International Journal of Electronics and Communications System
Volume 4, Issue 2, 99-111.
ISSN: 2798-2610
http://ejournal.radenintan.ac.id/index.php/IJECS/index
DOI: 10.24042/ijecs.v4i2.24419

Human Sperm Morphology Classification Using YOLOv5 Deep Learning Algorithm

Aristoteles
Universitas Lampung, Lampung, Indonesia

Ridho Sholehurrohman*
Universitas Lampung, Lampung, Indonesia

Nasywa Nathania Wirawan
Universitas Lampung, Lampung, Indonesia

Article Info

Article history:

Received: September 29, 2024
Revised: November 10, 2024
Accepted: December 07, 2024
Published: December 15, 2024

Keywords:

Classification;
Deep Learning;
Identification;
Sperm;
YOLOv5.

Abstract

Over the past two decades, one of the most common reproductive problems has been infertility. Infertility is a disease of the reproductive system characterized by the inability to conceive after 12 months or more of regular unprotected sexual intercourse. According to the World Health Organization (WHO), the reproductive rate has declined drastically, with male infertility now accounting for 36% of cases, often due to abnormalities in sperm production. Currently, infertility screening is still done manually, evaluating sperm samples using a microscope, which often produces inconsistent results. However, advances in computer technology have resulted in significant research aimed at improving the analysis of male sperm infertility. This study utilized deep learning technology to identify sperm using the YOLOv5 method. This study involved several stages with data collection using 1330 photos with 2 classes, namely sperm and non-sperm in video format. The second stage involved preprocessing the dataset, which included data extraction, cropping, resizing, and labeling the data for training. The final stage involved testing the trained model to detect and classify sperm based on morphology. The experimental results show that the proposed method is effective in accurately classifying sperm images and analyzing motion videos from recorded video data with a mAP result of 73.1%. Therefore, in the context of this study, the YOLOv5 model is considered efficient in detecting and classifying objects, both sperm and non-sperm.

To cite this article: Aristoteles, R. Sholehurrohman, and N. Nathania W. "Human Sperm Morphology Classification Using YOLOv5 Deep Learning Algorithm," Int. J. Electron. Commun. Syst., vol. 4, no. 2, pp. 99-111, 2024.

INTRODUCTION

Reproduction represents a biological process through which organisms generate new offspring, ensuring the continuity and preservation of their species. Reproduction falls into two categories: sexual and asexual. In humans, it occurs through a sexual process involving the interaction of male and female reproductive systems, with sperm cells merging with egg cells during fertilization to form a totipotent diploid zygote [1]. Preserving reproductive health holds significant importance, particularly for adolescents, as they represent the future generation and the nation's successors.

The Health Law guarantees reproductive health rights and aims to maintain and improve the health of the reproductive system to create a healthy and quality generation [2], [3]. In the last two decades, the most common reproductive problem is infertility, which is the inability to get pregnant after 12 months of sexual intercourse without contraception. The World Health Organization (WHO) reports that reproductive fertility rates have decreased drastically, with infertility in men reaching 36% due to abnormalities in sperm output [4]. Dr. Sudraji Sumapraja also said that the current infertility rate in Indonesia is around 12-15% and this figure is expected to continue to increase every year and continue to increase. Infertility in Indonesia itself is often caused by the narrowing of the seminal ducts due to congenital infections,

• **Corresponding author:**
ridho.sholehurrohman, Lampung University, Indonesia ✉ ridho.sholehurrohman@fmipa.unila.ac.id

immunological factors/antibodies, smoking and also consuming alcohol [5].

Traditional infertility checks using microscopes are often inconsistent. With the development of technology, infertility analysis can be improved using technologies such as deep learning [6], [7], [8], [9]. Detection of human sperm morphology using deep learning technology has been studied [10], [11], using the YOLOv5 method to detect human sperm morphology from the VISEM dataset of 85 videos, with an mAP accuracy of 72.15%. This study shows that deep learning technology can simplify and accelerate infertility analysis. Based on this background, this study aims to overcome the challenges that still exist in conducting human infertility research using deep learning technology in object detection. The main focus of this study is the use of the YOLOv5 method in identifying and classifying human sperm based on morphology. where the YOLOv5 offers excellent execution speed, with the ability to reach 10 frames per second (fps) when detecting objects. This makes it an ideal choice for real-time applications, such as performing detection on sperm and non-sperm objects to be performed. YOLOv5 is written using the PyTorch framework, which provides greater flexibility in model development and implementation. This is in contrast to previous versions that used Darknet, making it easier to customize and integrate into various systems [12] Furthermore, it was analyzed using a confusion matrix to determine the performance of the classification results of the method.

In the Study of Sperm Algorithm Identification Using machine learning for efficient classification of human sperm based on morphology. To develop an object detection algorithm based on artificial neural networks that have good accuracy and is in line with fast computing will support the identification and classification of human sperm based on morphology. This requires high time and cost. To overcome these problems, sophisticated technology is needed, namely by using deep learning. Research on deep learning has been widely conducted [10], [13], [14], [15], [16]. Sperm or spermatozoa, which is a cell that will no longer undergo division or growth. Sperm is an elongated cell produced by the male reproductive system in humans. The purpose of sperm is to transfer DNA from the male to the egg [14]. Sperm cells are distinguished by 3 parts, namely the head, neck, and tail which are used for motility. In the book entitled Plant Morphology [15], the definition of morphology according to the branch of biology is the science that studies the shape and structure of organisms and their unique structural characteristics. Morphology studies the size and shape of plants, animals, and microbes, as well as the interactions of their parts. In the context of this study, sperm morphology is a branch of sperm analysis that studies the shape and structure of spermatozoa. The morphological characteristics of normal spermatozoa have an ideal neck, head, and tail (not branched). Normal sperm morphology includes an oval head morphology with an acrosome covering one-third of the sperm head, a body morphology that is less than the width of the head, and a tail morphology that is longer than the sperm head [16]. Digital images are images stored in digital file formats. Mathematically, digital images are a collection of pixels, which are unit values as continuous functions that represent the color intensity value at a point on the object image, where the image is a three-dimensional projection form converted to 2 dimensions [17]. Processing digital images or image pixels to produce understandable information is called digital image processing.

