University of Industry, Vietnam. His areas of interest are Image Processing, Neural Networks, Pattern Recognition, Fuzzy Logic, and Machine Learning.

Assoc. Prof. Dr. Viet-Vu Vu received a B.S. degree in Computer Science from Ha Noi University of Education in 2000, a M.S. degree in Computer Science from Ha Noi University of Technology in 2004, and a Doctor Degree in Computer Science from Paris 6 University in 2011. He is a researcher at the Information Technology Institute, Vietnam National University, Hanoi. His research interests include clustering, active learning, semisupervised clustering, and E-government applications.

Assoc. Prof. Dr. Hong-Seng Gan received his BEng and PhD in Biomedical Engineering from Universiti Teknologi Malaysia in 2012 and 2016, respectively. He is a faculty member of the School of AI and Advanced Computing, Xi'an Jiaotong – Liverpool Univesity, Suzhou, China. His research areas focus on machine learning, computer vision and medical image processing.

Associate Professor, Dr. Doan-Vinh Tran received a PhD in Informatics Education from the Russian Academy of Educational Sciences, Russia in 1997. Now, he is a lecturer at the Faculty of Educational Technology, University of Education, Vietnam National University, Hanoi. His research concentrates primarily on Informatics Education, Digital Education, applications VR/AR/MR/XR in Digital Education, Clustering, and Image processing.