

Article

Sperm Abnormality Detection Using Sequential Deep Neural Network

Suleman Shahzad ¹, Muhammad Ilyas ¹, M. Ikram Ullah Lali ², Hafiz Tayyab Rauf ^{3,*} , Seifedine Kadry ^{4,5,6} 
and Emad Abouel Nasr ⁷ 

¹ Department of CS & IT, University of Sargodha, Sargodha 40100, Pakistan

² Department of Information Sciences, University of Education Lahore, Lahore 54770, Pakistan

³ Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent ST4 2DE, UK

⁴ Department of Applied Data Science, Noroff University College, 94612 Kristiansand, Norway

⁵ Artificial Intelligence Research Center (AIRC), Ajman University, Ajman 346, United Arab Emirates

⁶ Department of Electrical and Computer Engineering, Lebanese American University, Byblos 1102 2801, Lebanon

⁷ Department of Industrial Engineering, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

* Correspondence: hafiztayyabrauf093@gmail.com

Abstract: Sperm morphological analysis (SMA) is an essential step in diagnosing male infertility. Using images of human sperm cells, this research proposes a unique sequential deep-learning method to detect abnormalities in semen samples. The proposed technique identifies and examines several components of human sperm. In order to conduct this study, we used the online Modified Human Sperm Morphology Analysis (MHSMA) dataset containing 1540 sperm images collected from 235 infertile individuals. For research purposes, this dataset is freely available online. To identify morphological abnormalities in different parts of human sperm, such as the head, vacuole, and acrosome, we proposed sequential deep neural network (SDNN) architecture. This technique is also particularly effective with low-resolution, unstained images. Sequential deep neural networks (SDNNs) demonstrate high accuracy in diagnosing morphological abnormalities based on the given dataset in our tests on the benchmark. Our proposed algorithm successfully detected abnormalities in the acrosome, head, and vacuole with an accuracy of 89%, 90%, and 92%, respectively. It is noteworthy that our system detects abnormalities of the acrosome and head with greater accuracy than current state-of-the-art approaches on the suggested benchmark. On a low-specification computer/laptop, our algorithm also requires less execution time. Additionally, it can classify photos in real time. Based on the results of our study, an embryologist can quickly decide whether to use the given sperm.

Keywords: neural networks; sperm abnormality; sequential deep neural network; fertility prediction

MSC: 68T07



Citation: Shahzad, S.; Ilyas, M.; Lali, M.I.U.; Rauf, H.T.; Kadry, S.; Nasr, E.A. Sperm Abnormality Detection Using Sequential Deep Neural Network. *Mathematics* **2023**, *11*, 515. <https://doi.org/10.3390/math11030515>

Academic Editors: María Del Mar Rueda and Beatriz Cobo Rodríguez

Received: 2 January 2023

Revised: 13 January 2023

Accepted: 16 January 2023

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Infertility means lack of pregnancy with unprotected intercourse after 12 months. A total of 40–50% of reported cases suffer from male factor infertility. Globally, 15% of people suffer from infertility problems [1,2]. Research shows that semen quality in men is decreasing gradually with time [3]. More than 60 papers were reviewed, and it was found that the seminal fluid number of sperms and quality in a given sample have declined in the past 50 years [4]. To seek the reason behind this, scientists have performed semen analyses according to World Health Organization (WHO) rules, assessing semen volume, sperm concentration, total sperm count, sperm morphology, sperm vitality, and sperm motility [5]. However, manual semen analysis is challenging and time-consuming even for a medical expert [6] Therefore, researchers have been developing automatic systems for semen analysis for several decades. After the digitization of images, computer-aided

sperm analysis (CASA) was announced in the 1980s, making it possible to analyze images using computer systems [7]. CASA provides a rapid and objective assessment of sperm motility and concentration but with low accuracy due to the presence of other particles in sperm samples. So, CASA is not suggested for medical use.

In past research, researchers proposed an automatic sperm-tracking technique called a fully automated multi-sperm-tracking algorithm, which trails hundreds of spermatozoa at a given time [8]. In another study, they proposed a method that uses a convolutional neural network (CNN) that classifies a given sample into normal and abnormal sperms. Their results were evaluated on a close dataset, and they compared it with other approaches [9]. In previous research methodologies, the CASA tool was used with classic image processing and machine learning [9]. Deep learning, however, has been increasingly important in computer vision applications in recent years [10–12]. These deep-learning algorithms are state-of-the-art networks for automatic image analyses, such as VGG16, VGG19, DensNet, GoogleNet, AlexNet, ResNet, etc. [13–15].

In this research work, we introduced new sequential deep-learning architecture for the automatic analysis of sperm morphology structure. Furthermore, we evaluated the effectiveness of modern machine-learning and deep-learning techniques over sperm microscopic videos of human sperm microscopic datasets/videos and related parameters to automate the prediction of human sperm fertility. To perform this task, we used an online freely available dataset named MHSMA [16]. This dataset is derived from the Human Sperm Morphology Analysis (HSMA-DS) dataset [17]. Furthermore, experts annotated this dataset with the help of guidelines provided by [18,19].

This is a challenging task because of the below-mentioned reasons.

1. Images in the given dataset are very noisy.
2. The sperms are not stained.
3. Data samples were recorded with a low-magnification microscope, so images were unclear.
4. The images were captured with a low-magnification microscope and thus lacked clarity.
5. There is a severe imbalance between the normal and abnormal sperm classes.
6. There are not enough sperm images for the training phase.
7. For the analysis to be helpful for clinical purposes, it should be performed in real time.

We use data augmentation and sampling techniques to overcome the challenges mentioned above to resolve class imbalance issues and training image shortages. Then, we suggest architecture for a sequential deep neural network that can be taught to distinguish between normal and abnormal sperm heads, acrosomes, and vacuoles. Our proposed algorithm combines Conv2d (2D convolution layer), BatchNorm2d (two-dimensional batch normalization), ReLU (rectified linear activation unit), MaxPool2d (max pool two-dimensional), and flattened layers. It shows impressive results on non-images. Furthermore, our approach can be used for medical objectives. Additionally, each sperm may be checked in as little as 25 milliseconds, allowing our approach to function in real time. Experimental findings demonstrate the efficiency of our approach in terms of accuracy, precision, recall, and F1 score, which is the most advanced way for this dataset.

Section 2 reviews previous work on sperm morphology analysis using different deep-learning techniques. Section 3 describes the dataset and attributes of the proposed research work. Section 4 presents our proposed model with a deep-learning algorithm. Sections 5–8 include our experimental and training setup details, evaluation matrices, result comparison, and discussions. Finally, we summarize conclusions and future directions in Section 9.

2. Related Work

In the literature, a vast body of knowledge is available on automated sperm selection. In one of these studies, the proportion of boar spermatozoa heads was assessed and a characteristic intracellular density distribution pattern was found [20,21]. A deviation model was established and calculated for each sperm's head in this procedure. Then, an ideal value was taken into account for each sperm categorization. Sperm tails were then

taken out, and the gaps in the outlines of the heads were closed using morphological closure. Finally, 60 sperm heads were isolated from the background using Otsu's technique [22].