METHOD

Figure 1 illustrates the research workflow, which consists of several structured steps to address the research problem effectively. The method begins with the problem formulation step, which involves explicitly defining and scoping the research challenge to guarantee focus and alignment with the study's objectives. The next step, literature research, entails acquiring cutting-edge insights and discovering holes in existing approaches. This step not only acts as a foundation for system design, but also identifies chances to incorporate novel tactics such as efficient preprocessing techniques and hyperparameter configurations. In the pre-processing stage, images are prepared using complex algorithms such as resizing tailored for the model architecture-specific color adjustments to improve feature recognition.

These refinements assure cleaner and higher-quality input data, which is crucial for developing the overall performance of the model [15], [16], [17].

The next step, data training, involves an advanced hyperparameter tuning procedure in which configurations such as learning rates, batch sizes, and training epochs are optimized through experimentation. In comparison to conventional methods, our methodology ensures higher accuracy and faster convergence. The trial simulation stage carefully examines the model under varied settings, utilizing measures including accuracy, precision, recall, and F1-score to assess its robustness and performance compared to baseline approaches. Finally, the report writing stage thoroughly describes the findings, offering a comparative analysis that emphasizes the method's quantitative advantages in accuracy, efficiency, and system robustness. This workflow presents an organized and unique methodology for object detection by streamlining data preparation, improving model training strategies, and assuring rigorous performance validation, Advancing the state of research in object detection [10,20].

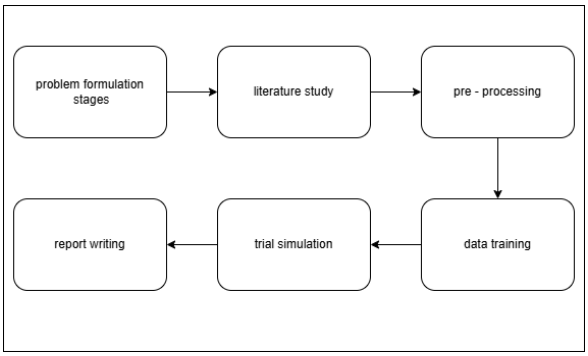

Figure 1. Research Flow

Sperm

Sperm or spermatozoa, is a cell that will no longer undergo division or growth. Sperm are elongated cells and are produced by the male reproductive system in humans. The purpose of sperm is to transfer DNA from the male to the egg [14]. Sperm cells are distinguished by 3 parts, namely the head, neck, and tail which are used as motility.

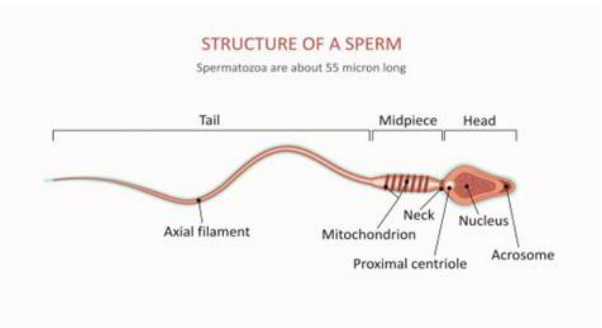

Figure 2. Sperm Morphology (https://doktersehat.com/)

Morphology

In a book entitled Plant *Morphology* [8], the definition of *morphology* according to the branch of biology is the study of the shape and structure of organisms and their unique structural characteristics. Morphology studies the size and shape of plant, animal, and microbial forms, as well as the interaction of their parts.

In the context of this study, sperm morphology is a branch of sperm analysis that studies the shape and structure of spermatozoa, the morphological characteristics of normal spermatozoa have an ideal neck, head, and tail (not branched). Normal sperm morphology includes oval head morphology with acrosomes covering one-third of the sperm head, body morphology less than the width of the head, and tail morphology that is longer than the spermatozoa head [16].

Digital Image Processing

Digital images are images stored in digital file formats. Mathematically, a digital image is a collection of pixels, which is a unit value as a continuous function that represents the color intensity value at a point in the object image, where the image is a three-dimensional projection form that is converted to 2 dimensions [17]. Processing digital images or image pixels to produce information that can be understood is called digital image processing. Digital image representations are generally in the form of M×N, where the row and column indices state a coordinate point in the image, while the matrix elements state the gray level of the image. An image with a matrix size of Mx N can be expressed as a matrix of size M-1 x N-1, namely [18].

$$\begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & f(1 \ddots,1) & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

Digital images can be grouped into three types based on their pixel values, [18]:

a. Binary Image: This image has only two intensity values, namely black and white. Black color is represented by pixels with a value of 1, while white color is represented by pixels with a value of 0 [19];
b. Grayscale Image: Grayscale images have a single gray layer where the pixel values range from 0 to 255. This means that this image has a gray scale that varies from black (0) to white (255) [20];
c. Color (RGB) Image: RGB images consist of three layers, namely Red, Green, and Blue, where each layer has a value range between 0 to 255. This image allows combining the three basic colors to create various colors and shades [21].

Video Digital

Video is a collection of frames recorded in a certain unit of time, usually 1 second of video contains 25-30 image frames. Digital video is the representation of information through a series of images captured and displayed according to the standard scanning system, frame rate, and image size used by video technology. Digital video images can be taken using a digital video camera or a device equipped with a digital video camera [22]. Mathematically, digital video is a function f (n, m, t), where n = {1, 2, ..., N} represents the nth row of the image, then m = {1, 2, ..., M} represents the mth pixel of the image, and t represents the image frame at time t, while f represents the color intensity value of the pixel at position n, m of the image frame at time t [17].

Resize Image

Resizing an image is the act of intentionally resizing an image, either for technical or aesthetic reasons. This process can be done online using various software or applications. Resizing allows users to change the width, height, or diagonal size of an image, as well as change the aspect ratio of the image. Common reasons for resizing include reducing image size and optimizing resolution, aspect ratio, and image file size. YOLOv5 uses an adjustment method called "proportional scaling" to maintain the aspect ratio of the image when resizing it [23].

Random Sampling

Random Sampling is a method in image sampling where each data is given the same opportunity to be selected as a sample [21]. The random sampling method is more efficient in separating image data compared to other methods because the samples taken randomly in using this method are more representative of the population and can produce more accurate conclusions. Random sampling does not require the process of dividing the data into parts that are used as training and testing each time. This results in more efficient time and maximizes model performance [24].

