COMPARATIVE ANALYSIS OF DIFFERENT MACHINE LEARNING MODELS FOR DETECTION AND CLASSIFICATION OF BULL SPERM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ANKARA UNIVERSITY

by

Mustafa AYDOGAN

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOMEDICAL ENGINEERING

ANKARA
2024

All rights reserved

ABSTRACT

Masters Thesis

COMPARATIVE ANALYSIS OF DIFFERENT MACHINE LEARNING MODELS FOR DETECTION AND CLASSIFICATION OF BULL SPERM

Mustafa AYDOGAN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Biomedical Engineering

Supervisor: Assist. Prof. Hasan Onur KELES

Researchers have employed machine learning models in a series of studies, specifically focusing on human fertility, to prevent potential issues leading to human infertility. However, there is a lack of research focusing on machine learning algorithms for the animal infertility. In this thesis, we investigated and compared the different machine learning approach for facilitating the sperm selection process in the animal infertility, especially bull infertility. Additionally, we examined the potential effects of the electric field on bull sperm cells. The relationship between the applied electric field on the bull sperm and the performance of machine learning algorithms is an important research topic. In this thesis, data were collected from bull sperm images obtained under different electric fields, and then the comparison and an assessment of the effectiveness of machine learning algorithms employed in classification and object detection was conducted. We observed significant differences based on performance between various algorithms under the various electric field.

January 2024, 94 pages

Key Words: machine learning, deep learning, convolutional neural networks, assisted reproductive technologies, electric fields

ÖZET

Yüksek Lisans Tezi

BOĞA SPERMİNİN TESPİTİ VE SINIFLANDIRILMASI İÇİN FARKLI MAKİNE ÖĞRENİMİ MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

Mustafa AYDOĞAN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Hasan Onur KELEŞ

Araştırmacılar, kısırlığa yol açan potansiyel sorunları önlemek için, özellikle insan doğurganlığına odaklanan bir dizi çalışmada, makine öğrenimi modellerini kullandılar. Ancak, hayvan kısırlığına odaklanan makine öğrenimi algoritmaları üzerinde yeterli sayıda araştırma bulunmamaktadır. Bu tezde, özellikle boğa kısırlığı üzerine sperm seçim sürecine yardımcı olmak amacıyla farklı makine öğrenimi yaklaşımlarını inceleyip karşılaştırdık. Ayrıca, boğa sperm hücreleri üzerinde elektrik alanının potansiyel etkileride incelendi. Uygulanan elektrik alanının boğa spermleri üzerindeki etkisi ile makine öğrenimi ve derin öğrenme algoritmalarının performansı arasındaki ilişki önemli bir araştırma konusudur. Bu tez kapsamında, farklı elektrik alanları altında elde edilen boğa sperm görüntülerinden veri toplandı. Ardından, sınıflandırma ve nesne tespiti için kullanılan makine öğrenimi algoritmalarının etkinliğini karşılaştırmak ve değerlendirmek amacıyla bir analiz gerçekleştirildi. Farklı elektrik alanları altında, çeşitli algoritmalar arasında önemli performans farklılıkları gözlemledik.

Ocak 2024, 94 sayfa

Anahtar Kelimeler: makine öğrenmesi , derin öğrenme , evrişimsel sinir ağları , yardımcı üreme teknolojileri , elektrik alanlar

FOREWORD AND ACKNOWLEDGEMENTS

I sincerely thank Assist. Prof. Dr. Hasan Onur Keles for his invaluable guidance and insightful contributions to developing the knowledge and perspectives presented in this thesis. Without his wise counsel and unwavering support, this thesis would not carry the same significance. I would like to thank Assist. Prof. Dr. Mehmet Yuksekkaya and Assoc. Prof. Dr. Koray Tekin for all support.

I also owe a special thank you to my dear brother, Mert Aydoğan for being a constant source of support throughout this challenging process. Your presence motivated me, and I felt your unwavering belief in my success. Sharing this journey with you has been a profound enrichment.

Lastly, I want to express my sincere thanks to my family and friends. Without your love and support, this thesis would not have been possible. I am truly grateful to all of you.

With respect

May God bless all of you.

Mustafa AYDOGAN
Ankara, January 2024

TABLE OF CONTENTS

THESIS APPROVAL
ETHIC i
ABSTRACT ii
ÖZET iii
FOREWORD AND ACKNOWLEDGEMENTS iv
ABBREVIATIONS vii
LIST OF FIGURES viii
LIST OF TABLES xii
1. INTRODUCTION 1
1.1 Animal Infertility 1
1.2 Sperm Selection / In vivo - In vitro 2
1.3 Effect of Electric Field on the Sperm 4
1.4 Machine Learning and Deep Learning in Sperm Selection 6
1.5 Purpose of Thesis 8
2. MATERIAL AND METHOD 10
2.1 Preparation of Sperm Sample 10
2.2 Preparation of Setup 10
2.2.1 Applying electric field to sperm 11
2.2.2 Image collection and sperm analysis software 11
2.2.3 Transfer of sperm data 12
2.3 Preparation of Data for Models 13
2.3.1 Image extraction from video 13
2.3.2 Preprocessing 14
2.3.3 Labeling data 16
2.3.3.1 Labeling data for object detection 16
2.3.3.2 Labeling data for classification 21
2.4 Determination of the Algorithms to Be Used and the Training Phase 24
2.4.1 YOLO algorithm 24
2.4.1.1 Training the model with YOLO 26
2.4.2 State of the art models and training phase 28
2.4.2.1 VGG16 model and training process 29
2.4.2.2 VGG19 model and training process 30
2.4.2.3 ResNet50 model and training process 30
2.4.2.4 ResNet101 model and training process 32
2.4.2.5 ResNet152 model and training process 32
3. RESEARCH FINDINGS 33
3.1 Application of Electric Field 33
3.2 Results of Classification Models 36
3.2.1 Results for VGG-16 38

3.2.2 Results for VGG-19 39
3.2.3 Result for ResNet50 40
3.2.4 Result for ResNet101 41
3.2.5 Result for ResNet152 42
3.3 Results for YOLO Models 44
3.3.1 Results of YOLO models for head-only labeled data 45
3.3.1.1 YOLOv5N 45
3.3.1.2 YOLOv5S 48
3.3.1.3 YOLOvM 50
3.3.1.4 YOLOv5L 53
3.3.1.5 YOLOv5X 55
3.3.2 Results of YOLO models for all sperm labeled data 59
3.3.2.1 YOLOv5N 59
3.3.2.2 YOLOv5S 61
3.3.2.3 YOLOv5M 64
3.3.2.4 YOLOv5L 66
3.3.2.5 YOLOv5X 69
3.3.3 Evaluation of YOLO models performance in different electric fields 73
3.3.3.1 Head-labeled model 73
3.3.3.2 All sperm labeled model 76
4. DISCUSSION 80
5.CONCLUSION 85
REFERENCES 86
CURRICULUM VITAE 94

ABBREVIATIONS

AVI	Auido Video Interleave
CASA	Computer-Assisted Sperm Analysis
DL	Deep Learning
IOU	Intersection Over Union
ML	Machine Learning
PBS	Phosphate-buffered saline
ResNet	Residual Networks
SCA	Sperm Class Analyzer
SOTA	State of the Art
VGG	Visual Geometry Group
YOLO	You Only Look Once

LIST OF FIGURES

Figure 2.1 Setup mechanism used for electric field application 11
Figure 2.2 An example of a report from SCA software 12
Figure 2.3 Sample sperm images 14
Figure 2.4 Data before preprocessing / Data after preprocessing 15
Figure 2.5 Data before preprocessing / Data after preprocessing 16
Figure 2.6 An example of the YOLO data label format and label file 17
Figure 2.7 An example of makesense.ai 18
Figure 2.8 Data labeled only head of sperm 18
Figure 2.9 Labeled data based on all sperm 19
Figure 2.10 Graph of labels and label sizes for fully labeled sperm data 19
Figure 2.11 Pairs chart of label sizes for fully labeled sperm data 20
Figure 2.12 Graph of labels and label sizes for head-labeled sperm 20
Figure 2.13 Graph of labels and label sizes for head-labeled sperm 21
Figure 2.14 Classification data for the normal class 22
Figure 2.15 Classification data for the abnormal class 22
Figure 2.16 Data produced using the method 1 23
Figure 2.17 Data produced using the method 2 23
Figure 2.18 Graph of labels and label sizes for head-labeled sperm 24
Figure 2.19 Architecture of YOLOv5 25
Figure 2.20 Architecture of VGG16 30
Figure 2.21 Architecture of ResNet50 31
Figure 3.1 Pie charts of sperm groups in percentages for sperm groups 34
Figure 3.2 Graph of speed values for each sperm in the fast progressing group 36
Figure 3.3 Basic of confusion matrix 37
Figure 3.4 Normalized and non-normalized confusion matrices for VGG16 39
Figure 3.5 Normalized and non-normalized confusion matrices for VGG19 40
Figure 3.6 Normalized and non-normalized confusion matrices for ResNet50 41
Figure 3.7 Normalized and non-normalized confusion matrices for ResNet101 42
Figure 3.8 Normalized and non-normalized confusion matrices for ResNet152 43
Figure 3.9 Comparison of accuracy values of state-of-the-art models 43

Figure 3.10 Visualization for IOU 44
Figure 3.11 Normalized Confusion matrix for YOLOv5N head labeled data 45
Figure 3.12 Results graph for YOLOv5N head labeled data 46
Figure 3.13 Precision Confidence Curve for YOLOv5N head labeled data 46
Figure 3.14 Precision-Recall Curve for YOLOv5N head labeled data 47
Figure 3.15 F1-Confidence Curve for YOLOv5N head labeled data 47
Figure 3.16 Normalized Confusion matrix for YOLOv5S head labeled data 48
Figure 3.17 Results graph for YOLOv5S head labeled data 48
Figure 3.18 Precision Confidence Curve for YOLOv5S head labeled data 49
Figure 3.19 Precision-Recall Curve for YOLOv5S head labeled data 49
Figure 3.20 F1-Confidence Curve for YOLOv5S head labeled data 50
Figure 3.21 Normalized Confusion matrix for YOLOv5M head labeled data 50
Figure 3.22 Results graph for YOLOv5M head labeled data 51
Figure 3.23 Precision Confidence Curve for YOLOv5M head labeled data 51
Figure 3.24 Precision-Recall Curve for YOLOv5M head labeled data 52
Figure 3.25 F1-Confidence Curve for YOLOv5M head labeled data 52
Figure 3.26 Normalized Confusion matrix for YOLOv5L head labeled data 53
Figure 3.27 Results graph for YOLOv5L head labeled data 53
Figure 3.28 Precision Confidence Curve for YOLOv5L head labeled data 54
Figure 3.29 Precision-Recall Curve for YOLOv5L head labeled data 54
Figure 3.30 F1-Confidence Curve for YOLOv5L head labeled data 55
Figure 3.31 Normalized Confusion matrix for YOLOv5X head labeled data 55
Figure 3.32 Results graph for YOLOv5X head labeled data 56
Figure 3.33 Precision Confidence Curve for YOLOv5X head labeled data 56
Figure 3.34 Precision-Recall Curve for YOLOv5X head labeled data 57
Figure 3.35 F1-Confidence Curve for YOLOv5X head labeled data 57
Figure 3.36 Normalized Confusion matrix for YOLOv5N all sperm labeled data 59
Figure 3.37 Results graph for YOLOv5N all sperm labeled data 59
Figure 3.38 Precision Confidence Curve for YOLOvN all sperm labeled data 60
Figure 3.39 Precision-Recall Curve for YOLOv5N all sperm labeled data 60
Figure 3.40 F1-Confidence Curve for YOLOv5N all sperm labeled data 61
Figure 3.41 Normalized Confusion matrix for YOLOv5S all sperm labeled data 61

Figure 3.42 Results graph for YOLOv5S all sperm labeled data....62
Figure 3.43 Precision Confidence Curve for YOLOvS all sperm labeled data 62
Figure 3.44 Precision-Recall Curve for YOLOv5S all sperm labeled data 63
Figure 3.45 F1-Confidence Curve for YOLOv5S all sperm labeled data....63
Figure 3.46 Normalized Confusion matrix for YOLOv5M all sperm labeled data 64
Figure 3.47 Results graph for YOLOv5M all sperm labeled data 64
Figure 3.48 Precision Confidence Curve for YOLOvM all sperm labeled data 65
Figure 3.49 Precision-Recall Curve for YOLOv5M all sperm labeled data....65
Figure 3.50 F1-Confidence Curve for YOLOv5M all sperm labeled data 66
Figure 3.51 Normalized Confusion matrix for YOLOv5L all sperm labeled data 66
Figure 3.52 Results graph for YOLOv5L all sperm labeled data 67
Figure 3.53 Precision Confidence Curve for YOLOvL all sperm labeled data 67
Figure 3.54 Precision-Recall Curve for YOLOv5L all sperm labeled data 68
Figure 3.55 F1-Confidence Curve for YOLOv5L all sperm labeled data....68
Figure 3.56 Normalized Confusion matrix for YOLOv5N all sperm labeled data....69
Figure 3.57 Results graph for YOLOv5X all sperm labeled data....69
Figure 3.58 Precision Confidence Curve for YOLOvN all sperm labeled data....70
Figure 3.59 Precision-Recall Curve for YOLOv5N all sperm labeled data....70
Figure 3.60 F1-Confidence Curve for YOLOv5N all sperm labeled data....71
Figure 3.61 Comparison for YOLOv5N....73
Figure 3.62 Comparison for YOLOv5S....74
Figure 3.63 Comparison for YOLOv5M....74
Figure 3.64 Comparison for YOLOv5L....74
Figure 3.65. Comparison for YOLOv5X....75
Figure 3.66 Comparison of average detected sperm numbers for 12V and Control Group....75
Figure 3.67 Comparison of average confidence for 12V and Control Group....76
Figure 3.68 Comparison for YOLOv5N....76
Figure 3.69 Comparison for YOLOv5S....77
Figure 3.70 Comparison for YOLOv5M....77
Figure 3.71 Comparison for YOLOv5L....77
Figure 3.72 Comparison for YOLOv5X....78

Figure 3.73 Comparison of average detected sperm numbers for 12V and Control Group 78
Figure 3.74 Comparison of average confidence for 12V and Control Group..... 78

LIST OF TABLES

Table 2.1 System features used for training 26
Table 2.2 Parameters used for training 29
Table 3.1 Results of velocity and progressivity on different groups 33
Table 3.2 Percentage of rapid progressive sperm 35
Table 3.3 Speed values for each sperm in the rapid progressive group 35
Table 3.4 Results for VGG16 39
Table 3.5 Results for VGG19 39
Table 3.6 Results for ResNet50 40
Table 3.7 Results for ResNet101 41
Table 3.8 Results for ResNet152 42
Table 3.9 The various YOLOv5 models exhibit distinct evaluation metrics and corresponding values. 72

1. INTRODUCTION

1.1 Animal Infertility

According to data from the Food and Agriculture Organization (FAO), the livestock industry plays a substantial role, contributing to almost 40% constituting about 80% of the total agricultural output in developed nations and roughly 20% in less economically advanced nations. Supporting the livelihoods of at least 1.3 billion people globally, this crucial sector contributes approximately 33% to the global food protein supply (Herrero et al., 2013; Yitbarek, 2020). In the livestock industry, the ability of animals to reproduce effectively is essential for maximizing productivity, maintaining sustainability, and promoting genetic diversity, all of which are vital elements for success in this industry (H.-Q. Wang et al., 2022). Therefore, the livestock industry significantly contributes to the economies of developing countries, forming a significant part of their overall economic structure (Sahoo et al., 2021).

Infertility refers to a transient disruption in the reproductive processes, preventing the animal from achieving pregnancy (Givens, 2006). Fertility challenges have persisted across generations, impacting various species, including humans and farm animals. Despite improvements in diagnostic methods and infertility treatments, data comparing 1990 and 2010 reveal marginal changes in human infertility rates, with a slight dip in primary (0.1%) and a low rise in secondary infertility (0.4%) (Mascarenhas et al., 2012). While comprehensive studies in farm animals are limited, available information suggests a growing prevalence of subfertility in livestock. This trend poses risks to animal well-being and farm economies due to extended intervals between births and the escalated removal of animals (Dobson et al., 2008). In the world of dairy farming, there has been a drop in the initial service pregnancy rate, going from 70% to 40% over the past 30 to 50 years. Infertility results from a combination of factors in both males and females, with male factor infertility accounting for 40–50% of case (Kumaresan et al., 2020). Various factors are related to infertility in animals, including inadequate care and nutrition, the presence of infectious diseases, unsuccessful fertilization, and instances of embryonic deaths, all of which contribute significantly to this condition (Kaltungo & Musa, 2013).

The artificial insemination (AI) sector places considerable emphasis on the financial implications of bull reproductive capabilities. While there have been substantial efforts to focus on and improve female fertility, the enhancement of bull fertility, unfortunately, has received less attention. According to the National Association of Animal Breeders (NAAB, 2018), approximately targeted collection involved 3.7 million units of reproductive fluid from male bovines., with unit of 4 million being dispensed across the United States in 2018. In that particular year, USA shipped bull semen unit of 3.9 million internationally , amounting to a total revenue of over $13 million(Butler et al., 2020).The most recent study conducted by the NAHMS revealed that 7.62% of producers involved in beef cattle had adopted the technology of Artificial Insemination (AI). Reasons cited for this low adoption rate included not just the investment of time and workforce but also the cost (USDA, 2009)(Butler et al., 2020). Improving bull fertility through strategic selection and management not only enhances conception rates but also positively impacts various economically significant production traits. Genetic correlations have been observed between bull fertility and attributes like average daily growth, heifer gestation, and calving span(Butler et al., 2020). At this stage, bull fertility stands out as a crucial factor in the context of the livestock industry(Park et al., 2013). Many cases of reproductive issues in livestock stem from bulls having problems with low semen amount, bad semen quality, or poor health.

We focused on the bull sperm instead of human sperm in this thesis. Bull sperm can be produced under more controlled conditions in a laboratory environment. This allows standardization and better management of experimental conditions, thus increasing the reliability and reproducibility of the data obtained (Perry, 2021). Bull sperm can be produced in larger quantities and more effectively than human sperm. This implies the capacity to better respond to the needs of large-scale experiments or applications, making it possible to obtain research data on a wide scope (Özbek et al., 2021).

1.2 Sperm Selection / In vivo - In vitro

Sperm selection for Assisted Reproductive Technologies (ART) involves the meticulous identification and isolation of spermatozoa with optimal characteristics to enhance the

chances of successful fertilization. The ejaculate comprises a diverse sperm population, including some that are unsuitable for fertilization and may not survive the female reproductive tract's selection process. Since only a fraction of the ejaculate is used in assisted reproduction, a high presence of abnormal or immature sperm can reduce the likelihood of a successful outcome. In assisted reproduction for animals, selecting spermatozoa with fertility-associated traits involves assessing normal morphology, acrosome integrity, motility, and chromatin integrity. Despite techniques like swim-up used to prepare sperm for in vitro fertilization, the quantity obtained is often insufficient for artificial insemination in most livestock species (Morrell, 2020). In cases where natural fertilization does not occur under in vivo conditions, in vitro ICSI (Intracytoplasmic Sperm Injection) and IVF (in-vitro fertilization) techniques can make fertilization possible (Vaughan & Sakkas, 2019). Sperm cells consist primarily of two distinct components: the head and the flagellum. The primary function of a sperm cell is to have a significant impact on the process of fertilization. Notably, there exists considerable diversity among sperm cells, encompassing variations in head size, tail length, and other characteristics. This diversity has the potential to influence the functionality and effectiveness of sperm cells (Teves & Roldan, 2022).

The process of sperm selection involves various techniques, such as DNA fragmentation analysis, gradient separation, hyaluronic acid binding, microfluidics, as well as evaluations based on morphology and motility. Each of these methods yields distinct outcomes in the selection of sperm for assisted reproductive procedures (Vaughan & Sakkas, 2019). These techniques are essential for isolating sperm cells from plasma in the context of in vitro fertilization (IVF).

The evaluation of sperm cells fundamentally determines the assessment of sperm quality as a measure of fertilization rate. In addition to traditional microscopic methods these evaluations also include various computer-aided assessment software known as "CASA."(Tanga et al., 2021). In the late 1980s, the development of Computer-Aided Sperm Analysis (CASA) technology was directed towards scrutinizing the locomotion characteristics or kinematics of sperm. CASA has demonstrated significant success in not only analyzing sperm movement but also in measuring various semen characteristics, like

sperm concentration and the percentages of progressive motility. This versatile technology has found widespread application across numerous animal species, playing a crucial role in laboratories dedicated to domesticated animal production and studies on reproductive toxicology (Mortimer et al., 2015). These types of software assess different parameters, including the tail beat frequency and speed of sperm cells. This study will involve a process based on morphological assessment. During the morphological assessment of sperm cells, there are parameters generally accepted by the World Health Organization (WHO). The acrosome, neck, midpiece, and tail segments of the sperm cell should have specific proportions with each other, covering certain measurements in the head and tail regions. These criteria are called as "Kruger criteria"(World Health Organization, 2010).

1.3 Effect of Electric Field on the Sperm

An electric field is generated by voltage differences, while a magnetic field arises from the flow of electric current. The general public is routinely exposed to electromagnetic fields (EMFs) in their daily lives, with increased prevalence due to modern technologies. Activities like using electricity, microwaves, and cell phones contribute to EMF exposure. Additionally, medical procedures such as X-rays, MRI, and ultrasounds introduce another set of potential sources for EMFs. The literature is very limited for the effect of electric field on the sperm. Most studies focused on the impact of electric fields on testis physiopathology, sperm parameters and DNA integrity in live human and animals. This study reports the existence of certain frequency windows for the resonance of the effects of the MF on human spermatozoa. Rapid motility was significantly affected by the exposure of spermatozoa to MF, but sperm structural parameter had remained intact (Górski et al., 2020)(Aslankoc et al., 2018).

The study by Zalata et al. (2015) aimed to investigate the effects of in vitro exposure to radiofrequency electromagnetic fields (RF-EMF) emitted from mobile phones on seminal clusterin (CLU) gene expression, sperm DNA fragmentation, and sperm motility index. Their findings revealed a significant decrease in sperm motility, linear velocity, and the sperm linearity index, alongside a notable increase in sperm DNA fragmentation (Zalata

et al., 2015). Similarly, Górski et al. (2020) conducted research to assess the impact of extremely low-frequency electromagnetic fields (ELF-EMF) at 50 Hz on human sperm motility. Their study discovered that ELF-EMF had a detrimental effect on sperm movement, with a significant decrease in spermatozoa motility speed and a notable increase in lateral head deviation values observed upon exposure to the electromagnetic field (Górski et al., 2020). Additionally, in a separate investigation by Ahmadi et al. (1997), electric fields were applied to human sperm cells, resulting in a significant rise in the proportion of semen undergoing acrosome loss (Ahmadi et al., 1997). In another study, Wu et al. (2017) examined the impact of long-term exposure to a Statical Electric Field (SEF) on the reproductive capability of male mice. They subjected a group of mice to environmental static electric field (SEF) using a high-voltage device, evaluating various parameters including sperm morphology and motility. However, no notable variations were observed among the group exposed to the environmental static electric field (SEF) and the other groups (Wu et al., 2017) . Moreover, in the study by Wagenaar et al. (2016), Segerink's team pioneered sperm assessment technology by developing a microfluidic system to trap and analyze individual sperm. Utilizing electric impedance monitoring, they investigated the flagellar beat frequency of sperm and its response to various stimuli. This groundbreaking work holds promise for advancing our understanding of sperm motility and could have significant implications for fertility assessments and reproductive health (Wagenaar et al., 2016).

The advancement of digital technologies has not only automated numerous established laboratory assays but has also led to the creation of innovative methods for evaluating semen sample quality. This automation has greatly streamlined laboratory processes, enhancing both efficiency and reproducibility. Furthermore, cutting-edge computer technologies have facilitated the design and implementation of novel laboratory analyses tailored specifically for assessing semen sample quality. This integration of technological advances into laboratory practices represents a significant leap forward, offering improved precision, efficiency, and a more comprehensive analysis of semen samples (Mocé & Graham, 2008). In a study conducted by Ariagno et al. (2017), Computer-aided sperm analysis (CASA) was utilized to examine sperm assays of infertile men with varicocele after varicocelectomy, aiming to assess any improvements in sperm

characteristics (Ariagno et al., 2017). The behavior of bull sperm cells can exhibit variations under the influence of different physical factors. Research on how sperm cells respond to different physical stimuli, whether there are any changes in their velocities, tail beat frequencies, and the potential positive outcomes of these effects, is an area of investigation. At this point, it is worth considering how the performance of machine and deep learning algorithms might be affected, despite the possibility of positive results from the mentioned influences, when evaluating them together.

1.4 Machine Learning and Deep Learning in Sperm Selection

Machine learning, a subset of artificial intelligence, enables computers to extract insights from accumulated experiences (Raschka et al., 2020).It possesses the capacity to manipulate data through self-generated algorithms and generate predictions (Janiesch et al., 2021). In the human context, this entails the capability to acquire knowledge autonomously, relying on accumulated experiences without external instructions (Parisi et al., 2019). It fundamentally revolves around the emulation and refinement of the human learning process (Deng et al., 2020). Machine learning, with its various types, has numerous and diverse applications. It can be categorized under three main algorithms: unsupervised, supervised, and reinforcement learning (Janiesch et al., 2021). Each method has specialized areas where the algorithm excels and delivers better results (LeCun et al., 2015). Face and voice recognition, autonomous vehicles, marketing, healthcare, education, language learning, sports, social media, gaming, and many other fields and purposes utilize machine learning (Sharma et al., 2021) (Patil et al., 2021). Deep learning, a subset of machine learning, represents an advanced approach in machine learning allows computers to autonomously analyze, understand, and extract valuable insights from unprocessed data. The outcomes derived from deep learning surpass those achieved by conventional machine learning methods (Chauhan & Singh, 2018). Deep learning has been applied across diverse research domains, including medicine, language recognition, text document processing, and more, owing to its capacity to effectively manage large-scale datasets (You et al., 2021).

The preference for deep learning in sperm analysis is justified by several key factors. Firstly, the inherent complexity of sperm data, which is challenging for traditional methods to analyze effectively, finds a suitable solution in the flexibility of deep learning algorithms (Janiesch et al., 2021). Additionally, deep learning models excel in automatically extracting meaningful features from datasets, proving particularly advantageous in discerning crucial information in sperm analysis (LeCun et al., 2015). Transfer learning capabilities further enhance the applicability of deep learning, enabling the transfer of acquired knowledge from one domain to another, thereby proving effective even with smaller datasets (Weiss et al., 2016). Moreover, the high performance exhibited by deep learning models, especially when trained on large datasets, contributes to the reliability and accuracy of results in processing extensive sperm datasets (Najafabadi et al., 2015).

In the process of sperm selection, the assessment of sperm morphology stands out as a significant parameter. Conventional methods for sperm selection and evaluation of morphology of sperm pose numerous challenges, such as unpredictability, time constraints, subjective interpretation, inaccuracy, and more. In the current context, machine learning and deep learning have the capability to alleviate all these disadvantages by facilitating more objective and swift selections (Iqbal et al., 2020; You et al., 2021).There are many studies on human sperm selection conducted using deep learning and machine learning algorithms. In the study conducted by (Riordon et al., 2019) ,to decrease infertility rates, the classification of human sperm cells has been carried out using the VGG algorithm. Proposed method can classify sperm cells at high accuracy level. In the study conducted by Michael and his team (Dobrovolny et al., 2023), a comparative analysis of deep learning algorithms was carried out with the aim of reducing subjective selection on human sperm cells. Potential problems that may arise during sperm selection were thoroughly addressed, and the development of deep learning algorithms aimed at solving these issues was the primary objective. Rather than classification, the focus was on direct object detection on human sperm. The proposed method demonstrated the capability to achieve high precision values. Javadi et al. suggested an approach for the identification of malformations in sperm morphology using images of human sperm cells. A deep neural network architecture was developed and

subsequently trained to identify abnormalities in various aspects of human sperm. It is important to highlight that the proposed algorithm demonstrates higher accuracy compared to current state-of-the-art methods in the detection of abnormalities (Javadi & Mirroshandel, 2019) . McCallum et al. demonstrated the training of DNN using roughly 1000 sperm cells with confirmed DNA quality, allowing precise prediction of DNA quality solely from direct illumunation images. The findings suggest a moderate association between the depiction of a sperm cell and its DNA quality, as well as the ability to differentiate cells with superior DNA integrity (McCallum et al., 2019).

Artificial intelligence algorithms have primarily been applied to the study of human sperm, with limited research extending to the analysis of animal sperm. Further exploration in this area is warranted for a comprehensive understanding of the broader implications and potential applications of AI in reproductive biology. For instance in the study conducted by Priyanto and colleagues (Hidayatullah et al., 2021), bull sperm cells were utilized. This research aimed to identify the reasons for potential errors during motility measurements of sperm cells employing Computer-Assisted Sperm Analysis softwares (CASA). This study performs sperm motility classification using Support Vector Machine (SVM) with specific CASA parameters. The proposed method showed higher accuracy values than other methods. Given the significant issue of infertility in farm animals, as previously emphasized, the use of machine learning and deep learning algorithms has the potential to actively contribute to efforts aimed at reducing infertility in livestock. Further investigation and practical application of these computational approaches are essential for a more thorough understanding and effective implementation in enhancing reproductive outcomes for agricultural animals.

1.5 Purpose of Thesis

The aim of this thesis is to examine and compare the effectiveness of various machine learning and deep learning models in classifying and detecting of bull sperm cells. This study specifically focuses on bull sperm behavior in electric fields. We compared various deep learning algorithm performances including YOLO, ResNet and VGG for sperm selection. While daily life electric fields' effects on human sperm have been studied

extensively, limited research exists on individual bull sperm. This study significantly contributes to the field of sperm selection by introducing a artificial intelligence-based approach for detection of sperm cells. The study presents a dataset comprising "73" video recordings, each lasting 50 seconds and containing 951 frames, specifically focusing on bull sperm. The thesis puts forth a more streamlined method for sperm selection in both IVF and ICSI reproductive phases.

Sperm selection is vital not just in humans but also in farm animals. Among the processes involved, classifying sperm cells poses one of the most challenging aspects. Evaluating sperm motility through manual observation using a microscope is a demanding task due to the rapid movement of spermatozoa within the visual field. (Thambawita et al., 2023)

Machine learning algorithms has the capacity to provide support embryologists and clinicians in the process of sperm selection. The classification performances of each machine learning algorithm may vary, presenting an opportunity for optimization. Enhancing these performances can be achieved through the application of an electric field. Given the rapid movement of sperm cells, tracking high-quality sperm cells can pose challenges for clinicians. Introducing a low-rate electric field to the sperm cells can reduce their velocity, facilitating a more manageable classification process for machine learning algorithms. This approach aims to improve the efficiency of sperm selection in reproductive procedures.

2. MATERIAL AND METHOD

This study specifically chose bull sperm as the target species for data collection. The sperm sample is commercially available for public use. The ethical approval document for the utilization of sperm cells and the execution of experiments has been obtained from the Ankara University Local Ethics Committee for Animal Experiments. The approval number is 2023-12-110.

2.1 Preparation of Sperm Sample

In this thesis, the bull sperm was used. The volume of the sampled sperm was determined as 10 microliters, and then 30 microliters of Phosphate Saline Solution (PBS) was added to dilute the sperm. This process aims to place the sample in a more homogeneous environment and bring it to a certain concentration. PBS was chosen as the added solution because it is a commonly used buffer solution to maintain and stabilize cells. This dilution process represents an important step to ensure working und more controllable conditions in the laboratory environment.

2.2 Preparation of Setup

In the course of the experimental setup preparation, a system was devised, facilitating the acquisition of sperm images through the Computer-Assisted Sperm Analysis (CASA) system and the facile application of an electric field. The experimental configuration consisted of two plates strategically positioned on either side of the sperm cells. Electric voltages of 3V, 6V, and 12V were sequentially applied to the plates with positive and negative charges. The plates were positioned with a 60mm gap between them.The acquisition of images was accomplished through the segmentation of videos into individual frames.

2.2.1 Applying electric field to sperm

Three different levels of electric fields, 50 V/m , 100 V/m and 200 V/m, were applied to the sperm sample through two plates. In this system, the distance between two plates is determined as 60mm (Fig. 2.1).

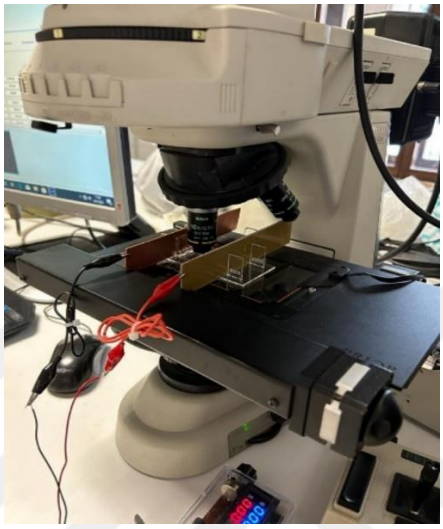
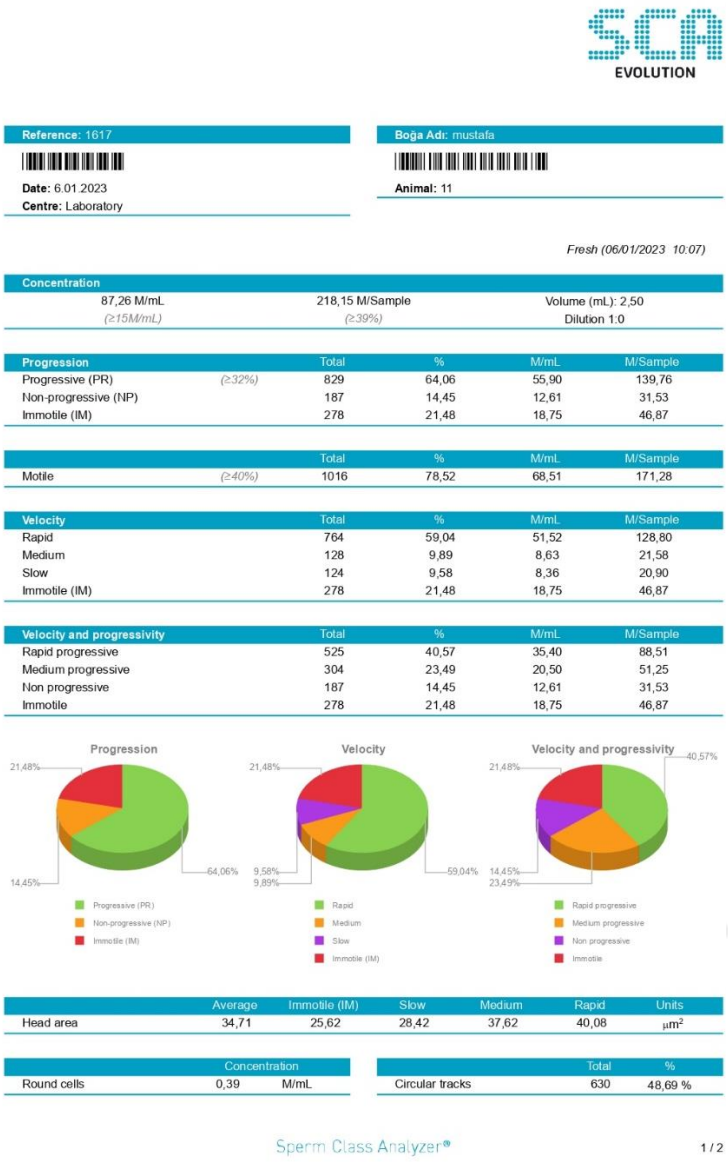


Figure 2.1 Setup mechanism used for electric field application

2.2.2 Image collection and sperm analysis software

SCA Evolution software has been used effectively as a Computer Aided Semen Analysis (CASA) program. This software includes the SCA® CASA System, which provides accurate, reproducible and automatic evaluation of a variety of important sperm parameters. These parameters include sperm motility, concentration, morphology, DNA fragmentation, viability, acrosome reaction and leukocytes (Mortimer et al., 2015; Tanga et al., 2021) Sperm videos were collected as 50-second videos using the microscope's own video recording software.