In another approach, normal and abnormal sperm groups were classified using four phases. (1) Pre-processing images: converting RGB images to grayscale and using a median filter to eliminate noise; (2) separating and identifying each sperm with the help of the Sobel edge-detection technique; (3) segmentation: the head, midpiece, and tail of the sperm were separated; and (4) statistical measurement: a categorization was carried out to distinguish between healthy and unhealthy sperm [23]. In another study, the authors compared sperm nuclear morphometric subpopulations in several animals, including pigs, goats, cows, sheep, etc. [24]. A total of 70 sperm images were processed using ImageJ software, and the findings were utilized for clustering [1]. This approach combines multivariate cluster analyses with computer-assisted sperm-morphology-analysis fluorescence (CASMA-F) technology. In the 1980s, computer-aided sperm analysis (CASA) was announced for sperm analysis. It was a very successful software used to measure sperm characteristics, such as sperm concentration and progressive and non-progressive motility in many animal species. However, in the case of human semen analysis, the CASA tool does not show good results due to complications in the fluid of human semen samples [25].

On the other hand, sperm classification using conventional machine-learning (ML) techniques has been quite successful. This accomplishment was made possible by the idea of an ML pipeline that can tell the difference between normal and abnormal sperms in different sperm portions with the help of a solo sperm image, which is made up of two parts. In the first section, shape-based descriptors were used as a feature extraction technique for the manual extraction of sperm cells. The second section uses these attributes to classify sperm images using a support vector machine [26].

In one of their investigations, a group of researchers benchmarked various combinations of descriptors/classifiers on sperm head classification into five categories: one normal class and four abnormal classes [26]. They used a mix of three distinct shape-based descriptors to extract appropriate sperm attributes from a sperm image. These characteristics were input into four different classifiers: decision tree; naive Bayes; nearest neighbor; and support vector machine (SVM). SVM achieved the greatest mean accurate classification of 49% among these classifiers. As is clear from this illustration, conventional learning algorithms mainly rely on the manual modelling of the data format and extracting the characteristics of human sperm cells. Because the data were complicated and people created the representations, the process took time and was prone to mistakes. As a result, current research has focused on "deep learning," a method that aims to minimize human biases.

Deep-learning algorithms have the potential to learn how to accurately categorize each sperm, as well as how to efficiently portray the data on their own, making them a viable option for overcoming these limitations. Convolutional neural networks, a subclass of deep-learning algorithms, are the most promising for image categorization tasks [27]. These neural network designs consist of several layers that may be divided into two successive sections based on their operation. The first section uses techniques, such as convolution and pooling layers, to learn an abstract representation (i.e., several valuable properties) from a given dataset. The second section, a multi-layer feed-forward neural network, receives these abstract representations and trains an approximation function to map them into the necessary categories [28].

These techniques have just lately begun to be used in sperm morphology studies. Early research focused on categorizing the whole sperm dataset as normal or abnormal [28–30]. The first pre-trained deep-learning model was used in 2018 to classify healthy sperm and identify whether sperm is normal or abnormal. Researchers developed a smartphone-based microscope to differentiate between normal and abnormal sperm images and learn a deep convolutional neural network. They argued that using their network with a smartphone-based microscope could evaluate human sperm at home [28].

Another study proposed a region-of-interest (ROI) segmentation methodology that automatically segments sperm images depending on sperm count, with the help of fuzzy

C-means clustering and modified overlapping group shrinkage (MOGS). The author used the Sperm Morphology Image Dataset (SMID) to categorize each sperm image into three groups: normal; abnormal; and non-sperm. Moreover, they also trained an efficient neural network (Mobile-Net) from scratch without pre-training on other datasets and obtained stunning results [29].

The author also proposed smartphone-based data acquisition and reporting techniques for sperm motility analysis. According to the author, this is the first smartphone technique in an expert system for semen analysis up to 2021. Furthermore, they introduced a multi-stage hybrid analyzing approach for video stabilization, sperm motility, and concentration analysis. They also used the Kalman filter for sperm tracking. Authors claim that this system can report more detailed outcomes in different situations and has more advantages than any other expert system previously used for semen analysis concerning cost, modularity, and portability [31].

In the following study, the author classifies sperm heads into four to five categories [32,33], given by HuSHeM [34] and SCIAN [35] datasets. They employed a pre-trained VGG19 model [33]. They fine-tuned the datasets mentioned above to classify sperm heads into four (HuSHeM dataset) and five (SCIAN dataset) different categories [36]. After this research, researchers manually created a unique convolutional neural network, increasing the accuracy of the prior study on both datasets [32]. Another field of study has focused on classifying the three components of human sperm images (the head, acrosome, and vacuole) into normal and abnormal sperms using freely available datasets (MHSMA) [16]. In their initial investigation, Javadi and Ghasemian used a deep-learning technique to address this issue. In their work, they created a deep-learning system from scratch by hand and trained it using this dataset. They classified the head, vacuole, and acrosome more accurately than prior research that used hand-crafted heuristics [37]. A more recent work, however, used a unique evolutionary technique to create the architecture of a CNN and achieved better results on the same dataset. This neural architecture method showed an accuracy of 77.33%, 77.66%, and 91.66% in the head, acrosome, and vacuole, respectively. This algorithm is called Genetic Neural Architecture Search (GeNAS) [37]. Similarly, in 2022, Chandra et al. also used the same dataset and applied all famous pre-trained deep-learning models, such as VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet, NASNetMobile, NASNetLarge, and Xception. According to experimental findings, the deep-learning-based system performs better than human specialists in classifying sperm with great accuracy, reliability, and throughput. By visualizing the feature activations of deep-learning models, they further analyzed the sperm cell data and offered a fresh viewpoint. Last but not least, a thorough study of the experimental findings revealed that ResNet50 and VGG19 had the highest accuracy rates of 87.33%, 71%, and 73% for the vacuole, acrosome, and head labels, respectively [38]. The following Table 1 shows the latest research and outcomes on the MHSMA dataset.

Table 1. MHSMA dataset with previously proposed techniques.

Ref	Dataset	Method of Classification	Network	Outcomes
[17]	HSMA-DS	SMA algorithm		The SMA algorithm was used for classification, and results show above 90% accuracy for sperm abnormality detection. It also shows high rates of true positive and true negative.
[16]	MHSMA	Deep CNN	Deep CNN	CNN was used to classify sperm image datasets into normal and abnormal sperms with the help of multiple morphological characteristics. As a result, CNN achieved an F 0.5 score of 94.65%, 83.86%, and 84.74% in the vacuole, head, and acrosome abnormality revealing, respectively.
[38]	MHSMA	Deep CNN	VGG19 and ResNet50	VGG19 and ResNet50 showed the best accuracy of 87.33%, 71%, and 73% for the vacuole, acrosome, and head label, respectively.
[37]	MHSMA	CNN	Genetic Neural Architecture Search	This neural architecture method showed an accuracy of 77.33%, 77.66%, and 91.66% in the head, acrosome, and vacuole, respectively.

3. Dataset Parameters and Distribution

This paper uses a freely available benchmarked MHSMA [16] derived from the HSMA-DS dataset [17]. Experts marked this dataset with the help of guidelines provided by [18,19]. A well-defined acrosome makes up 40 to 70% of the sperm head, and the sperm head’s length and width are between 3 and 5 μm and 2 and 3 μm , respectively. The presence of vacuoles shows that the sperm is abnormal. Axially, the midpiece is around 1 micrometer wide, making it one and a half times as big as the head. Additionally, the typical tail features, such as homogeneity, uncoiling, being thinner than the mid-piece, and having a length of 45 μm , must be discernible. Figure 1 shows a schematic of a sperm cell’s structure built on marked labels.