Artificial Intelligence

Artificial Intelligence is one of the fields in computer science aimed at creating software and hardware that can function as something that can think like a human [18]. Artificial Intelligence allows computers to learn from experience, identify patterns, make decisions, and complete complex tasks quickly and efficiently. Artificial Intelligence consists of several types of technology, such as machine learning and deep learning [22].

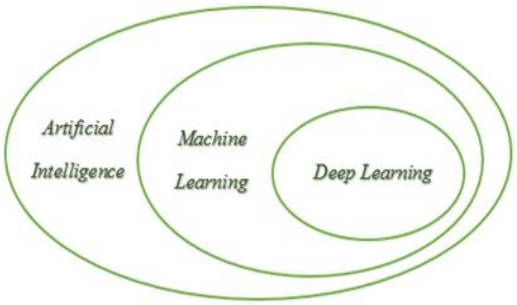

Figure 3. Linkages between AI, ML and Deep Learning

Deep Learning

Deep Learning is a category of machine learning that uses multiple layers of nonlinear data processing for feature extraction and transformation, both supervised and unsupervised, pattern analysis, and classification. Deep Learning is a method of applying machine learning that uses artificial neural networks to mimic the workings of the human brain [25]. Deep Learning enables the investigation of computational models consisting of multiple processing layers at different levels of data representation abstraction. These techniques have made significant progress in speech recognition,

visual object recognition, object recognition, and many other fields [26]. Deep Learning discovers complex structures in large data sets using backpropagation algorithms to show how the machine should change the internal parameters used to compute the representation of each layer from the representation of the previous layer [27].

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of neural network specifically designed to process visual data, such as images and videos. CNN is included in Deep Learning algorithms that work feed-forward. CNN can only be used on data that has a two-dimensional data structure such as images or sound, and uses convolution operations in matrices and four-dimensional weights which are a set of convolution kernels [20]. CNN has been widely applied to image and video data, such as image classification, object identification, image segmentation, and others [28].

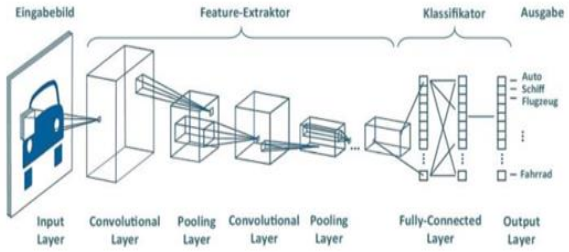

Figure 4. CNN architecture [29].

You Only Look Once (YOLOv5)

YOLOv5 is one of the advanced object detection algorithms owned by the YOLO method. The higher processing speed and accuracy make this machine-learning algorithm applicable in real-time applications [12], [30]. The YOLO V5 architecture consists of Backbone (CSPDarknet), Neck (PANet), and Head (YOLO Layer), as shown in Fig 5.

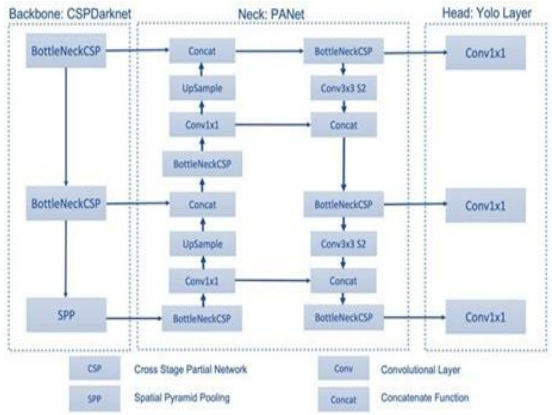

Figure 5. CNN architecture [25].

The backbone deals with the release of features for various levels in the Cross Stage Partial (CSP) network. Furthermore, the bottleneck formulates and sends image features to the Neck (PA-Net) and Special Pyramid Pooling (SPP) [26]. The Neck contains a series of BottleNeckCSP circuits and convolutional networks. In the last step, the head aggregates the image features using a series convolutional network to process the prediction boxes and their associated classes. Prediction box localization is based on an object tracking algorithm that iteratively corrects the position of the bounding box [31].

CSPDarknet53

The feature extraction that will be used in YOLOv5 is CSPDarknet53. This is a "backbone" structure in YOLOv5 developed from the Darknet53 algorithm in the YOLOv4 algorithm which is used as object detection [32]. With the addition of the Cross Stage Partial Connections (CSP) feature that integrates the feature map from the initial stage to the final stage of the network. This CSP implementation reduces computation by 20%, outperforming other state-of-the-art "backbone" architectures [33]. The CSPDarknet53 architecture can be seen in Figure 6.

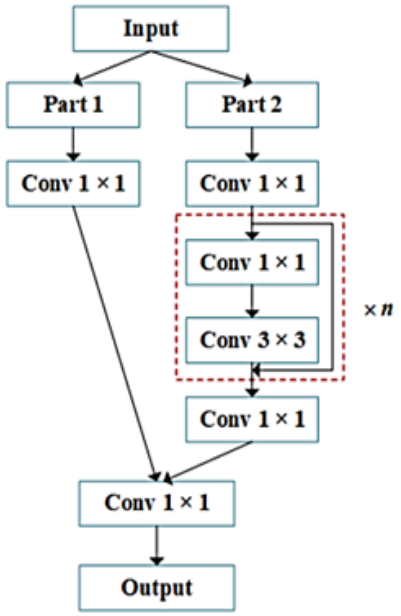

Figure 6. CSPDarknet53 architecture [20].

Confusion matrix

Confusion matrix is one of the methods used to evaluate the accuracy of a classification model. Confusion matrix is a table used to evaluate the performance of a classification model [29]. This table contains

information about the number of predictions made using a model. The categories of confusion prediction results in tabular form can be seen in Table 1.

Table 1. Table Confusion Matrix

Aktual / Predicted	positif	negatif
positif	True positive	False positif
negatif	False Negative	True Negatif

Model Prediction Categories on Confusion Matrix:
True Positive (TP) = Positive data detected as positive
False Positive (FP) = Positive data detected as negative
True Negative (TN) = Negative data detected negative
False Negative (FN) = Negative data detected positive

In confusion matrix, performance measurement is generally done by calculating the values of accuracy, precision, recall, and specificity [34].