A report obtained especially in an environment where a 3V electric field is applied serves as an example of the data provided by the system. This sample report, along with value outputs such as rate and progress, can contain important information for understanding the effects of electromagnetic interactions on sperm properties (Fig. 2.2).



SCA EVOLUTION

Reference: 1617

Date: 6.01.2023
Centre: Laboratory

Boğa Adı: mustafa

Animal: 11

Fresh (06/01/2023 10:07)

Concentration

87,26 M/mL	218,15 M/Sample	Volume (mL): 2,50
(≥15M/mL)	*(≥39%)*	Dilution 1:0

Progression		Total	%	M/mL	M/Sample
Progressive (PR)	*(≥32%)*	829	64,06	55,90	139,76
Non-progressive (NP)		187	14,45	12,61	31,53
Immotile (IM)		278	21,48	18,75	46,87

		Total	%	M/mL	M/Sample
Motile	*(≥40%)*	1016	78,52	68,51	171,28

Velocity	Total	%	M/mL	M/Sample
Rapid	764	59,04	51,52	128,80
Medium	128	9,89	8,63	21,58
Slow	124	9,58	8,36	20,90
Immotile (IM)	278	21,48	18,75	46,87

Velocity and progressivity	Total	%	M/mL	M/Sample
Rapid progressive	525	40,57	35,40	88,51
Medium progressive	304	23,49	20,50	51,25
Non progressive	187	14,45	12,61	31,53
Immotile	278	21,48	18,75	46,87

	Average	Immotile (IM)	Slow	Medium	Rapid	Units
Head area	34,71	25,62	28,42	37,62	40,08	μm^2

	Concentration	
Round cells	0,39	M/mL

	Total	%
Circular tracks	630	48,69 %

Sperm Class Analyzer®

1 / 2

Figure 2.2 An example of a report from SCA software

The camera's own video software was used directly to collect the sperm image, and videos were recorded.

2.2.3 Transfer of sperm data

Sperm videos of 50 seconds each were collected and transferred in AVI (Audio Video Interleave) format for analysis and storage purposes. The AVI format was preferred because it provides wide multimedia player support and stores video data without

compression. This preserves the original quality of the videos, ensuring they are kept in a suitable format for detailed analysis. Additionally, the wide availability of the AVI format simplifies the research process by increasing data compatibility across different analysis tools and software. (Malaga & Koppel, 2017; Situmorang & Panggabean, 2021)

2.3 Preparation of Data for Models

The data preparation process involves three essential steps: extracting images from videos, preprocessing, and labeling. The images were obtained by utilizing VLC software, processed frame-by-frame to enhance the quality of data. The preprocessing stage is deemed essential for achieving improved results, and additional software tools were employed for this purpose.

2.3.1 Image extraction from video

Training deep learning models for sperm analysis requires having sufficient and useful data to begin with. The dataset employed in this study consists of images obtained from sperm videos obtained through the VLC program. Each video was separated by 10 frames per second (FPS), and in this way, video length (seconds) x 10 sperm images were obtained for each video.

This data preparation process is an important step to increase and generalize The cognitive capacity of the deep learning framework. Separating the video into frames allows analyzing the movement and morphological characteristics of each sperm cell in more detail. Choosing the FPS value of 10 optimizes the amount of data processed while maintaining sufficient separation and information density.

The benefits of this data preparation process include increasing the generalization ability by increasing diversity during the training, better capturing the dynamic movements of sperm cells, and contributing to the model producing more reliable results. Additionally, obtaining 10 sperm images each video length x allows the model to be trained on more

samples in a learning set, which can help the model better understand various situations and improve overall sperm analysis performance.

The primary objective of this data preparation approach is to increase the accuracy and reliability in the sperm analysis process by making the deep learning model more specific and effective. This allows the model to be applied more successfully in real-world conditions and to produce more accurate results in reproductive health assessment.

Two examples of the data produced at the end of the above-mentioned process are presented in Fig.2.3.

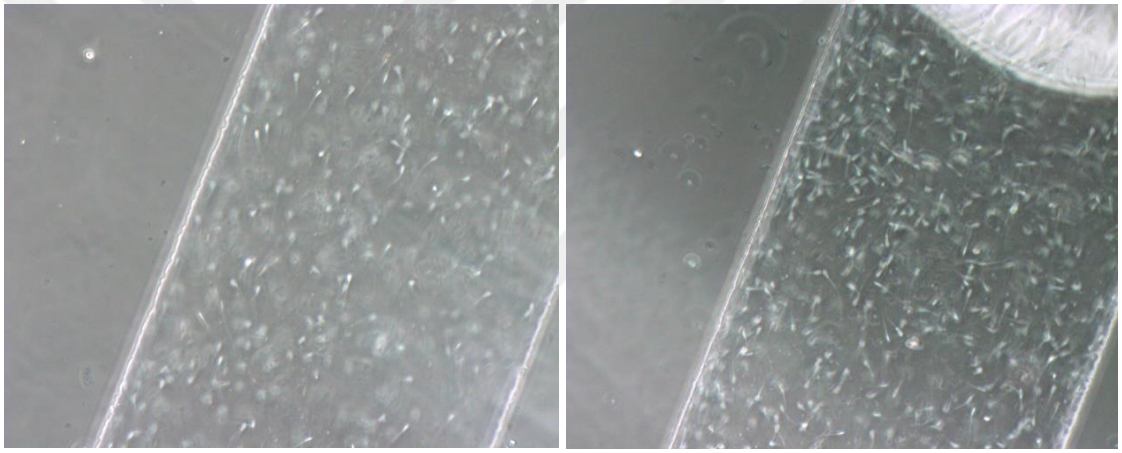


Figure 2.3 Sample sperm images

2.3.2 Preprocessing

Data preprocessing stands as a phase applied within the realm of both machine learning and data analytics applications and refers to the process of transforming the available data into a more meaningful, useful or more suitable format for the model.

Data preprocessing stands out as a crucial phase in the knowledge exploring process, demanding increased time and struggle throughout the entirety of the data analysis procedure. Numerous techniques exist for data preprocessing, encompassing activities such as data cleansing, data amalgamation, data reduction . Data preprocessing involves

correcting inconsistencies and removing noise through cleaning, integrating data from various sources, reducing redundancy, and scaling data for easier processing. These techniques are primarily based on programming models (Hassanien & Darwish, 2021). For example, preprocessing steps performed on sperm data may include: increasing contrast to reduce background noise, performing image corrections to obtain a homogeneous background, balancing image illumination with techniques such as histogram equalization, and resizing if necessary. These steps allow the model to focus on the clear and meaningful features necessary to identify and learn about sperm cells more accurately. At the same time, Preprocessing has made a great contribution to the understanding of sperm type for the taggers during the data tagging process.

During the data preprocessing process, contrast and highlights were increased by 100%, while the exposure parameter was reduced by 100%.These data preprocessing steps allows the model to be trained in a more robust, reliable and highly generalizable way. Additionally, data brought into a predetermined standard format can facilitate the process of combining data from different sources and performing a consistent analysis. Data preprocessing is a critical step to achieve successful results in machine learning projects.

Examples of images created before and after applying preprocessing to the data are given in Fig. 2.4 and Fig 2.5.

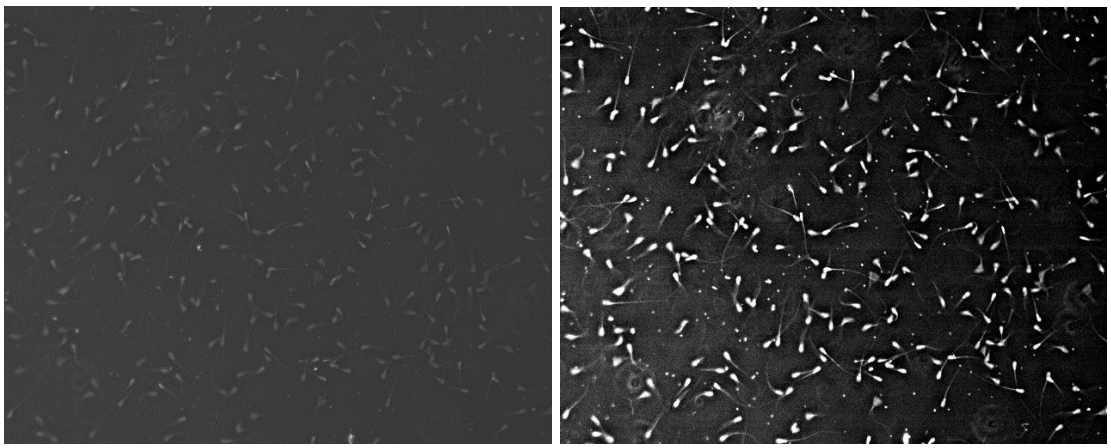


Figure 2.4 Data before preprocessing / Data after preprocessing

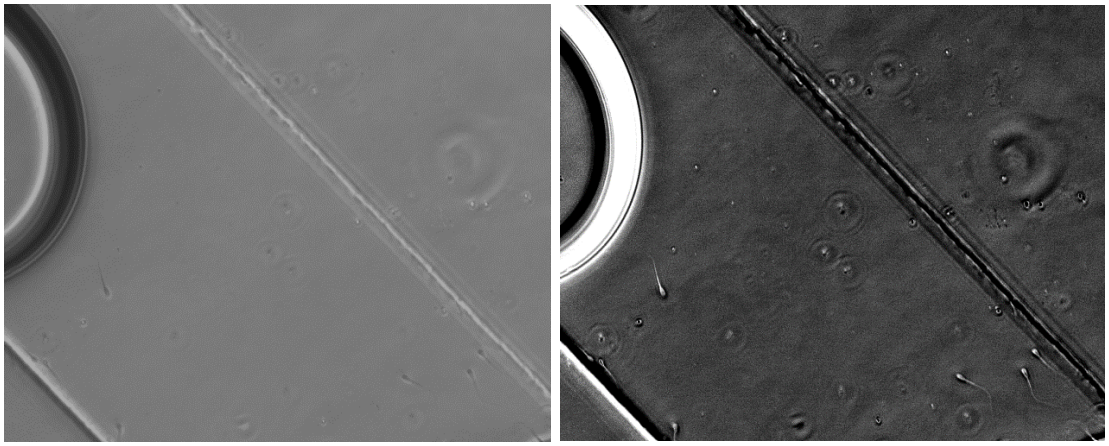


Figure 2.5 Data before preprocessing / Data after preprocessing

2.3.3 Labeling data

Data must be labeled in order for deep learning models to select sperm visually (Object Detection) and assign it to the correct class (Classification). The data labeling process was carried out based on the standardization parameters outlined in (Perry, 2021). The study conducted in Australia aims to enhance the reliability of sperm evaluations and reveal more meaningful sperm assessments.

Data labeling is a process by which human labelers add meaning to data sets used to train machine learning models. In this process, elements in the data set are tagged to be associated with a specific task. For example, if a dataset is used to train an image recognition model, labels representing specific objects or features are added to each image. These labels helps the machine learning model produce more accurate and meaningful results by enabling the model to learn specific classes or categories. Data labeling is a necessary process in supervised machine learning algorithms.

2.3.3.1 Labeling data for Object Detection

YOLO (You Look Only Once) Model is a specialized deep learning algorithm for multiclass object detection. Unlike traditional deep learning algorithms, YOLO works by dividing an image into multiple regions at the same time and predicting the class and location of objects in these regions. In this way, it is aimed to reduce working speed and

cost. It is considered a really fast method for detecting the target object and assigning it to the correct class using a single deep learning network. (Redmon et al., 2016)

The data labeling format of the YOLO algorithm is as follows:

The .txt file, which will contain the labels of the objects in the image, should contain a line for each object and its format should be class, x_center, y_center, width, height]. Box coordinates must be in xywh format, normalized between 0 and 1. If the boxes are in pixels, x_center and width should be divided by the image width, and y_center and height should be divided by the image height. Class numbers must be indexed starting from zero. For example, the label file in Figure 2.6 should be like the one on the right.

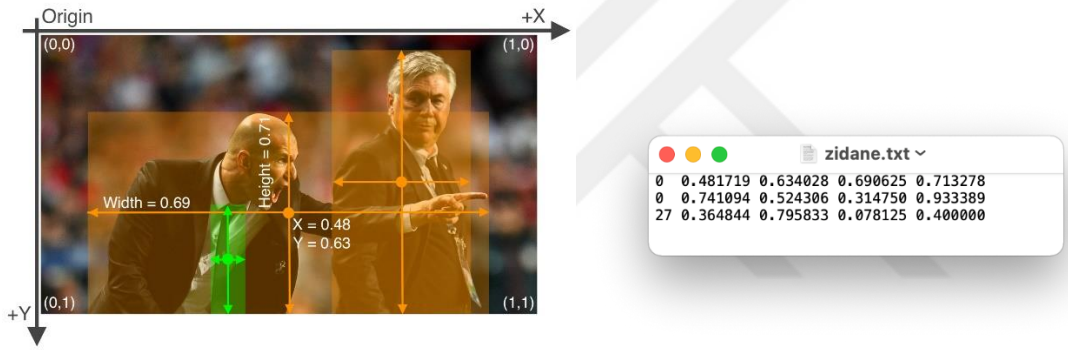


Figure 2.6 An example of the YOLO data label format and label file (Ultralytics, t.y.-c)

Here, index 0 represents the human class and 27 represents the tie class.

The data labeled for the YOLO algorithm was labeled using the labeling tool called “makesense” (https://makesense.ai). This tool is a customized online tool for quickly labeling data and obtaining output in the desired format.

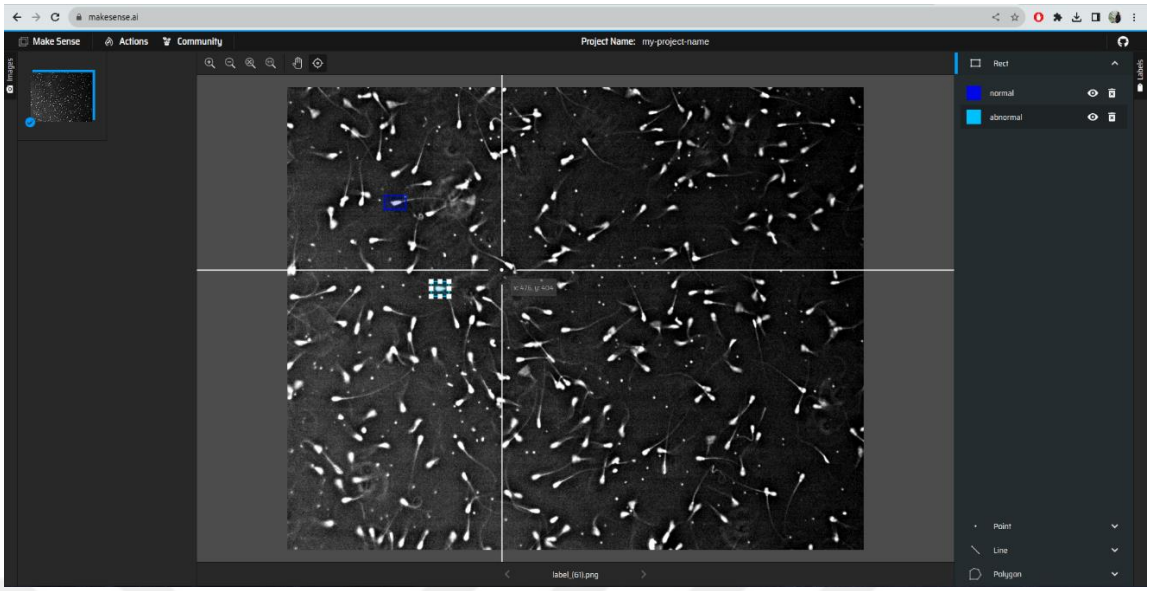


Figure 2.7 An example of makesense.ai

Sperms were initially labeled under 3 classes: normal, abnormal and unclassifiable. The Unclassifiable class was removed from the labels on the grounds that it created confusion in the model and was unnecessary.

In this process, sperms were labeled in two different types: completely rectangular (with their tails) and square only in the head area. The reason for this situation will be mentioned in the heading 4.2.3 Labeling the data for the Classification model. Two examples of these labels are given in Fig. 2.8 and Fig. 2.9.

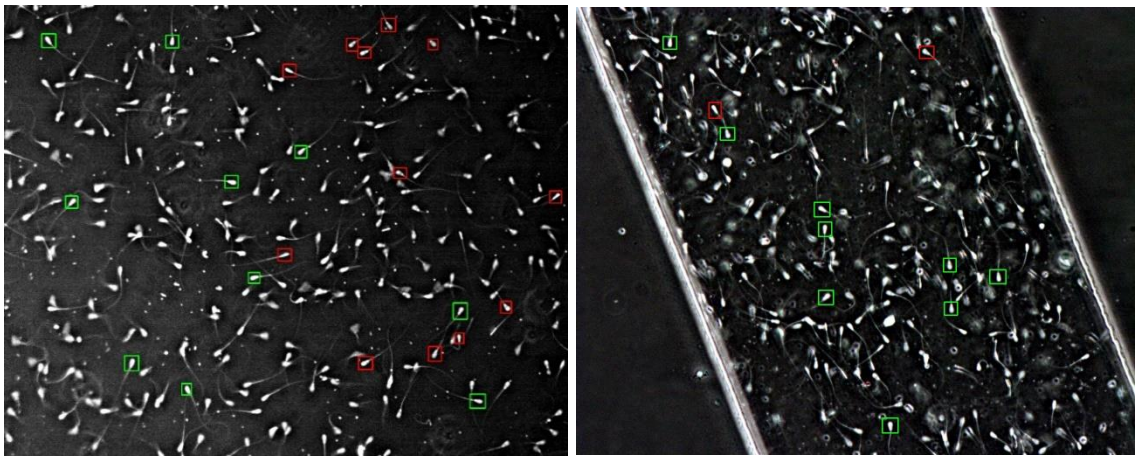


Figure 2.8 Data labeled only head of sperm

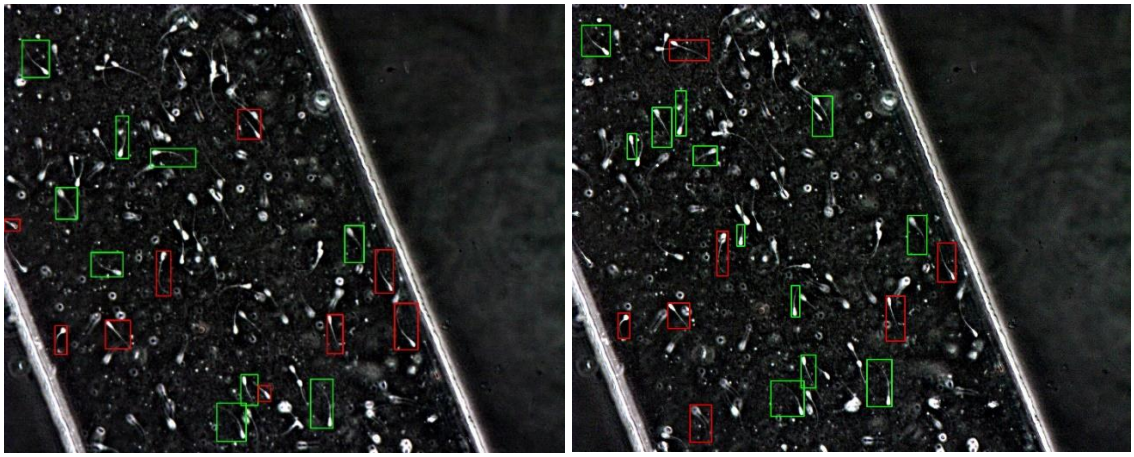


Figure 2.9 Labeled data based on all sperm
(Red labels: Abnormal sperm, Green labels: Normal sperm)

After the labeling process, the label distribution of the object detection data is plotted in Fig. 2.10, Fig. 2.11, Fig. 2.12, Fig. 2.13.

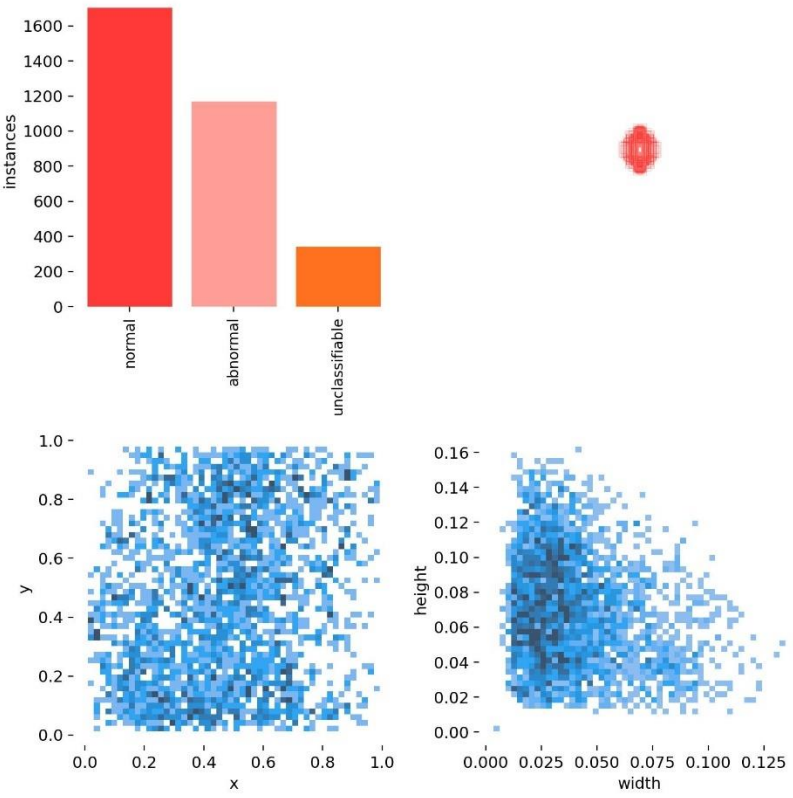


Figure 2.10 Graph of labels and label sizes for fully labeled sperm data

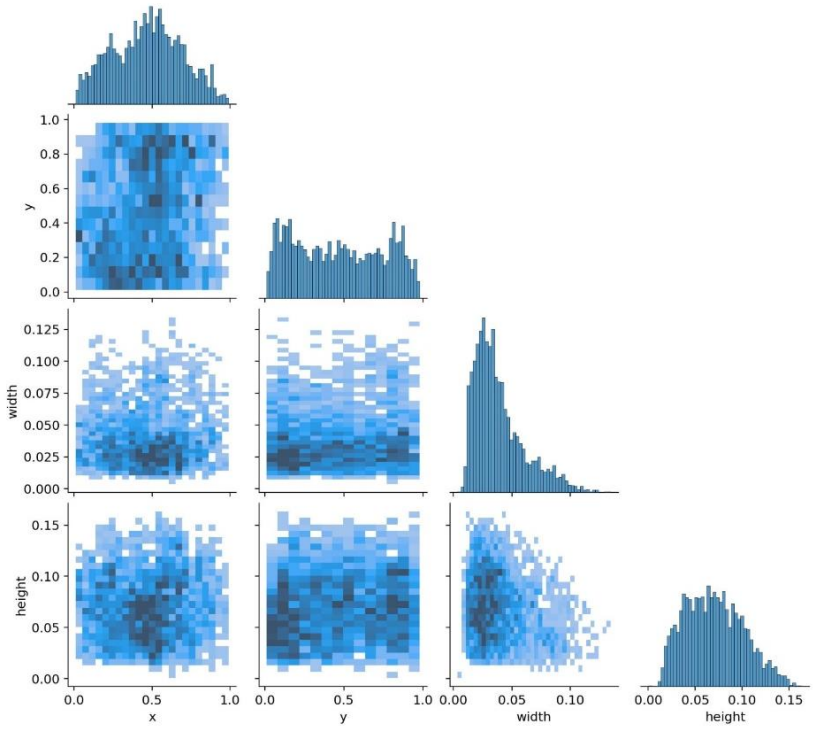


Figure 2.11 Pairs chart of label sizes for fully labeled sperm data

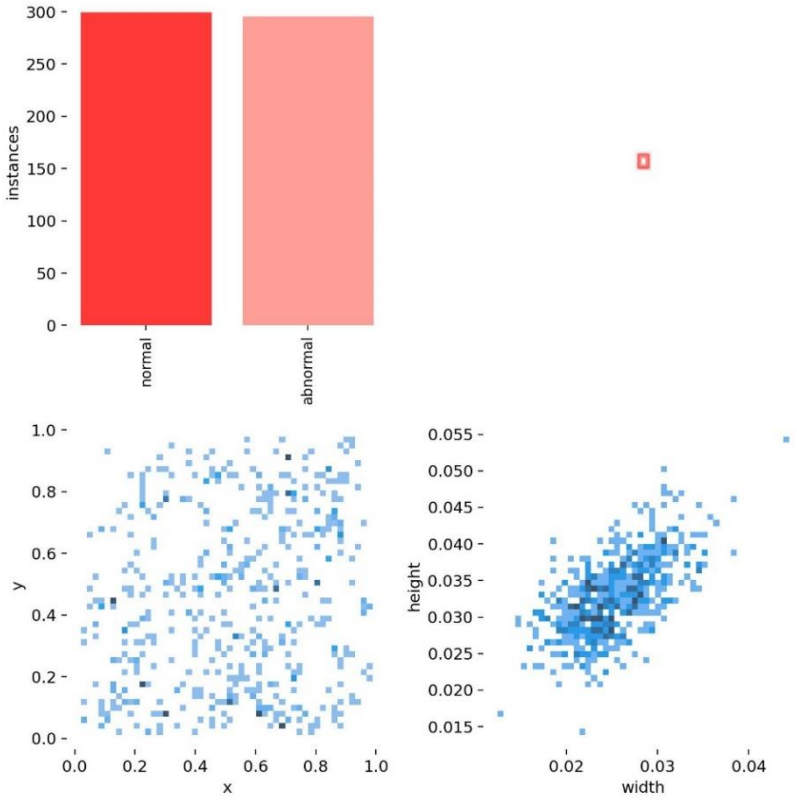


Figure 2.12 Graph of labels and label sizes for head-labeled sperm

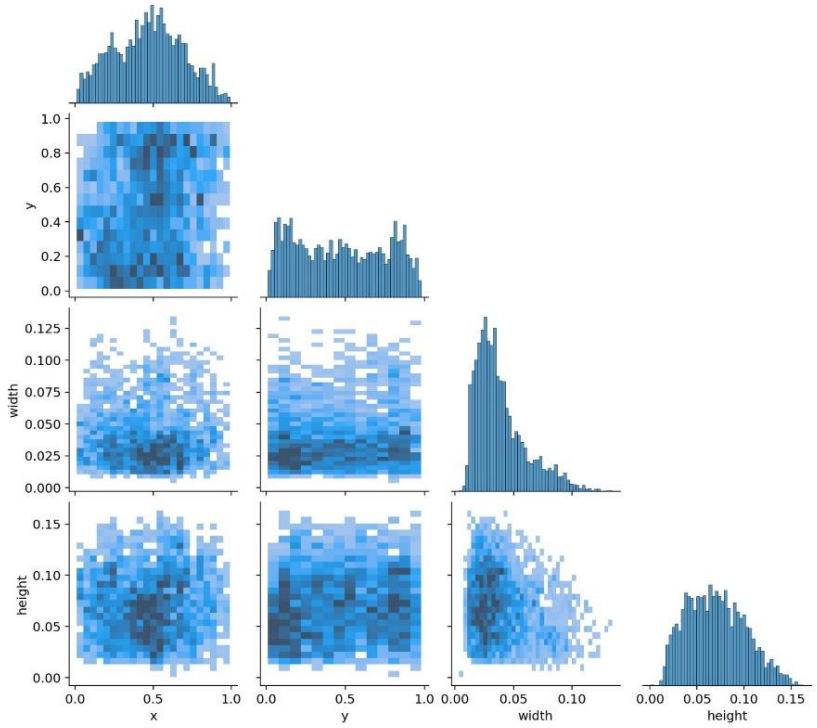


Figure 2.13 Graph of labels and label sizes for head-labeled sperm

2.3.3.2 Labeling data for classification

A classification model constitutes kind of machine learning model designed to determine whether a piece of data belongs to a particular class. This model usually relates given data to the probability values of a class and predicts the class exhibiting the greatest possibility. For example, a classification system running on an image can identify the content of an image and make predictions about whether the image belongs to a certain class, such as a cat or a dog.

Object detection, unlike classification, detects the location of the object on the image. In classification data, labels are not labeled based on a specific image. In classification data, labels are the images themselves and the class they belong to is the name of the folder in which they are located. All of the data must be the same size. 24 data samples for normal and abnormal classes are shown in Fig. 2.14, 2.15.

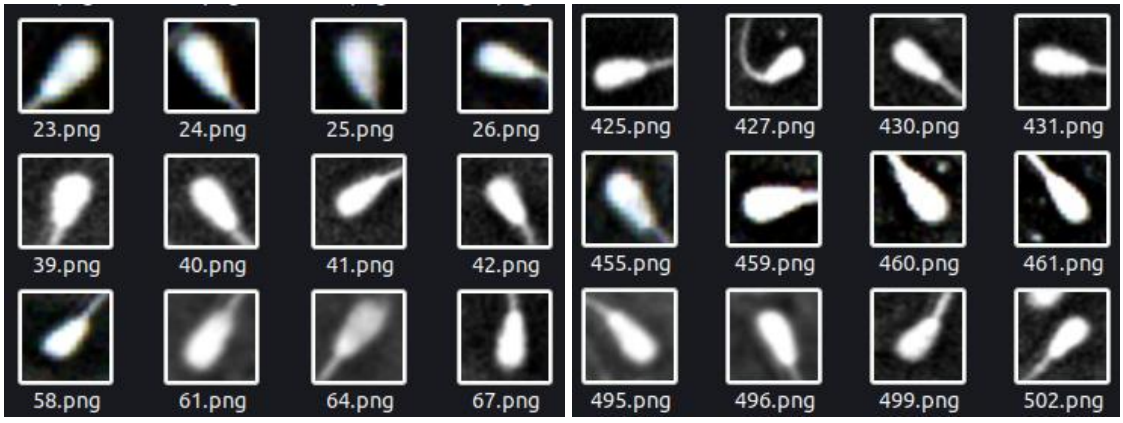


Figure 2.14 Classification data for the normal class

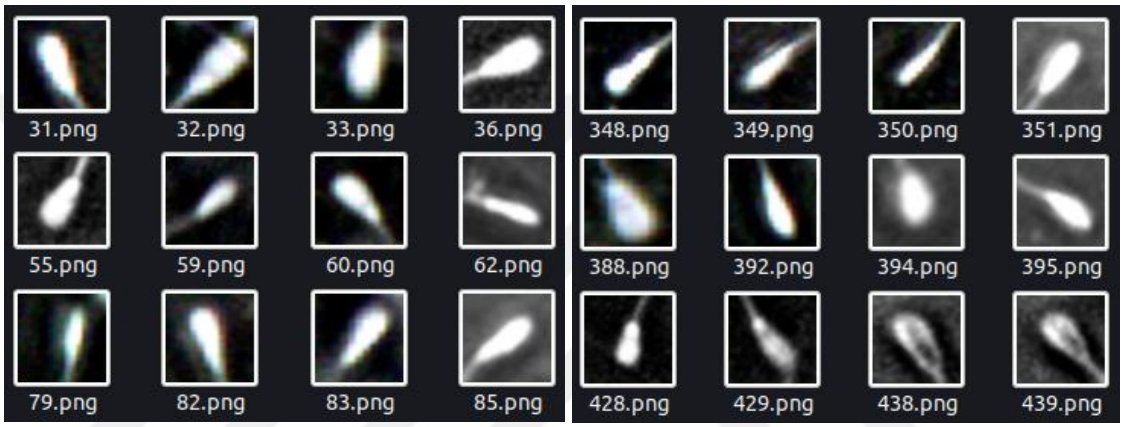


Figure 2.15 Classification data for the abnormal class

Classification data was created using the labels of object detection data. It was aimed to extract a single visual data for each sperm by cutting the labeled object detection data from the labeled points. In this process, two methods were followed for data labeled with the tail:

- Method 1: The label of the tail-labeled data is cut from the points it belongs to, and All images are adjusted to match the dimensions of the largest label.
- Method 2: Creating classification data without resizing the data by cutting the labels from the top-right corner of every individual label toward the leftward side and down to the largest label size so that each data is the same size.

Method 1 was not preferred because the structure of the sperm may be distorted when the images obtained are resized. Four examples of this are shown in Fig 2.16**.**

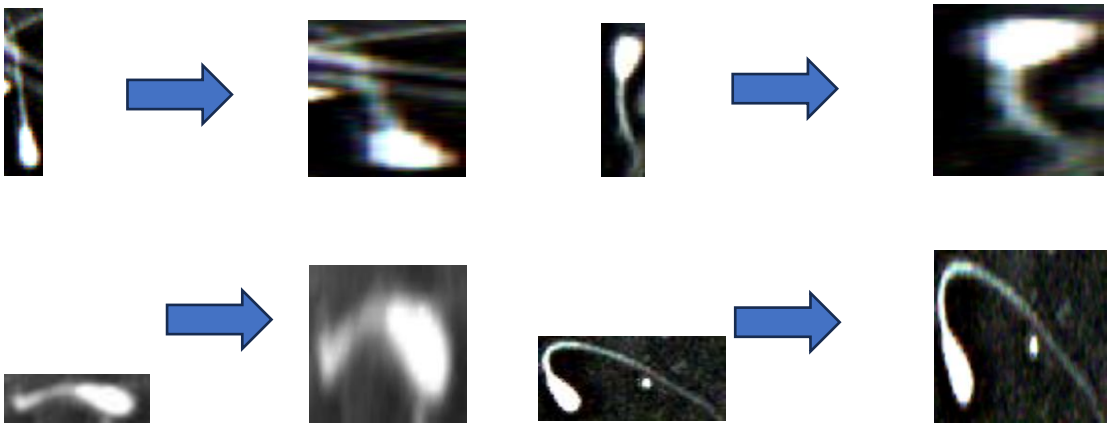


Figure 2.16 Data produced using the method 1

In the method 2, the data were not suitable for classification due to other sperm entering the background. Two examples for this situation are shown in Fig 2.17.

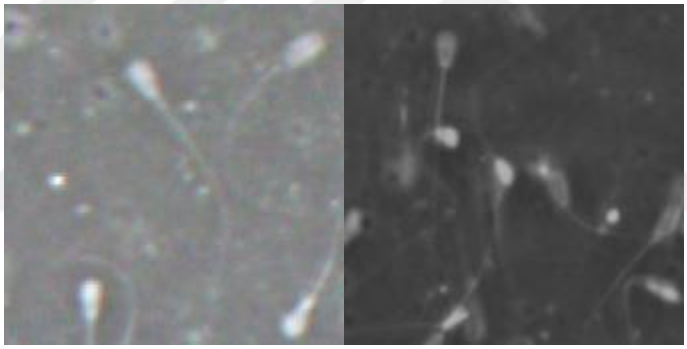


Figure 2.17 Data produced using the method 2

The first method was applied by labeling the sperm data quadratically only on the head, thus preventing the deterioration of the sperm structure.