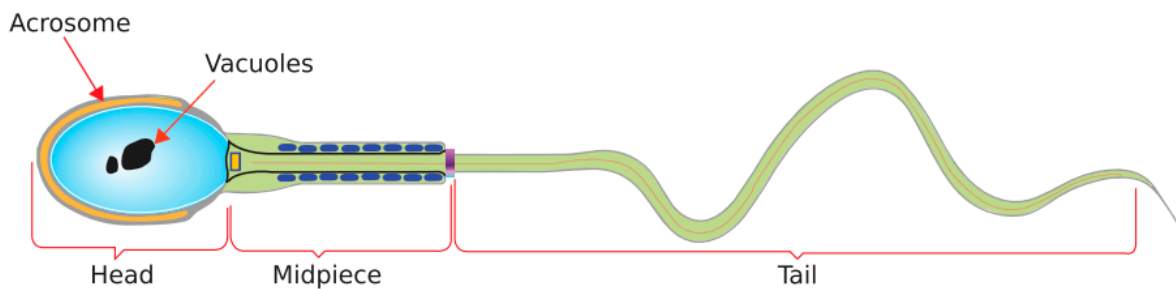


Figure 1. Sperm cell’s structure based on annotated labels.

This dataset used a low-magnification microscope (400 \times and 600 \times) to collect the 1540 grayscale non-stained sperm images. Each image has one sperm and is available in two distinct sizes, 64 \times 64 and 128 \times 128 pixels. In the field of sperm morphology analysis, specialists identified the head, vacuole, acrosome, neck, and tail of the sperm into normal and abnormal sperm images. Table 2 illustrates the distribution of samples from this collection. As can be seen from this chart, the vacuole, tail, and neck labels, in particular, suffer from the issue of data imbalance. Another concern is the limited amount of data samples because deep-learning algorithms need a lot of examples to create good function approximations for high-dimensional input-space problems [27]. The dataset’s normal and abnormal samples are displayed in Figure 2. Two distinct partitioning strategies were used in our research to assess each trained model. In the first technique, the training set, test set, and validation set are the three separate sets into which the MHSMA dataset is divided. Table 3 displays the division of the dataset into normal and abnormal samples in each section. Surprisingly, the technique’s distribution of training, test, and validation test samples match that of two other studies on this dataset, allowing for a fair comparison [16,37,38].

Table 2. MHSMA dataset samples details.

Label	#Normal	#Abnormal	%Normal
Acrosome	1086	454	70.52
Head	1122	418	72.86
Vacuole	1301	239	84.48
Tail and Neck	1471	69	95.52

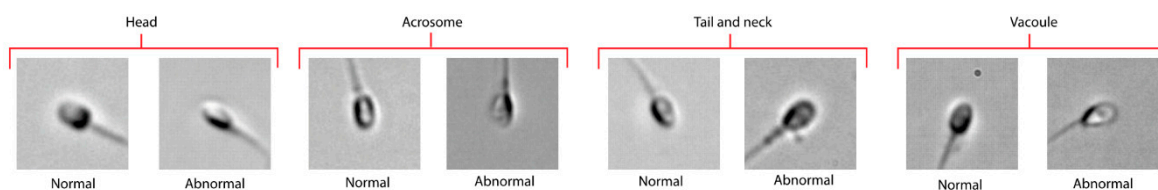


Figure 2. Division of MHSMA dataset.

Table 3. Division of MHSMA dataset in different partitions.

Set	Label	#Normal	#Abnormal	%Normal
Training set	Acrosome	699	301	69.90
	Head	727	237	72.70
	Vacuole	830	170	83.00
	Tail and Neck	954	46	95.40
Test set	Acrosome	213	87	71.00
	Head	219	81	73.00
	Vacuole	262	38	87.33
	Tail and Neck	284	16	94.67
Validation set	Acrosome	174	66	72.50
	Head	176	64	72.33
	Vacuole	209	31	87.08
	Tail and Neck	233	7	97.08

First, the entire MHSMA dataset is divided into five equal partitions. The training and validation sets comprise the left-behind 1240 samples. The test set, consisting of 300 sample photos, is chosen as one of these divisions. The validation set, composed of 240 samples drawn randomly from these four partitions, is used to test the training set, consisting of the remaining 1000 samples. The dataset split ratios for the training, testing, and validation sets are shown in Table 4.

Table 4. Dataset Distribution.

	Number of Images	Split Percentage
Training Set	1000	69.94%
Validation Set	240	15.58%
Test Set	300	19.48%

4. Proposed Method

This section introduces a new sequential deep-learning technique for automatically identifying a human sperm's head, acrosome, and vacuole from grayscale non-stained images. Next, we detail how our deep-learning network was developed and trained. In the end, we review the evaluation measures and procedures employed to rate our suggested model. It is important to emphasize that we do not investigate finding a solution to the categorization issue with the tail and neck labels because classifying the labels for the head, acrosome, and vacuole is challenging work for embryologists, and classifying the labels for the tail and neck is a very simple job.

In light of this, the main topic of this work is the three different classification challenges of identifying the head, acrosome, and vacuole as normal or abnormal. The beauty of our proposed algorithm is that we use the same number of layers to classify sperm cells into the head, vacuole, and acrosome. Moreover, our proposed model outperforms all previously proposed techniques on a given dataset by a significant margin. It should be highlighted that on the suggested benchmark, our system detects abnormalities of the head and acrosome with 12.67% and 11.34% greater accuracy than present modern approaches. Furthermore, our algorithm requires less execution time, even on a low-specification computer/laptop. Moreover, our algorithm classifies images in real-time environments. This enables an embryologist to quickly decide if a given sperm should be chosen or not.

Deep-Learning Model

The design of our deep neural network is based on the Sequential model. The Sequential model is a container class; we also call this a wrapper class used to compose a neural network. We can build any neural network with the Sequential model's help, which means we can combine different layers in our model, even though we can combine different networks in a single model. Our proposed sequential deep neural network contains three stacks of layers, and the layers present in each stack are explained below.

1. **Convolutional 2D layer:** This layer extracts image features from the input dataset with the help of the convolution kernel. Conv2d creates a convolution kernel, combining the input layer and producing a tensor of output. A bias vector is created and added to the outputs if use bias is True. Finally, if activation is not None, it is also applied to the outputs. Arguments of this layer include kernel size 4, stride 1, and padding 1.
2. **Batch normalization:** The output of Conv2d is used as the input to the batch norm2d layer, and the output of the batch norm2d layer is used as the input of the first ReLU. A ReLU activation function follows each convolutional layer.
3. **Max Pooling2d:** Max pooling reduces feature map size through downsampling. Pooling is normally conducted with the help of two common techniques. The first one is max pooling, and the second one is average pooling. However, max pooling is mostly recommended for image feature processing because it retains maximum output samples in the rectangular region. So, we also adopted the max pooling technique in this research with a max pooling stride of 2 that ensures late downsampling.

Similarly, we created three stacks of layers with minor changes. Details of our proposed sequential model are given below in Figure 3.