Research Road Map

Figure 7 is a road map for research on detection and classification of traffic prohibition signs using YOLOv5 in real time in Bandar Lampung City.

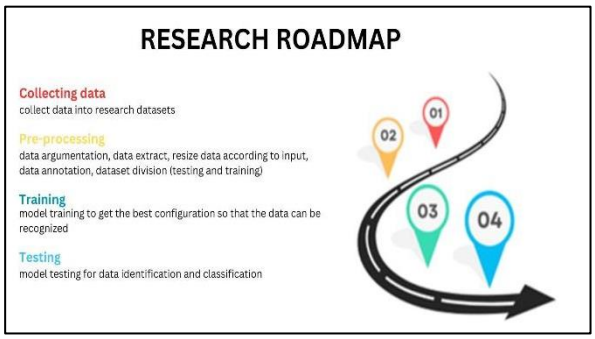

Figure 7. Research Road Map

The dataset in this study was obtained from the open dataset https://datasets.simula.no/visem/. The VISEM dataset contains data from 85 male participants aged 18 and above. This dataset contains more than 35 gigabytes of video, with each video lasting between two and seven minutes. The dataset used is in the form of 2 classes, namely sperm and non-sperm. Furthermore, data pre-processing is carried out by extracting videos into images, resizing according to method input, training images to get the best configurations, and testing models using the best configurations so that the implementation of human sperm identification and classification results based on morphology will be obtained.

RESULTS AND DISCUSSION

In this study, the implementation of sperm classification based on morphology is described in detail using the YOLOv5 algorithm. Analysis of training model processing, object detection and interpretation of sperm morphology identified and classified by the system. Analysis of performance evaluation on the implementation of YOLOv5 proposed in the study using confusion matrix so that the performance results of the YOLOv5 algorithm accuracy can be seen.

Dataset Collection

The dataset used in this study comes from the open source VISEM of human sperm with a total size of 35.2 gigabytes, which was originally an AVI video. From the dataset, 1330 images were taken and divided into two classes: sperm based on morphology and non-sperm. The entire dataset will be divided into three parts for model training purposes, namely training, validation, and testing data, with the aim of identifying sperm and non-sperm objects. The following is an image of the original VISEM data which is still in the form of an AVI video.

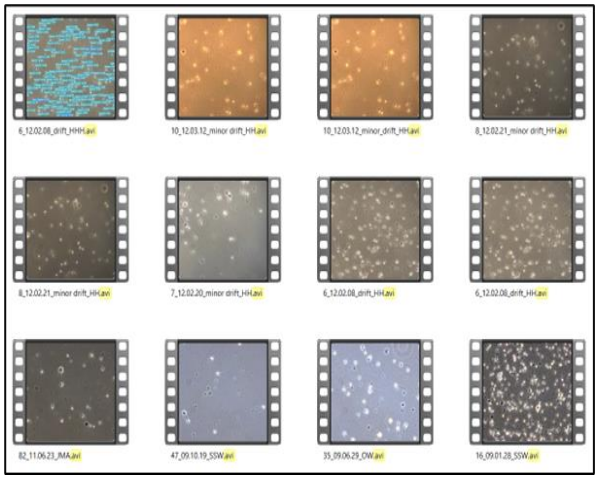

Figure 8. VISEM dataset with AVI extension

Pre-processing Dataset

There are several data pre-processing processes carried out in this study. 4.2.1

Video to Image Extraction The dataset taken on the human sperm VISEM is still in the form of an .AVI extension video. Performing video to image extraction aims to make the dataset used easier to annotate data. The following are the results of the extracted images.

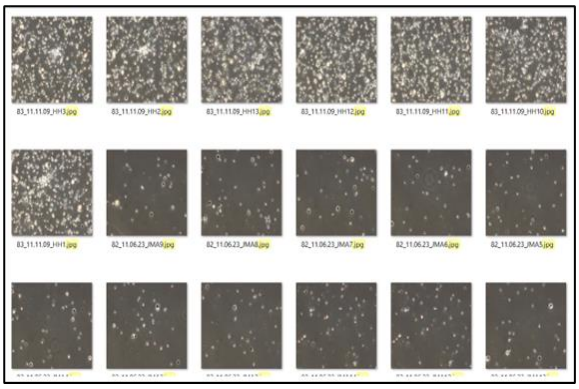

Figure 9. Iamge VISEM dataset

Resize Image

The image data that will be used, namely 1330 image data, will be cut and adjusted in size by resizing the extracted image. In the process of resizing data measuring 640 x 480 will be changed to 640 x 640 using the PIL Library which aims to make the data more ready for processing. The following is the result of a dataset that has been resized image.

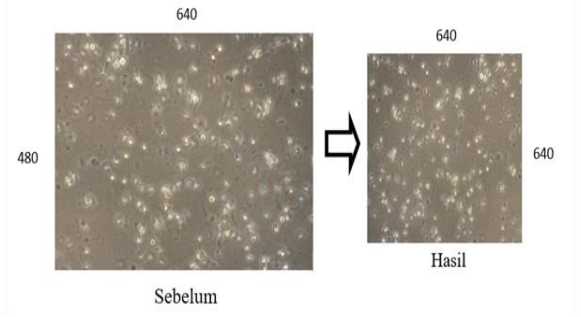

Figure 10. Resize Image Illustration

Data Labelling

This process is done to label the objects contained in the image dataset. The label given has two annotation classes, namely sperm and non-sperm. Each frame is annotated in the form of a bounding box on the morphological part of the sperm head and on objects that are not sperm. For this annotation use Yolo_mark, sourced from the GitHub link listed: https://github.com/AlexeyAB/Yolo_mark.git. The display of the data annotation can be seen in the following image.