A total of 768 data were created for classification. Of these data 376 are classified as part of the abnormal class while 392 fall within the normal category,
The data generated is based on the largest tag and its dimensions are 56x56 pixels.

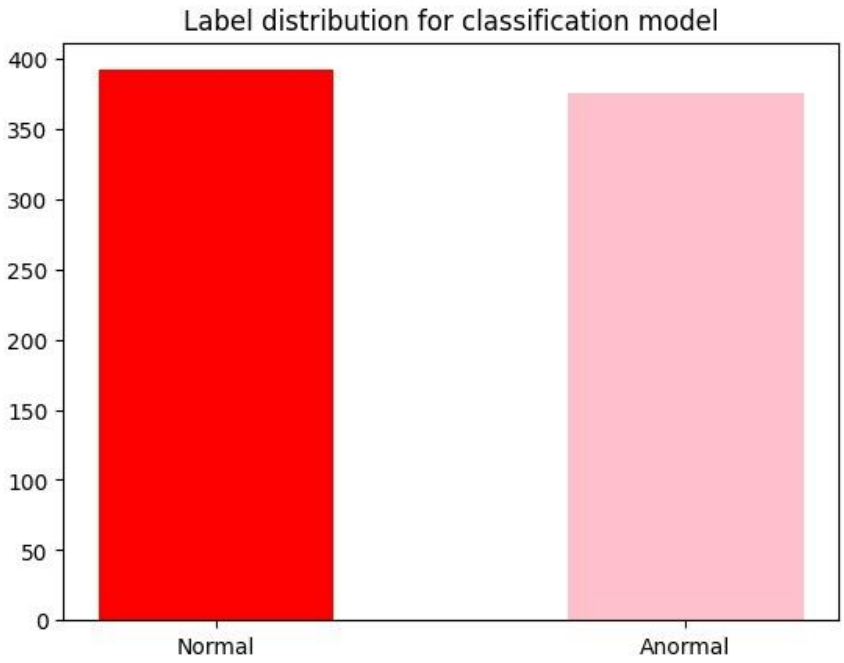


Figure 2.18 Graph of labels and label sizes for head-labeled sperm

2.4 Determination of the Algorithms to Be Used and the Training Phase

2.4.1 YOLO algorithm

YOLOv5 includes 5 different models. These are x(xlarge) , l(large) , m(medium) , s(small) and n(nano), depending on the number of layers and architectural size of the models.

YOLOv5 architecture consists of 3 main parts. These sections are called Backbone, Neck, and Head.

- Backbone: The backbone part of YOLOv5 uses the New CSP-Darknet53 structure. The basis of this structure is the modification of the Darknet architecture used in previous versions. The backbone contains convolution layers to extract feature maps from images.

- Neck: In YOLOv5, neck section is used to improve object detection performance by combining and enriching feature maps. For this purpose, SPPF (Spatial

Pyramid Pooling Fusion) and New CSP-PAN (Path Aggregation Network) structures have been integrated.

- Head: The head section creates the final output and can also be called the output layer. In this section, YOLOv5 performs object detection and classification tasks using the YOLOv3 Head. This is a structure used to predict the location and class of objects. (Ultralytics, t.y.-a)

Figure 2.19 shows the architecture of the convolutional neural network.

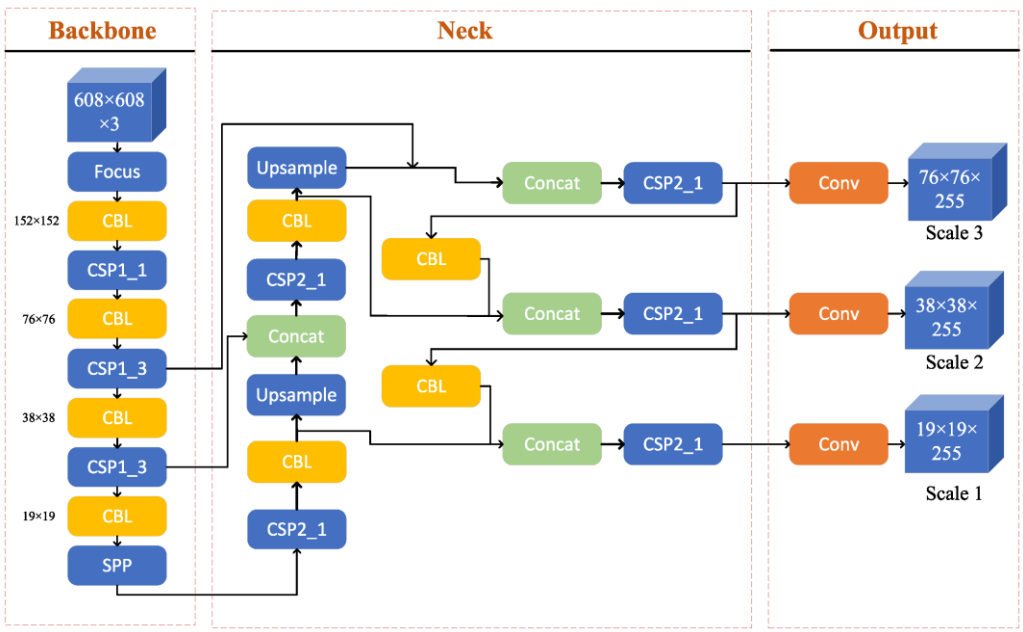


Figure 2.19 Architecture of YOLOv5 (Zhu et al., 2021)

The system, whose features are listed, was utilized in the training process for all models presented in Table 2.1.

Table 2.1 System features used for training

Proccessor	Intel i3 12100F
RAM	16GB 3200 Mhz
Graphic Card	Nvidia RTX 3060 8GB 256 Bit
Operating System (OS)	Ubuntu 20.04 LTS

2.4.1.1 Training the model with YOLO

During the model training process, the provided data is segmented into three partitions.: train, test ,validation. Sharing rate was selected as 80% for the Train data and 10% for the Test and Validation set.

Images during the training process with YOLO are 1024x1280 in size. Training was performed by resizing the data to 640x640 pixels. This reduces the number of cells entering the input layer, allowing the training process to be completed in a shorter time. YOLO carries out this process herself. (Ultralytics, t.y.-b)

The training process was completed within 100 epochs. An epoch refers to the entire training data set being passed by the model once. That is, when an epoch is completed, the model has seen the entire training data set. Epoch is often a determining factor in how long training will take. Each epoch completes in the same amount of time on average. Keeping the epoch number low may cause the model to learn adequately, while keeping it high may cause the model to overfit. (Brownlee, 2022) For this reason, 100 epochs were chosen to ensure that the model learned the data set at an optimum level.

During the training process, the batch size parameter was determined as 16 for the nano small and medium models, 8 for the large model, and 4 for the xlarge model. This parameter varies inversely to the architecture size and training time, and in direct proportion to the hardware power used. Batch size can be thought of as the data set used in a training process divided into small pieces. These chunks represent the quantity of

specimens in the dataset participating during training process simultaneously. As an instance, if a batch size value is 4, the model is updated by taking 4 samples from the training data set each time. For example, an image with dimensions 1024x1028 typically includes three spectral bands: Green, Red, Blue. These spectral bands are layers that carry the color information of the image, each of which can be thought of as a separate matrix. In this way, an image consists of 1024x1280x3 cells in total, and the value in each cell takes different integer values depending on the color depth of the image. 4 batch size means that 4x1024x1280x3 values are processed by the system at a time. (Brownlee, 2022)

Optimization challenges are omnipresent and materialize in virtually every sector of both scientific and engineering domains including machine learning.

Gradient descent (GD) stands out as a classic and widely employed method for addressing this particular problem (Netrapalli, 2019). But the conventional gradient descent method has become less effective in addressing large-scale machine learning problems(Tian et al., 2023). Due to various considerations, Stochastic Gradient Descent (SGD) has emerged as the predominant approach for addressing optimization challenges in machine learning scenarios, encompassing tasks (Newton et al., 2018).

The optimization algorithm employed was the stochastic gradient descent (SGD) algorithm. The primary aim of employing the optimization algorithm is to minimize the model's loss. The formulation of this optimization algorithm is in Eq. 2.1

$$\theta t + 1 = \theta t - \alpha \nabla J(\theta t; xi, yi) \tag{2.1}$$

Here; θt : represents the model parameters at the t-th iteration. α: represents the learning rate, controlling how much of a "step" is taken in each update $J(\theta t; xi, yi)$: represents the loss function computed by the current metrics of the model. xi denotes the input features, and yi represents the actual output value. $\nabla J(\theta t; xi, yi)$: represents the slope of the loss function concerning the parameters.

i: denotes a specific example in the dataset.

Throughout the training procedure, learning rate is an important coefficient and hyperparameter that determines the amount of updating of model parameters. This value, initially chosen by trial and error, affects the speed and stability.

A learning rate value that is excessively low may provide stable and regular training by updating the model parameters in small steps, but may prolong the time. On the other hand, choosing a learning rate value that is too large may cause excessive updates and instability of the training process (Blier et al., 2020). The effect of the learning rate is important as it controls the size of the update steps. Learning rate strategies can adapt over time to better approximate optimal values.

During the training process with YOLO, training was performed with both head-labeled data and tail-labeled data.

2.4.2 State of the art models and training phase

State of the art (SOTA) refers to the most advanced, most innovative and most effective techniques or models available in a field or subject. SOTA models contain certain architectures that have entered this field by showing high performance in contemporary research within the realm of deep learning. These models are generally constantly updated. (Deci, 2023)

VGG (Visual Geometry Group) and ResNet (Residual Network) are approaches among the State-of-the-art architectures.

VGG and ResNet are two prominent architectures in the field of image classification. (Tulasi Krishna & Kalluri, 2019) These architectures come in different depths. For example, VGG19 has an architecture consisting of 19 layers.(Carvalho et al., 2017) Model performance does not always increase in direct correlation with the quantity of layers. The parameters were employed in the trainings for all models in Table 2.2.

Table 2.2 Parameters used for training,

Batch size	8
Epoch	100
Learning Rate	0.03
Optimizer	Adam
Activation Function	Softmax
Loss Function	Categorical cross entropy
Train data (%) / image	80 / 632
Validation data (%) / image	10 / 79
Test data (%) / image	10 / 79

2.4.2.1 VGG16 model and training process

VGG16 represents a type of convolutional neural network (CNN) and a simple and widely used architecture for the ImageNet which is a large visual object recognition software research project. It was conceived and presented by Karen Simonyan and Andrew Zisserman. (Carvalho et al., 2017; Simonyan & Zisserman, 2014). The architecture consists of five convolutional blocks, with each block's input being the output of the previous block.(Tammina, 2019) On a dataset featuring over 14 million images across 1000 object classes in ImageNet, the VGG16 model accomplished an initial top-5 test accuracy of 92.7%.(Guan et al., 2019)

The presence of the numeral 16 in the VGG name indicates that it is a neural network with a depth of precisely 16 layers. (VGG). (R & Patilkulkarni, 2021) This signifies that VGG16 is a notably complex network, involving a total of approximately 138 million parameters.(Munir et al., 2019) Even according to modern standards, it is a huge network. The appeal of the VGGNet16 architecture is rooted in its simplicity, enhancing the overall attractiveness of the network.(Younis et al., 2022)

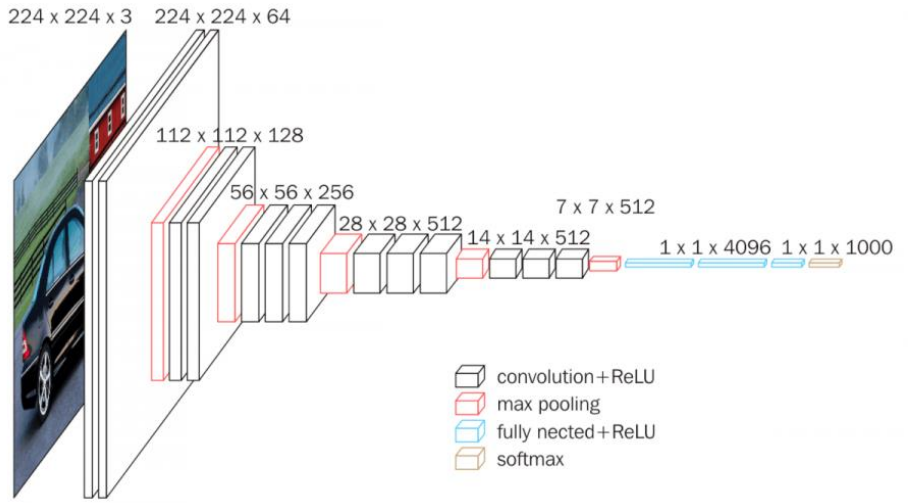


Figure 2.20 Architecture of VGG16 (Hassan, 2018)

Python language and TensorFlow library were used for training. The training procedure was conducted with the systems mentioned above. The training process lasted 32 minutes.

2.4.2.2 VGG19 model and training process

The VGG19 model, also known as VGG19, shares the fundamental concept with VGG16, differing only in its capacity for 19 layers. The numerical designations "16" and "19" represent the quantity of weight layers in the model, specifically referring to convolutional layers. Consequently, VGG19 incorporates three additional convolutional layers compared to VGG16 (Panigrahi et al., 2023).

The training process lasted 36 minutes.

2.4.2.3 ResNet50 model and training process

ResNet is an architecture designed to deal with the vanishing gradient problem, a problem encountered in training deep learning models. This architecture is especially famous for a version known as ResNet-50. ResNet makes it possible to add layers to convolutional neural networks to enable deeper and more effective learning through an innovation called shortcut connections (Alzubaidi et al., 2021).

Within layers, shortcut connections facilitate cross-layer connectivity, and they are devoid of parameters while being independent of data.(Alzubaidi et al., 2021)This makes the training of deeper networks more stable and makes it possible to achieve better performance.

Traditional networks, such as VGG neural networks, increase depth by using 3x3 filters at each layer (Suryawati et al., 2018). However, the uniqueness of ResNet comes from the fact that it creates deeper networks with less complexity than VGG by using fewer filters (Nasir et al., 2023). At the same time, even a smaller 18-layer ResNet can run significantly faster than VGG-19, reaching 1.8 billion FLOPs(Victor Ikechukwu et al., 2021).

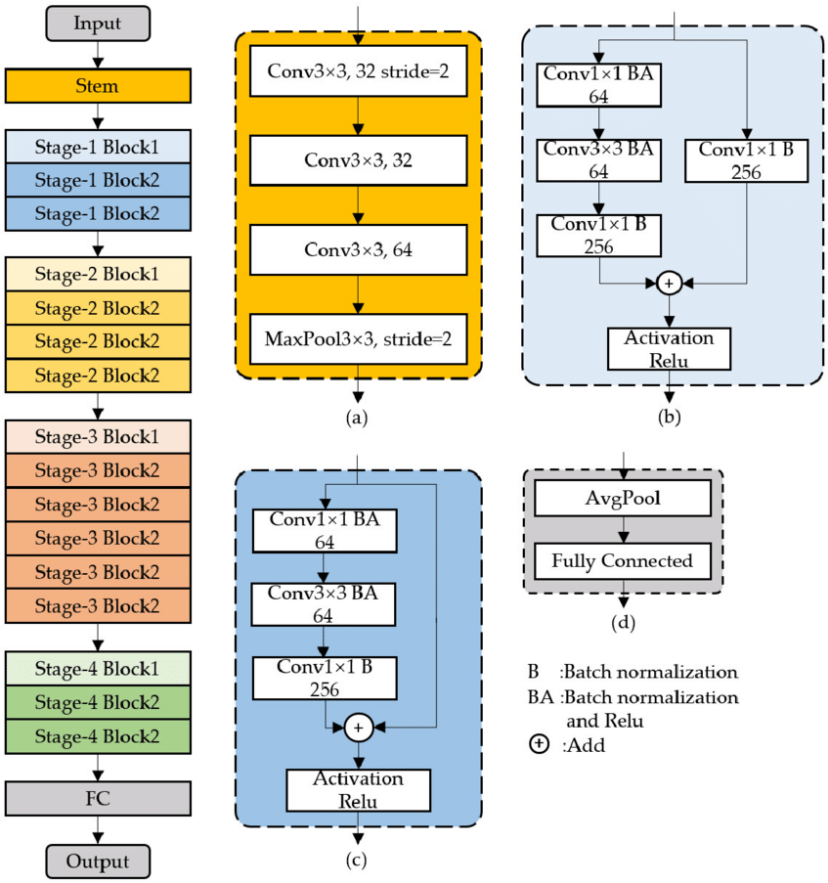


Figure 2.21 Architecture of ResNet50 (S. Wang et al., 2021)

Python language and tensorflow library were used for training. The training process lasted 39 minutes.

2.4.2.4 ResNet101 model and training process

Resnet101 has 101 layers and has a deeper and more detailed network structure than the Resnet50 architecture. This case has a larger size and learning curve, similar to the VGG16 and VGG19 connection.

The training process lasted 43 minutes.

2.4.2.5 ResNet152 model and training process

ResNet152 architecture has the largest ResNet architecture and has 152 layers in total. These layers increase detailing and model depth.

The training process lasted 51 minutes.

3. RESEARCH FINDINGS

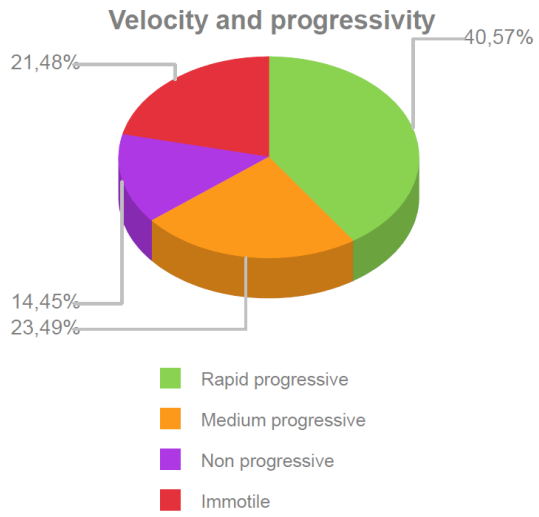
3.1 Application of Electric Field

Measurements were made and compared regarding the sperm data to which electric field was applied and the efficiency of sperm in the control group without electric field. The values resulting from these operations are stated in Table 3.1.

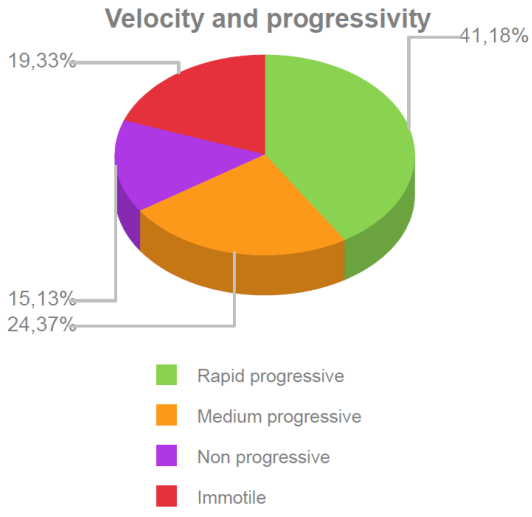
Table 3.1 Results of velocity and progressivity on different groups

Electric field	Type	Total	%	M/mL	M/Sample
3V (50V/m)	Rapid prog.	525	40.57	35.40	88.51
	Medium prog.	304	23.49	20.50	51.25
	Non prog.	187	14.45	12.61	31.53
	Immotile	278	21.48	18,75	46,87
6V (100V/m)	Rapid prog.	98	41.18	33.48	83.69
	Medium prog.	58	24.37	19.81	49.53
	Non prog.	36	15.13	12.30	30.74
	Immotile	46	19.33	15.71	39.28
12V (200V/m)	Rapid prog.	104	26.40	31.27	78.17
	Medium prog.	126	31.98	37.88	94.71
	Non prog.	92	23.35	27.66	69.15
	Immotile	72	18.27	21.65	54.12
Control Group	Rapid prog.	100	27.17	30.82	77.06
	Medium prog.	118	32.07	36.37	90.93
	Non prog.	92	25.00	28.36	70.89
	Immotile	58	15.76	17.88	44.69

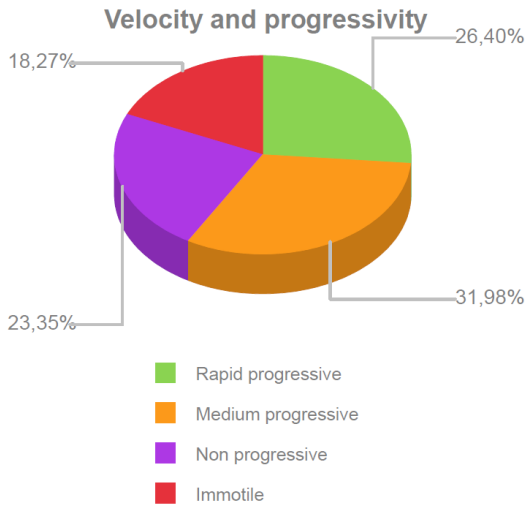
Pie charts showing the percentages of "velocity & progessivity" values according to sperm yield for the 3V , 6V , 12V sperm groups and the control group are shown.



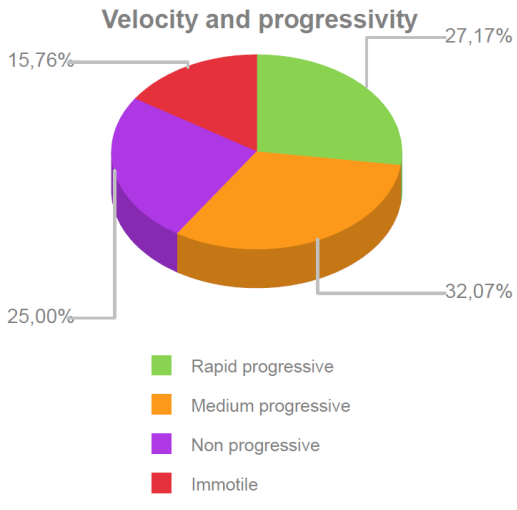
A) Pie chart for Group 3V



B) Pie chart for Group 6V



C) Pie chart for Group 12V



D) Pie chart for Control Group

Figure 3.1 Pie charts of sperm groups in percentages for sperm groups

The percentage values of rapid progressive sperm values are shown in Table 3.2.

Table 3.2 Percentage of rapid progressive sperm

3V Group	40.57
6V Group	41.18
12V Group	26.40
Control Group	27.17

Linear, Curve and average velocity values for sperm in the rapid progressive group were calculated for each group.

The voltage and electric field values specified here were obtained by applying different magnitudes to sperm cells. The voltage was adjusted until a significant difference in the motility values of sperm cells was observed. Simultaneously, considering the possibility of an impact on the DNA of sperm cells, the existing electric field values were determined.

The values for this situation clarified in Table 3.3.

Table 3.3 Speed values for each sperm in the rapid progressive group

	Curve Speed	**Linear Speed**	**Avg value**
3V (50V/m)	90.66	55.57	62.80
6V (100V/m)	81.32	49.87	58.21
12V (200V/m)	90.24	47.80	55.79
Control Group	95.40	47.94	56.87

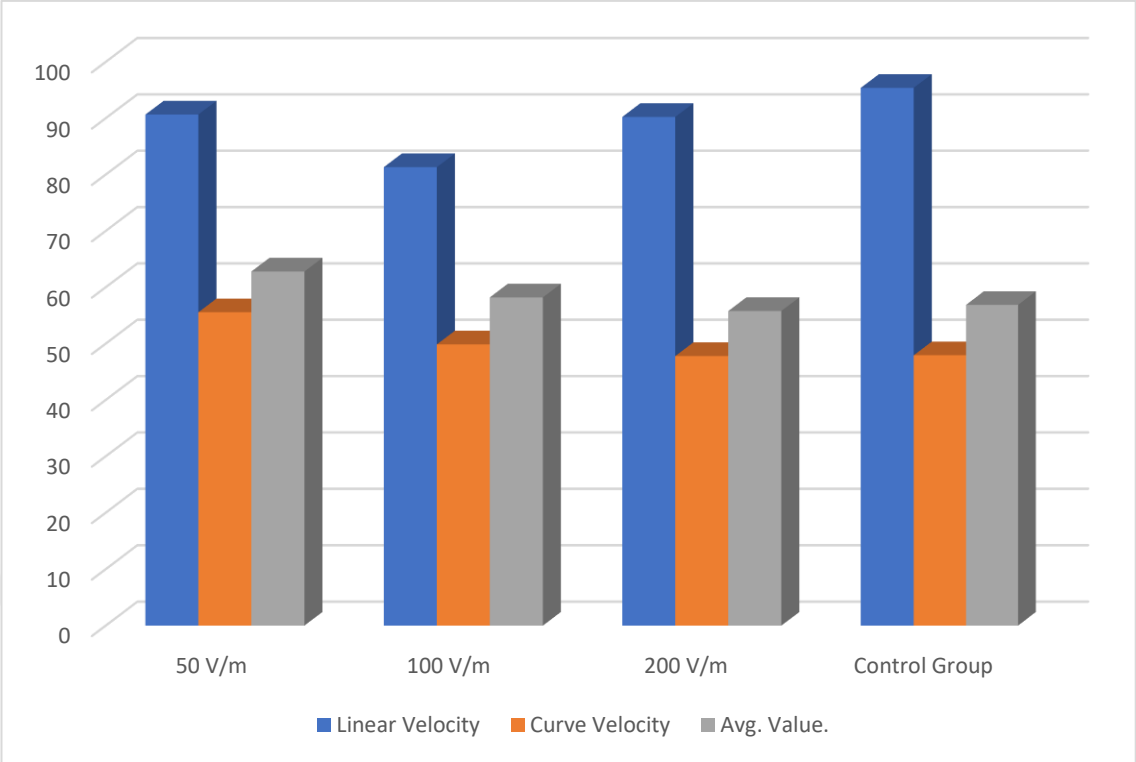


Figure 3.2 Graph of speed values for each sperm in the fast-progressing group

3.2 Results of Classification Models

Within the scope of this thesis study, Accuracy, Recall, Precision and F1-score are among the important metrics used in the evaluation of classification models. These metrics were employed for evaluating and analyzing the model's performance. The confusion matrix is utilized to compute these values.

- **Confusion Matrix:** It serves as a cross table, recording occurrences between two assessors, specifically the actual or true predicted classification and classification (Grandini et al., 2020) , as shown in Figure 4.3. The model predictions are denoted by the rows , whereas the true classifications are displayed in columns.

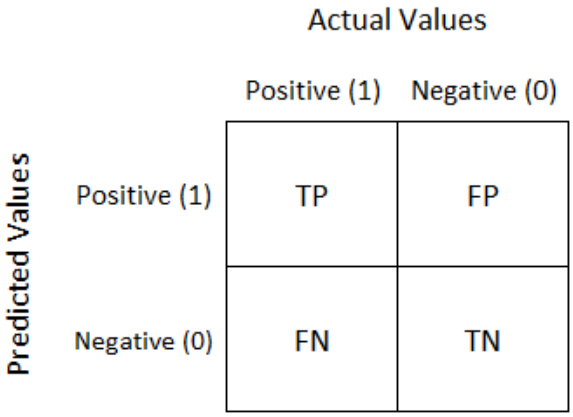


Figure 3.3 Basic of confusion matrix (Cochard, 2021)

- **Accuracy:** In classification Accuracy is one of the most used metrics, It originates from confusion mat.

  Formulation of accuracy is stated in Eq 4.1

  The precision of a classification model is computed through a formula that involves the sum of True Positive and True Negative values in the numerator, divided by the total of all entries in the confusion matrix in the denominator. True Positives and True Negatives denote elements accurately identified by the model, positioned on the primary diagonal of the confusion matrix. Conversely, the denominator encompasses all elements outside this diagonal that were misclassified by the model. In essence, accuracy signifies the probability of the model making accurate predictions when randomly selecting a unit and forecasting its class.(Grandini et al., 2020)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

- **Recall**:, Recall, referred to as the genuine positive ratio or sensitivity, represents the test's proficiency in discerning affirmative results. It gauges the efficiency of a classification model in identifying occurrences associated with a particular class in a given dataset. A minimal occurrence of false negatives corresponds to

heightened recall, whereas an elevated count of false negatives leads to diminished recall. (Alsalem et al., 2018)

Enhanced recall signifies improved performance in the classification model. Its formulation is stated in Eq 3.2 (Grandini et al., 2020)

$$Recall = \frac{TP}{TP+FN} \tag{3.2}$$

- **Precision:** Precision (P) is determined by dividing the number of true positives (TP) by the total of true positives and false positives.

  In multiclass classification tasks, precision is assessed at both the micro and macro levels based on the outcomes derived from the confusion matrix.(Alsalem et al., 2018)

  Precision indicates the trustworthiness of model during identifying as positive. Its formulation was stated in Eq 3.3 (Grandini et al., 2020)

$$Precision = \frac{TP}{TP+FP} \tag{3.3}$$

- **F1-Score:** It gauges the performance of a classification model by amalgamating Precision and Recall measures through the harmonic mean, derived from the confusion matrix. Its formulation is stated in Eq 4.4 (Grandini et al., 2020)

$$F1 - Score = 2 \times \left(\frac{precision \times recall}{precision + recall}\right) \tag{3.4}$$

3.2.1 Results for VGG-16

Results of VGG-16 are as follows.

Table 3.4 Results for VGG16

	Precision	Recall	F1-score	Support
Abnormal	0.77	0.69	0.73	39
Normal	0.73	0.82	0.76	40
Accuracy			0.75	79
Macro av	0.75	0.75	0.75	79
Weighted av	0.75	0.75	0.75	79

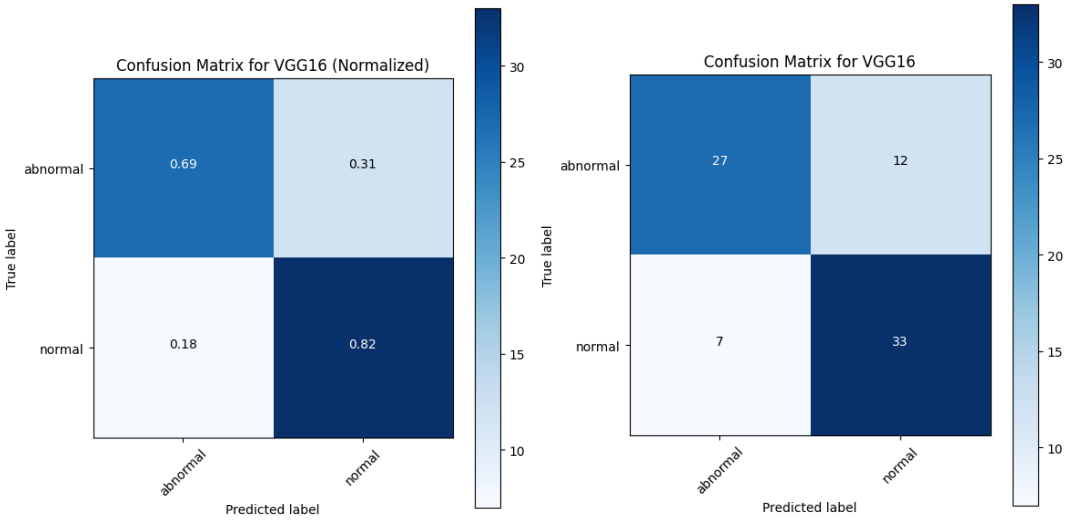


Figure 3.4 Normalized and non-normalized confusion matrices for VGG16

3.2.2 Results for VGG-19

Results of VGG-19 are as follows.

Table 3.5 Results for VGG19

	Precision	Recall	F1-score	Support
Abnormal	0.81	0.67	0.73	39
Normal	0.73	0.85	0.76	40
Accuracy			0.76	79
Macro av	0.77	0.76	0.76	79
Weighted av	0.77	0.76	0.76	79

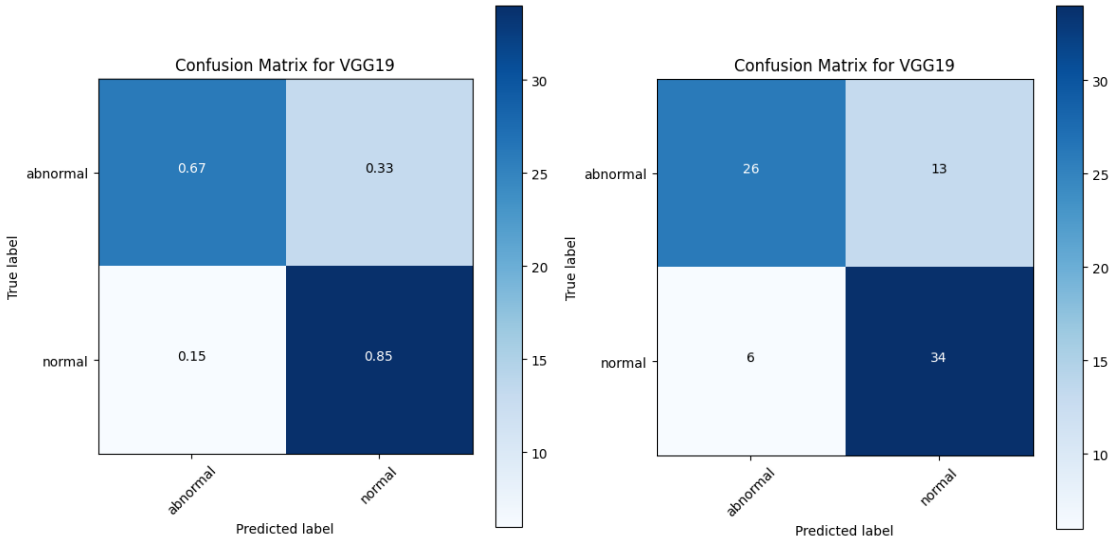


Figure 3.5 Normalized and non-normalized confusion matrices for VGG19

3.2.3 Result for ResNet50

Results of ResNet50 are as follows.

Table 3.6 Results for ResNet50

	Precision	Recall	F1-score	Support
Abnormal	0.81	0.67	0.73	39
Normal	0.72	0.85	0.78	40
Accuracy			0.76	79
Macro avg	0.77	0.76	0.76	79
Weighted avg	0.77	0.76	0.76	79

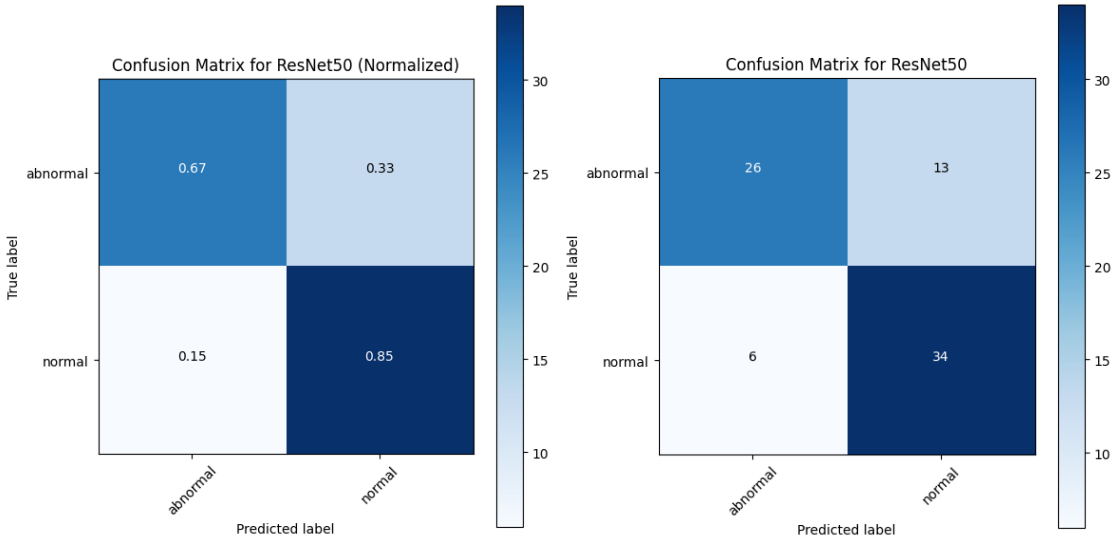


Figure 3.6 Normalized and non-normalized confusion matrices for ResNet50

3.2.4 Result for ResNet101

Results of ResNet101 are as follows.