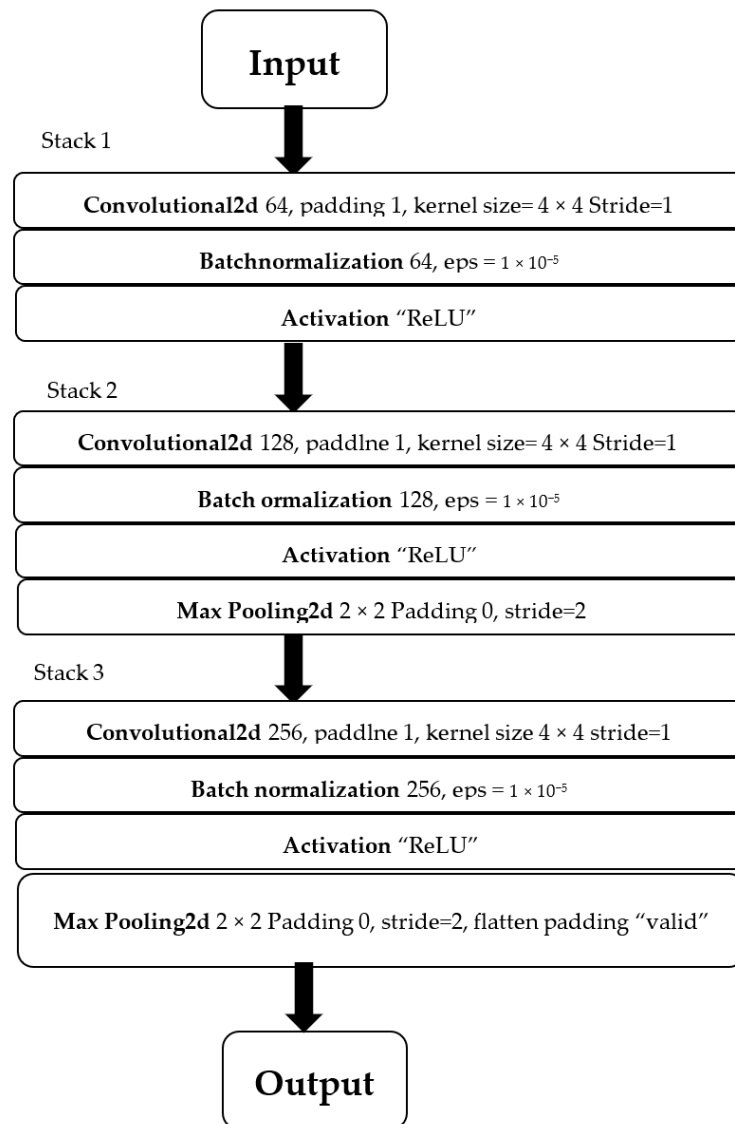


Figure 3. Proposed sequential deep neural network.

5. Training Configurations

In this section, we trained our proposed model to achieve predictions with maximum accuracy and minimum loss. To achieve this, we employed the cross-entropy loss function. This function is a metric used in machine learning to assess how well a categorization model performs. Loss or error is represented between 0 and 1 and the ideal model represents a 0 reading. Generally speaking, the objective is to bring your model as close to 0 as possible.

This function is estimated by subtracting the actual output from the predicted output. Then, we multiply that amount by “ $-y \times \ln(y)$ ”. In other words, we start with a negative number, raise it to the power of the y -positive logarithm, and then deduct it from our initial calculation. The cross-entropy loss calculation formula is as below:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (1)$$

We trained our model through images available in the training set, with the help of the hyperparameters given in Table 5.

Table 5. Details of hyperparameters for proposed model.

Label	Head	Acrosome	Vacuole
Optimizer	Adam	Adam	Adam
Learning-rate	1×10^{-4}	1×10^{-4}	0.001
WEIGHT_DECAY	0.01	0.01	0.01
Loss Function	Cross-entropy loss function	cross-entropy loss function	cross-entropy loss function
Batch size	64	64	64
eps	EPS	EPS	EPS

When our model gave the best validation accuracy, we marked its checkpoints throughout training, which will be documented. If more than one point gave us the same high validation accuracy, then we chose the highest accuracy with the lowest loss. Then, we deployed this model on the test set.

The beauty of our proposed model is that we used the same number of stack layers for the testing and validation set of the head, acrosome, and vacuole with different hyperparameters. On the training set, the parameters of the entire model was trained and adjusted for 100 iterations. The performance of our models was enhanced by obtaining many useful features for each of the three jobs.

To implement our proposed model, we used Python programming language with TORCH framework and Tensorflow backend. Furthermore, all our experiments were performed on Kaggle using NVIDIA Tesla P100 GPU.

6. Evaluation Metrics and Methods

We utilized numerical and visual performance indicators to assess the effectiveness of the models introduced in this paper. The following equations present these numerical evaluation metrics: (2)–(6). The symbols TP, FP, TN, and FN stand for true positive (normal sperms correctly recognized), false positive (abnormal sperms wrongly identified), true negative (abnormal sperms correctly identified), and false negative (normal sperms incorrectly detected) in these equations, respectively. Similarly, Equations (5) and (6) describe the $f0.5$ and $f1$ scores, respectively. Figure 4 displays the confusion matrix for different labels of the given dataset.

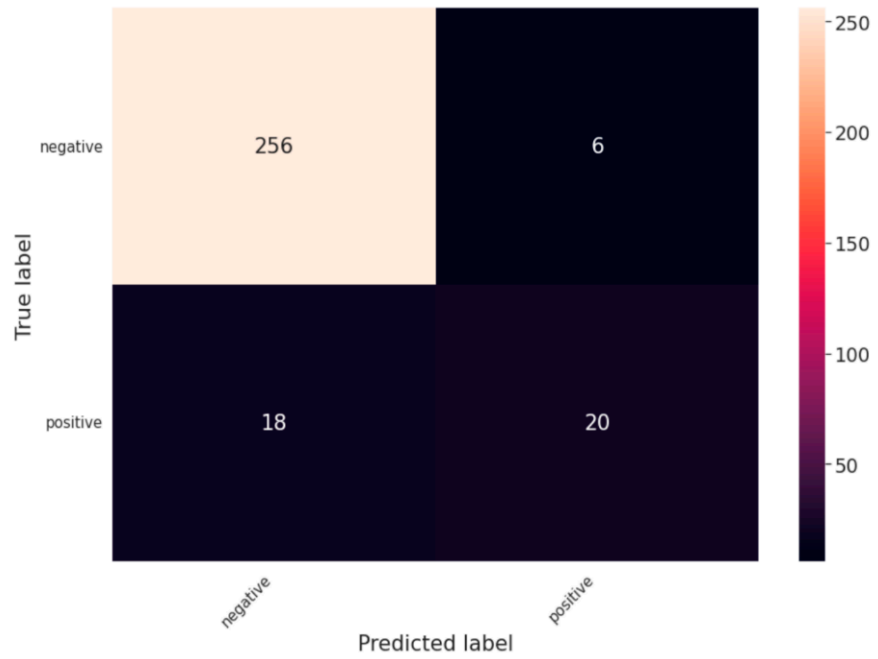
$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