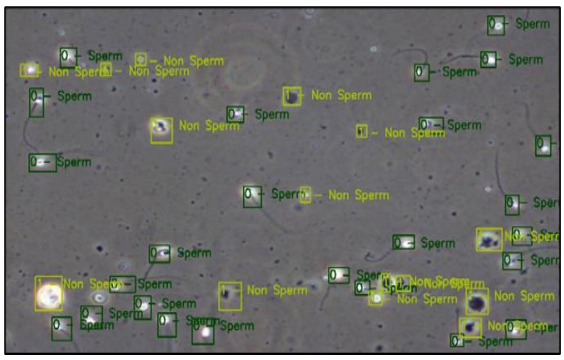

Figure 11. Display of Data Annotation Results

Dataset division

The division of datasets in both models is done by dividing the data into 3 parts, namely training data, validation data, and testing data. Where the third division of data is done using the random sampling method by adding the Sklearn library model and train_test_split coding. The random sampling method is used to separate data in the form of images with its easier process compared to other methods. Three scenarios of splitting data from a total of 1330 images were performed with different ratios: first, 80% for training (1065 data), 10% for validation (133 data), and 10% for testing (133 data); second, 75% for training (931 data), 25% for validation (332 data), and 5% for testing (67 data); third, 60% for training (798 data), 20% for validation (266 data), and 20% for testing (266 data). This image data is then stored in folders named train, valid, and test, with a random division.

Training Model

In conducting both training models, the training process is carried out using the same notebook on the Kaggle platform with a P100 GPU accelerator. The training process is carried out using data that has previously been separated from the entire data with a total of 1330 image data. Several training schemes have been carried out at this stage to determine the best identification of sperm and non-sperm objects based on morphology.

In the context of YOLO (You Only Look Once), Obj.data and Obj.names are two files used in model configuration and training. The Obj.data file contains information related to the dataset used for training the YOLO model. In training to be able to identify sperm and non-sperm image data. As for Obj.

```
classes= 2
train  = data/train.txt
valid  = data/train.txt
names = data/obj.names
backup = backup/
```

```
Sperm
Non Sperm
```

name contains a list of class or object names that will be identified by the YOLO model. Each line in this file contains one class or object name. The structure of the **obj.data** and **obj.name** files is illustrated in the accompanying diagrams. Additionally, there are hyperparameters in configuring the YOLOv5 training method. Hyperparameters aim to control the training process and the characteristics of the model that will be generated. This is very influential on the results of training a model. The table below is a scheme conducted using the same hyperparameters to see better results between the two methods.

Table 2. Table Hyperparameter

Parameter Name	Value
Batch size	16
Number of Classes	2
Learning Rate	0,002, 0,0002, 0,00002
Epoch	150
Momentum	0,937
Scale	0,9

Training Model YOLOv5

In training on YOLOv5 the batch size used is 16 with epoch 150. The following are the results obtained in training on the YOLOv5 model.

a. YOLOv5 mAP Accuracy Results

The results of testing the YOLOv5 model are obtained based on testing 3 different dataset scenarios and different learning rates. The size of the data division has been mentioned in the dataset division subchapter. As for the use of learning rates, namely 0.002, 0.0002, 0.00002. The following results are obtained from various scenarios in the following table 3.

Table 3. Accuracy results of train data obtained from learning rate 0.002.

No	Dataset division	AP		mAP
		Sperm	*Non Sperm*	
1	Train 60%, Validation 20%, Test 20%	82,9%	62%	72,4%
2	Train 70%, Validation 25%, Test 5%	83,7%	62,6%	73,1%
3	Train 80%, Validation 10%, Test 10%	82,7%	61,2%	72%

In Table 3, it can be seen that the largest mAP result at a learning rate of 0.002 is in the second scenario with a dataset division of 70% training, 25% validation and 5% testing. The largest mAP value is 73.1%. With an AP value of 83.7%% sperm and 62.6% non sperm. The AP value of each class has a not too large difference in accuracy value. The sperm class has the largest accuracy value in various scenarios compared to the non-sperm class.

Table 4. Accuracy results of train data obtained from learning rate 0.0002.

No	Dataset division	AP		mAP
		Sperm	*Non Sperm*	
1	Train 60%, Validation 20%, Test 20%	77,5%	50%	63,8%
2	Train 70%, Validation 25%, Test 5%	78,2%	53%	65,6%
3	Train 80%, Validation 10%, Test 10%	74,2%	41,9%	58%

In Table 4, it can be seen that the largest mAP result at a learning rate of 0.0002 is in the second scenario. The largest mAP value is 65.6%. Has a comparison of more than 2% compared to the first scenario. The AP value of each class has a not too large difference in accuracy value. The sperm class has the largest accuracy value in various scenarios compared to the non-sperm class.

Table 5. Accuracy results of train data obtained from learning rate 0.00002

No	Dataset division	AP		mAP
		Sperm	*Non Sperm*	
1	Train 60%, Validation 20%, Test 20%	61,7%	17,3%	39,6%
2	Train 70%, Validation 25%, Test 5%	61,9%	18,6%	40,2%
3	Train 80%, Validation 10%, Test 10%	59,1%	14,3%	36,7%

Table 5 shows the accuracy results of the experiment using a learning rate of 0.00002. In this experiment, the highest result was obtained in the second scenario, which was 40.2%. While the lowest result was obtained in the third scenario with a value of 36.7%. These results have a fairly large difference of 3.5%. In this experiment, the sperm class has a greater AP value compared to the non-sperm class. Table 3, Table 4, and Table 5 show that the highest accuracy (73.1%) is achieved with a learning rate of

0.002, where the data is divided into 70% for training, 25% for validation, and 5% for testing. In contrast, the lowest accuracy (36.7%) was obtained with a learning rate of 0.00002 with data distribution of 60% for training, 20% for validation, and 20% for testing. In addition, the best average precision in identifying sperm objects is 83.7%, while for non-sperm objects it is 62.6%. From these results, it can be concluded that the learning rate greatly affects the training of the model, and the data sharing scenario also affects the accuracy, especially in supporting the use of the right learning rate. The right combination of learning rate and data sharing scenario will improve the accuracy of the model.

b. Precision, Recall and F1-Score Results of YOLOv5 Model

In addition to AP and mAP, this research calculates precision, recall and F1-Score. These values are calculated based on the dataset division scenario and the type of learning rate. The dataset sharing scenario is made into three scenarios, while the learning rate uses three types as well, namely 0.002, 0.0002, 0.00002. The following is a display of the results of precision, recall, and F1-Score.