Table 3.7 Results for ResNet101

	Precision	Recall	F1-score	Support
Abnormal	0.54	0.85	0.66	39
Normal	0.73	0.80	0.76	40
Accuracy			0.57	79
Macro avg	0.60	0.57	0.54	79
Weighted avg	0.60	0.57	0.54	79

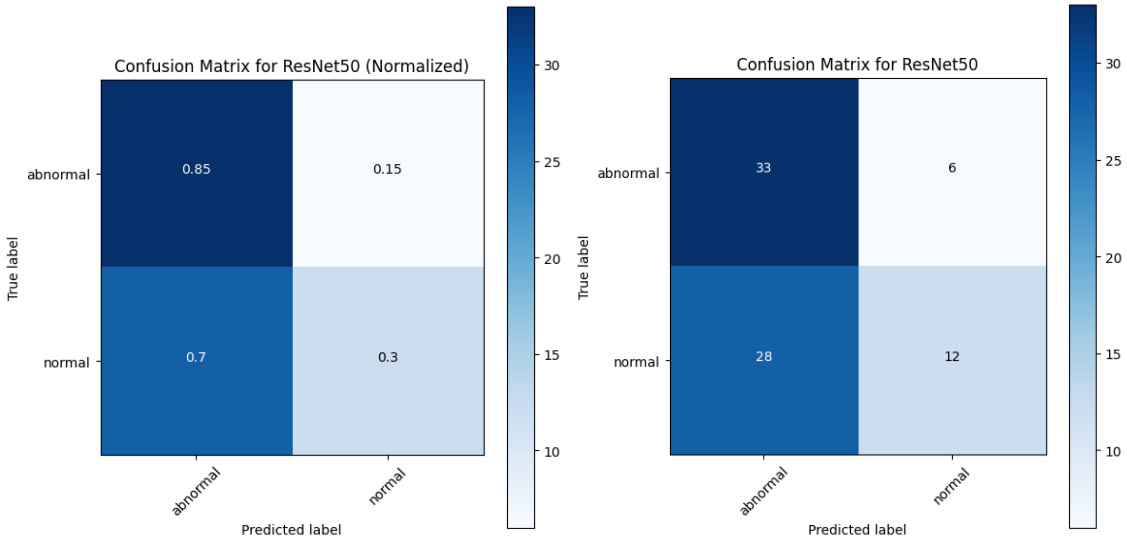


Figure 3.7 Normalized and non-normalized confusion matrices for ResNet101

3.2.5 Result for ResNet152

Results of ResNet152 are as follows.

Table 3.8 Results for ResNet152

	Precision	Recall	F1-score	Support
Abnormal	0.77	0.69	0.73	39
Normal	0.73	0.80	0.76	40
Accuracy			0.75	79
Macro avg	0.75	0.75	0.75	79
Weighted avg	0.75	0.75	0.75	79

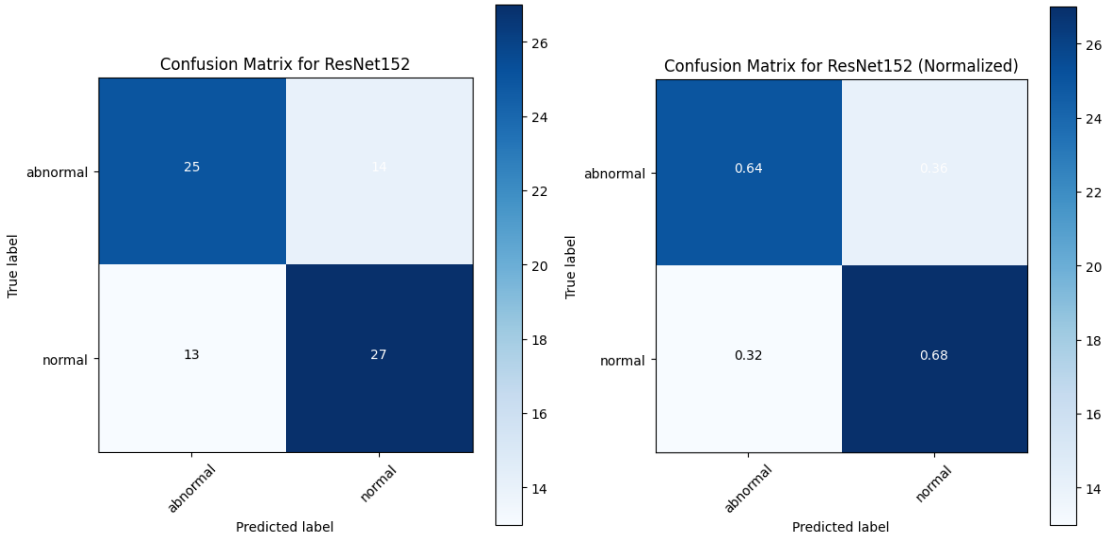


Figure 3.8 Normalized and non-normalized confusion matrices for ResNet152

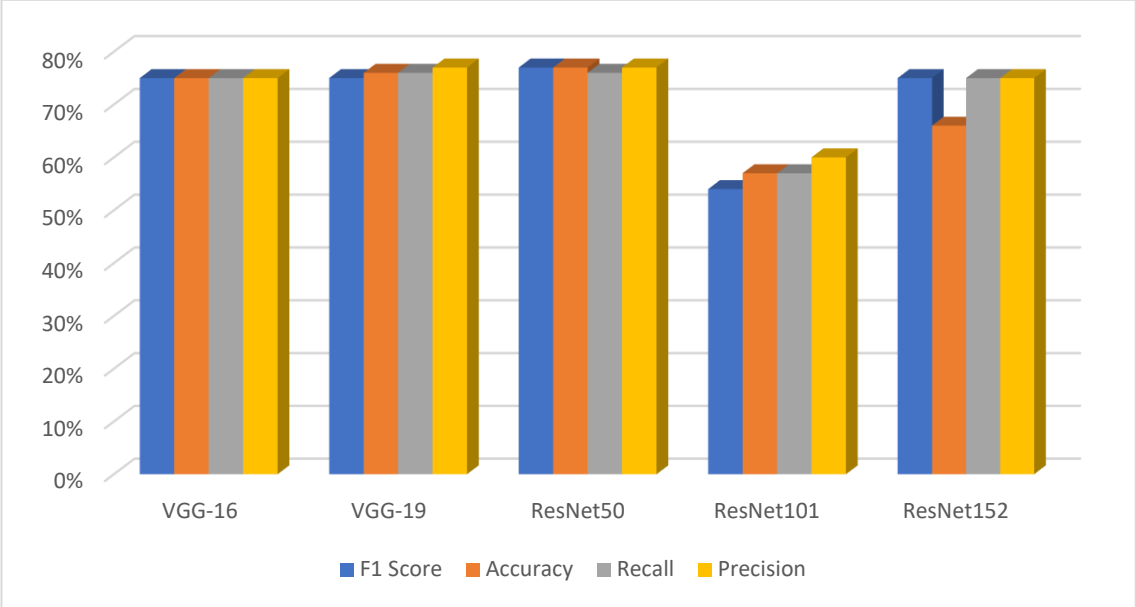


Figure 3.9 Comparison of accuracy values of state-of-the-art models

When comparing classification models in this context, it has been observed that the model trained with the ResNet50 architecture achieved the highest performance across all parameters. All parameters were taken into consideration in order to evaluate the different aspects of the models' distinct features. Among the contributing factors, the significant impact of the architecture's structure stands out prominently. The use of specifically labeled sperm head data has significantly contributed to achieving more successful results in algorithms trained with this dataset. The labeling of only the sperm head has allowed

sperm classification models to make a simpler and less complex choice. These factors have led to noticeable differences in the results obtained by models classifying sperm cells.

3.3 Results for YOLO Models

Precision , mAP , Recall and F1-Score and values were used in the evaluation of YOLO models.For the computation of mAP (mean Average Precision), IOU, Recall, Precision, AP and Precision-Recall Curve are necessary components.

In this process, intersection over Union (IoU) is a measure that quantifies the proportion of the overlapping area relative to the total area encompassed by two shapes, such as bounding boxes.(Zhou et al., 2019) A substantial IoU value suggests accurate predictions of object bounding boxes by the model, while a lower value may suggest areas for model improvement. Hence, IoU stands as a crucial metric for assessing and enhancing the efficacy of object detection models.

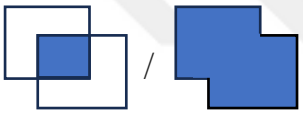
$$IOU = \frac{area\ of overlap}{area\ of\ union} =$$ 

Figure 3.10 Visualization for IOU (Cochard, 2021)

The mAP value exhibits internal variability. the mAP50 value accurately reflects predictions when there is a 50% overlap between the predicted and actual area ratio (IOU). Consequently, it is reasonable to anticipate a higher mAP75 parameter, indicating a 75% IOU overlap.

Utilizing the Precision-Recall curve aids in evaluating the effectiveness of an object detector. A curve is drawn for each object class by changing the confidence level. A well-performing object detector for a specific class is characterized by high precision and an increasing level of recall. So adjusting the confidence level should not undermine the high precision and recall of the object detector; both metrics should remain consistently high.

An effective object detector should have the capability to exclusively identify relevant objects. (GitHub - rafaelpadilla/Object-Detection-Metrics: Most popular metrics used to evaluate object detection algorithms., t.y.)

3.3.1 Results of YOLO models for head-only labeled data

Each graph enables the evaluation of the characteristics obtained from the model training from different perspectives. The correlation between each metric obtained as a result of the training is also depicted using various graphs.

3.3.1.1 YOLOv5N

Results of YOLOv5N are as follows.

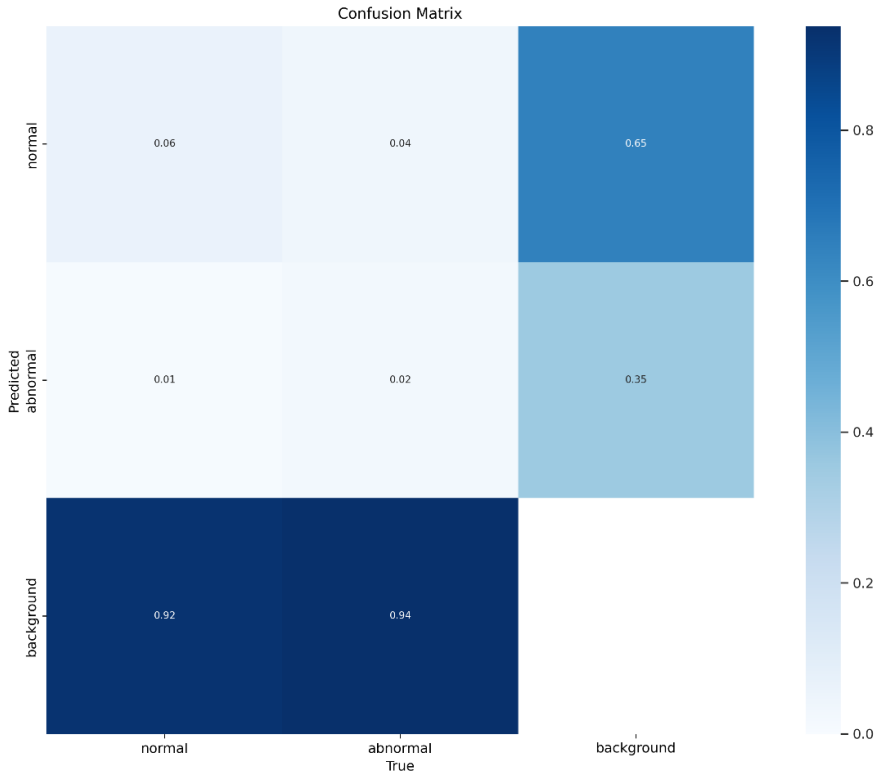


Figure 3.11 Normalized Confusion matrix for YOLOv5N head labeled data

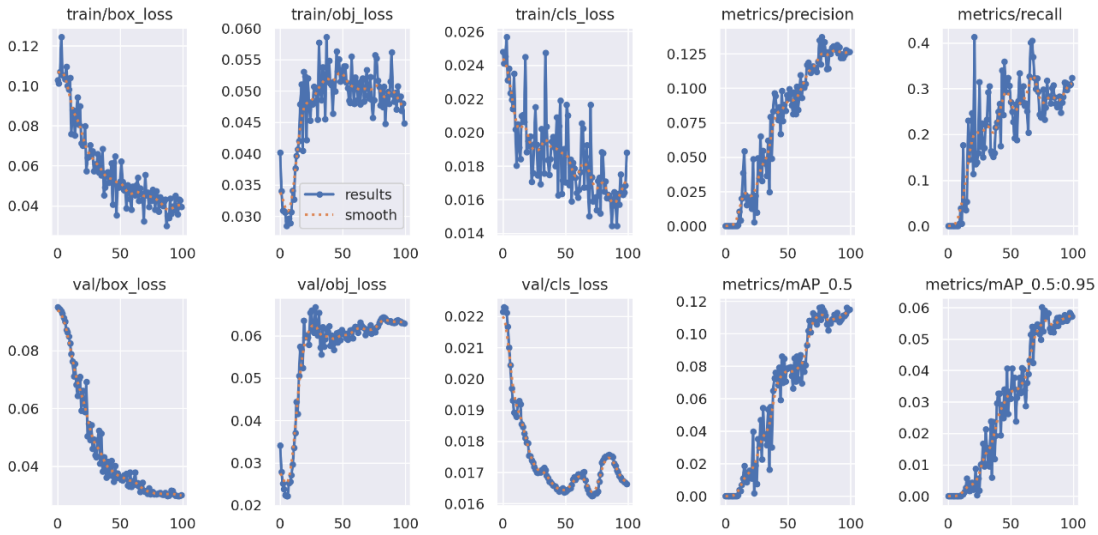


Figure 3.12 Results graph for YOLOv5N head labeled data

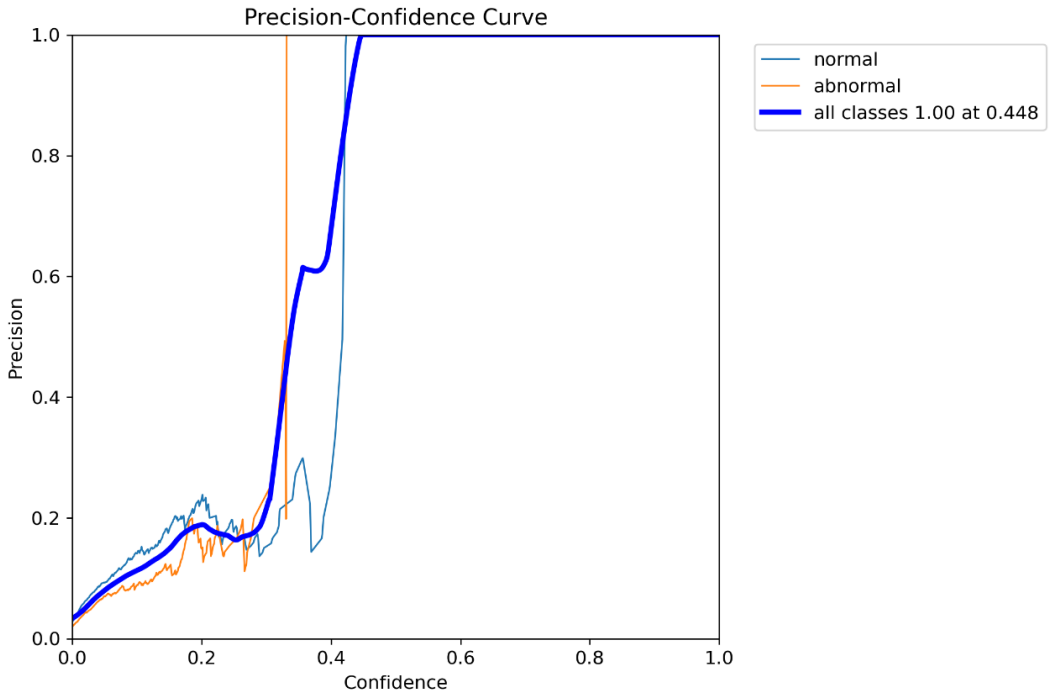


Figure 3.13 Precision Confidence Curve for YOLOv5N head labeled data

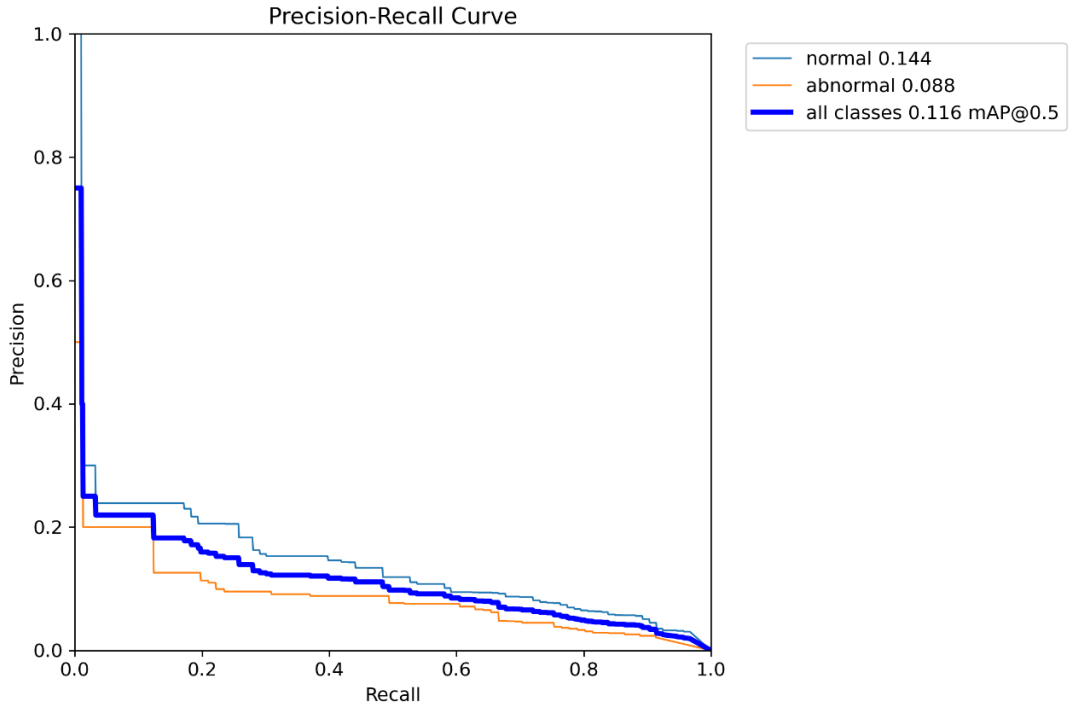


Figure 3.14 Precision-Recall Curve for YOLOv5N head labeled data

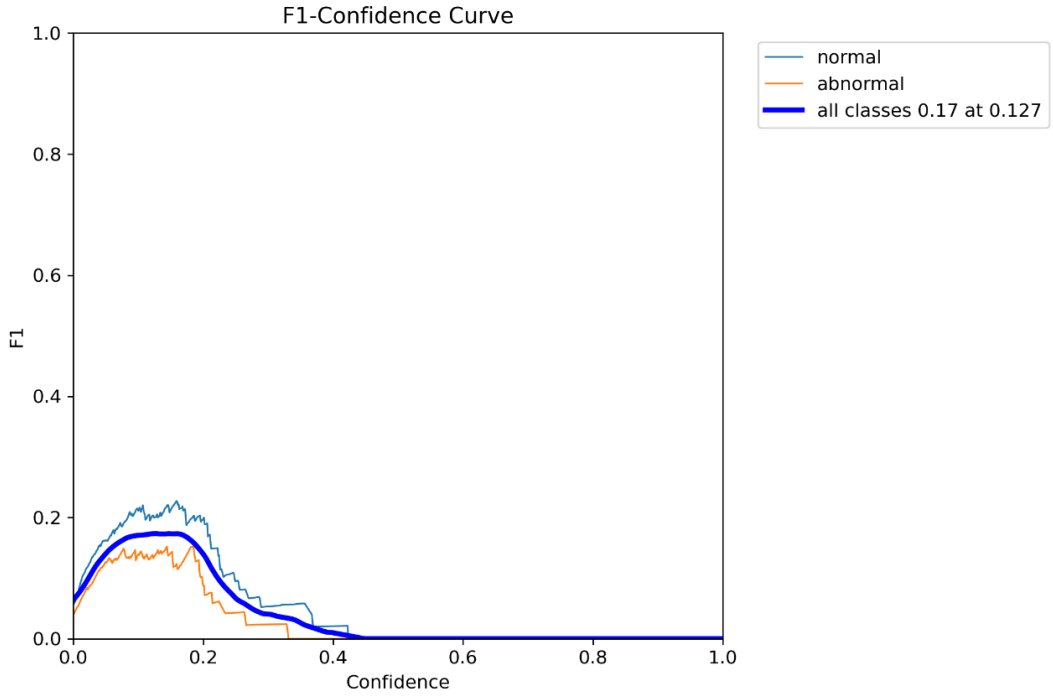


Figure 3.15 F1-Confidence Curve for YOLOv5N head labeled data

3.3.1.2 YOLOv5S

Results of YOLOv5S are as follows.

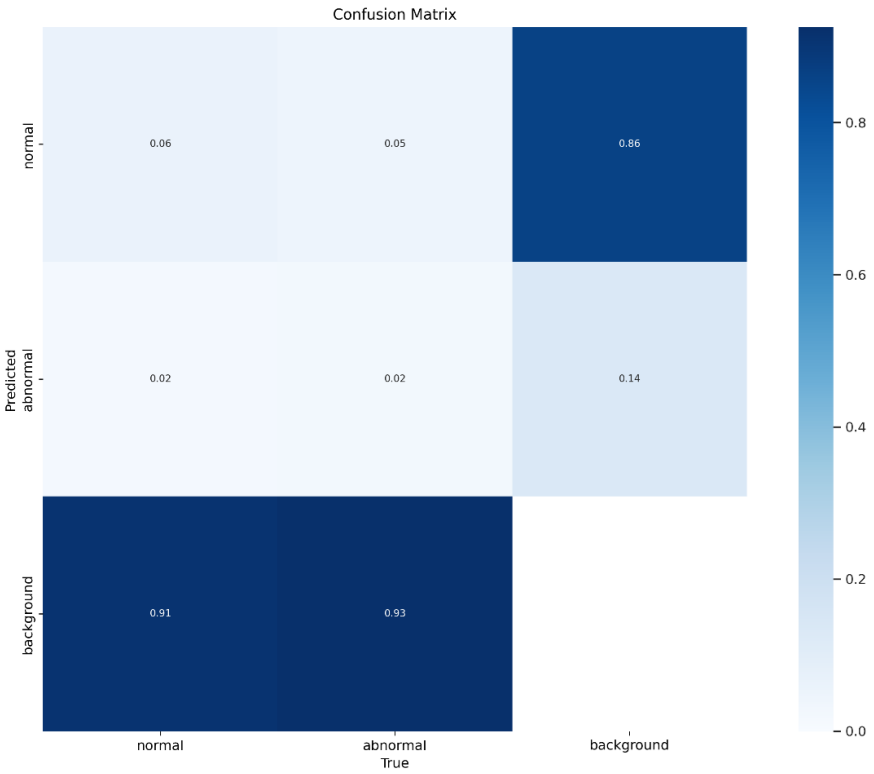


Figure 3.16 Normalized Confusion matrix for YOLOv5S head labeled data

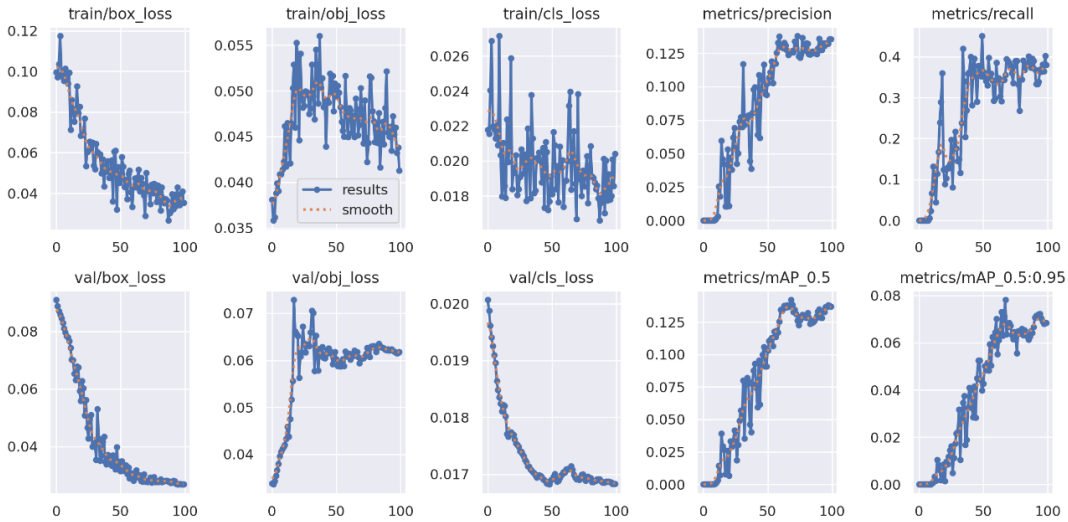


Figure 3.17 Results graph for YOLOv5S head labeled data

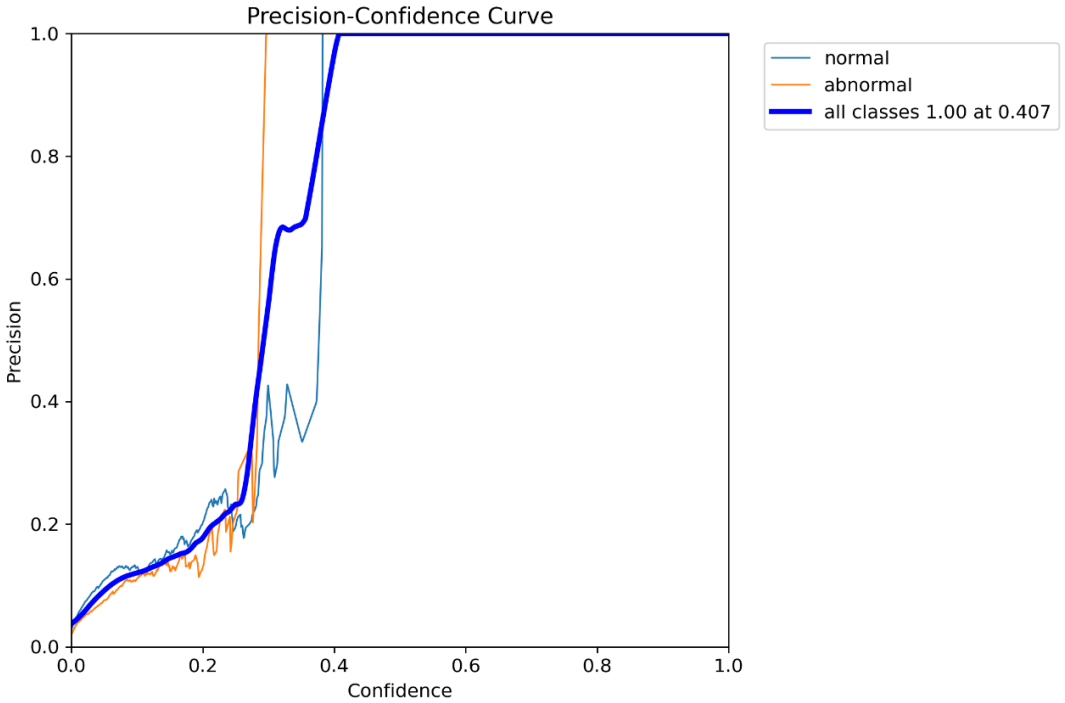


Figure 3.18 Precision Confidence Curve for YOLOv5S head labeled data

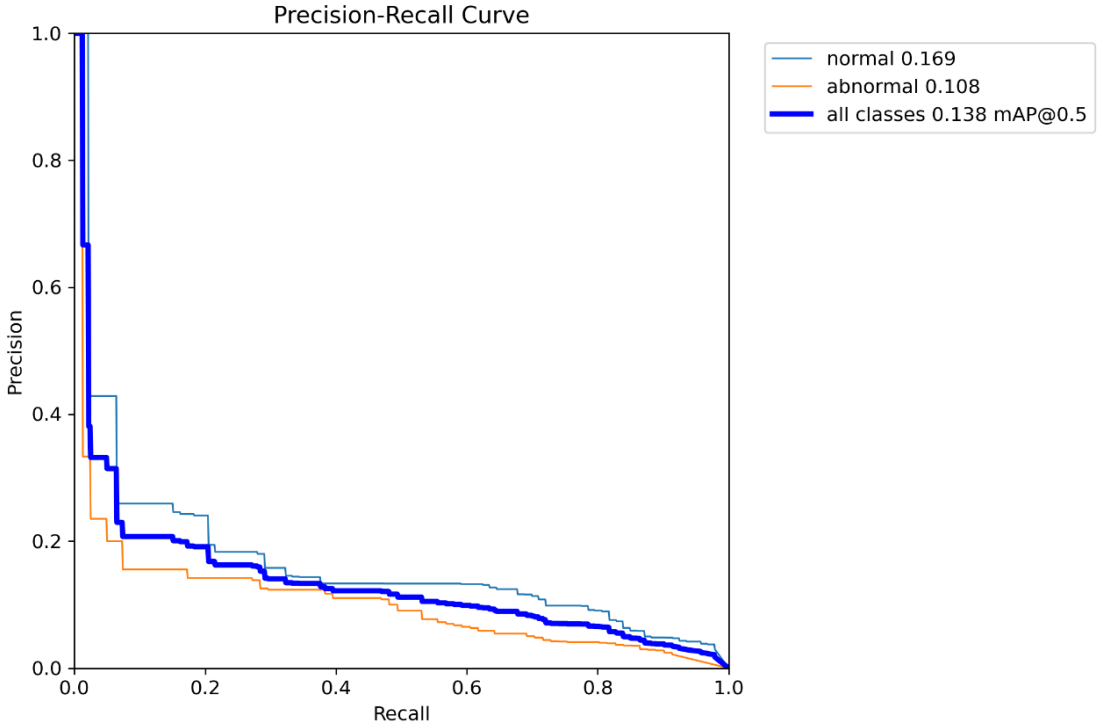


Figure 3.19 Precision-Recall Curve for YOLOv5S head labeled data

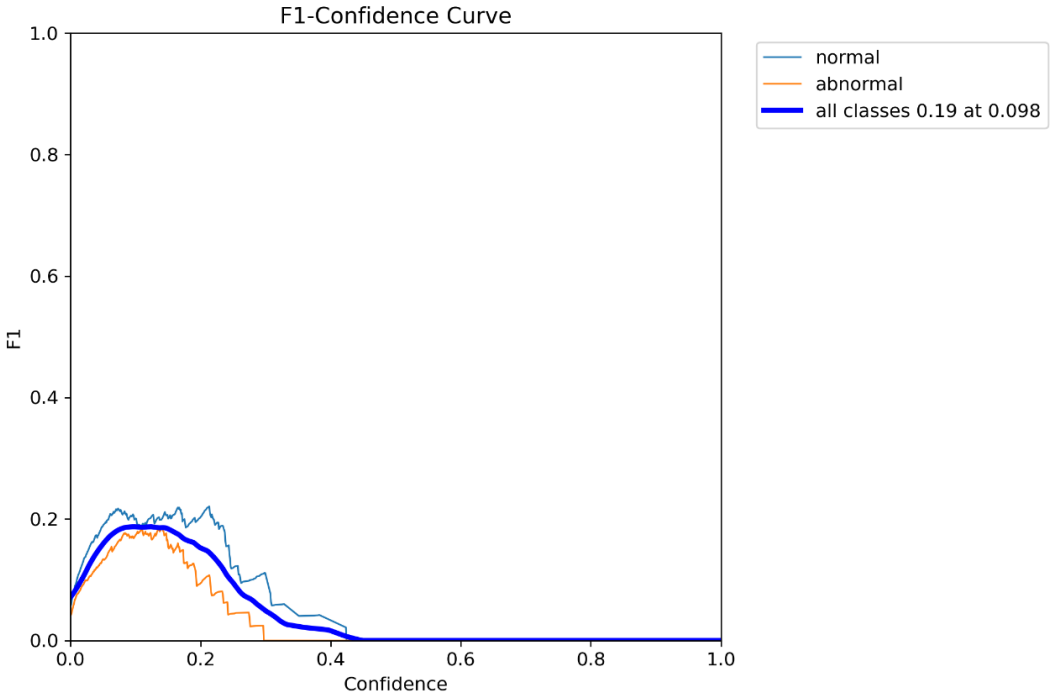


Figure 3.20 F1-Confidence Curve for YOLOv5S head labeled data

3.3.1.3 YOLOv5M

Results of YOLOv5M are as follows.