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

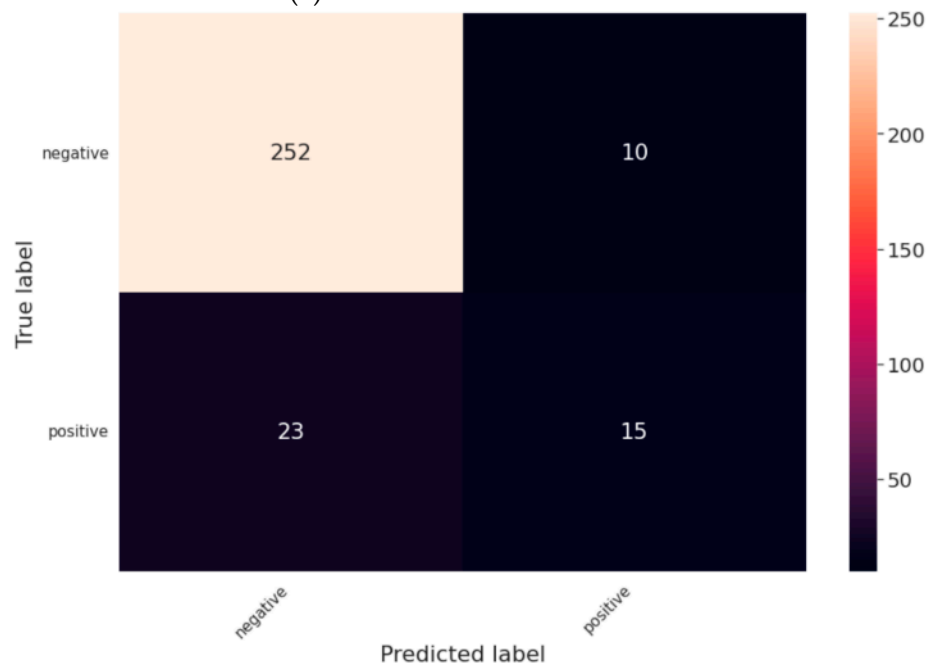
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$f_{\beta} \text{ Score} = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \tag{5}$$

$$f1 - \text{score} = \frac{2 \times P \times R}{P + R} \tag{6}$$

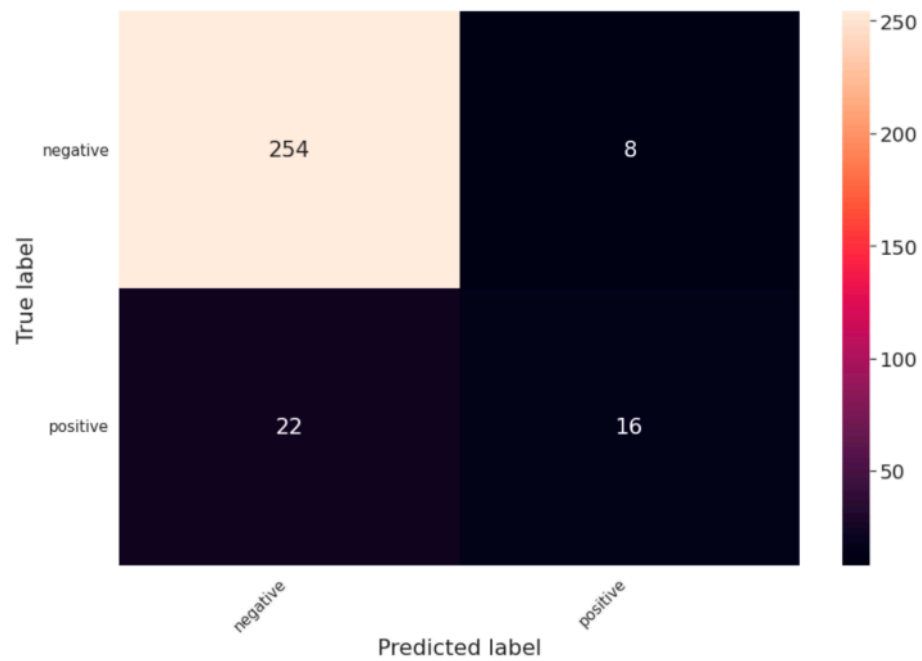


(a) Confusion matrix for vacuole



(b) Confusion matrix for acrosome

Figure 4. Cont.

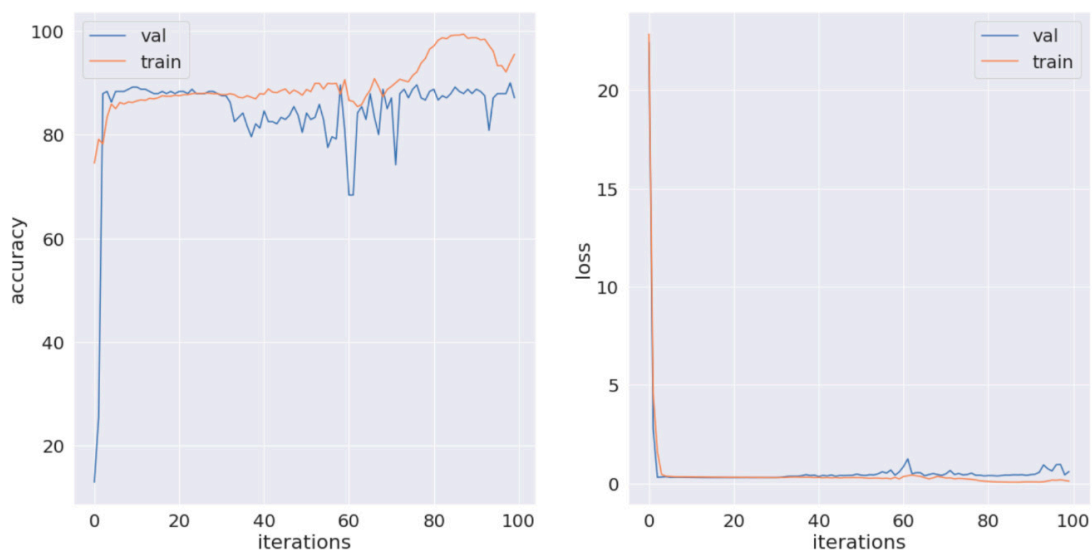


(c) Confusion matrix for head

Figure 4. Confusion matrix for different labels of given dataset.

7. Results Comparison

We separately trained the model to classify the sperm’s head, acrosome, and vacuole. We introduced our model with 1000 samples on 100 iterations (epochs) because we achieved the highest accuracy on 100 iterations. The validation set (240 pieces) determines the loss value after each iteration, and the checkpoint with the lowest validation loss is stored. We assessed the saved checkpoint on the held-out test set after training (300 samples). Figure 5 displays the accuracy and loss during the training and validation for each label over training iterations. Our proposed technique can independently predict abnormality in the sperm head, acrosome, and vacuole with an accuracy of 88.6%, 89%, and 92%, respectively. Table 6 compares our proposed model with pre-trained modern deep-learning architectures.



(a) Vacuole

Figure 5. Cont.