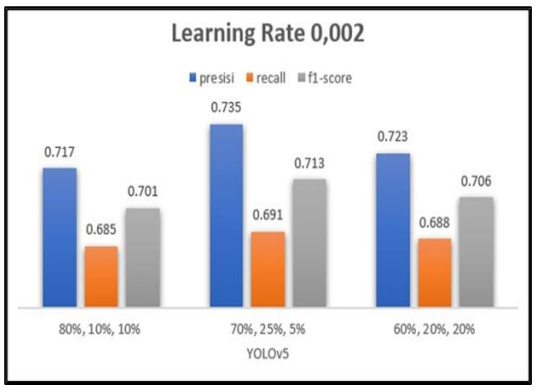

Figure 12. Graph of confusion value results from Precision, Recall, F1-Score at learning rate 0.002

In Fig. 12 is the value of the results of precision, recall and F1-Score at a learning rate of 0.002. The results obtained from the whole are not much different. The highest value is in the second data sharing scenario, the value obtained is for precision of 0.735, recall of 0.691, and F1-Score of 0.713.

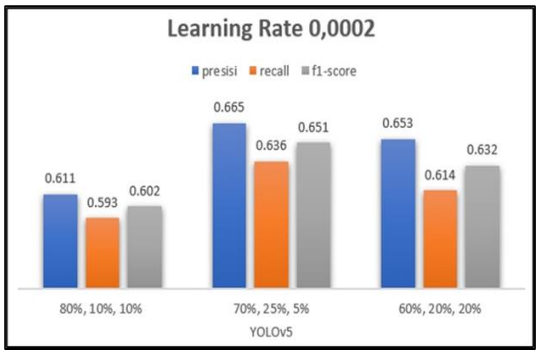

Figure 13. Graph of confusion value results from Precision, Recall, F1-Score at learning rate 0.0002.

Next is Fig. 13, the results of precision, recall and F1-Score at a learning rate of 0.0002. It can be seen that the biggest result is in the second dataset sharing scenario. The values obtained are all the same with a precision value of 0.655, recall 0.636, and F1-Score 0.651. In this case, the values obtained from the first scenario with the second and third scenarios are quite far away.

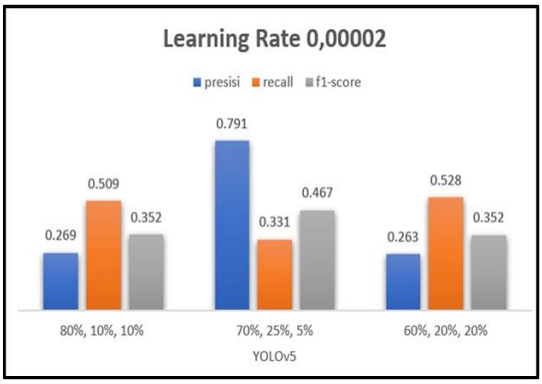

Figure 14. Graph of confusion value results from Precision, Recall, F1-Score at a learning rate of 0.00002.

Next is Fig. 14, which is the results of precision, recall and F1-Score at a learning rate of 0.00002. It can be seen that the biggest results are in the second dataset sharing scenario. The values obtained are all the same with the highest precision value of 0.791, with a recall of 0.331 and F1-Score of 0.467. Based on the appearance of Fig 12, Fig. 13 and Fig 14, it has the best results in the 0.002 scenario. Because the results obtained in each data division scenario are constant and have a very small difference value. While the smallest precision, recall and F1-Score

values are at a learning rate of 0.00002. This happens because the resulting value is low and inconsistent. So that it affects the accuracy of the results in predicting objects.

Testing Model YOLOv5.

This model detection test is carried out using images from one of the video datasets that have been extracted into photos from the VISEM dataset. This model detects sperm and non-sperm objects in the form of images so that the results of object checking by the model will be displayed based on the frames detected from the tested photos. Fig. 15 is a display of the best results with a learning rate of 0.002 detection from the YOLOv5 model that has been created.

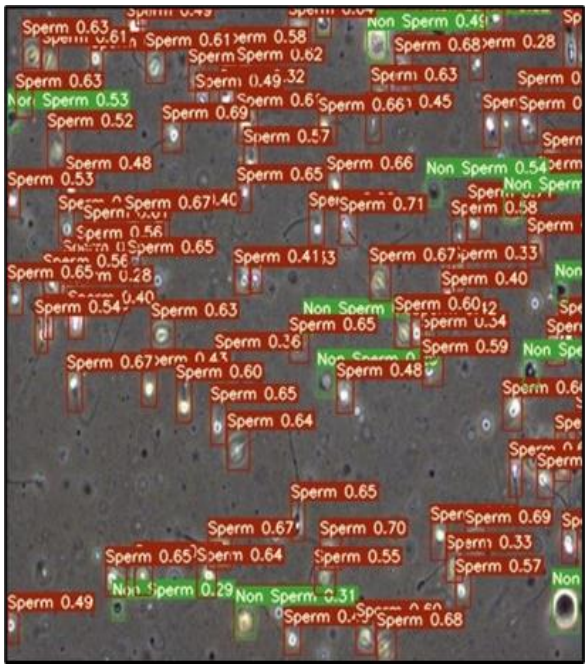

Figure 15. YOLOv5 Running Results

Analysis of Model Evaluation Results

Evaluation of the YOLOv5 model was carried out on the detection and classification process by calculating the mean Average Precision (mAP) and accuracy using a confusion matrix. The mAP calculation gives an idea of how good the model is at detecting and classifying objects across different scenarios, while the accuracy from the confusion matrix helps in assessing how precise the model is in distinguishing between different classes. Table 6 shows the highest average accuracy achieved by YOLOv5, which is an important indicator in assessing the effectiveness of this model in object detection and classification tasks.