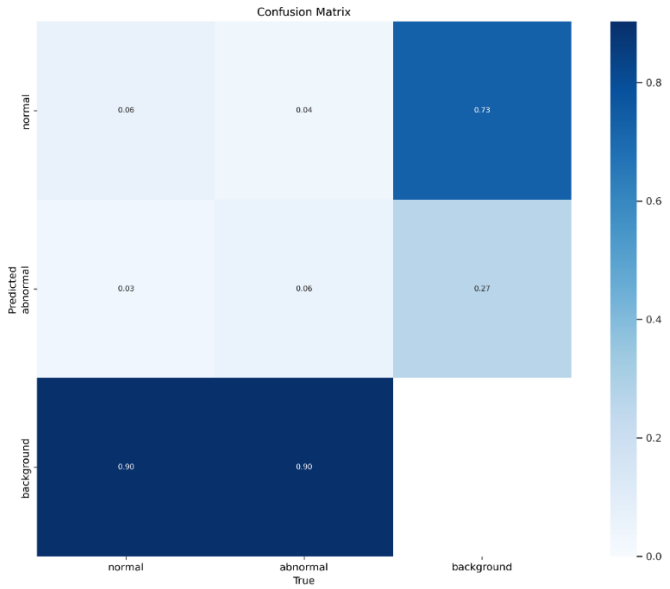


Figure 3.21 Normalized Confusion matrix for YOLOv5M head labeled data

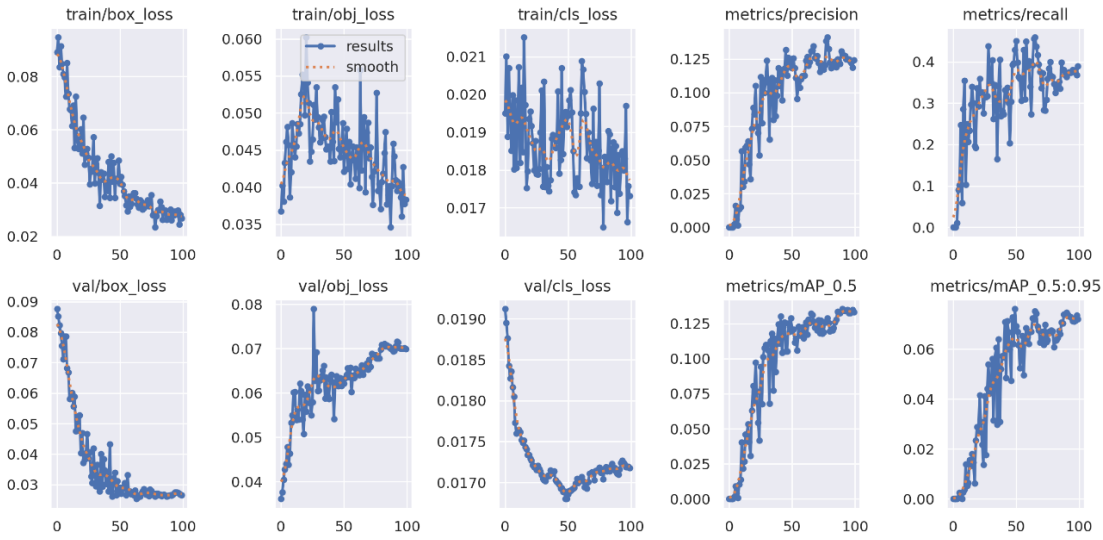


Figure 3.22 Results graph for YOLOv5M head labeled data

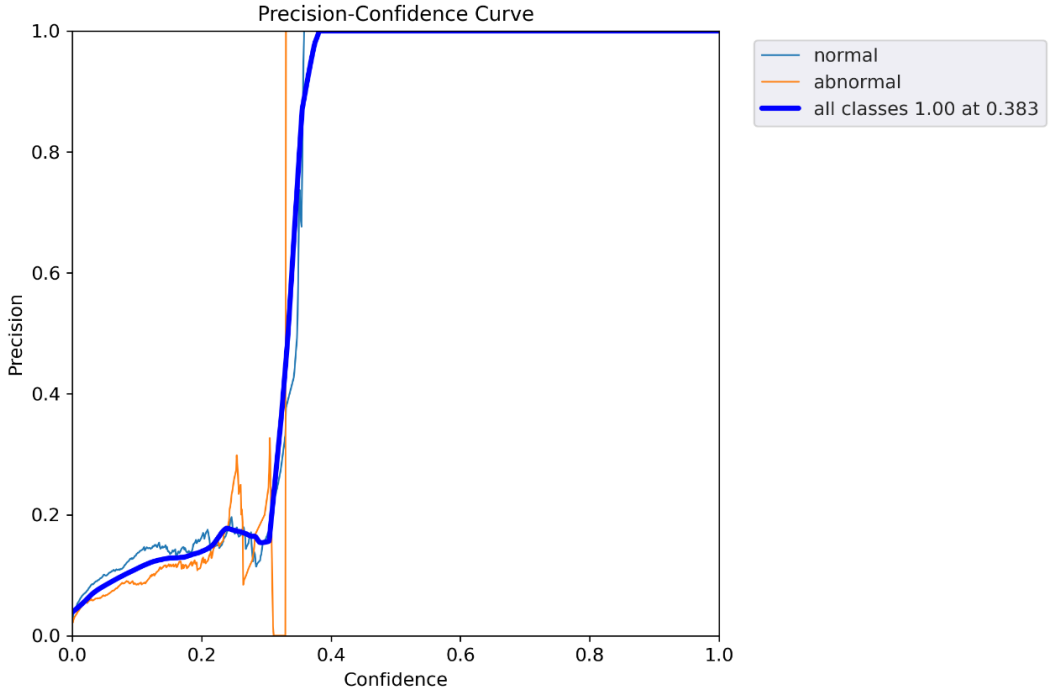


Figure. 3.23 Precision Confidence Curve for YOLOv5M head labeled data

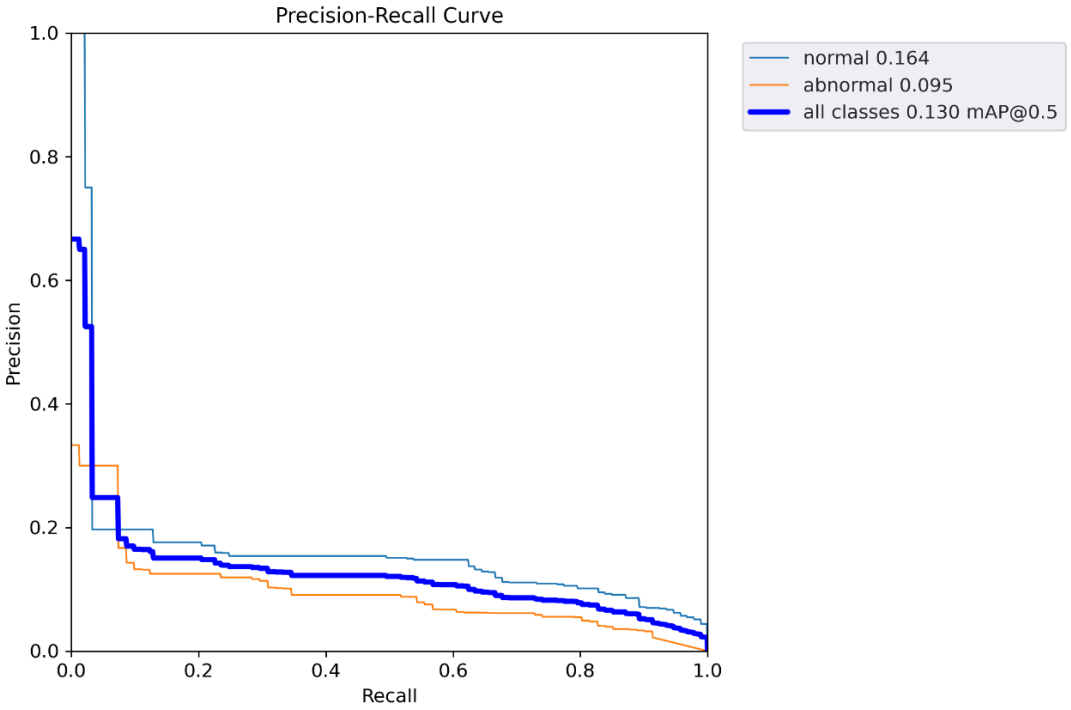


Figure 3.24 Precision-Recall Curve for YOLOv5M head labeled data

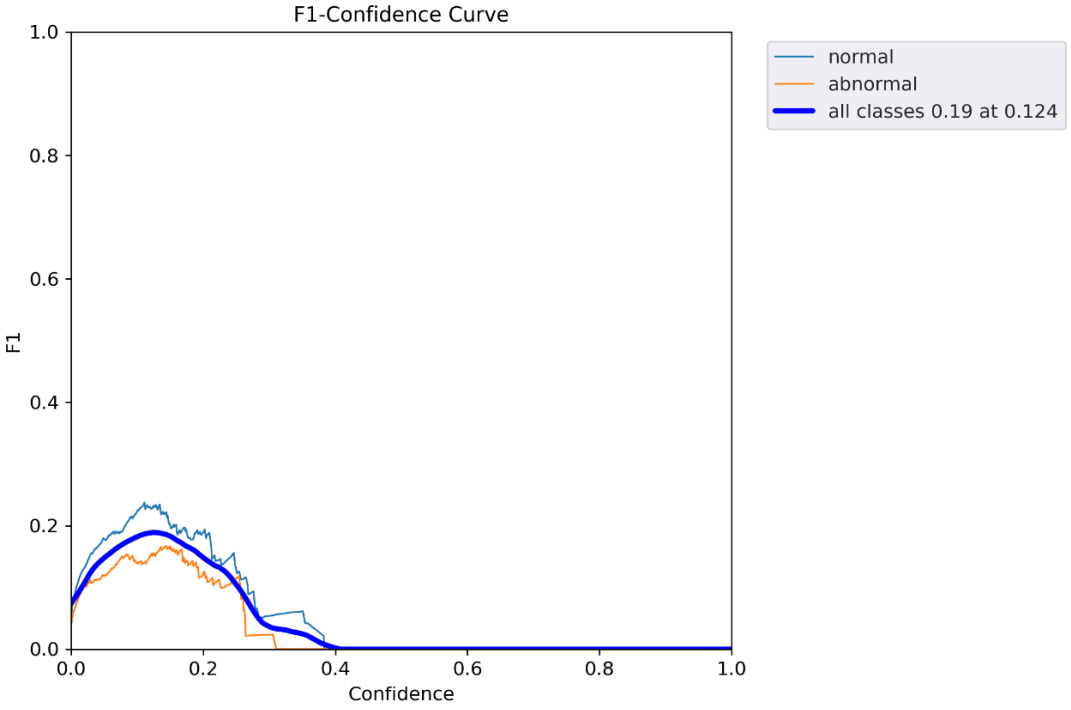


Figure 3.25 F1-Confidence Curve for YOLOv5M head labeled data

3.3.1.4 YOLOv5L

Results of YOLOv5L are as follows.

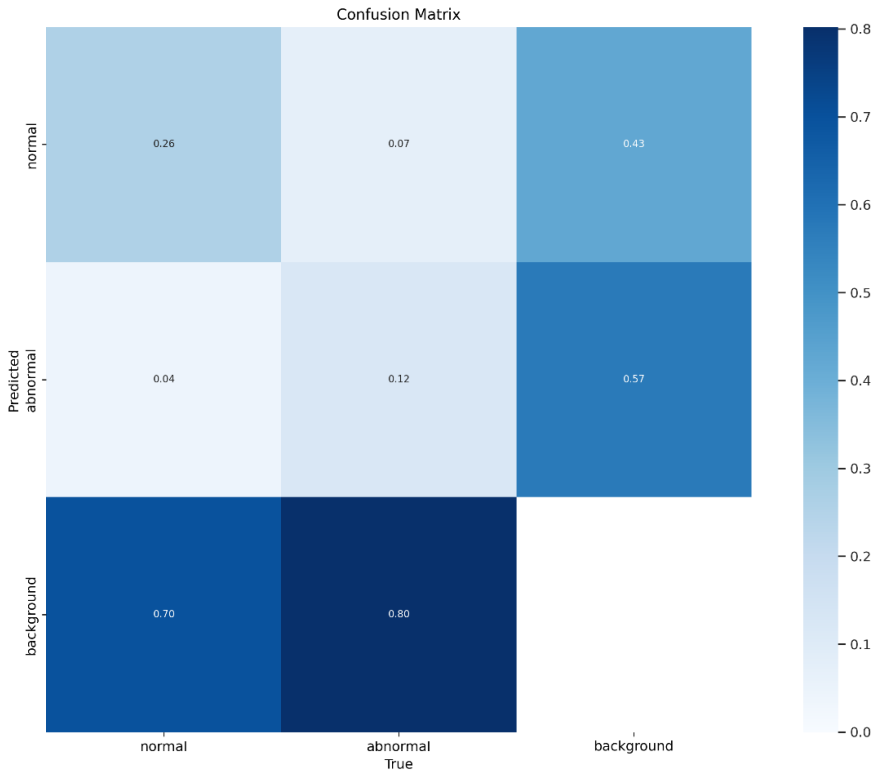


Figure 3.26 Normalized Confusion matrix for YOLOv5L head labeled data

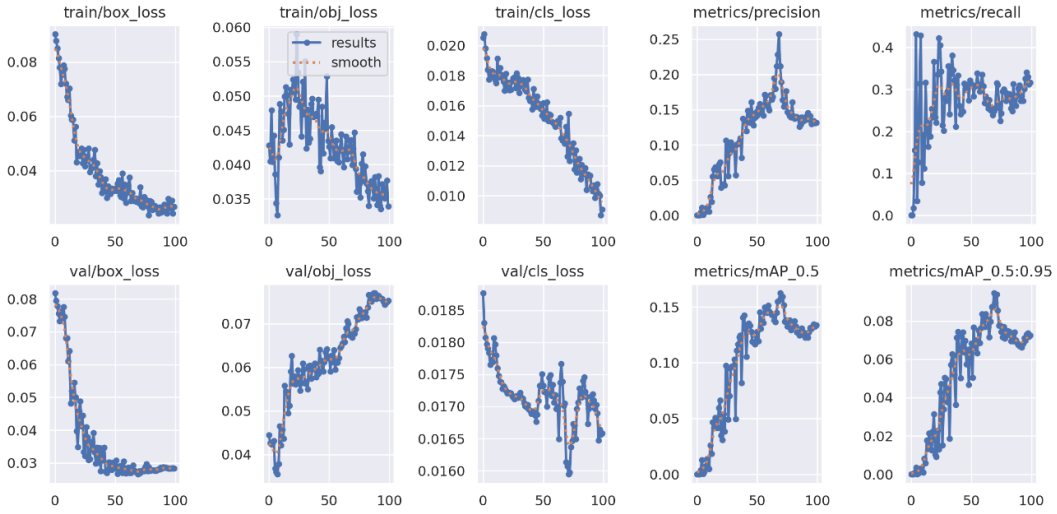


Figure 3.27 Results graph for YOLOv5L head labeled data

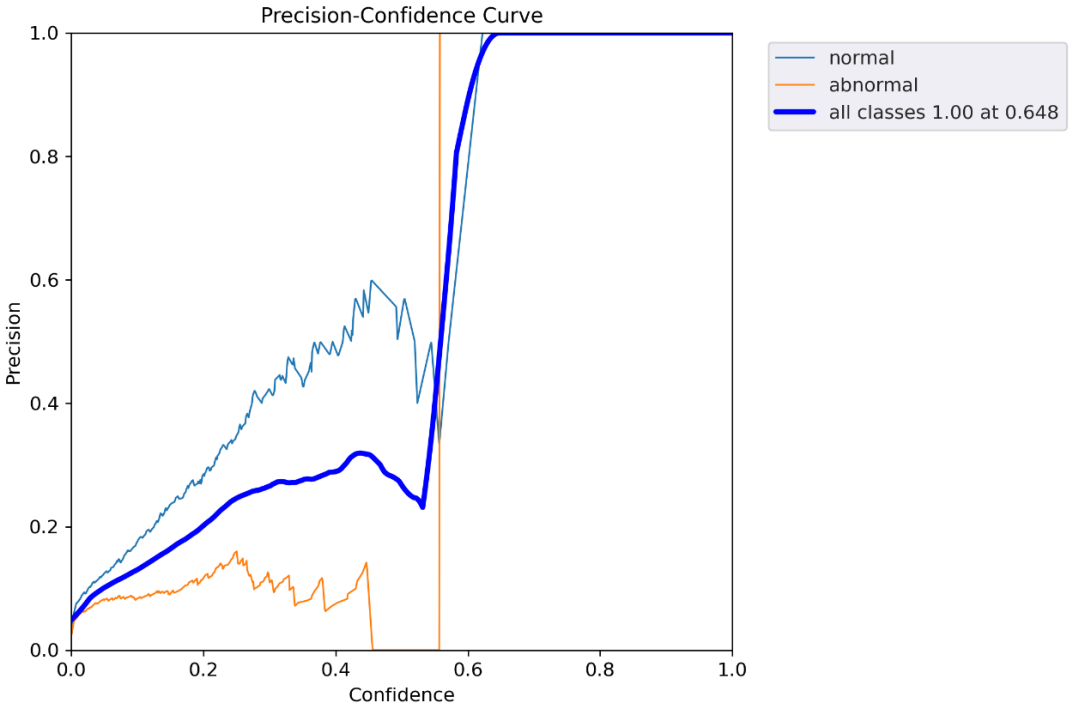


Figure 3.28 Precision Confidence Curve for YOLOv5L head labeled data

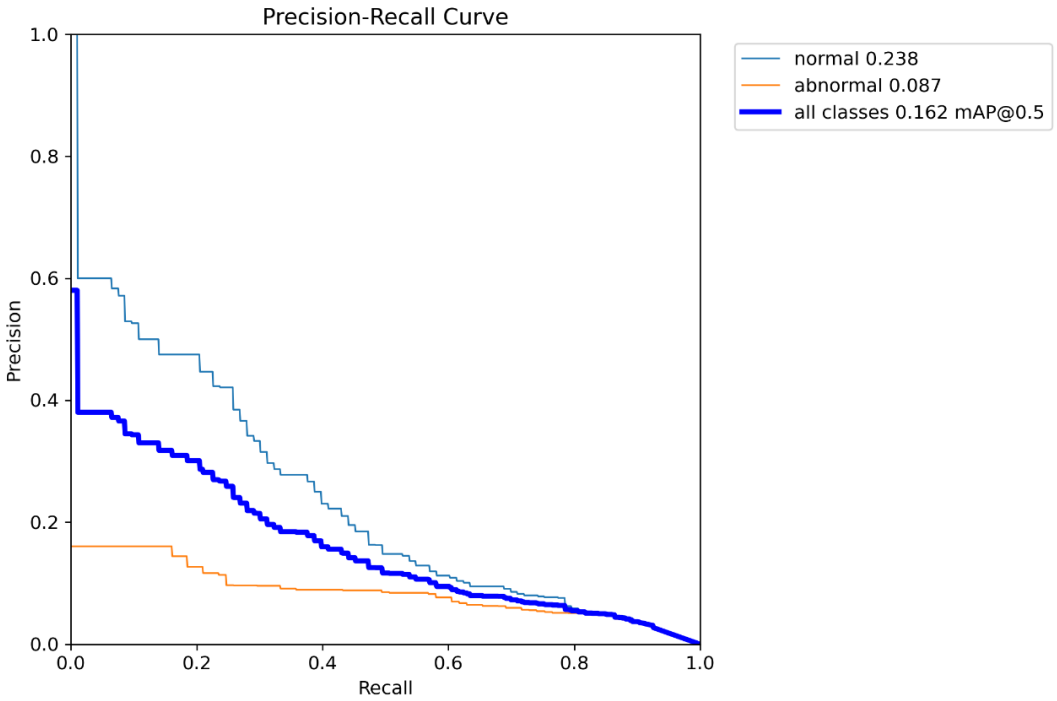


Figure 3.29 Precision-Recall Curve for YOLOv5L head labeled data

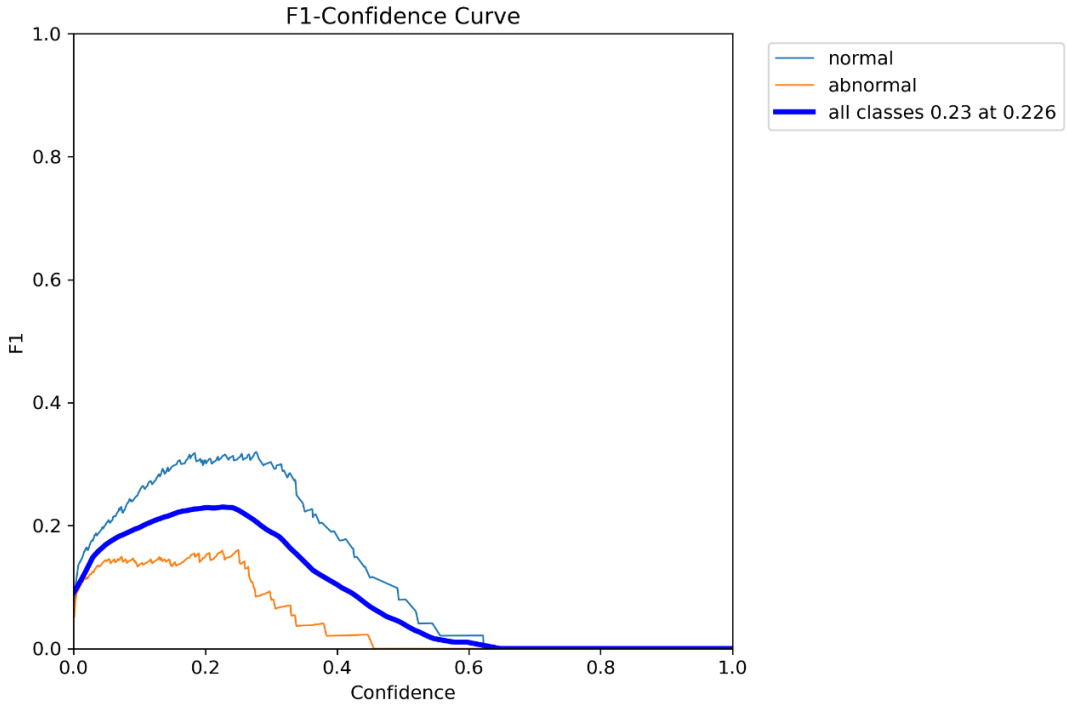


Figure 3.30 F1-Confidence Curve for YOLOv5L head labeled data

3.3.1.5 YOLOv5X

Results of YOLOv5X are as follows.

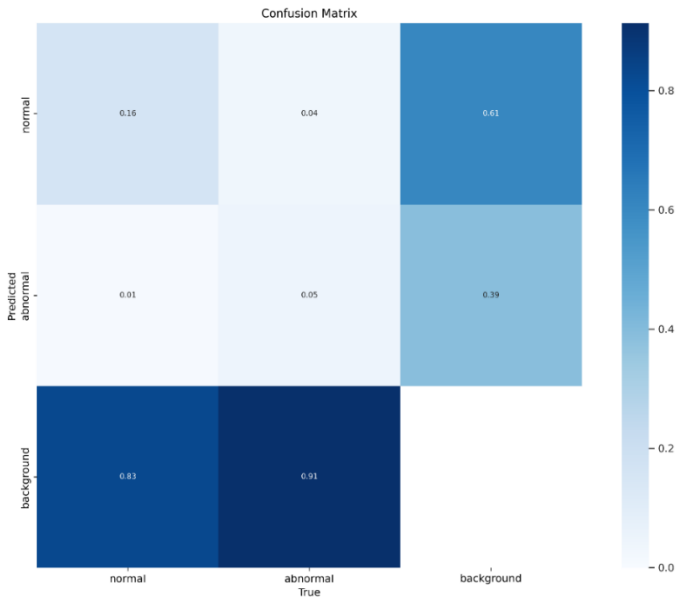


Figure 3.31 Normalized Confusion matrix for YOLOv5X head labeled data

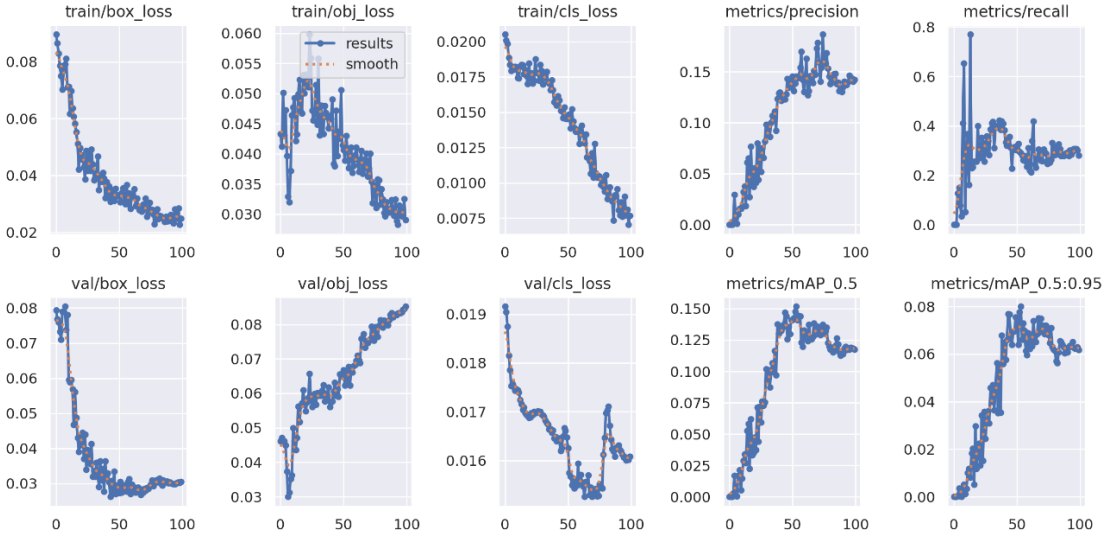


Figure 3.32 Results graph for YOLOv5X head labeled data

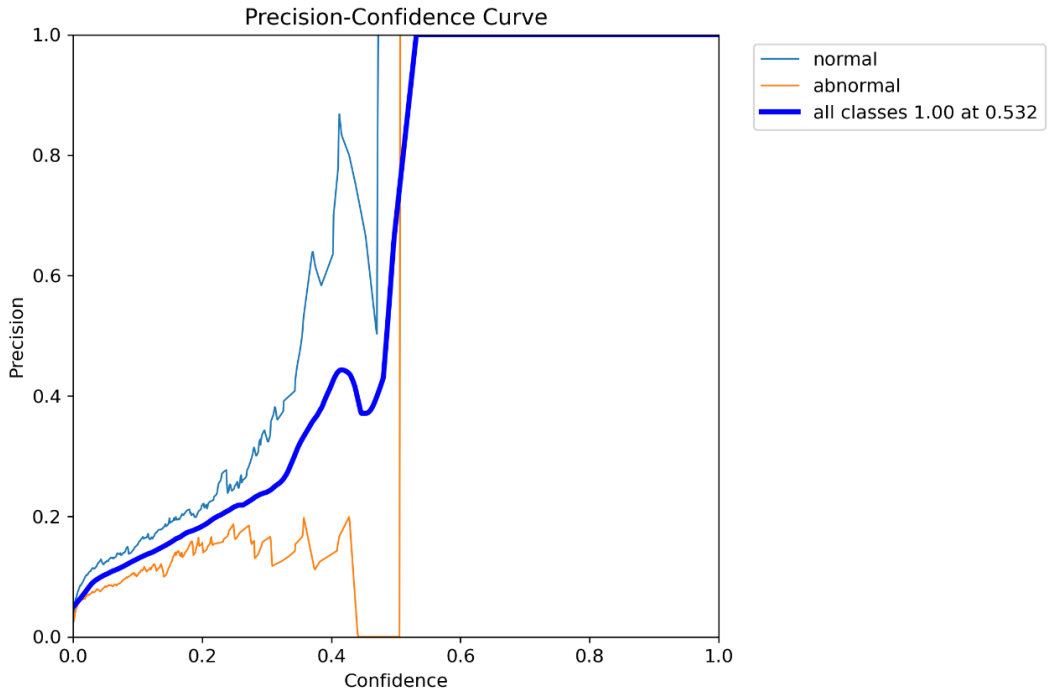


Figure 3.33 Precision Confidence Curve for YOLOv5X head labeled data

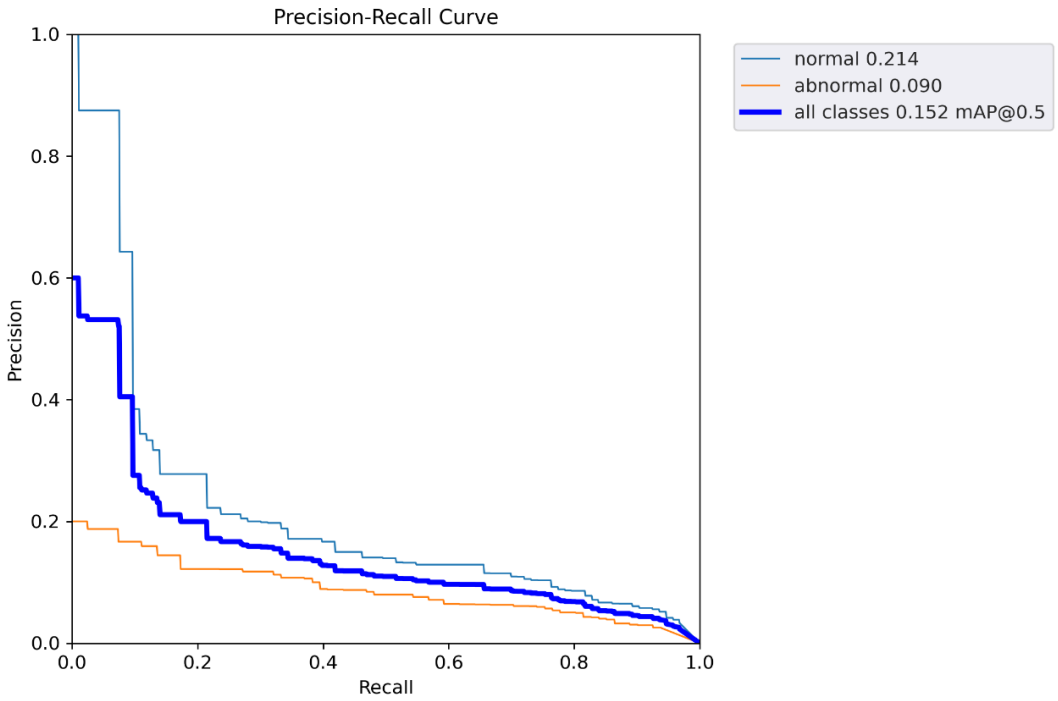


Figure 3.34 Precision-Recall Curve for YOLOv5X head labeled data

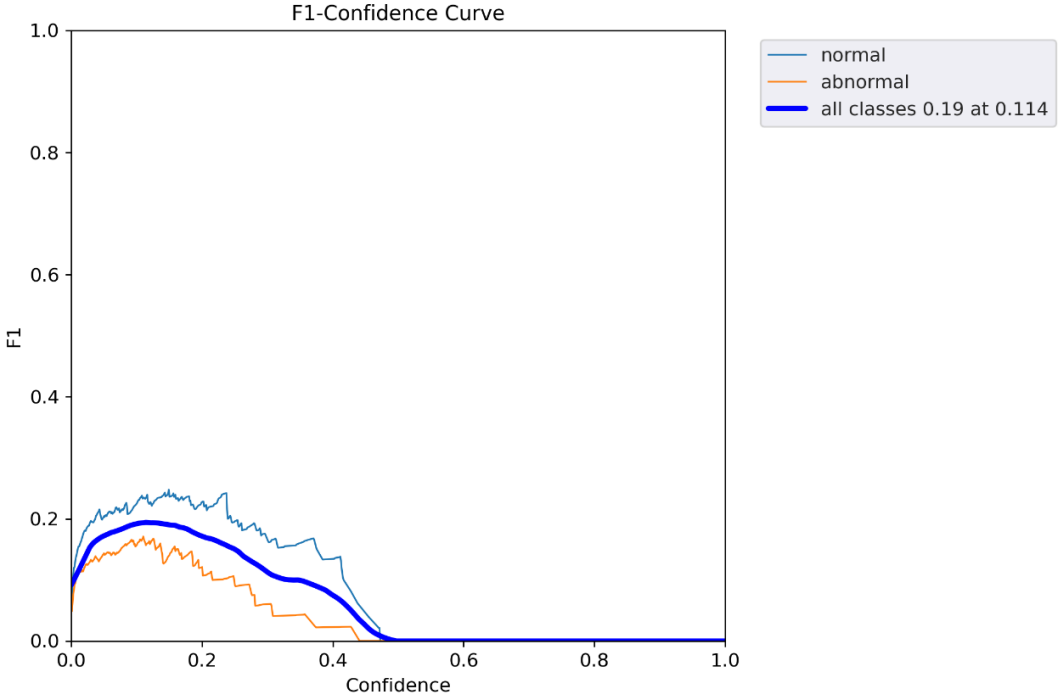


Figure 3.35 F1-Confidence Curve for YOLOv5X head labeled data

For each model, Precision-Recall, Precision-Confidence, and F1-Confidence curves have been examined**. Precision-Recall curves** for models trained with head-only labeled data have been separately drawn for normal and abnormal classes. Additionally, an average curve has been plotted for all classes. The curve for the normal class exhibits high precision and low recall values, indicating that the model excels at correctly classifying normal examples but performs poorly in classifying abnormal examples accurately. The average curve falls between the curves for normal and abnormal classes, demonstrating the overall good performance of the model. However, it is crucial to note that the model does not classify both classes equally well.

In the same models, **Precision-Confidence curves** have been separately plotted for normal and abnormal classes. When considering the AUC value, the models exhibit AUC values between 0.90 and 1. Consequently, based on the evaluation criteria for machine learning models, it has been determined that the models perform successfully. However, it is noteworthy that the curve for the normal class (depicted by the orange line) is higher than the curve for the abnormal class (depicted by the green line). This indicates that the models classifies the normal class better than the abnormal class.

F1 Score-Confidence curves have been separately plotted for normal and abnormal classes. Here, on average, for all models, the F1 score of the model increases as confidence levels rise. This indicates that the model performs better in classifications with higher confidence levels. Additionally, the F1 score for the normal class is higher than that for the abnormal class. This suggests that the model can distinguish the normal class better than the abnormal class.

3.3.2 Results of YOLO models for all sperm labeled data

3.3.2.1 YOLOv5N

Results of YOLOv5N are as follows.