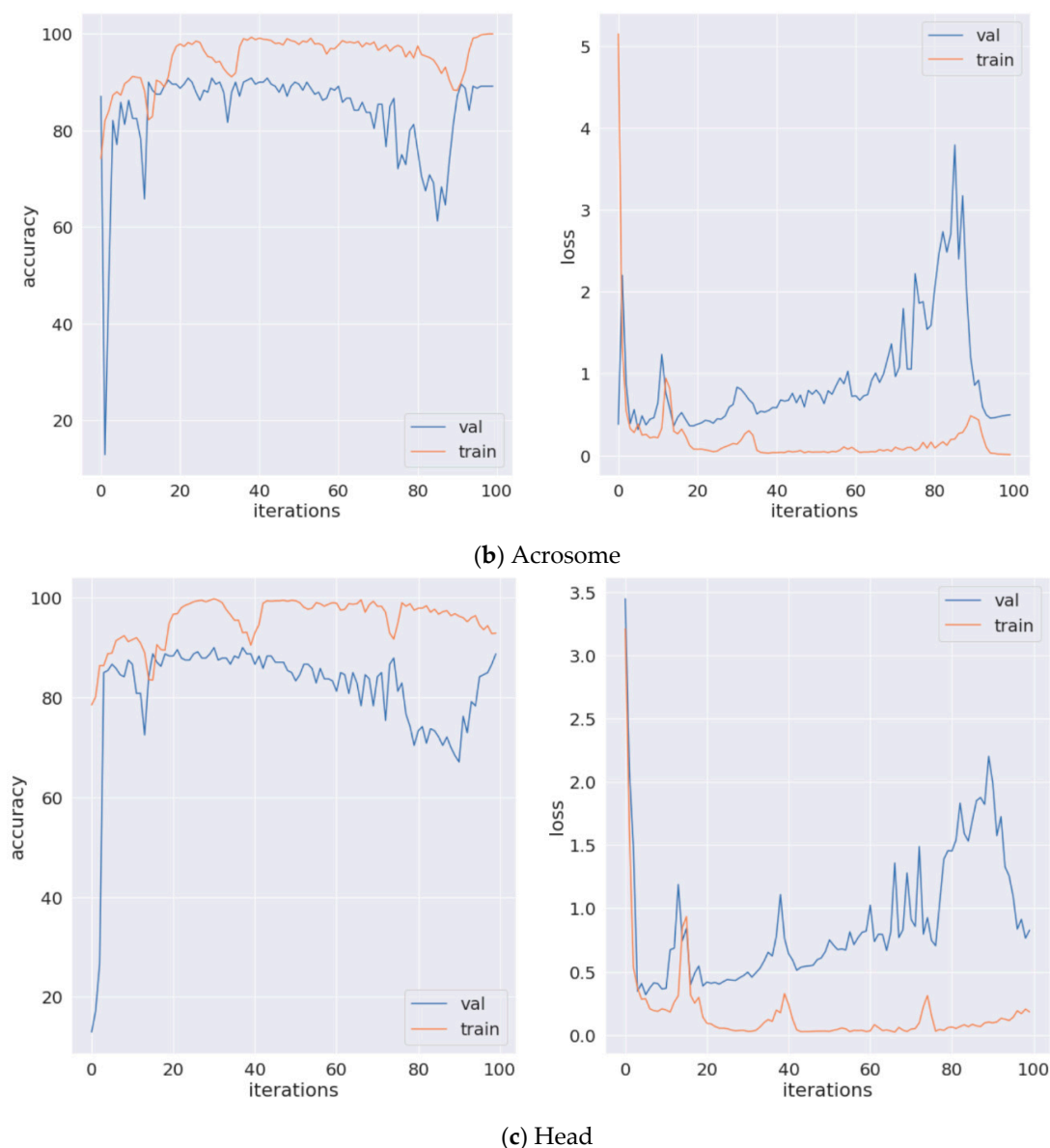


Figure 5. Progression of accuracy and loss on the validation set during sequential deep-learning model training. In this graph, the orange line represents the highest validation accuracy achieved, and the blue line represents the highest training accuracy achieved in the (a) vacuole, (b) acrosome, and (c) head. Similarly, the right side of the graph shows maximum validation loss with the blue line and maximum training loss with the orange line.

We evaluated the efficiency of our proposed technique against 11 pre-trained state-of-the-art deep-learning models [38] tested with the same dataset. Our proposed model outperformed all modern deep-learning models by a significant margin. Table 6 presents the big picture of these models, besides our model, concerning accuracy, F1 score, recall, and precision in different parts of sperm cells.

Similarly, after comparing with pre-trained modern deep-learning models, we also compared our results with the previously proposed technique by different researchers with the same dataset. As a result, we demonstrated that, when applied to all three head, vacuole, and acrosome labels of the MHSMA dataset, our proposed model can identify better than modern architecture previously proposed models by different researchers, such as manually designed CNN architecture, random search and image processing approaches, in terms of accuracy, precision, and $f0.5$ [37]. Eventually, as shown in Table 7, our proposed model shows higher accuracy, precision, recall, and $f0.5$ on the test set for all three labels.

Table 6. (a) Comparison with modern deep-learning models for acrosome sperm cells. (b) Comparison with modern deep-learning models for vacuole sperm cells. (c) Comparison with modern deep-learning models for head sperm cells.

(a)				
Model	Accuracy	Precision	Recall	F1 Score
SDNN	89.00	96.18	91.63	95.23
VGG16	70.67	84.15	72.30	77.78
VGG19	71.00	83.87	73.24	78.20
ResNet.50	71.00	84.24	72.77	78.09
Inception V3	61.00	78.57	61.97	69.29
InceptionResNet	70.33	84.44	71.36	77.35
MobileNet	71.00	85.00	71.83	77.86
MobileNetV2	69.00	81.58	72.77	76.92
DenseNet	66.00	81.71	67.14	73.71
NASNetMobile	71.00	83.51	73.71	78.30
NASNetLarge	63.00	80.36	63.38	70.87
Xception	55.00	73.78	56.81	64.19
(b)				
Model	Accuracy	Precision	Recall	F1 Score
SDNN	92.00	97.70	93.43	95.52
VGG16	87.33	96.28	88.93	92.46
VGG19	87.33	95.90	89.31	92.49
ResNet.50	87.33	95.53	89.69	92.52
Inception V3	82.67	94.49	85.11	89.56
InceptionResNet	84.33	95.36	86.26	90.58
MobileNet	77.67	94.12	79.39	86.13
MobileNetV2	86.67	95.87	88.55	92.06
DenseNet	85.67	95.44	87.79	91.45
NASNetMobile	86.00	95.45	88.17	91.67
NASNetLarge	28.00	73.47	27.48	40.00
Xception	48.00	86.30	48.09	61.76
(c)				
Model	Accuracy	Precision	Recall	F1 Score
SDNN	90.00	96.94	92.02	95.92
VGG16	73.00	86.70	74.43	80.10
VGG19	73.33	87.17	74.43	80.30
ResNet.50	73.67	87.23	74.89	80.59
Inception V3	70.00	84.49	72.15	77.83
InceptionResNet	73.00	87.10	73.97	80.00
MobileNet	73.00	86.32	74.89	80.20
MobileNetV2	62.67	81.29	63.47	71.28
DenseNet	72.33	86.56	73.52	79.51
NASNetMobile	71.67	86.02	73.06	79.01
NASNetLarge	69.67	84.41	71.69	77.53
Xception	66.00	81.97	68.49	74.63

Table 7. Comparison of the latest proposed models with the SDNN technique.