Table 6. YOLOv5 model average accuracy results

No	Dataset division	AP		mAP
		Sperm	*Non Sperm*	
1	Train 60%, Validation 20%, Test 20%	82,9%	62%	72,4%
2	Train 70%, Validation 25%, Test 5%	83,7%	62,6%	73,1%
3	Train 80%, Validation 10%, Test 10%	82,7%	72,4%	72%

Based on the table analysis results, the YOLOv5 model shows consistent performance in recognizing Sperm objects with an average accuracy above 82% on various dataset shares. In contrast, the accuracy for Non-Sperm objects is significantly lower, ranging from 62% to 72.4%. Increasing the proportion of training data from 60% to 80% showed a slight increase in Non-Sperm accuracy but did not consistently improve the overall mAP. This indicates that the model can learn better with more training data, the balance between training, validation, and test data is crucial to achieve optimal generalization on new data. Furthermore, in the computational evaluation, it can be seen that the model requires a slightly longer computation time for training. Still, the overall mAP can be seen in the following Fig. 16.

Figure 16. YOLOv5 Computation Time Results

The figure above shows a bar graph depicting the computation time required by the YOLOv5 algorithm, labeled "COMPUTATION." From the graph, it can be seen that YOLOv5 requires 36 hours of computation time. YOLOv5 is a deep learning model used for object detection and is well-known for its speed and accuracy. However, the computation time of 36 hours shows that the training or inference process of this model is quite intensive. Some factors that may affect the length of computation time include the large dataset size, the hardware used, the complexity of the model, and the hyperparameters settings such as batch size, learning rate, and number of epochs. In this situation, using better hardware or adjusting the hyperparameters can help reduce the computation time.

LIMITATIONS

This study is limited to using the YOLOv5 algorithm to detect human sperm morphology. The dataset used is exclusively from the VISEM dataset, which consists of 85 movies divided into two categories: sperm and non-sperm. There will be no additional or changed datasets included in this study. The evaluation is mainly focused on specific performance criteria such as mean average precision (mAP), precision, and recall. This study is computational in nature, hence it excludes direct clinical testing on patients or the collecting of physical samples.

CONCLUSION

Based on the results, explanations and tests that have been carried out, there are several conclusions, namely in this study the results of the YOLOv5 model were obtained using 3 learning rates, namely 0.002, 0.0002, and 0.000002. As well as using 3 data sharing scenarios. Based on the scenarios and learning rates used, the biggest results obtained from each learning rate are 0.002 with an mAP of 73.1%, 0.0002 with an mAP of 65.6% and 0.00002 with an mAP of 40.2% with a computation time of 36 hours for the entire data training process. Therefore, in the context of this study, the YOLOv5 model is considered efficient in detecting and classifying objects, both sperm and non-sperm.

It is recommended that further research explore the application of this method by increasing the number of class variants in the objects used, such as detecting defective sperm and normal sperm using the latest version of the YOLO method.

AUTHOR CONTRIBUTIONS

We express our gratitude to all parties involved in this research for their valuable support and contributions. Special thanks to the Department of Computer Science, University of Lampung, for providing assistance with laboratory facilities and data analysis.

ACKNOWLEDGMENT

The experiment in this research used NVIDIA Tesla K80 and Tesla K20 provided by Department of Computer Science, University of Lampung.

REFERENCES

[1] K. K. Siu, V. H. B. Serrão, A. Ziyyat, and J. E. Lee, "The cell biology of fertilization: Gamete attachment and fusion," *J Cell Biol*, vol. 220, no. 10, p. e202102146, Aug. 2021, doi: https://doi.org/10.1083/jcb.202102146.

[2] X. Hu and B. Huang, "Face Detection based on SSD and CamShift," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Dec. 2020, pp. 2324–2328. doi: https://doi.org/10.1109/ITAIC49862.2020.9339094.

[3] C. Shalev, "Rights to Sexual and Reproductive Health: The ICPD and the Convention on the Elimination of All Forms of Discrimination against Women," *Health and Human Rights*, vol. 4, no. 2, pp. 38–66, 2000, doi: https://doi.org/10.2307/4065196.

[4] F. F. Rashidi, J. L. Grantner, and I. Abdel-Qader, "Box-Trainer Assessment System with Real-Time Multi-Class Detection and Tracking of Laparoscopic Instruments, using CNN," *Acta Polytechnica Hungarica*, vol. 19, no. 2, pp. 7–27, 2022, doi: https://doi.org/10.12700/APH.19.2.2022.2.1.

[5] Q.-C. Mao, H.-M. Sun, Y.-B. Liu, and R.-S. Jia, "Mini-YOLOv3: Real-Time Object Detector for Embedded Applications," *IEEE Access*, vol. 7, pp. 133529–133538, 2019, doi: https://doi.org/10.1109/ACCESS.2019.2941547.

[6] I. de Santiago and L. Polanski, "Data-Driven Medicine in the Diagnosis and Treatment of Infertility," *Journal of Clinical Medicine*, vol. 11, no. 21, Art. no. 21, Jan. 2022, doi: https://doi.org/10.3390/jcm11216426.

[7] K. A. L. Abdullah, T. Atazhanova, A. Chavez-Badiola, and S. B. Shivhare, "Automation in ART: Paving the Way for the Future of Infertility Treatment," *Reprod. Sci.*, vol. 30, no. 4, pp. 1006–1016, Apr. 2023, doi: https://doi.org/10.1007/s43032-022-00941-y.

[8] S. R., N. B.r., R. Radhakrishnan, and S. P., "Computational intelligence for early detection of infertility in women,"

Engineering Applications of Artificial Intelligence, vol. 127, no. 1, p. 107400, Jan. 2024, doi: https://doi.org/10.1016/j.engappai.2023.107400.

[9] S. A. Hicks *et al.*, "Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction," *Sci Rep*, vol. 9, no. 1, p. 16770, Nov. 2019, doi: https://doi.org/10.1038/s41598-019-53217-y.

[10] M. Dobrovolny, J. Benes, O. Krejcar, and A. Selamat, "Sperm-cell Detection Using YOLOv5 Architecture," in *Bioinformatics and Biomedical Engineering*, I. Rojas, O. Valenzuela, F. Rojas, L. J. Herrera, and F. Ortuño, Eds., Cham: Springer International Publishing, 2022, pp. 319–330. doi: https://doi.org/10.1007/978-3-031-07802-6_27.