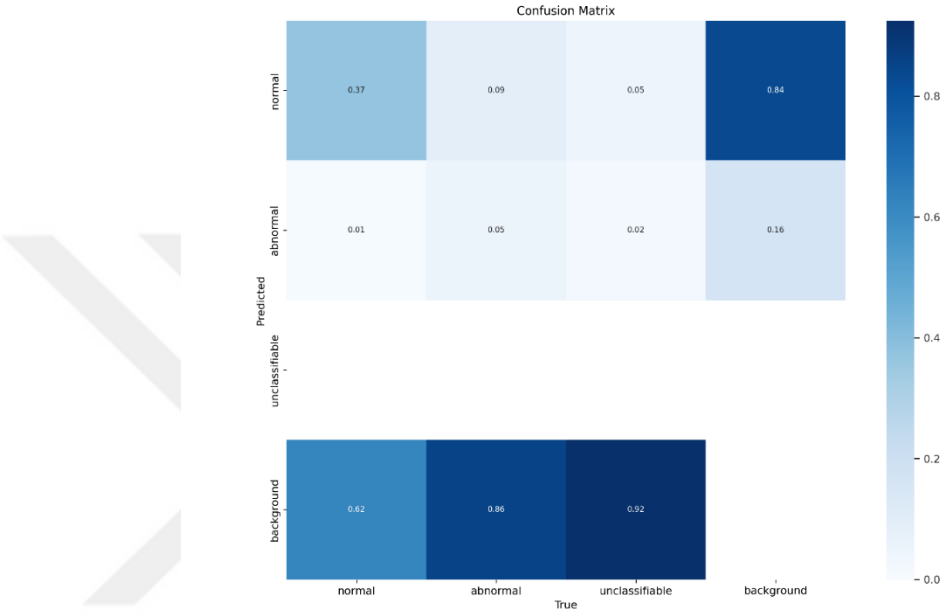


Figure 3.36 Normalized Confusion matrix for YOLOv5N all sperm labeled data

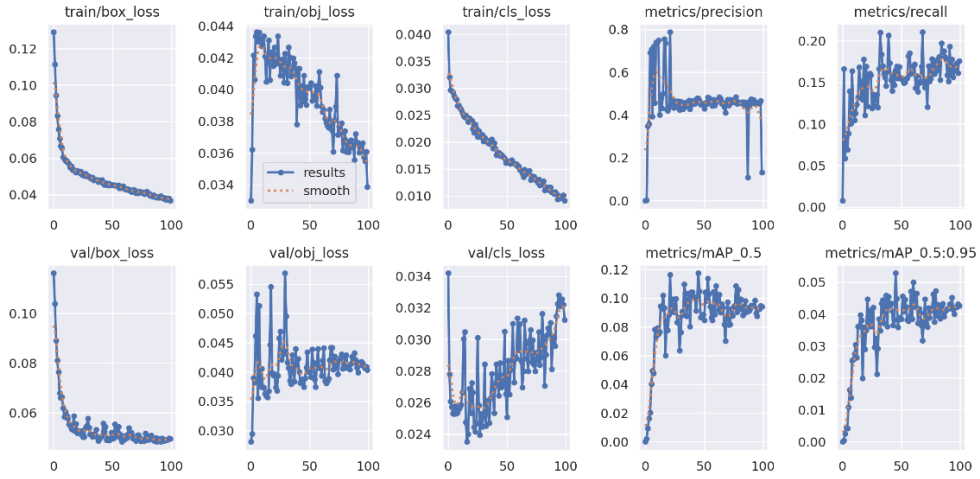


Figure 3.37 Results graph for YOLOv5N all sperm labeled data

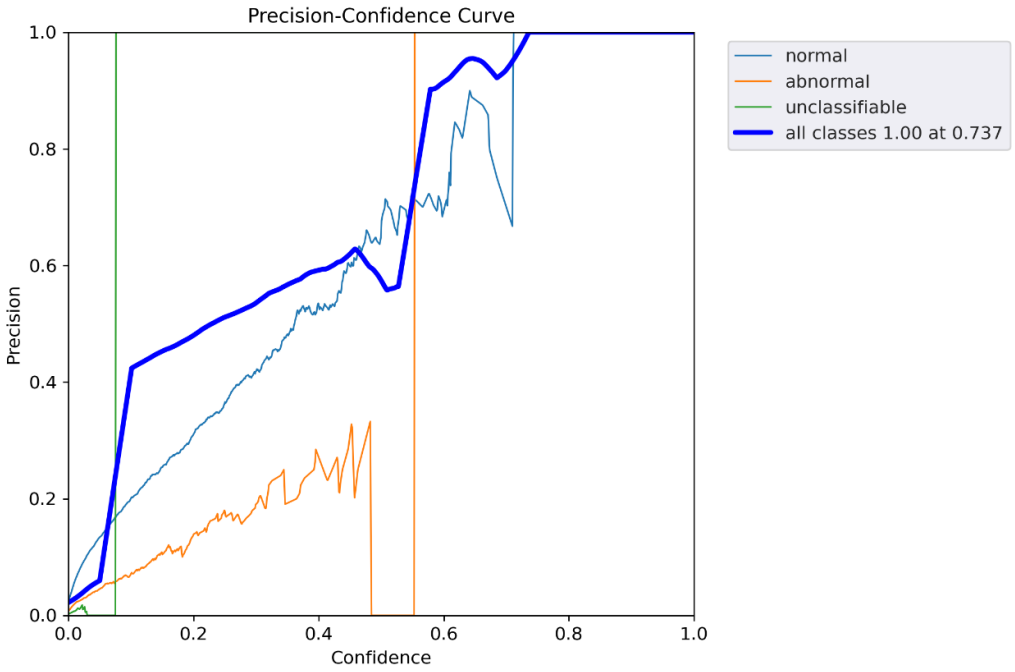


Figure 3.38 Precision Confidence Curve for YOLOv5N all sperm labeled data

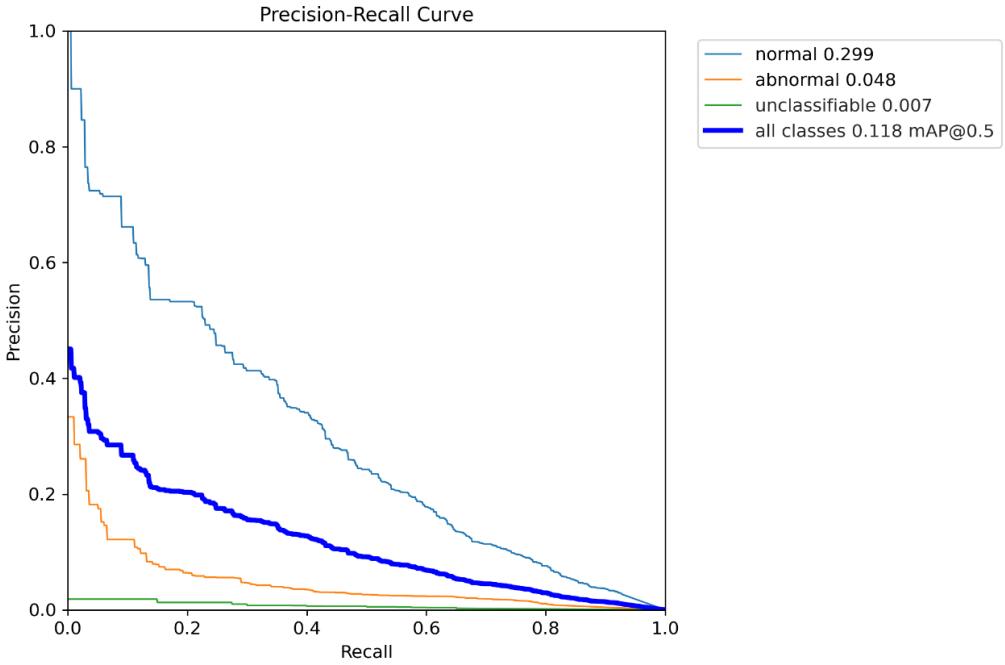


Figure 3.39 Precision-Recall Curve for YOLOv5N all sperm labeled data

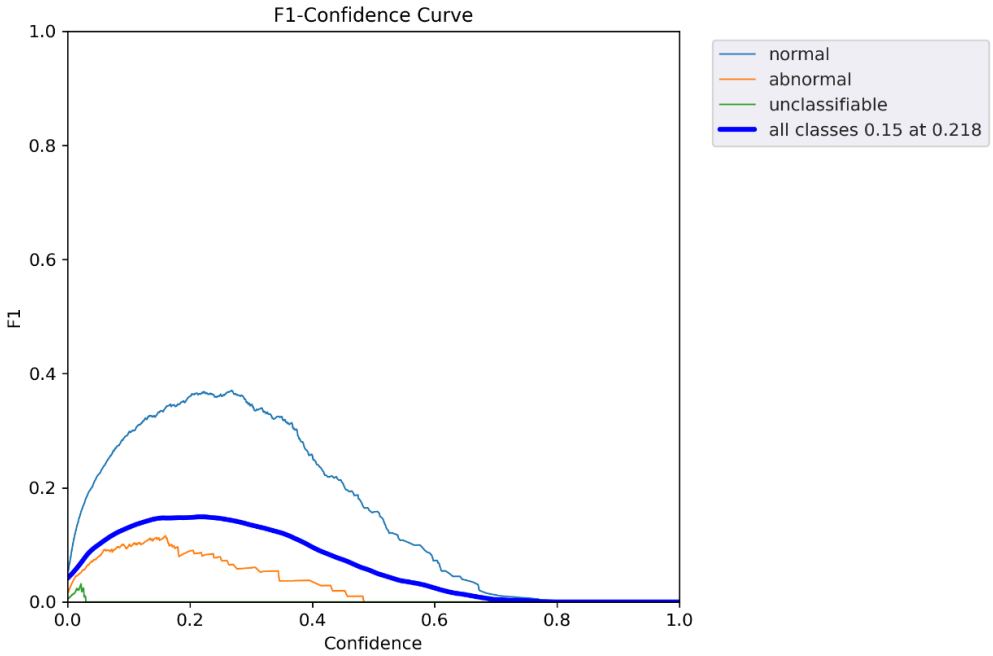


Figure 3.40 F1-Confidence Curve for YOLOv5N all sperm labeled data

3.3.2.2 YOLOv5S

Results of YOLOv5S are as follows.

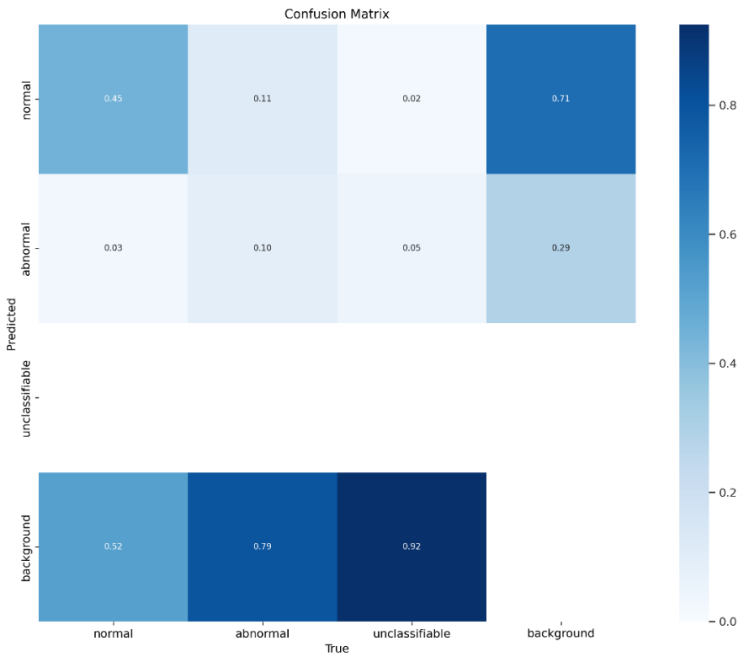


Figure 3.41 Normalized Confusion matrix for YOLOv5S all sperm labeled data

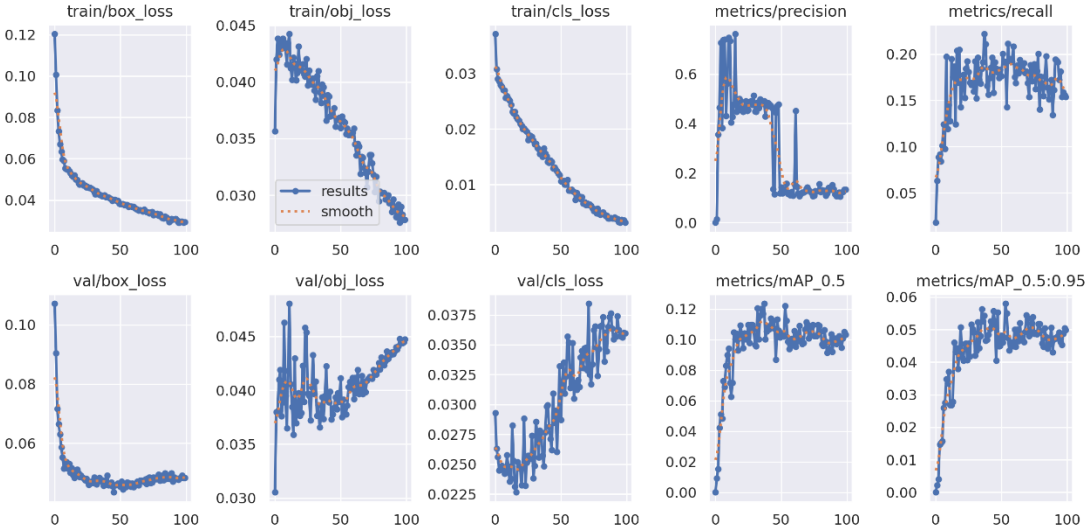


Figure 3.42 Results graph for YOLOv5S all sperm labeled data

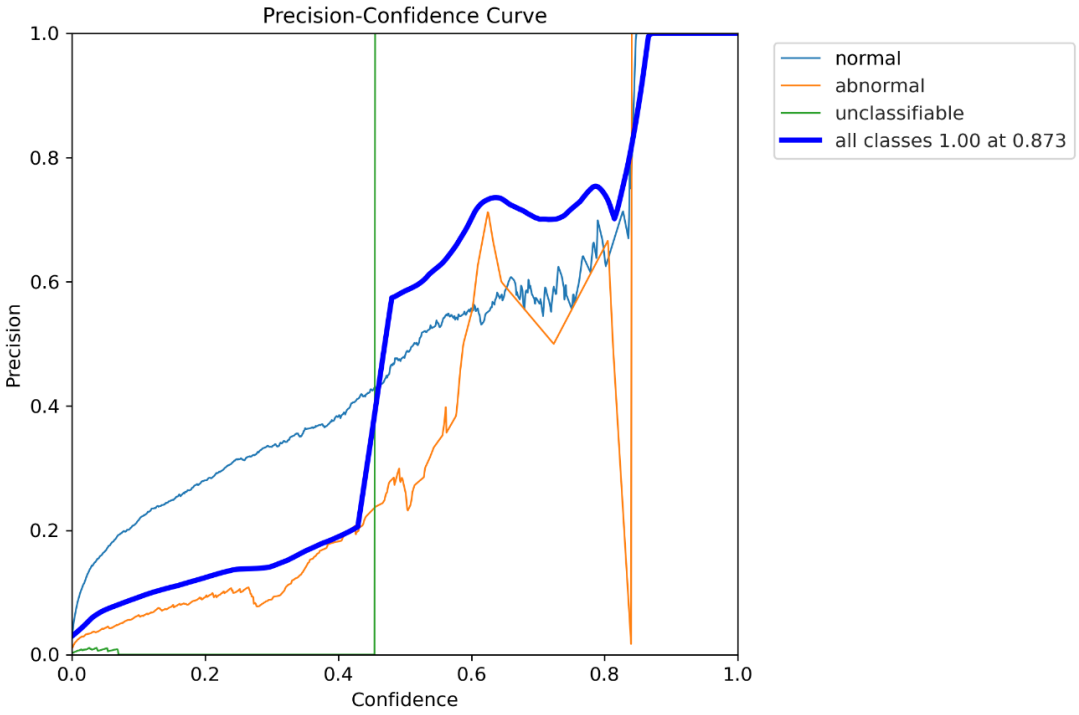


Figure 3.43 Precision Confidence Curve for YOLOvS all sperm labeled data

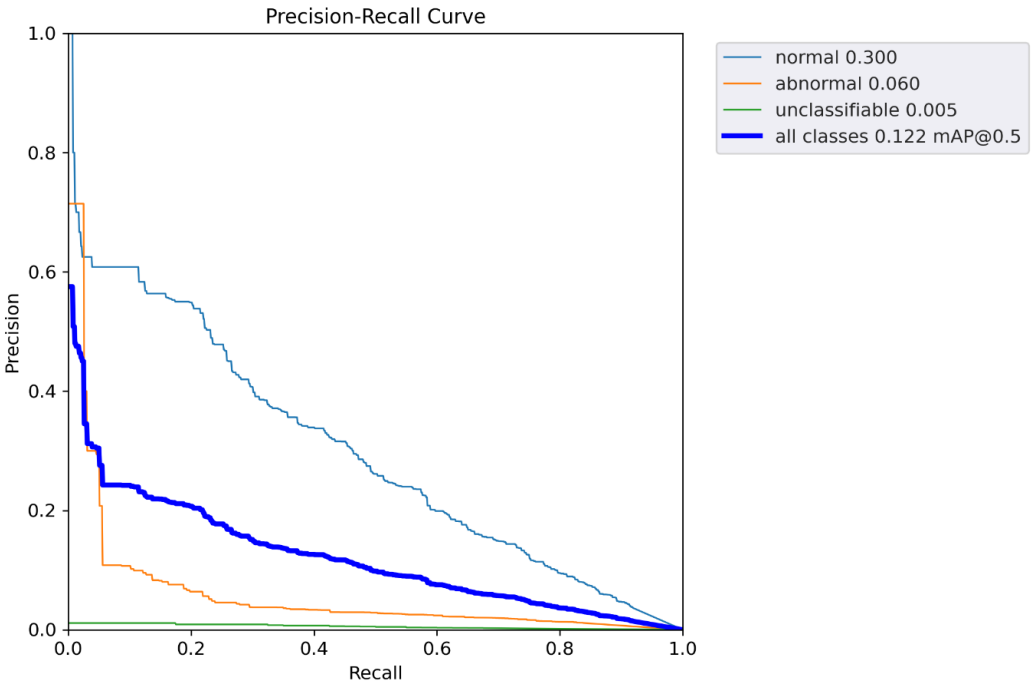


Figure 3.44 Precision-Recall Curve for YOLOv5S all sperm labeled data

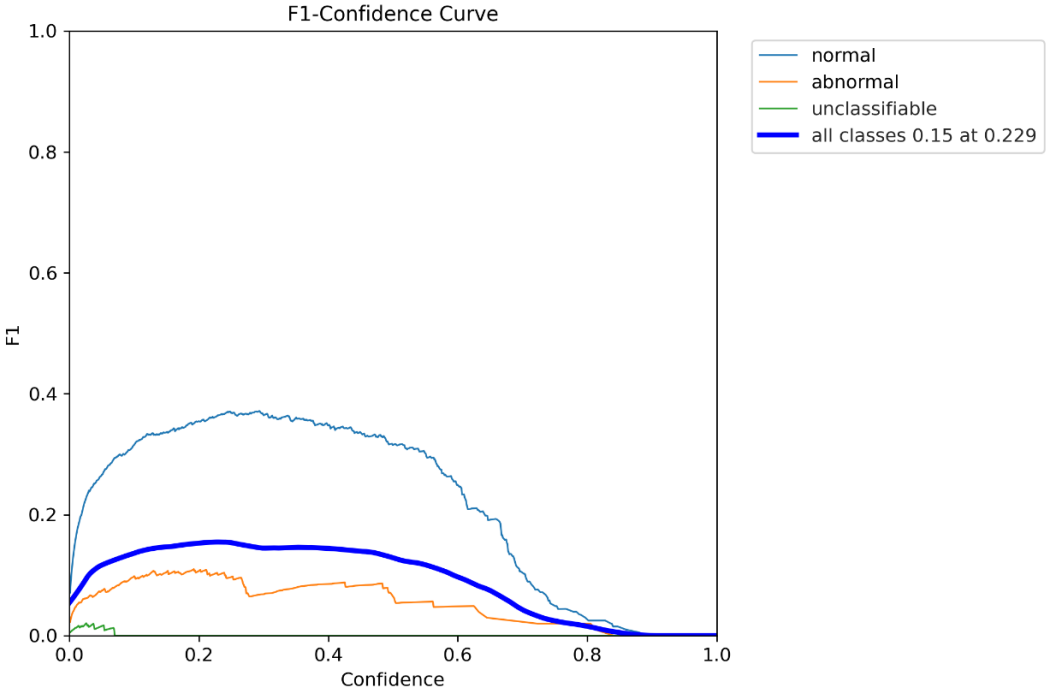


Figure 3.45 F1-Confidence Curve for YOLOv5S all sperm labeled data

3.3.2.3 YOLOv5M

Results of YOLOv5M are as follows.

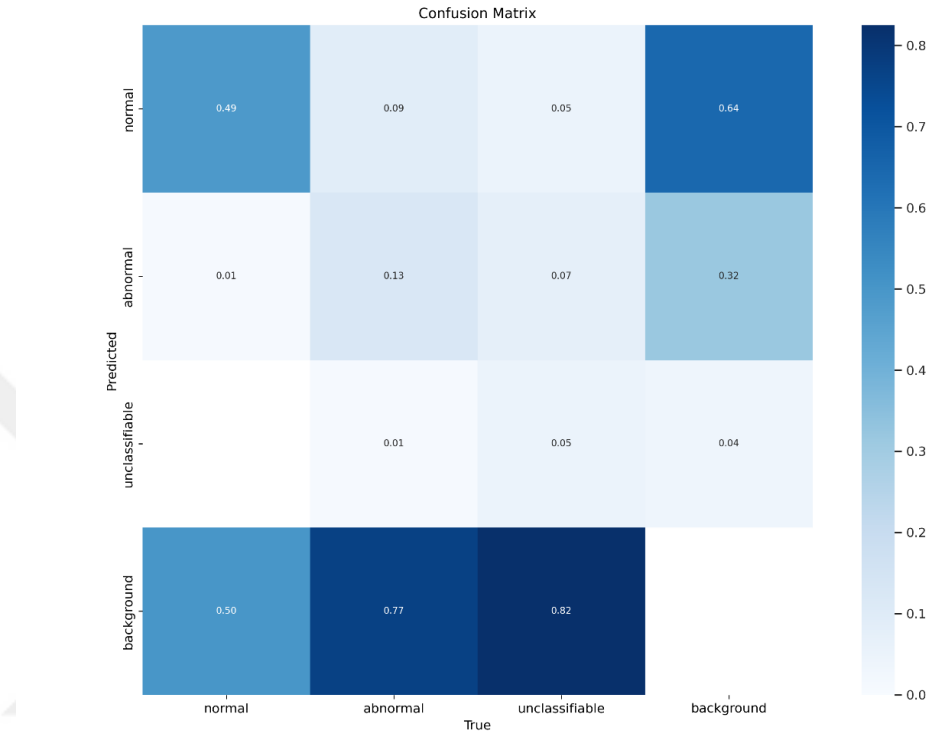


Figure 3.46 Normalized Confusion matrix for YOLOv5M all sperm labeled data

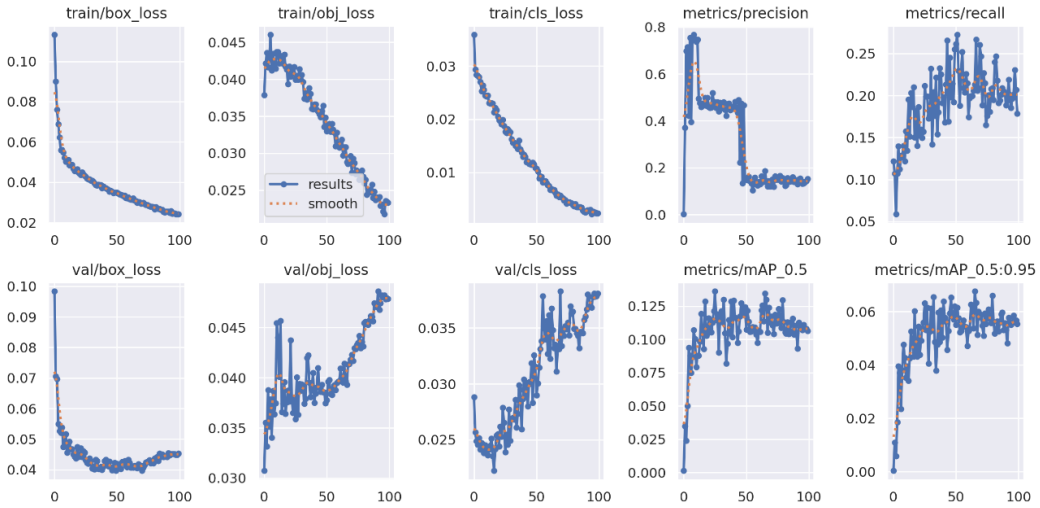


Figure 3.47 Results graph for YOLOv5M all sperm labeled data

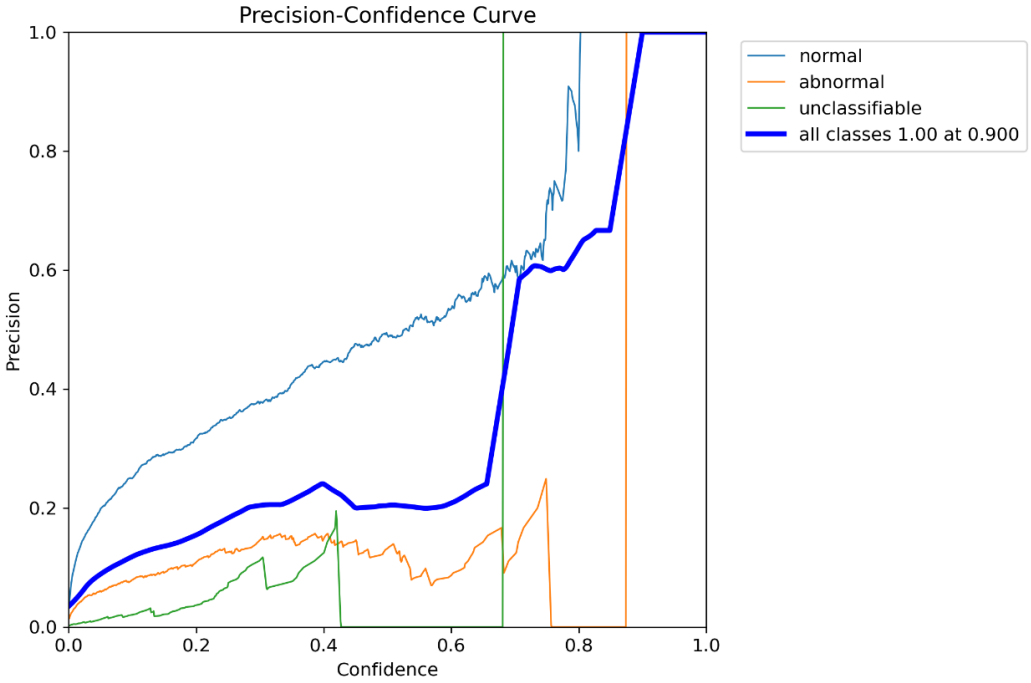


Figure 3.48 Precision Confidence Curve for YOLOvM all sperm labeled data

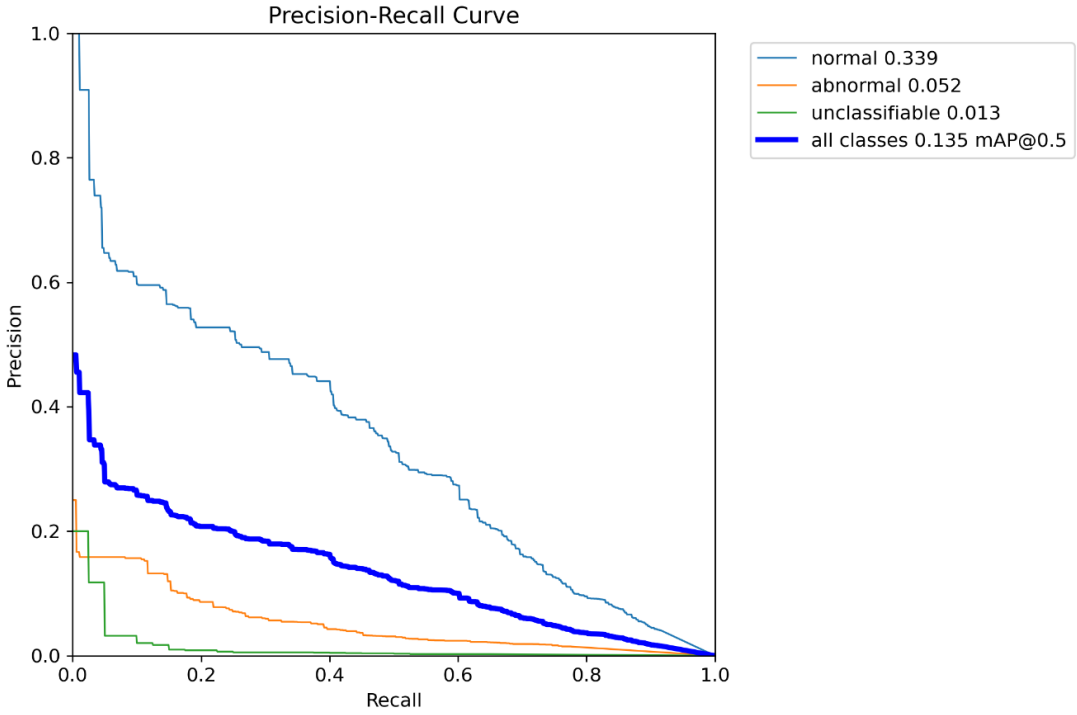


Figure 3.49 Precision-Recall Curve for YOLOv5M all sperm labeled data

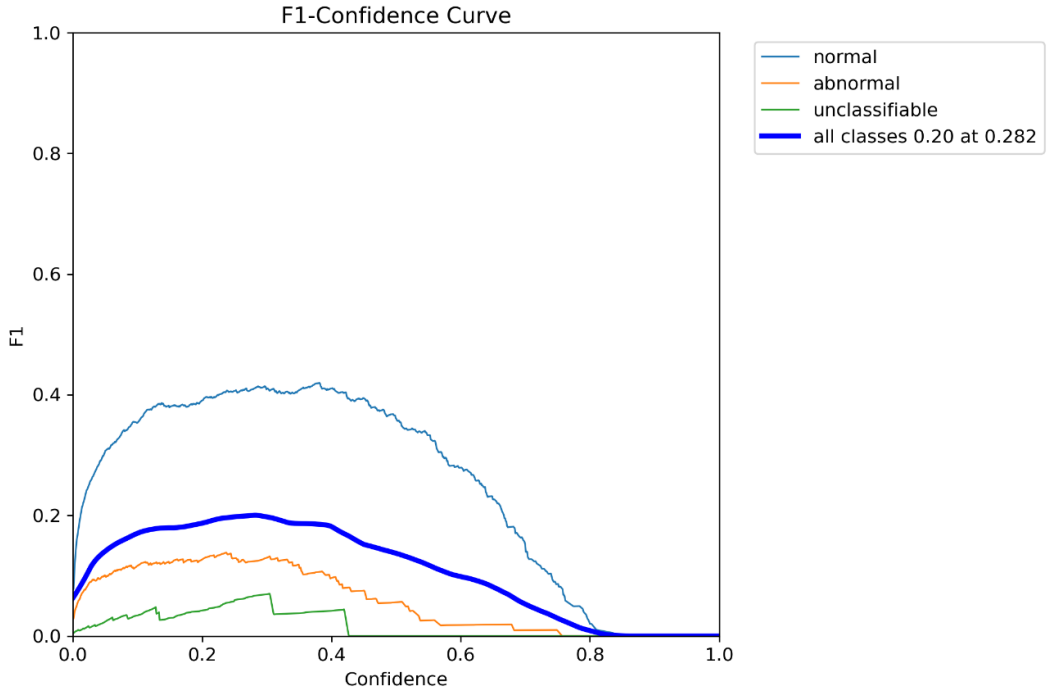


Figure 3.50 F1-Confidence Curve for YOLOv5M all sperm labeled data

3.3.2.4 YOLOv5L

Results of YOLOv5L are as follows.

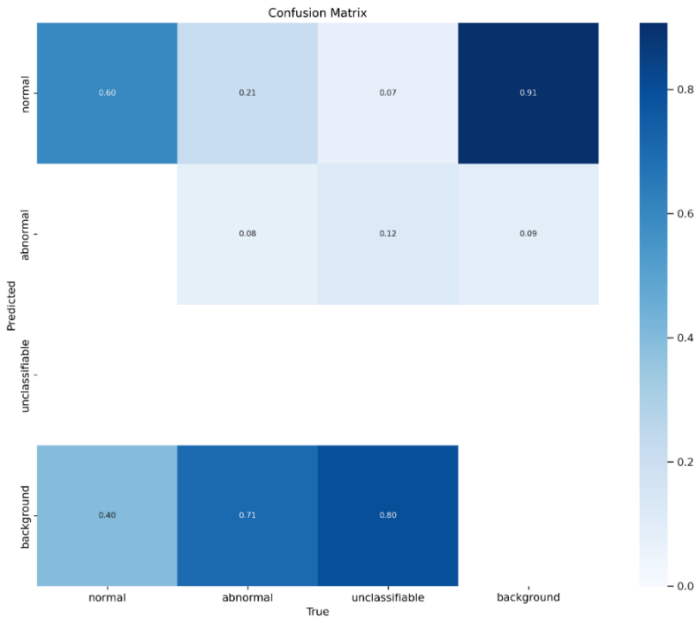


Figure 3.51 Normalized Confusion matrix for YOLOv5L all sperm labeled data

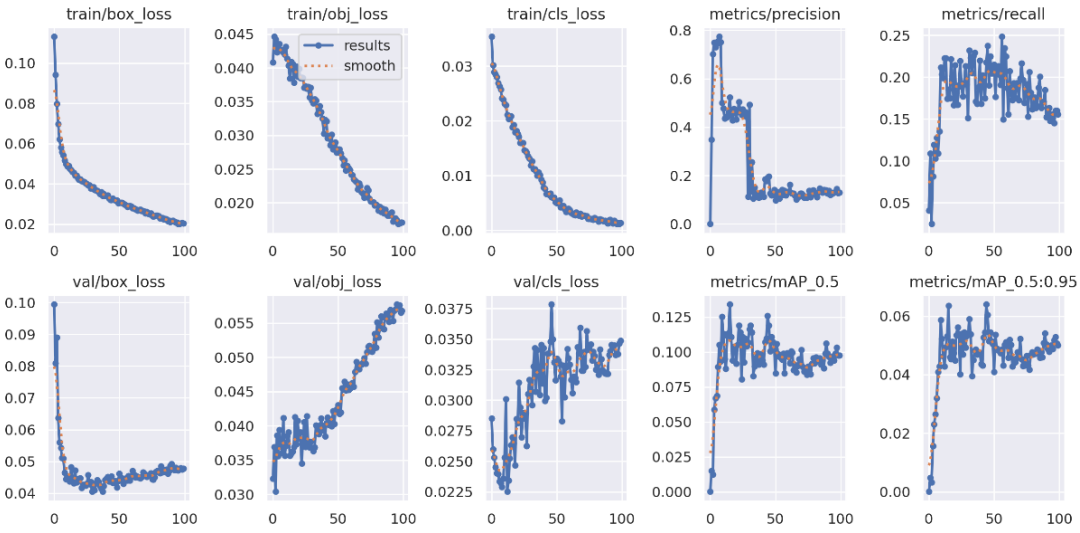


Figure 3.52 Results graph for YOLOv5L all sperm labeled data

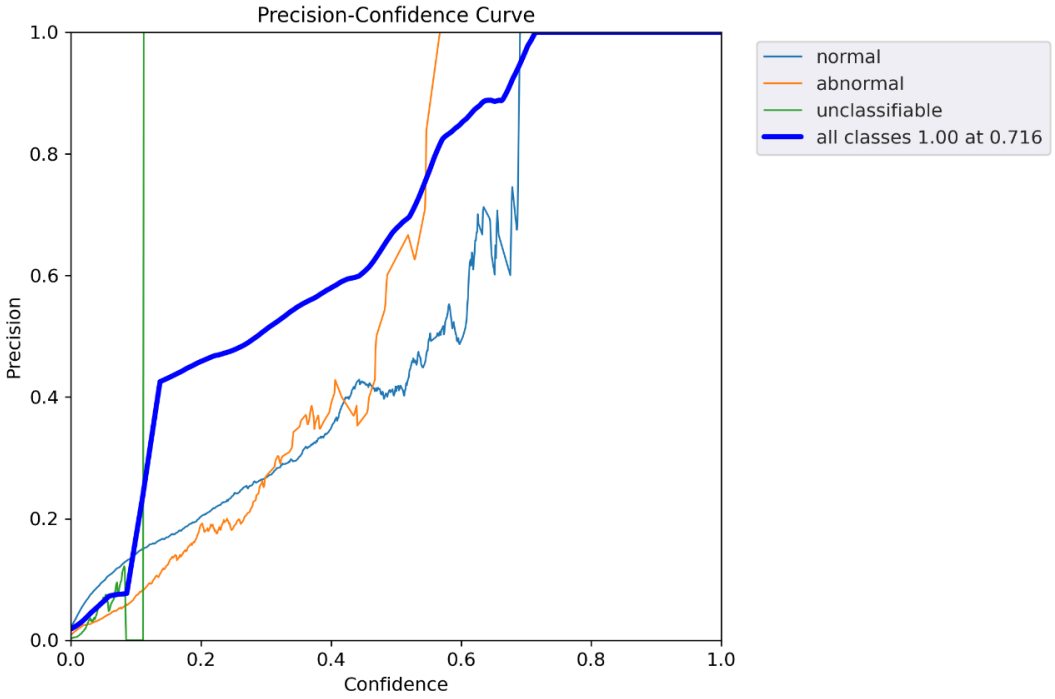


Figure 3.53 Precision Confidence Curve for YOLOv5L all sperm labeled data

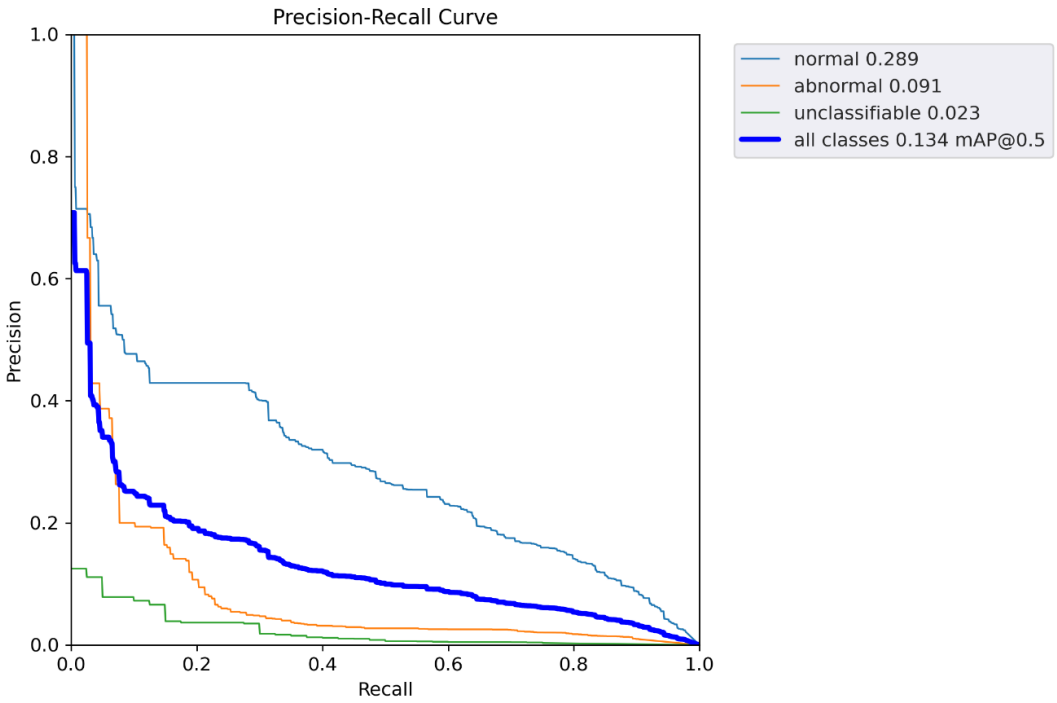


Figure. 3.54 Precision-Recall Curve for YOLOv5L all sperm labeled data

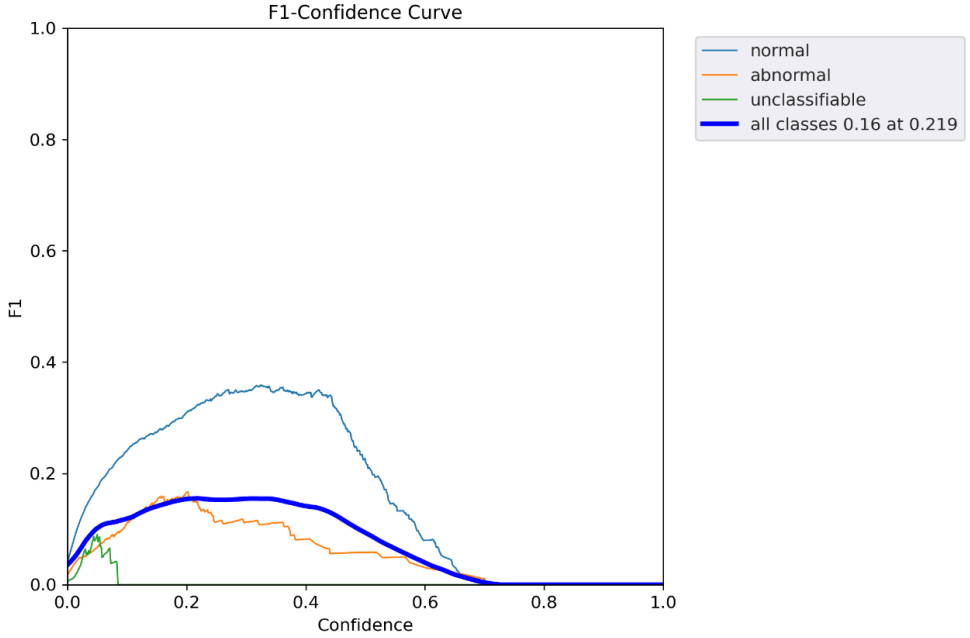


Figure 3.55 F1-Confidence Curve for YOLOv5L all sperm labeled data

3.3.2.5 YOLOv5X

Results of YOLOv5X are as follows.