Label	Method	Accuracy	Precision	Recall	f0.5 Score
Acrosome	SDNN	89.00	96.18	91.63	95.23
	GeNAS	77.66	84.76	83.56	84.42
	Random Search	69.66	74.5	86.8	76.67
	Javadi et al.	76.67	85.93	80.02	84.74
	Ghasemian et al.	N/A	N/A	N/A	N/A
Head	SDNN	90.00	96.94	92.02	95.92
	GeNAS	77.33	84.47	84.47	84.47
	Random Search	76.00	80.49	88.58	81.98
	Javadi et al.	77.00	83.48	85.93	83.86
	Ghasemian et al.	61.00	76.71	71.79	75.68

Table 7. Cont.

Label	Method	Accuracy	Precision	Recall	<i>f</i> 0.5 Score
Vacuole	SDNN	92.00	97.70	93.43	96.82
	GeNAS	91.66	95.40	95.03	95.32
	Random Search	89.00	94.20	93.12	93.98
	Javadi et al.	91.33	94.36	95.80	94.65
	Ghasemian et al.	80.33	83.21	93.56	85.09

8. Discussion

Our proposed deep-learning technique shows excellent accuracy on the MHSMA dataset, as we have *f*0.5 and *f*1 scores. These outcomes prove that our SDNN technique is highly recommended for solving sperm classification problems. This approach is more efficient than 11 pre-trained deep-learning models. It also outperformed the latest deep-learning techniques proposed by researchers, including automatically generated deep networks, including the GeNAS algorithm [16,37]. On the head and acrosome labels, our proposed model increases their accuracy by 12.67% and 11.34%, respectively.

After a detailed study of the literature, we concluded that deep-learning models are the finest solution that quickly solve image classification problems compared to conventional approaches, especially for sperm image datasets, such as MHSMA, HuSHeM, and SCIAN. It is interesting to know whether our algorithm can be implemented in real time, such as in fertility clinics. The answer is yes; we can use these deep-learning models in fertility clinics because their accuracy, precision, and other metrics are remarkably greater than even expert manual sperm abnormality identification. We can improve our algorithms by applying our strategies in hospitals and fertility clinics. Two methods can enhance their effectiveness. First, as we train deep-learning algorithms on larger amounts of data, their performance and dependability tend to improve. The images we encounter daily are simple to comprehend, which means they are easy to learn, but analyzing and interpreting images in the medical field requires a great deal of experience. This is particularly true for the classification issue with sperm abnormalities and many other image segmentation and classification issues in medical imaging.

As a result, we conclude that MHSMA dataset magnitude is modest compared to other freely available online datasets. Therefore, we can improve the accuracy and efficiency of our model by gathering more data. To overcome this problem, the first solution is to resolve the dataset's shortfall and carry out a particular procedure to collect more sperm images for the training set. However, this is a complex, costly, and time-consuming procedure. However, if we paid closer attention, we would see that daily operations of this nature occur in hospitals and clinics. In light of this, we advise that medical facilities create guidelines to systematically compile information from people who refer them for an evaluation of their sperm. Here, another point to be noted is that each sperm dataset differs from the others because of the unique equipment and procedures used to gather each dataset. So, supposing one model works well on a given (MHSMA) dataset, in that case, it cannot provide similar or impressive results on other datasets, such as SCIAN and HuSHeM [35], that record sperm images with a different microscope. This means that while creating deep-learning models, dataset variability should be taken into account. Additionally, supposing we wish to apply deep-learning techniques in fertility clinics, in that case, we will ensure that we use the same tools and methods as those who collect the dataset (especially for the test set partition of the dataset).

9. Conclusions

In this research, our proposed sequential deep-learning approach significantly outperformed existing sperm morphology analysis (SMA) methods regarding accuracy, precision, and recall, as well as the *f*0.5 and *f*1 score. Notably, the accuracy of the head, acrosome, and vacuole labels using our SDNN technique was 90%, 92%, and 89%, respectively. Significantly, sequential deep-learning has never been used to examine sperm morphology.

Mainly, when combined, stack network-based deep learning is used. Furthermore, we hypothesize that our ground-breaking SDNN technique can also be applied to address issues related to SMA in the real world, such as fertility clinics, etc. In addition, although our dataset contains labels associated with four separate sperm components and is one of the largest datasets currently available, we need to increase the dataset size, specifically the test set, to conduct additional empirical testing of our state-of-the-art model. Nevertheless, our research demonstrates that this is the most effective method. Therefore, we are hopeful that the fertility departments of healthcare institutions will adopt our suggested deep-learning technique for fertility prediction and analyses.

Author Contributions: Conceptualization, S.S., M.I., M.I.U.L., H.T.R., S.K. and E.A.N.; methodology, S.S., M.I. and M.I.U.L.; software, M.I.U.L., H.T.R. and S.K.; validation, S.S., M.I., M.I.U.L. and H.T.R.; formal analysis, S.S. and M.I.; investigation, M.I.U.L., H.T.R., S.K. and E.A.N.; resources, M.I., M.I.U.L. and H.T.R.; writing—original draft preparation, S.S. and M.I.; writing—review and editing, S.S., M.I., M.I.U.L., H.T.R., S.K. and E.A.N.; supervision, M.I., M.I.U.L. and H.T.R.; funding acquisition, E.A.N. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP2023R164), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP2023R164), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abramoff, M.D.; Magalhães, P.J.; Ram, S.J. Image processing with ImageJ. *Biophotonics Int.* **2004**, *11*, 36–42.
2. Stouffs, K.; Tournaye, H.; van der Elst, J.; Liebaers, I.; Lissens, W. Is there a role for the nuclear export factor 2 gene in male infertility? *Fertil. Steril.* **2008**, *90*, 1787–1791. [[CrossRef](#)]
3. Swan, S.H.; Elkin, E.P.; Fenster, L. The question of declining sperm density revisited: An analysis of 101 studies published 1934–1996. *Environ. Health Perspect.* **2000**, *108*, 961. [[CrossRef](#)]
4. Carlsen, E.; Giwercman, A.; Keiding, N.; Skakkebaek, N.E. Evidence for decreasing quality of semen during past 50 years. *BMJ Br. Med. J.* **1992**, *305*, 609. [[CrossRef](#)] [[PubMed](#)]
5. Cooper, T.G.; Noonan, E.; von Eckardstein, S.; Auger, J.; Baker, H.W.G.; Behre, H.M.; Haugen, T.B.; Kruger, T.; Wang, C.; Mbizvo, M.T.; et al. World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update* **2010**, *16*, 231–245. [[CrossRef](#)] [[PubMed](#)]
6. Menkveld, R.; Holleboom, C.A.G.; Rhemrev, J.P.T. Measurement and significance of sperm morphology. *Asian J. Androl.* **2011**, *13*, 59. [[CrossRef](#)]
7. Amann, R.P.; Waberski, D. Computer-assisted sperm analysis (CASA): Capabilities and potential developments. *Theriogenology* **2014**, *81*, 5–17.e3. [[CrossRef](#)] [[PubMed](#)]
8. Urbano, L.F.; Masson, P.; Vermilyea, M.; Kam, M. Automatic Tracking and Motility Analysis of Human Sperm in Time-Lapse Images. *IEEE Trans. Med. Imaging* **2017**, *36*, 792–801. [[CrossRef](#)]
9. Dewan, K.; Dastidar, T.; Ahmad, M. Estimation of Sperm Concentration and Total Motility from Microscopic Videos of Human Semen Samples. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 2380–23807. [[CrossRef](#)]
10. Li, W.; Zhang, L.; Wu, C.; Cui, Z.; Niu, C. A new lightweight deep neural network for surface scratch detection. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 1999–2015. [[CrossRef](#)]
11. Sahlol, A.T.; Abd Elaziz, M.; Tariq Jamal, A.; Damaševičius, R.; Hassan, O.F. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. *Symmetry* **2020**, *12*, 1146. [[CrossRef](#)]
12. Banan, A.; Nasiri, A.; Taheri-Garavand, A. Deep learning-based appearance features extraction for automated carp species identification. *Aquac. Eng.* **2020**, *89*, 102053. [[CrossRef](#)]
13. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [[CrossRef](#)] [[PubMed](#)]