[11] M. Dobrovolny, J. Benes, J. Langer, O. Krejcar, and A. Selamat, "Study on Sperm-Cell Detection Using YOLOv5 Architecture with Labaled Dataset," *Genes*, vol. 14, no. 2, Art. no. 2, Feb. 2023, doi: https://doi.org/10.3390/genes14020451.

[12] Z. Xiaoping, J. Jiahui, W. Li, H. Zhonghe, and L. Shida, "People's Fast Moving Detection Method in Buses Based on YOLOv5," *Int. J. Sens. Sens. Netw.*, vol. 9, no. 1, Art. no. 1, May 2021, doi: https://doi.org/10.11648/j.ijssn.20210901.15.

[13] Zhang Xiaoping, Ji Jiahui, Wang Li, He Zhonghe, and Liu Shida, "People's Fast Moving Detection Method in Buses Based on YOLOv5, International Journal of Sensors and Sensor Networks, Science Publishing Group," *International Journal of Sensors and Sensor Networks*, vol. 9, no. 1, pp. 30–37, 2021, doi: https://doi.org/10.11648/j.ijssn.20210901.15.

[14] P. R. Nasrullah *et al.*, "Effect of Ajwa date pits powder (Phoenix dactylifera L.) on body composition, lipid profile and blood pressure in patients with hyperlipidemia: A randomized clinical trial," *Avicenna J Phytomed*, vol. 13, no. 2, pp. 143–152, 2023, doi: 10.22038/AJP.2022.21316.

[15] R. Claßen-Bockhoff, "Plant Morphology: The Historic Concepts of Wilhelm Troll, Walter Zimmermann and Agnes Arber," *Annals of Botany*, vol. 88, no. 6, pp. 1153–1172, Dec. 2001, doi: https://doi.org/10.1006/anbo.2001.1544.

[16] S. N. Aliyah, H. Santoso, and H. Zayadi, "Analisis Normalitas dan Abnormalitas Spermatozoa Segar Sapi Limousin (Bos taurus) dan Sapi Bali (Bos sondaicus) Sebelum Proses Pembekuan Di Balai Besar Inseminasi Buatan Singosari Malang," *SCISCITATIO*, vol. 3, no. 1, Art. no. 1, Apr. 2022, doi: 10.21460/sciscitatio.2022.31.85.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd edition. Upper Saddle River, NJ: Pearson, 2007. Available: https://sde.uoc.ac.in/sites/default/files/sde_videos/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods-ilovepdf-compressed.pdf.

[18] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, Art. no. 1, May 2020, doi: https://doi.org/10.31294/ijcit.v5i1.795.

[19] Z. Z. R. Permana and I. Puspasari, "Echocardiogram Image Quality Enhancement using Upsampling and Histogram Matching Methods," *Jurnal Teknik Elektro*, vol. 14, no. 2, Art. no. 2, 2022, doi: https://doi.org/10.15294/jte.v14i2.42081.

[20] Y. Zhou, S. Liu, N. Wang, and Z. Zhao, "Leveraging an Asymmetric Convolutional Neural Network for Power System Object Intelligent Recognition," *J. Phys.: Conf. Ser.*, vol. 2777, no. 1, p. 012003, Jun. 2024, doi: https://doi.org/10.1088/1742-6596/2777/1/012003.

[21] S. Noor, O. Tajik, and J. Golzar, "Simple Random Sampling," *International Journal of Education & Language Studies*, vol. 1, no. 2, pp. 78–82, Dec. 2022, doi: 10.22034/ijels.2022.162982.

[22] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.

[23] Y. Guo *et al.*, "Improved YOLOV4-CSP Algorithm for Detection of Bamboo Surface Sliver Defects With Extreme Aspect Ratio," *IEEE Access*, vol. 10, Jan. 2022, doi: 10.1109/access.2022.3152552.

[24] S. K. Ahmed, "How to choose a sampling technique and determine sample size for research: A simplified guide for researchers," *Oral Oncology Reports*, vol. 12, p. 100662, Dec. 2024, doi: 10.1016/j.oor.2024.100662.

[25] John D. Kelleher, *Deep Learning*. The MIT Press, 2019. Accessed: Dec. 22, 2024, doi: https://doi.org/10.7551/mitpress/11171.001.0001.

[26] K. Rybczynski and Mustapha Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," *IEEE Access*, 2023, Accessed: Dec. 22, 2024. Available: https://ieeeaccess.ieee.org/featured-articles/cnnsandvision_objectdetection/.

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE COMPUTER SOCIETY. doi: https://doi.org/10.1109/CVPR.2016.91.

[28] W. S. Eka Putra, "Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101," *JTITS*, vol. 5, no. 1, Mar. 2016, doi: https://doi.org/10.12962/j23373539.v5i1.15696.

[29] R. N. Reinaldo and S. Dwiasnati, "Prediction of Customer Data Classification by Company Category Using Decision Tree Algorithm (Case Study: PT. Teknik Kreasi Solusindo)," *International Journal of Advanced Multidisciplinary*, vol. 2, no. 2, pp. 229–238, Jul. 2023, doi: https://doi.org/10.38035/ijam.v2i2.285.

[30] Z. N. Khudhair *et al.*, "Color to Grayscale Image Conversion Based on Singular Value Decomposition," *IEEE Access*, vol. 11, pp. 54629–54638, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3279734.

[31] Y. Egi, M. Hajyzadeh, and E. Eyceyurt, "Drone-Computer Communication Based Tomato Generative Organ Counting Model Using YOLO V5 and Deep-Sort," *Agriculture*, vol. 12, no. 9, Art. no. 9, Sep. 2022, doi: https://doi.org/10.3390/agriculture12091290.

[32] Y. Shen, D. Liu, F. Zhang, and Q. Zhang, "Fast and accurate multi-class geospatial object detection with large-size remote sensing imagery using CNN and Truncated NMS," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 191, pp. 235–249, Sep. 2022, doi: https://doi.org/10.1016/j.isprsjprs.2022.07.019.

[33] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 23, 2020, *arXiv*: arXiv:2004.10934. doi: https://doi.org/10.48550/arXiv.2004.10934.

[34] E. Alpaydın, *Introduction to Machine Learning*. The MIT Press, 2009. Accessed: Dec. 22, 2024. [Online]. Available: https://mitpress.mit.edu/9780262012430/introduction-to-machine-learning/.