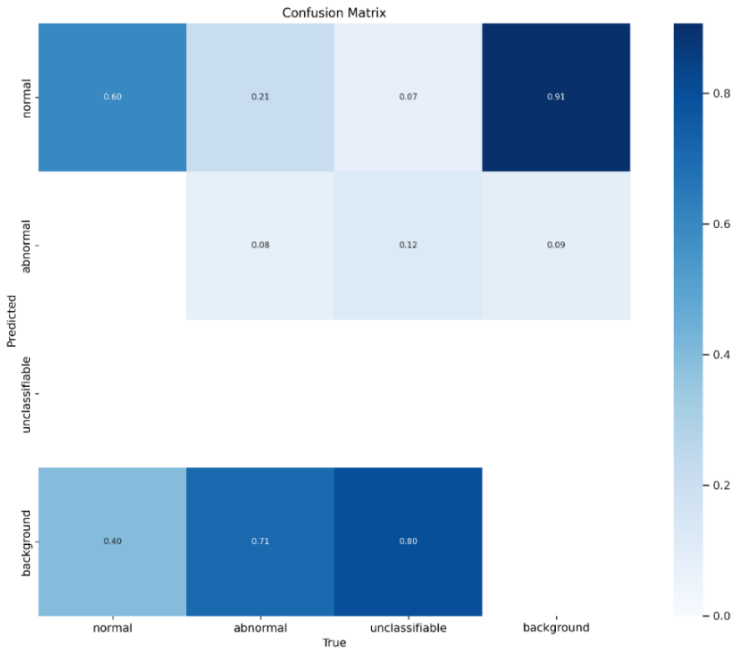


Figure 3.56 Normalized Confusion matrix for YOLOv5N all sperm labeled data

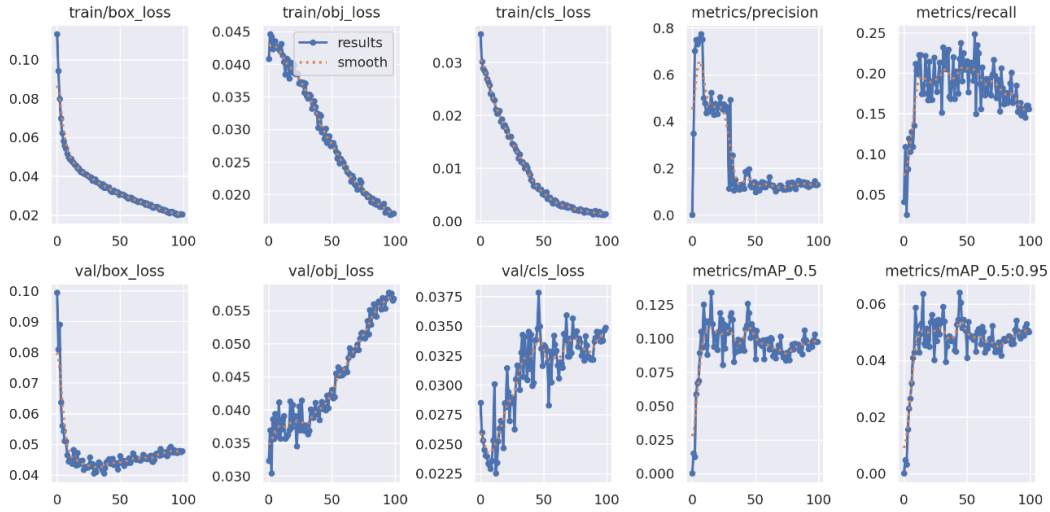


Figure 3.57 Results graph for YOLOv5X all sperm labeled data

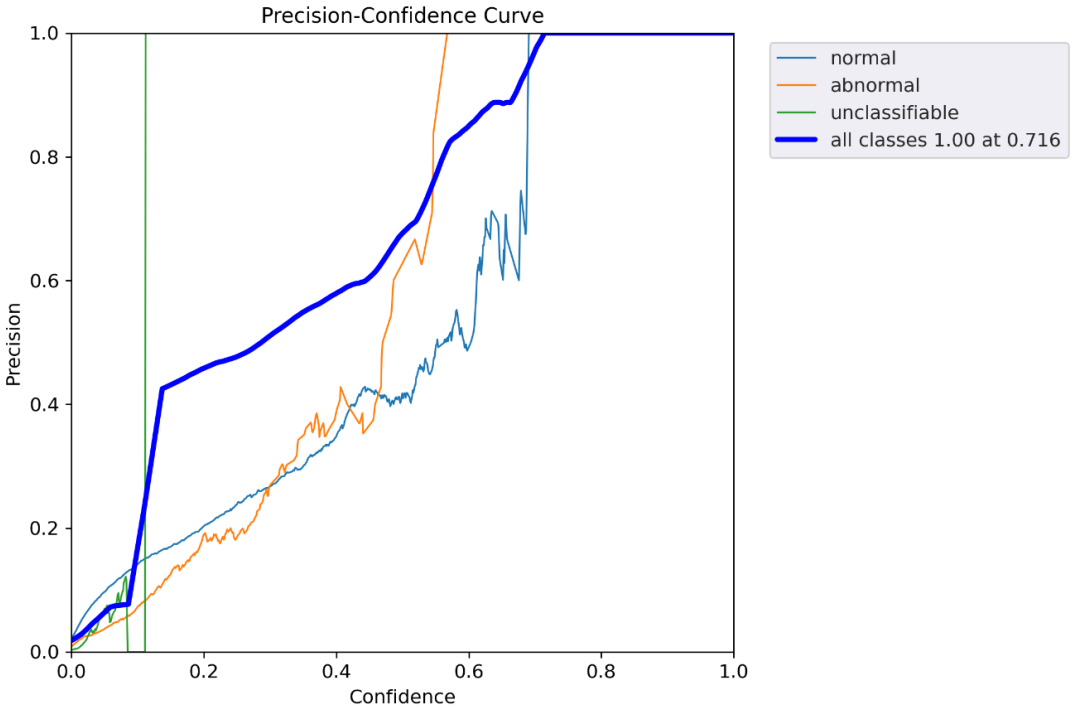


Figure 3.58 Precision Confidence Curve for YOLOv5N all sperm labeled data

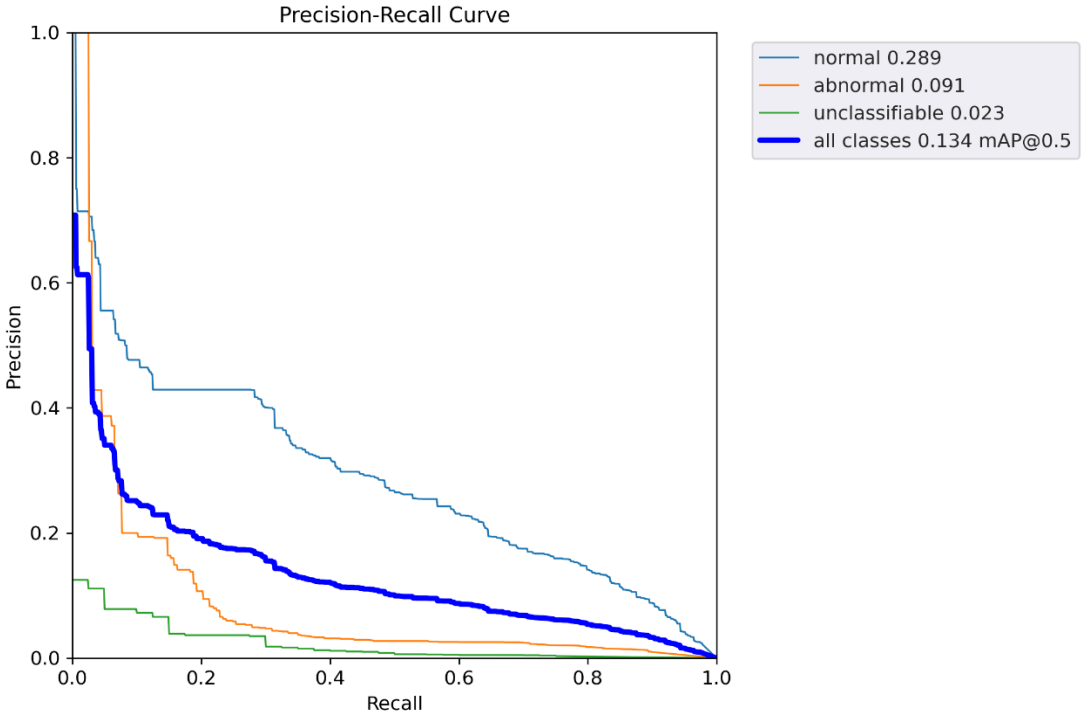


Figure 3.59 Precision-Recall Curve for YOLOv5N all sperm labeled data

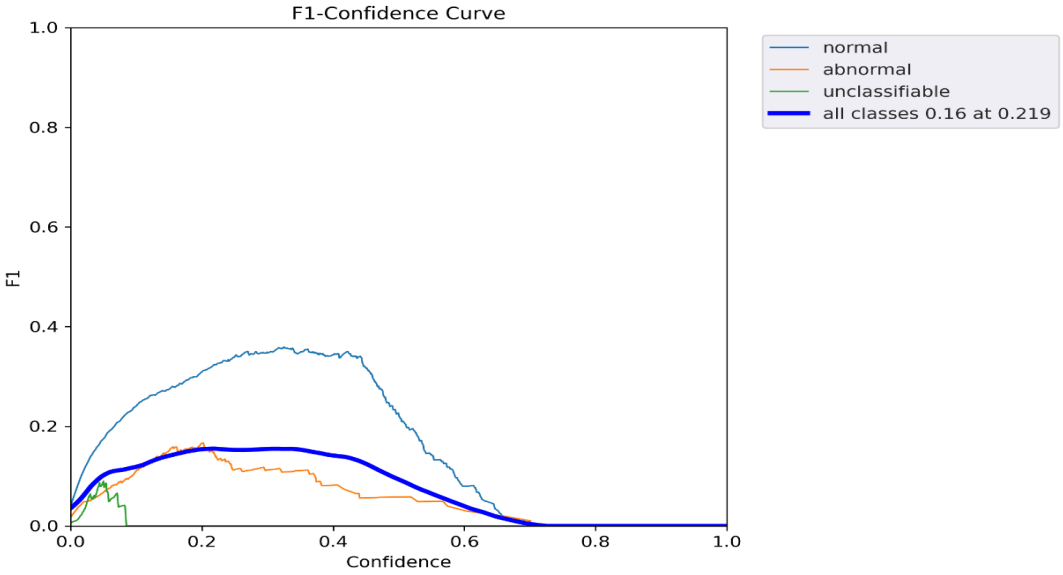


Figure 3.60 F1-Confidence Curve for YOLOv5N all sperm labeled data

Precision-Recall curves have been plotted separately for normal and abnormal classes. Additionally, an average curve has been generated for all classes. The curve for the normal class exhibits high precision and low recall values, indicating the model's success in accurately classifying normal examples but displaying weaker performance in accurately classifying abnormal examples. The average curve is positioned in the middle of the curves drawn for normal and abnormal classes, indicating overall good model performance. However, it is essential to note that the model does not equally classify both classes successfully.

Precision-Confidence curves for normal and abnormal classes have been separately plotted using the same models. The AUC values range between 0.90 and 1, indicating successful performance according to machine learning evaluation criteria. However, the curve for the normal class, represented by the orange line, is positioned higher than the green line for the abnormal class. This suggests that the model more effectively classifies the normal class compared to the abnormal class.

F1 Score - Confidence curves for normal and abnormal classes have been separately plotted using the same models. Here, on average, the F1 score increases as the confidence level rises for all models. This indicates that the model exhibits superior performance in

classifications with high confidence levels. Additionally, the F1 score for the normal class is higher compared to the abnormal class. This implies that the model can more effectively distinguish the normal class from the abnormal class.

Table 3.9 The various YOLOv5 models exhibit distinct evaluation metrics and corresponding values

Sperm-Head-labeled	mAP_0.5	mAP_0.5:0.95	Precision	Recall
YOLOv5x	0.11748	0.061681	0.14222	0.28017
YOLOv5l	0.13371	0.072399	0.13171	0.31653
YOLOv5m	0.13313	0.072097	0.12428	0.38929
YOLOv5s	0.13679	0.068471	0.13553	0.37953
YOLOv5n	0.11501	0.05723	0.1265	0.0301
All Sperm-labeled	**mAP_0.5**	**mAP_0.5:0.95**	**Precision**	**Recall**
YOLOv5x	0.10353	0.05324	0.15126	0.12746
YOLOv5l	0.097478	0.049989	0.13023	0.15549
YOLOv5m	0.10648	0.055445	0.15255	0.17836
YOLOv5s	0.10326	0.049893	0.13376	0.15376
YOLOv5n	0.094109	0.042556	0.13263	0.17584

In the present study, model parameters were evaluated using two different labeling types for Detection and Classification models. It has been observed that there is a limited number of reference studies available to compare the obtained results in the literature. Literature reviews revealed studies comparing YOLO model parameters for human sperm selection. When these studies were compared, it was found that the values obtained for both labeling types were not exceptionally high. The primary reason for this observation is the limited availability of the dataset. Additionally, it was observed that the characteristics of the utilized data and the resolution ratio during video recording contributed to the lower results in the obtained parameter outcomes.

Algorithms trained with both labeling types yielded lower Precision, Recall, and mAP values compared to results from other studies. However, algorithms trained with "all sperm labeled" data obtained even lower parameter values. The inclusion of the tail part during labeling appears to be a significant factor in the relatively lower outcomes.

3.3.3 Evaluation of YOLO models performance in different electric fields

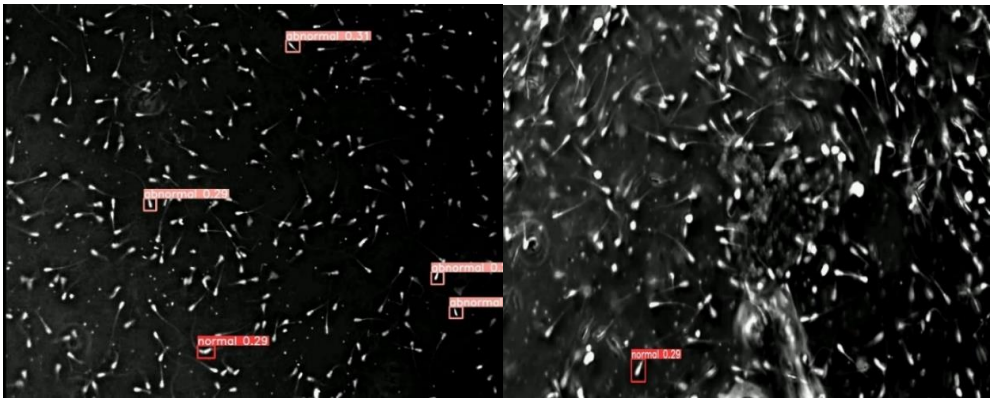
In this process, a comparison was made between the video of the sperm sample to which 12V was applied and the video of the sperm sample without external intervention. Accordingly, the results are stated visually and numerically.

During this evaluation process; The average number of sperm detections in the video and the average confidence values of the detections were taken as basis.

The average number of object detections was computed by dividing the overall count of sperm detections by the product of the video duration and the FPS (Frame per second) value of the video. The average confidence value in the video was computed by dividing the total of all sperm detection values by the total number of detected objects.

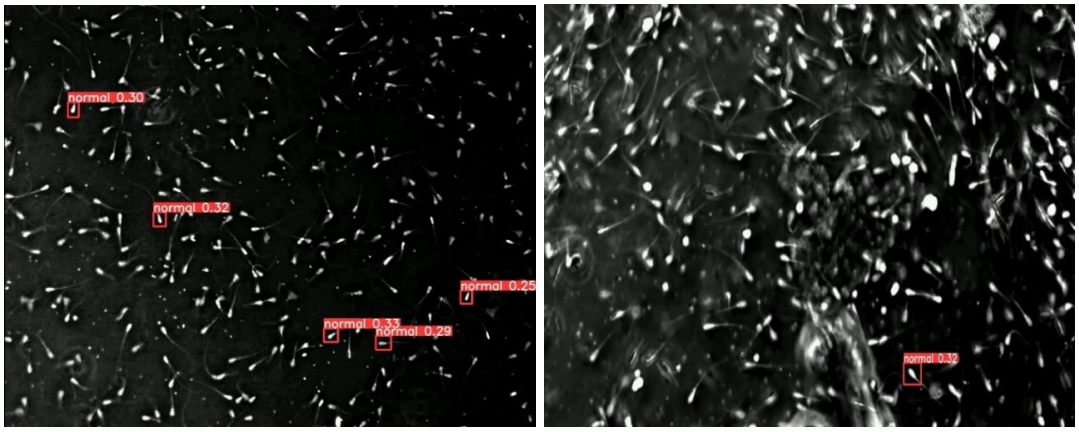
3.3.3.1 Head-labeled model

The videos acquired from sperm cells exposed to electric field of 200 V/m and annotated with the "head" tag exhibited considerable distinctions between the experimental and control groups. Remarkably, the utilization of the YOLOv5L model on the videos obtained under the 200 V/m condition resulted in markedly enhanced outcomes with respect to analysis and detection. The results are as follows :



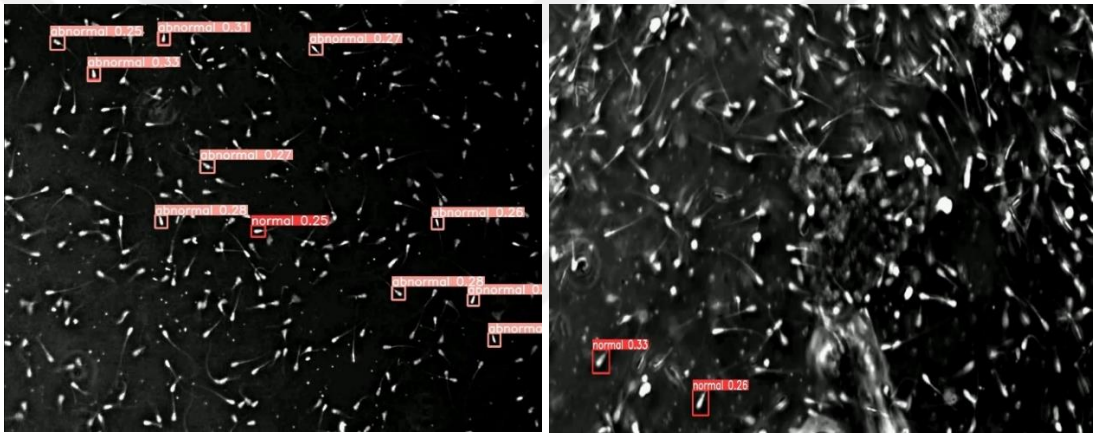
A) 200V/m Electric Field B) No electric field

Figure 3.61 Comparison for YOLOv5N



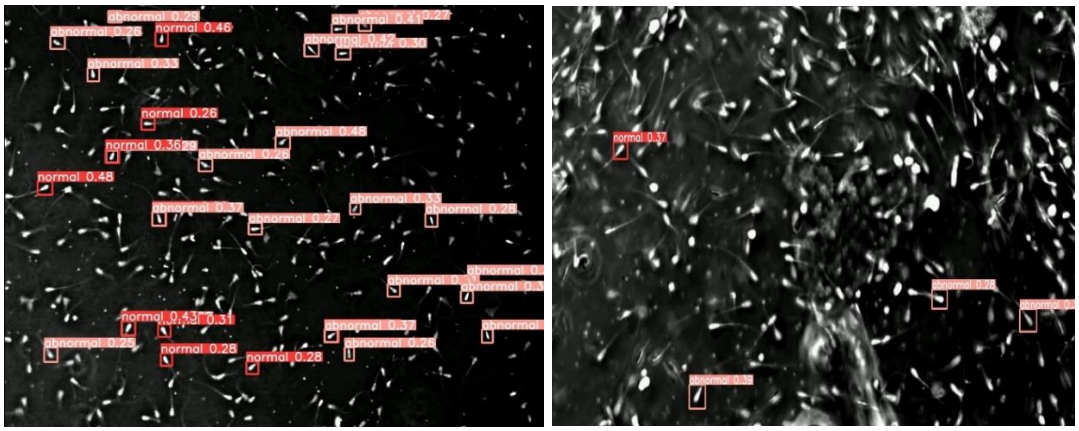
A)200V/m Electric Field B) No electric field

Figure 3.62 Comparison for YOLOv5S



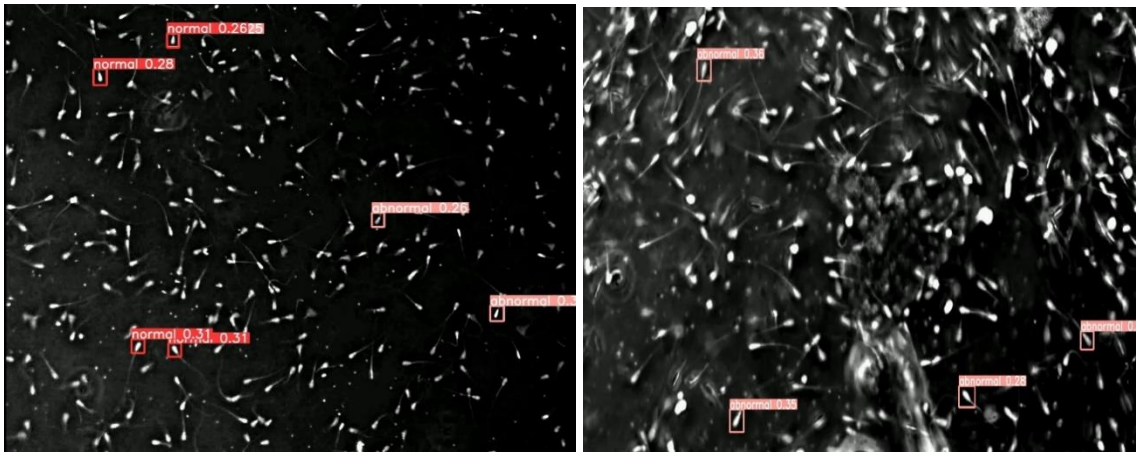
A) 200V/m Electric Field B) No electric field

Figure 3.63 Comparison for YOLOv5M



A) 200V/m Electric Field B) No electric field

Figure3.64 Comparison for YOLOv5L



A) 200V/m Electric Field B) No electric field

Figure 3.65. Comparison for YOLOv5X

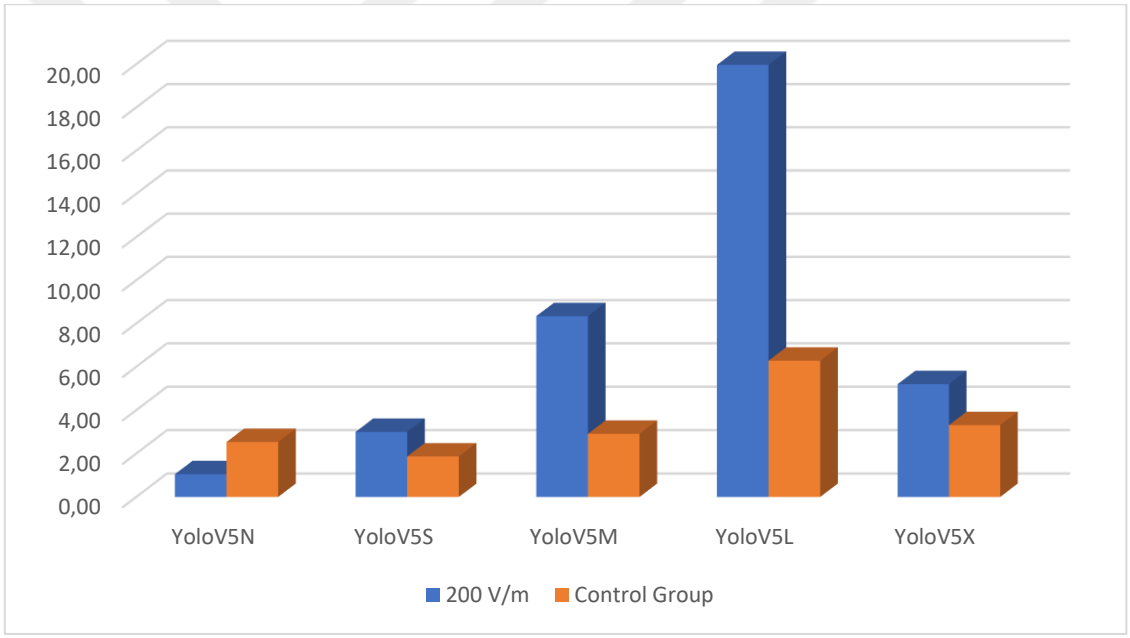


Figure 3.66 Comparison of average detected sperm numbers for 12V and Control Group

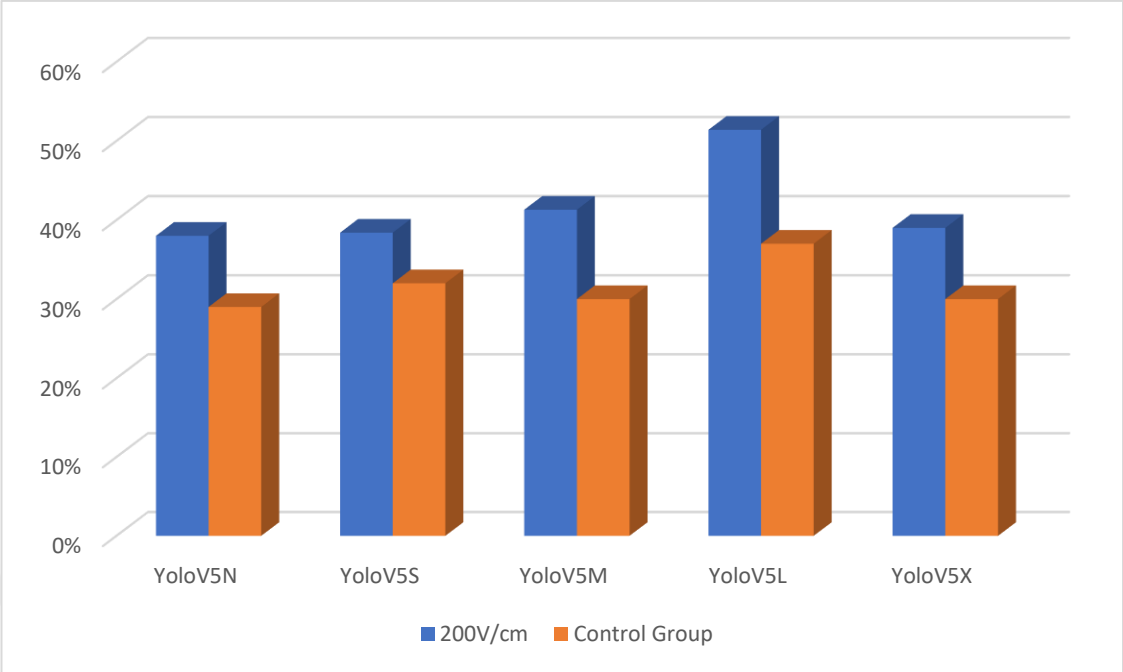
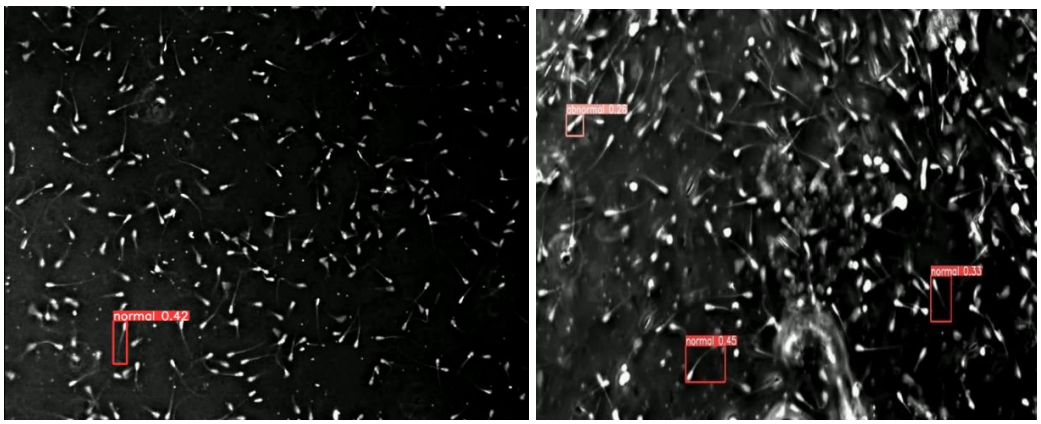


Figure 3.67 Comparison of average confidence for 12V and Control Group

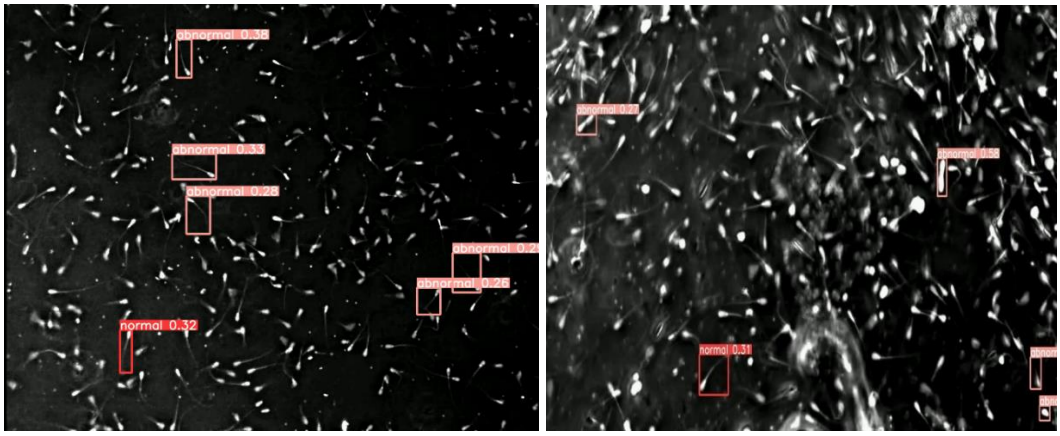
3.3.3.2 All sperm labeled model

The YOLOv5 detections and the control groups exhibit no significant differences. The results obtained with the 'All Sperm Labeled' models are as follows:



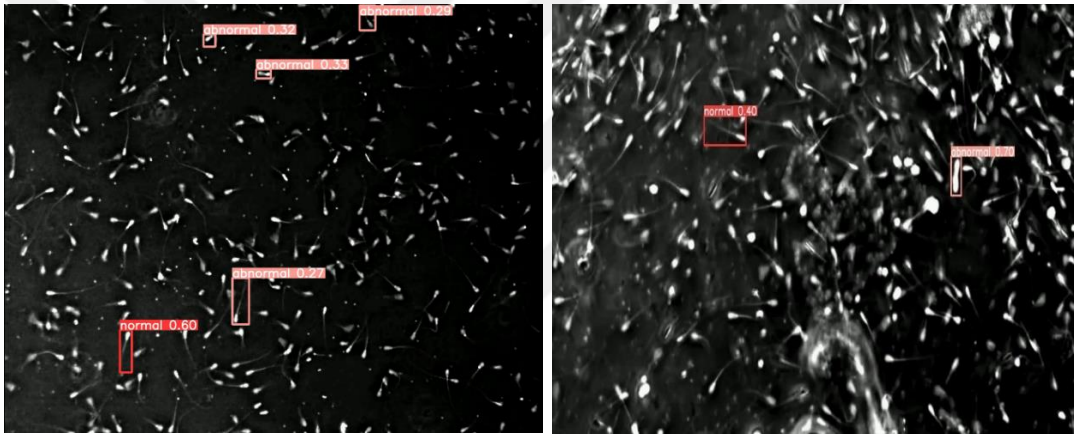
A)200V/m Electric Field B) No electric field