14. Fan, Y.; Xu, K.; Wu, H.; Zheng, Y.; Tao, B. Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition, MLP and LSTM network. *IEEE Access* **2020**, *8*, 25111–25121. [[CrossRef](#)]
15. Afan, H.A.; Osman, A.I.A.; Essam, Y.; Ahmed, A.N.; Huang, Y.F.; Kisi, O.; Sherif, M.; Sefelnasr, A.; Chau, K.-W.; El-Shafie, A. Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. *Eng. Appl. Comput. Fluid Mech.* **2021**, *15*, 1420–1439. [[CrossRef](#)]
16. Javadi, S.; Mirroshandel, S.A. A novel deep learning method for automatic assessment of human sperm images. *Comput. Biol. Med.* **2019**, *109*, 182–194. [[CrossRef](#)]
17. Ghasemian, F.; Mirroshandel, S.A.; Monji-Azad, S.; Azarnia, M.; Zahiri, Z. An efficient method for automatic morphological abnormality detection from human sperm images. *Comput. Methods Programs Biomed.* **2015**, *122*, 409–420. [[CrossRef](#)]
18. Menkveld, R.; Stander, F.S.; Kotze, T.J.; Kruger, T.F.; Zyl, J.A. The evaluation of morphological characteristics of human spermatozoa according to stricter criteria. *Hum. Reprod.* **1990**, *5*, 586–592. [[CrossRef](#)]
19. World Health Organization (WHO). *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 6th ed.; World Health Organization, Department of Reproductive Health and Research: Geneva, Switzerland, 2021. Available online: <https://www.who.int/publications/i/item/9789240030787> (accessed on 25 June 2022).
20. Sánchez Lidia Petkov, N.; Alegre, E. Statistical approach to boar semen head classification based on intracellular intensity distribution. In Proceedings of the 11th International Conference, CAIP 2005, Versailles, France, 5–8 September 2005; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Volume 3691 LNCS, pp. 88–95. [[CrossRef](#)]
21. Sánchez, L.; Petkov, N.; Alegre, E. Statistical approach to boar semen evaluation using intracellular intensity distribution of head images. *Cell. Mol. Biol.* **2006**, *52*, 38–43. [[CrossRef](#)]
22. Otsu, N. A Threshold Selection Method from GRAY-level Histograms. *IEEE Trans Syst. Man. Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
23. Abbiramy, V.; Shanthi, V. Spermatozoa Segmentation and Morphological Parameter Analysis Based Detection of Teratozoospermia. *Int. J. Comput. Appl. IJCA* **2010**, *3*, 19–23. [[CrossRef](#)]
24. Vicente-Fiel, S.; Palacín, I.; Santolaria, P.; Yániz, J.L. A comparative study of sperm morphometric subpopulations in cattle, goat, sheep and pigs using a computer-assisted fluorescence method (CASMA-F). *Anim. Reprod. Sci.* **2013**, *139*, 182–189. [[CrossRef](#)] [[PubMed](#)]
25. Mortimer, S.T.; van der Horst, G.; Mortimer, D. The future of computer-aided sperm analysis. *Asian J. Androl.* **2015**, *17*, 545–553. [[CrossRef](#)] [[PubMed](#)]
26. Chang, V.; Garcia, A.; Hitschfeld, N.; Härtel, S. Gold-standard for computer-assisted morphological sperm analysis. *Comput. Biol. Med.* **2017**, *83*, 143–150. [[CrossRef](#)] [[PubMed](#)]
27. Goodfellow, I.; Bengio, Y.; Courville, Y. Deep Learning. Available online: <https://mitpress.mit.edu/9780262035613/> (accessed on 15 June 2022).
28. Thirumalaraju Prudhvi Kanakasabapathy, M.K.; Bormann, C.L.; Kandula, H.; Sai Pavan, S.K.; Yarravarapu, D.; Shafiee, H. Human sperm morphology analysis using smartphone microscopy and deep learning. *Fertil. Steril.* **2019**, *112*, e41. [[CrossRef](#)]
29. Ilhan, H.O.; Serbes, G.; Aydin, N. Automated sperm morphology analysis approach using a directional masking technique. *Comput. Biol. Med.* **2020**, *122*, 103845. [[CrossRef](#)] [[PubMed](#)]
30. Thirumalaraju, P.; Bormann, C.L.; Kanakasabapathy, M.; Doshi, F.; Souter, I.; Dimitriadis, I.; Shafiee, H. Automated sperm morphology testing using artificial intelligence. *Fertil. Steril.* **2018**, *110*, e432. [[CrossRef](#)]
31. Ilhan, H.O.; Yuzkat, M.; Aydin, N. Sperm Motility Analysis by using Recursive Kalman Filters with the smartphone based data acquisition and reporting approach. *Expert Syst. Appl.* **2021**, *186*, 115774. [[CrossRef](#)]
32. Abbasi, A.; Miah, E.; Mirroshandel, S.A. Effect of deep transfer and multi-task learning on sperm abnormality detection. *Comput. Biol. Med.* **2021**, *128*, 104121. [[CrossRef](#)]
33. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. In Proceedings of the 2nd International Conference on Learning Representations, ICLR 2014-Workshop Track Proceedings, Banff, AB, Canada, 14–16 April 2014; pp. 1–8.
34. Shaker, F. *Human Sperm Head Morphology Dataset (HuSHeM)*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 3. [[CrossRef](#)]
35. Shaker, F.; Monadjemi, S.A.; Alirezaie, J.; Naghsh-Nilchi, A.R. A dictionary learning approach for human sperm heads classification. *Comput. Biol. Med.* **2017**, *91*, 181–190. [[CrossRef](#)]
36. Riordon, J.; McCallum, C.; Sinton, D. Deep learning for the classification of human sperm. *Comput. Biol. Med.* **2019**, *111*, 103342. [[CrossRef](#)]
37. Miah, E.; Mirroshandel, S.A.; Nasr, A. Genetic Neural Architecture Search for automatic assessment of human sperm images. *Expert Syst. Appl.* **2022**, *188*, 115937. [[CrossRef](#)]
38. Chandra, S.; Gourisaria, M.K.; Gm, H.; Konar, D.; Gao, X.; Wang, T.; Xu, M. Prolificacy Assessment of Spermatozoan via State-of-the-Art Deep Learning Frameworks. *IEEE Access Pract. Innov. Open Solut.* **2022**, *10*, 13715–13727. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.