Figure 3.68 Comparison for YOLOv5N



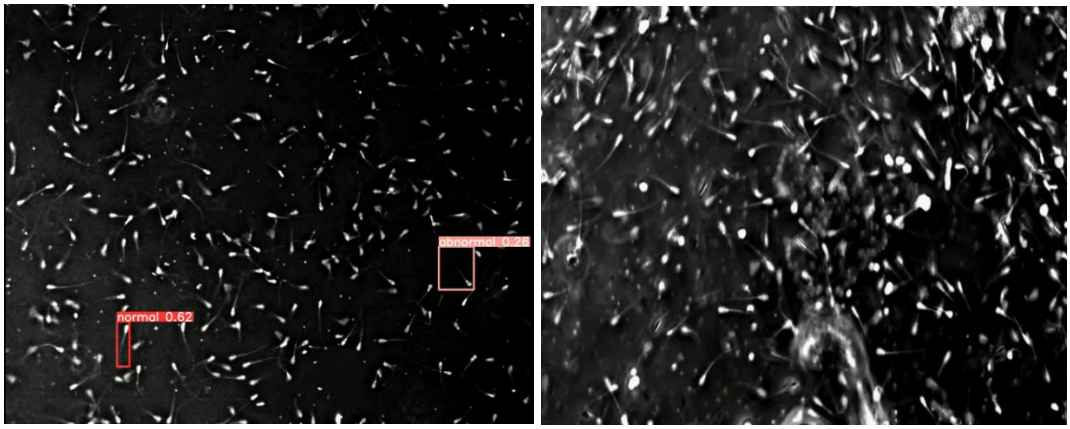
A)200V/m Electric Field B) No electric field

Figure 3.69 Comparison for YOLOv5S



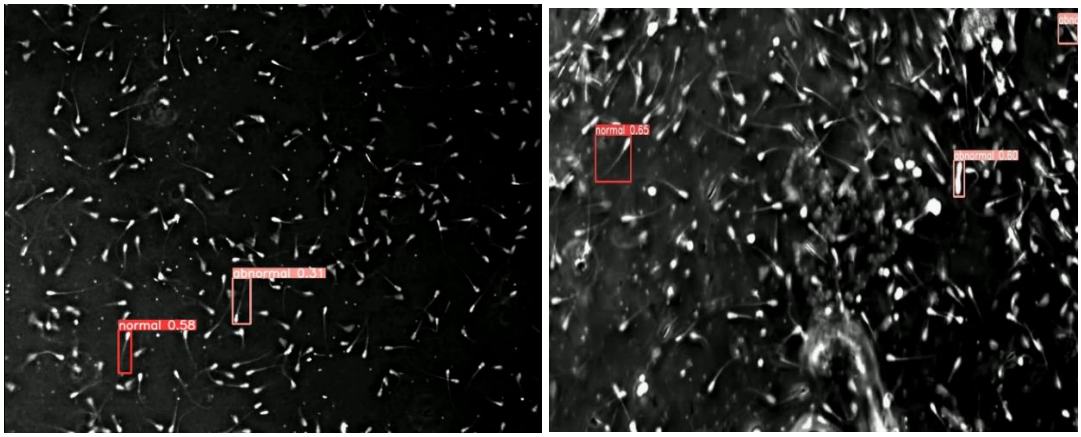
A)200V/m Electric Field B) No electric field

Figure 3.70 Comparison for YOLOv5M



A)200 V/m Electric Field B) No electric field

Figure 3.71 Comparison for YOLOv5L



A)200V/m Electric Field B) No electric field

Figure 3.72 Comparison for YOLOv5X

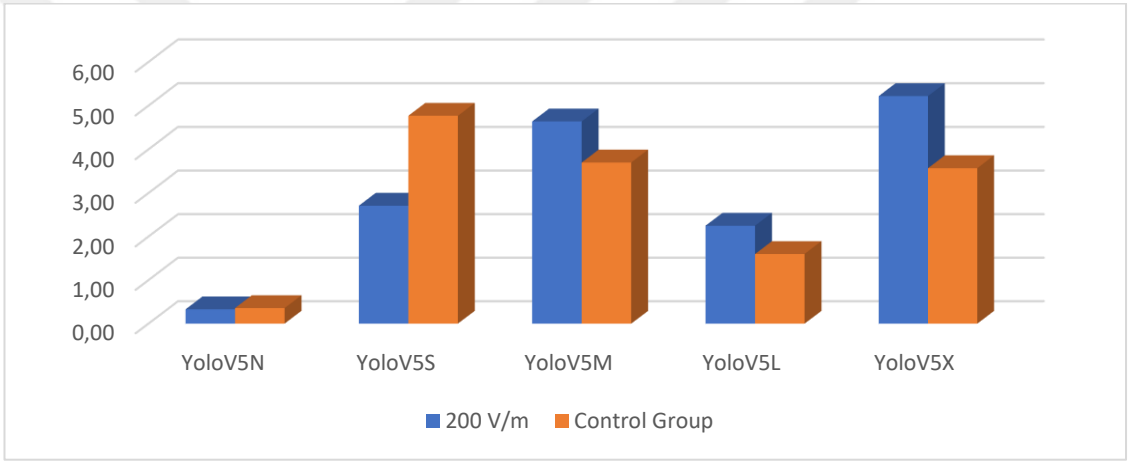


Figure 3.73 Comparison of average detected sperm numbers for 12V and Control Group

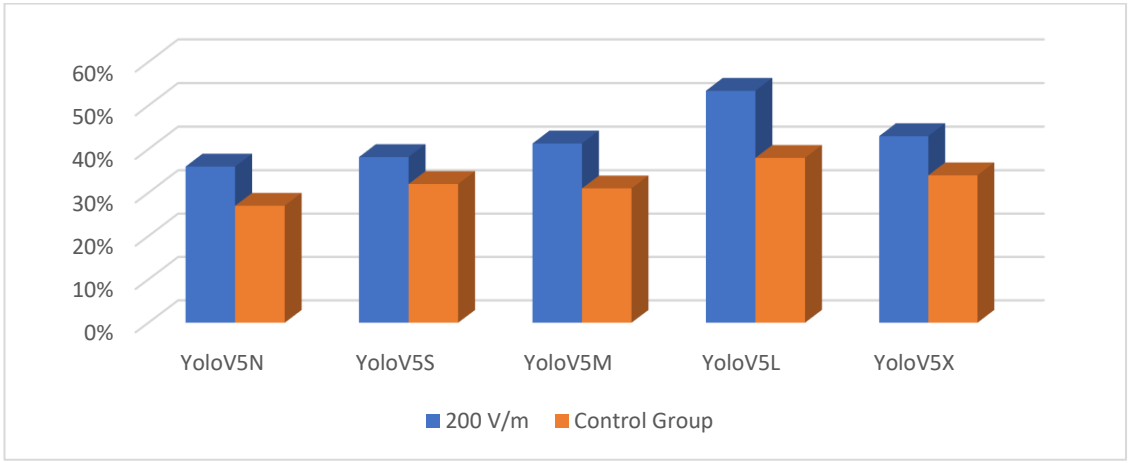


Figure 3.74 Comparison of average confidence for 12V and Control Group

Models trained with head-labeled datasets have been observed to exhibit superior performance compared to models labeled with whole-cell annotations. Several reasons account for this situation. In the head-labeling process, fewer categories are utilized in evaluation criteria due to the absence of the tail section. The exclusion of the tail component contributes to more effective outcomes in the selection process, as it involves fewer components. Labeling only the head portion of sperm cells, coupled with categorization considering anomalies in the head section, optimizes labeling and training processes, positively influencing the model's detection and classification performance.

The detrimental impact of electric fields on the movement speed of sperm cells is a well-established fact. The expected effect of daily exposure to electric fields on bull sperm motility has been emphasized. The performance of YOLO algorithms on videos obtained under this influence has been examined. It has been observed that the YOLOv5L model, trained only with head-labeled data, achieves significantly higher success compared to other models. In this context, the architecture employed by the YOLOv5L model plays a crucial role in this success.

4. DISCUSSION

The primary objective of this investigation was to evaluate the capability of machine learning algorithms in bull sperm selection and simultaneously investigate the effects of electric field on bull sperm. The results obtained from the study demonstrate that both modern technologies, represented by machine learning algorithms, and physical factors, particularly the electric field, exert significant effects on artificial intelligence models for sperm detection. This enables the enhancement of reproductive success thereby allowing for more effective genetic improvements. The findings indicate that these factors could play a significant role in bull production and provide a new perspective on genetic improvements and reproductive strategies in the agricultural industry. In particular, the ability of these algorithms to analyze complex relationships within large datasets can provide insights beyond traditional methods.

The effects of the electric field on bull sperm represent another crucial point in the study. Existing studies have showed a pronounced impact of the electric field on sperm motility, morphology, and genetic integrity(Bernabò et al., 2007; Górski et al., 2020). It was assumed that the electric field can directly influence the sperm selection processes and that this effect could enhance the quality in the agricultural industry. This, in turn, offers farmers the potential to optimize environmental conditions and increase reproductive success.

The study demosrated the effectiveness of the electric field in influencing machine learning algorithms for sperm selection. Before evaluating models performance, a comprehensive examination of the outcomes resulting from the application of the electric field was conducted. CASA systems were employed during the assessment of the electric field's impact on sperm cells. CASA, an abbreviation for Computer-Assisted Sperm Analysis, stands out as one of the most widely used techniques for a thorough analysis of sperm cells across various parameters. It serves as a valuable tool, demonstrating the capability to provide prompt, reliable, and objective quantitative assessments of sperm quality(Horst et al., 2018). In addition to assessing sperm concentration and sperm motility, modern Computer-Assisted Sperm Analysis (CASA) systems the automate

evaluation of morphology of sperm, motility, integrity of spermatozoa , and the acrosome reaction(van der Horst, 2020).

Throughout the course of these examinations, comprehensive assessments were conducted, encompassing diverse parameters such as velocity and progressivity.Sperm cells were categorized into four distinct groups: non-progressive , medium and rapid progressive, and immotile. The evaluation of sperm cell velocity involved an analysis across various groups, considering curve, linear speed, and average speed, particularly focusing on the rapid progressive group.The obtained results from this study are anticipated to provide a foundation for future investigations. Following the application of 50 V/m and 100 V/m electric fields, a noticeable increase in linear speed was identified compared to the condition where no electric field was applied (control group). However, with the application of a 200 V/m electric field, a marked difference in linear speed was not evident within the control group.The application of the 200 V/m electric field resulted in a decrease in sperm velocity.As a consequence of the application of 50 V/m and 100 V/m electric fields, there was an approximately 1.5-fold increase in the percentage of Rapid Progressive sperm compared to the scenario in which no electric field was applied (control group). These variances in sperm velocities may provide machine learning algorithms with the capacity to attain heightened precision and accuracy values in the realm of sperm detection.These findings highlight the potential role of the electric field as a factor in sperm selection and reproductive success.

During evaluation of performances of models large differences have been observed in the performance of deep learning models on different electric fields and control groups. To evaluate the models , different parameters was used. Distinct parameters were employed for classification and detection models.Various metrics were employed in the performance evaluation process of classification algorithms. In this context, VGG-16, VGG-19, ResNet50, ResNet101, and ResNet152 models were utilized for the classification of head-labeled data. F1-score, Accuracy, Recall, Precision values were employed to objectively examine the classification efficacy of the algorithms. Accuracy measures correct predictions relative to total data points. Precision gauges the ratio of correctly predicted positive points, crucial in high false positive cost scenarios. Recall

signifies the proportion of actual positive points predicted correctly, particularly valuable in high false negative cost scenarios. The F1 Score finds widespread use in the comprehensive assessment of classification models.(Banerjee, 2023; Hand et al., 2021; Yacouby & Axman, 2020) The study demonstrated that the highest accuracy value was achieved in the ResNet50 class. No significant differences were observed among the F1-score, Precision ,Recall values for the VGG-16, VGG-19, ResNet50, and ResNet152 models. In a conducted study by Iqbal et al. in 2020, a proposed model was presented and compared with various models using a head-labeled dataset for classification(Iqbal et al., 2020). Yüzkat et al. (2020) also conducted a study where different classification algorithms were compared, specifically utilizing low-quality sperm images. The reported recall and precision values for VGG-19 were 0.37 and 0.74, respectively.In that study the results of different epoch sizes were clarified(Yüzkat et al., 2020).In our research, the precision of VGG-19 was found to be 0.77. This outcome may be attributed to factors such as the volume and characteristics of the data, as well as parameters employed during training, including batch size and epoch size.

We used YOLOv5 models for object detection, conducting the training process using head-labeled data and all-sperm-labeled data. Precision, Recall, and mAP (mean average precision) metrics were utilized for evaluating the YOLO models. Additionaly the curves were used to evaluate the performances of YOLO models. The Precision-Recall curves reveal that the models excels in accurately classifying normal examples but struggles with abnormal examples, as evidenced by high precision and low recall for the normal class. Precision-Confidence curves, AUC values ranging from 0.90 to 1 indicate successful model performance. However, the normal class curve surpasses the abnormal class curve, implying that the model is more effective at classifying normal examples than abnormal ones. Lastly, the F1 Score-Confidence curves demonstrate that the model's overall performance improves with higher confidence levels. Moreover, the F1 score for the normal class is higher than that for the abnormal class, indicating the model's superior ability to distinguish between normal and abnormal examples. The algorithms trained with both labeling types exhibited lower Precision, Recall, and mean Average Precision (mAP) values compared to results reported in other studies. Particularly noteworthy was the finding that algorithms trained with "all sperm labeled" data achieved even lower

parameter values. The inclusion of the tail part during labeling emerged as a significant factor influencing these relatively lower outcomes. Magnitude of the dataset is an important parameter effecting the outcomes of the training process. In addition to the data's precision and reliability, the volume and balance of the data used also play a significant role in this context (Yüzkat et al., 2020).

Following model training, testing was conducted on videos subjected to electric fields, with algorithms tested using a video featuring a 200 V/m electric field. Average detected sperm and average confidence values were compared against the reference. In the YOLOv5L model trained with the head-labeled dataset, the number of detected sperm in the 200 V/m was tripple observed in the control group, while for YOLOv5M model, there was a difference of nearly 2 times. A 14% difference in average confidence values was observed in the YOLOv5L model in the 200 V/m group. In this study, different versions of YOLOv5 have yielded varying results, and there could be multiple reasons for this discrepancy. At this juncture, the primary factor is likely the dataset. The size, quality, and characteristics of the dataset, along with the architectural choices within the model, can be considered the major contributors. The fundamental reason for different outcomes among YOLO versions in various studies is primarily attributed to the employed dataset. In the study (Thambawita et al., 2023) labeled data was utilized to train YOLO models. In contrast to our study, the mentioned research indicates that YOLOv5m demonstrates maximum precision. Additionally, aside from other factors, the differences could be attributed to the hyperparameters employed in the respective studies.

In this context, the study demonstrated that the YOLOv5L model yielded significantly more efficient results on average when exposed to a video with a 200 V/m electric field.However, in models trained with all-sperm labeled data, the results in terms of the average detected sperm count varied significantly, making it challenging to assert a correlation among the models. The study revealed YOLOv5S exhibited the lowest performance, while YOLOv5X achieved the highest performance, showing a 46% difference in contrast with the reference group.In terms of average confidence values, the models exhibited an average success of approximately 10% relative to the reference group. It was concluded that the number of labels and data should be increased to increase

efficacy of YOLO and classification algorithms. Increasing the number of labeled data in future studies will increase model performance. Additionally, in the future , to test the model more precisely, employing systems that can apply an electric field will yield more detailed and meaningful results.

The application of an electric field facilitated the ease of model selection in conjunction with a decrease in bull sperm cell motility. However, a crucial aspect that warrants investigation pertains to the potential effects of electric field application on sperm cell DNA. Simultaneously, further exploration into the feasibility and industrial applicability of this system emerges as another imperative area for research. The potential collection of a more extensive dataset during the course of the study could have had a significant impact on the outcomes. Factors such as the quantity and quality of labeled data, the resolution of videos obtained during recording, and the number of repetitions performed represent limitations that could influence the results of the study. Consideration of these variables in future investigations holds the potential to augment the efficacy of the obtained findings.

5. CONCLUSION

This thesis presents a comprehensive investigation into the effectiveness of various machine learning models and their response to electric fields, with a focus on their impact on bovine sperm cells. The findings validate that applying electric fields to bull sperm cells significantly improves the performance of AI models, resulting in superior outcomes. Moreover, the research conducts a thorough comparative analysis of sperm selection performances across cutting-edge models. The primary objective of this thesis is to systematically address infertility issues in bulls. The research outcomes not only affirm the potential for enhancing AI model performance through the application of electric fields to bull sperm cells but also highlight the importance of comparative assessments in understanding the sperm selection capabilities of modern models.

REFERENCES

Ahmadi, A., Bongso, A., & Ng, S. C. (1997). Induction of Acrosome Reaction in Human Sperm by Exposing to an Electrical Field. *Archives of Andrology*. https://doi.org/10.3109/01485019708988532

Alsalem, M. A., Zaidan, A. A., Zaidan, B. B., Hashim, M., Albahri, O. S., Albahri, A. S., Hadi, A., & Mohammed, K. I. (2018). Systematic Review of an Automated Multiclass Detection and Classification System for Acute Leukaemia in Terms of Evaluation and Benchmarking, Open Challenges, Issues and Methodological Aspects. *Journal of Medical Systems*, *42*(11), 204. https://doi.org/10.1007/s10916-018-1064-9

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Ariagno, J. I., Mendeluk, G. R., Furlan, M. J., Sardi, M., Chenlo, P., Curi, S. M., Pugliese, M. N., Repetto, H. E., & Cohen, M. (2017). Computer-aided sperm analysis: A useful tool to evaluate patient's response to varicocelectomy. *Asian Journal of Andrology*, *19*(4), 449-452. https://doi.org/10.4103/1008-682X.173441

Aslankoc, R., Gumral, N., Saygin, M., Senol, N., Asci, H., Cankara, F. N., & Comlekci, S. (2018). The impact of electric fields on testis physiopathology, sperm parameters and DNA integrity-The role of resveratrol. *Andrologia*, *50*(4), e12971. https://doi.org/10.1111/and.12971

Banerjee, C. (2023, Haziran 25). What is and When to use precession, recall and f1 score. *Medium*. https://medium.com/@chandradip93/what-is-and-when-to-use-precession-recall-and-f1-score-884403ee525e

Bernabò, N., Tettamanti, E., Pistilli, M. G., Nardinocchi, D., Berardinelli, P., Mattioli, M., & Barboni, B. (2007). Effects of 50Hz extremely low frequency magnetic field on the morphology and function of boar spermatozoa capacitated in vitro. *Theriogenology*, *67*(4), 801-815. https://doi.org /10.1016/ j .theriogenology.2006.10.014

Blier, L., Wolinski, P., & Ollivier, Y. (2020). Learning with Random Learning Rates. Içinde U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, & C. Robardet (Ed.), *Machine Learning and Knowledge Discovery in Databases* (ss. 449-464). Springer International Publishing. https://doi.org/10.1007/978-3-030-46147-8_27

Brownlee, J. (2022, Ağustos 9). Difference Between a Batch and an Epoch in a Neural Network. *MachineLearningMastery.Com*. https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/

Butler, M. L., Bormann, J. M., Weaber, R. L., Grieger, D. M., & Rolf, M. M. (2020). Selection for bull fertility: A review. *Translational Animal Science*, *4*(1), 423-441. https://doi.org/10.1093/tas/txz174

Carvalho, T., de Rezende, E. R. S., Alves, M. T. P., Balieiro, F. K. C., & Sovat, R. B. (2017). Exposing Computer Generated Images by Eye's Region Classification via Transfer Learning of VGG19 CNN. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 866-870. https://doi.org/10.1109/ICMLA.2017.00-47

Chauhan, N. K., & Singh, K. (2018). A Review on Conventional Machine Learning vs Deep Learning. *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, 347-352. https://doi.org/10.1109/GUCON.2018.8675097

Cochard, D. (2021, Ekim 6). mAP: Evaluation metric for object detection models. *Axinc-Ai*. https://medium.com/axinc-ai/map-evaluation-metric-of-object-detection-model-dd20e2dc2472

Deci. (2023, Ekim 18). *An Overview of State of the Art (SOTA) DNNs*. Deci. https://deci.ai/blog/sota-dnns-overview/

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating Machine Learning with Human Knowledge. *iScience*, *23*(11). https://doi.org/10.1016/j.isci.2020.101656

Dobrovolny, M., Benes, J., Langer, J., Krejcar, O., & Selamat, A. (2023). Study on Sperm-Cell Detection Using YOLOv5 Architecture with Labaled Dataset. *Genes*, *14*(2), 451. https://doi.org/10.3390/genes14020451

Dobson, H., Walker, S. L., Morris, M. J., Routly, J. E., & Smith, R. F. (2008). Why is it getting more difficult to successfully artificially inseminate dairy cows? *Animal: An International Journal of Animal Bioscience*, *2*(8), 1104-1111. https://doi.org/10.1017/S175173110800236X

GitHub—Rafaelpadilla/Object-Detection-Metrics: Most popular metrics used to evaluate object detection algorithms. (t.y.). Geliş tarihi 18 Aralık 2023, gönderen https://github.com/rafaelpadilla/Object-Detection-Metrics

Givens, M. D. (2006). A clinical, evidence-based approach to infectious causes of infertility in beef cattle. *Theriogenology*, *66*(3), 648-654. https://doi.org/10.1016/j.theriogenology.2006.04.021

Górski, R., Kotwicka, M., Skibińska, I., Jendraszak, M., & Wosiński, S. (2020). Effect of low-frequency electric field screening on motility of human sperm. *Annals of Agricultural and Environmental Medicine*, *27*(3), 427-434. https://doi.org/10.26444/aaem/116019

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview* (arXiv:2008.05756). arXiv. https://doi.org/10.48550/arXiv.2008.05756

Guan, Q., Wang, Y., Ping, B., Li, D., Du, J., Qin, Y., Lu, H., Wan, X., & Xiang, J. (2019). Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *Journal of Cancer*, *10*(20), 4876-4882. https://doi.org/10.7150/jca.28769

Hand, D. J., Christen, P., & Kirielle, N. (2021). F*: An interpretable transformation of the F-measure. *Machine Learning*, *110*(3), 451-456. https://doi.org/10.1007/s10994-021-05964-1

Hassan, M. ul. (2018, Kasım 20). *VGG16—Convolutional Network for Classification and Detection*. https://neurohive.io/en/popular-networks/vgg16/

Hassanien, A. E., & Darwish, A. (Ed.). (2021). *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges* (C. 77). Springer International Publishing. https://doi.org/10.1007/978-3-030-59338-4

Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S., & Rufino, M. C. (2013). The roles of livestock in developing countries. *Animal*, *7*(s1), 3-18. https://doi.org/10.1017/S1751731112001954

Hidayatullah, P., Mengko, T. L. E. R., Munir, R., & Barlian, A. (2021). Bull Sperm Tracking and Machine Learning-Based Motility Classification. *IEEE Access*, *9*, 61159-61170. https://doi.org/10.1109/ACCESS.2021.3074127

Horst, G. van der, Maree, L., & Plessis, S. S. du. (2018). Current perspectives of CASA applications in diverse mammalian spermatozoa. *Reproduction, Fertility and Development*, *30*(6), 875-888. https://doi.org/10.1071/RD17468

Iqbal, I., Mustafa, G., & Ma, J. (2020). Deep Learning-Based Morphological Classification of Human Sperm Heads. *Diagnostics*, *10*(5), Article 5. https://doi.org/10.3390/diagnostics10050325

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695. https://doi.org/10.1007/s12525-021-00475-2

Javadi, S., & Mirroshandel, S. A. (2019). A novel deep learning method for automatic assessment of human sperm images. *Computers in Biology and Medicine*, *109*, 182-194. https://doi.org/10.1016/j.compbiomed.2019.04.030

Kaltungo, B. Y., & Musa, I. W. (2013). A Review of Some Protozoan Parasites Causing Infertility in Farm Animals. *ISRN Tropical Medicine*, *2013*, 1-6. https://doi.org/10.1155/2013/782609

Kumaresan, A., Das Gupta, M., Datta, T. K., & Morrell, J. M. (2020). Sperm DNA Integrity and Male Fertility in Farm Animals: A Review. *Frontiers in Veterinary Science*, *7*, 321. https://doi.org/10.3389/fvets.2020.00321

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Malaga, R. A., & Koppel, N. B. (2017). A Comparison Of Video Formats For Online Teaching. *Contemporary Issues in Education Research (CIER)*, *10*(1), Article 1. https://doi.org/10.19030/cier.v10i1.9876

Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., & Stevens, G. A. (2012). National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys. *PLOS Medicine*, *9*(12), e1001356. https://doi.org/10.1371/journal.pmed.1001356

McCallum, C., Riordon, J., Wang, Y., Kong, T., You, J. B., Sanner, S., Lagunov, A., Hannam, T. G., Jarvi, K., & Sinton, D. (2019). Deep learning-based selection of human sperm with high DNA integrity. *Communications Biology*, *2*(1), 250. https://doi.org/10.1038/s42003-019-0491-6

Mocé, E., & Graham, J. K. (2008). In vitro evaluation of sperm quality. *Animal Reproduction Science*, *105*(1-2), 104-118. https://doi.org/10.1016/j.anireprosci.2007.11.016

Morrell, J. M. (2020). Sperm Selection in Animal-Assisted Reproduction. İçinde *Emerging Technologies in Biophysical Sciences: A World Scientific Reference* (1-3, ss. 143-162). WORLD SCIENTIFIC. https://doi.org/ 10. 1142/9789811226106_0008

Mortimer, S. T., van der Horst, G., & Mortimer, D. (2015). The future of computer-aided sperm analysis. *Asian Journal of Andrology*, *17*(4), 545-553. https://doi.org/10.4103/1008-682X.154312

Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers*, *11*(9), Article 9. https://doi.org/10.3390/cancers11091235

NAAB. 2018. Semen production. National Association of Animal Breeders. https://www.naab-css.org/semen-sales [accessed February 6, 2019].

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1. https://doi.org/10.1186/s40537-014-0007-7

Nasir, N., Kansal, A., Barneih, F., Al-Shaltone, O., Bonny, T., Al-Shabi, M., & Al Shammaa, A. (2023). Multi-modal image classification of COVID-19 cases using computed tomography and X-rays scans. *Intelligent Systems with Applications*, *17*, 200160. https://doi.org/10.1016/j.iswa.2022.200160

Netrapalli, P. (2019). Stochastic Gradient Descent and Its Variants in Machine Learning. *Journal of the Indian Institute of Science*, *99*(2), 201-213. https://doi.org/10.1007/s41745-019-0098-4

Newton, D., Yousefian, F., & Pasupathy, R. (2018). Stochastic Gradient Descent: Recent Trends. İçinde *Recent Advances in Optimization and Modeling of Contemporary Problems* (ss. 193-220). INFORMS. https://doi.org/10.1287/educ.2018.0191

Özbek, M., Hitit, M., Kaya, A., Jousan, F. D., & Memili, E. (2021). Sperm Functional Genome Associated With Bull Fertility. *Frontiers in Veterinary Science*, *8*. https://www.frontiersin.org/articles/10.3389/fvets.2021.610888

Panigrahi, S., Nanda, B. S., Bhuyan, R., Kumar, K., Ghosh, S., & Swarnkar, T. (2023). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. *Heliyon*, *9*(3). https://doi.org/10.1016/j.heliyon.2023.e13444

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, *113*, 54-71. https://doi.org/10.1016/j.neunet.2019.01.012

Park, Y.-J., Kim, J., You, Y.-A., & Pang, M.-G. (2013). Proteomic Revolution to Improve Tools for Evaluating Male Fertility in Animals. *Journal of Proteome Research*, *12*(11), 4738-4747. https://doi.org/10.1021/pr400639x

Patil, T., Pandey, S., & Visrani, K. (2021). A Review on Basic Deep Learning Technologies and Applications. Içinde K. Kotecha, V. Piuri, H. N. Shah, & R. Patel (Ed.), *Data Science and Intelligent Applications* (ss. 565-573). Springer. https://doi.org/10.1007/978-981-15-4474-3_61

Perry, V. E. A. (2021). The Role of Sperm Morphology Standards in the Laboratory Assessment of Bull Fertility in Australia. *Frontiers in Veterinary Science*, *8*, 672058. https://doi.org/10.3389/fvets.2021.672058

R, S., & Patilkulkarni, S. (2021). Visual speech recognition for small scale dataset using VGG16 convolution neural network. *Multimedia Tools and Applications*, *80*(19), 28941-28952. https://doi.org/10.1007/s11042-021-11119-0

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, *11*(4), Article 4. https://doi.org/10.3390/info11040193

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. https://doi.org/10.1109/CVPR.2016.91

Riordon, J., McCallum, C., & Sinton, D. (2019). Deep learning for the classification of human sperm. *Computers in Biology and Medicine*, *111*, 103342. https://doi.org/10.1016/j.compbiomed.2019.103342

Sahoo, B., Choudhary, R. K., Sharma, P., Choudhary, S., & Gupta, M. K. (2021). Significance and Relevance of Spermatozoal RNAs to Male Fertility in Livestock.

Frontiers in Genetics, *12*. https://www.frontiersin.org/articles/ 10. 3389/ fgene. 2021.768196

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, *2*(1), 24-28. https://doi.org/10.1016/j.gltp.2021.01.004

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. İçinde *arXiv e-prints*. https://doi.org/10. 48550/ arXiv.1409.1556

Situmorang, S., & Panggabean, J. F. (2021). Design and Development of Video and Audio File Converter Applications. *Jurnal ICT : Information and Communication Technologies*, *12*(2), Article 2.

Suryawati, E., Sustika, R., Yuwana, R. S., Subekti, A., & Pardede, H. F. (2018). Deep Structured Convolutional Neural Network for Tomato Diseases Detection. *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 385-390. https://doi.org/10.1109/ICACSIS.2018.8618169

Tammina, S. (2019). *Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images*. *9*, p9420. https://doi.org/10. 29322/IJSRP.9.10.2019.p9420

Tanga, B. M., Qamar, A. Y., Raza, S., Bang, S., Fang, X., Yoon, K., & Cho, J. (2021). Semen evaluation: Methodological advancements in sperm quality-specific fertility assessment — A review. *Animal Bioscience*, *34*(8), 1253-1270. https://doi.org/10.5713/ab.21.0072

Teves, M. E., & Roldan, E. R. S. (2022). Sperm bauplan and function and underlying processes of sperm formation and selection. *Physiological Reviews*, *102*(1), 7-60. https://doi.org/10.1152/physrev.00009.2020

Thambawita, V., Hicks, S. A., Storås, A. M., Nguyen, T., Andersen, J. M., Witczak, O., Haugen, T. B., Hammer, H. L., Halvorsen, P., & Riegler, M. A. (2023). VISEM-Tracking, a human spermatozoa tracking dataset. *Scientific Data*, *10*(1), 260. https://doi.org/10.1038/s41597-023-02173-4

Tian, Y., Zhang, Y., & Zhang, H. (2023). Recent Advances in Stochastic Gradient Descent in Deep Learning. *Mathematics*, *11*(3), Article 3. https://doi.org/10.3390/ math11030682

Tulasi Krishna, S., & Kalluri, H. kumar. (2019). *Deep Learning and Transfer Learning Approaches for Image Classification*.

Ultralytics. (t.y.-a). *Architecture Summary*. Geliş tarihi 17 Aralık 2023, gönderen https://docs.ultralytics.com/yolov5/tutorials/architecture_description

Ultralytics. (t.y.-b). *Augment*. Geliş tarihi 17 Aralık 2023, gönderen https://docs. ultralytics.com/reference/data/augment

Ultralytics. (t.y.-c). *Object Detection Datasets Overview*. Geliş tarihi 19 Aralık 2023, gönderen https://docs.ultralytics.com/datasets/detect

USDA. 2009. Part II: reference of beef cow-calf management practices in the United States. United State Department of Agriculture National Animal Health Systems Monitoring System Beef Cow-Calf studies. Ft. Collins (CO). https:// www.aphis.usda.gov/animal_health/nahms/beefcowcalf/ downloads/beef0708/Beef0708_dr_PartII_1.pdf [accessed February 6, 2019].

van der Horst, G. (2020). Computer Aided Sperm Analysis (CASA) in domestic animals: Current status, three D tracking and flagellar analysis. *Animal Reproduction Science*, *220*, 106350. https://doi.org/10.1016/j.anireprosci.2020.106350

Vaughan, D. A., & Sakkas, D. (2019). Sperm selection methods in the 21st century. *Biology of Reproduction*, *101*(6), 1076-1082. https://doi.org/10. 1093/biolre/ioz032

Victor Ikechukwu, A., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global Transitions Proceedings*, *2*(2), 375-381. https://doi.org/10.1016/j.gltp.2021.08.027

Wagenaar, B. de, Geijs, D. J., Boer, H. de, Bomer, J. G., Olthuis, W., Berg, A. van den, & Segerink, L. I. (2016). Spermometer: Electrical characterization of single boar sperm motility. *Fertility and Sterility*, *106*(3), 773-780.e6. https://doi.org/10.1016/j.fertnstert.2016.05.008

Wang, H.-Q., Wang, T., Gao, F., & Ren, W.-Z. (2022). Application of CRISPR/Cas Technology in Spermatogenesis Research and Male Infertility Treatment. *Genes*, *13*(6), 1000. https://doi.org/10.3390/genes13061000

Wang, S., Xia, X., Ye, L., & Yang, B. (2021). Automatic Detection and Classification of Steel Surface Defect Using Deep Convolutional Neural Networks. *Metals*, *11*, 388. https://doi.org/10.3390/met11030388

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

World Health Organization. (2010). *WHO laboratory manual for the examination and processing of human semen*. 271.

Wu, S., Di, G., & Li, Z. (2017). Does static electric field from ultra-high voltage direct-current transmission lines affect male reproductive capacity? Evidence from a laboratory study on male mice. *Environmental Science and Pollution Research*, *24*(22), 18025-18034. https://doi.org/10.1007/s11356-017-9229-5

Yitbarek, M. (2020). *Livestock and Livestock products by 2050*.

You, J. B., McCallum, C., Wang, Y., Riordon, J., Nosrati, R., & Sinton, D. (2021). Machine learning for sperm selection. *Nature Reviews Urology*, *18*(7), 387-403. https://doi.org/10.1038/s41585-021-00465-1

Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., & Kawuwa, H. B. (2022). Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches. *Applied Sciences*, *12*(14), Article 14. https://doi.org/10.3390/app12147282

Yüzkat, M., Ilhan, H. O., & Aydın, N. (2020). Morphological Classification of Low Quality Sperm Images Using Deep Learning Networks. *2020 Medical Technologies Congress (TIPTEKNO)*, 1-4. https://doi.org/10.1109/TIPTEKNO50054.2020.9299318

Zalata, A., El-Samanoudy, A. Z., Shaalan, D., El-Baiomy, Y., & Mostafa, T. (2015). In Vitro Effect of Cell Phone Radiation on Motility, DNA Fragmentation and Clusterin Gene Expression in Human Sperm. *International Journal of Fertility & Sterility*, *9*(1), 129-136.

Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., & Yang, R. (2019). IoU Loss for 2D/3D Object Detection. *2019 International Conference on 3D Vision (3DV)*, 85-94. https://doi.org/10.1109/3DV.2019.00019

Zhu, L., Geng, X., Li, Z., & Liu, C. (2021). Improving YOLOv5 with Attention Mechanism for Detecting Boulders from Planetary Images. *Remote Sensing*, *13*(18), Article 18. https://doi.org/10.3390/rs13183776