MULTIMODAL CONVOLUTIONAL NEURAL NETWORKS FOR SPERM MOTILITY AND CONCENTRATION PREDICTIONS

GOH VOON HUEH

UNIVERSITI TEKNOLOGI MALAYSIA

ACKNOWLEDGEMENT

In pursue of this master's degree, I was indebted to countless people who have supported me all the way until it was completed. My supervisor, Ts Dr Muhammad Amir As'ari has provided me invaluable feedback on my results, and that I am benefited greatly from his wealth of knowledge and patience editing. It is my honour to have the opportunity to learn so much from Dr Amir.

A very special note to my very dear friend Ying Hui, for her immense emotional support and company everyday even though we were 4500km apart from each other, not to mentioned her technical suggestions which ease so much of my burden. Also, I extend my thanks to Chee Ching, Rebecca, and Jolene for being supportive and cheering me up for times I am in doubt.

Last but not least, I am grateful for my parents and brother whose constant love and support keep me motivated and confident. I could not have come so far without them. Thank you all.

ABSTRACT

Manual semen analysis is a conventional method to assess male infertility which includes laboratory technicians examining on parameters such as sperm motility and concentration. Manual evaluation is prone to human errors that causes precision and accuracy issues. The purpose of this research study is to adopt computer vision deep learning techniques and multimodal learning approach in sperm parameters prediction using video-based and image-based input. Convolutional neural network (CNN) has benefited technology industry in recent years, and it has been widely applied in computer vision research tasks as well. Most of the well-established model were designed and pretrained for image-based input, whereas temporal information of video-based input might not be extracted properly using these architectures. Three-dimensional CNN (3DCNN) would be an alternative as it was designed to extract motion and temporal features, which are vital for sperm motility prediction. For sperm concentration, since two-dimensional CNN (2DCNN) is efficient in recognizing and extracting spatial features, Residual Network (ResNet) could be adopted for sperm concentration prediction with minimal modification on the original architecture. On the other hand, multimodal learning approach is a technique to aggregate learnt features from different deep learning architecture that adopted other forms of modalities, and provide deep learning model better insights on their tasks. Hence, multimodal learning has been introduced in this research study, where the finalized deep learning architecture received both image-based (frames extracted from video samples) and video-based (stacked frames pre-processed from video samples) input that could provide well-extracted spatial and temporal features for sperm parameters prediction. In this research study, VISEM dataset has been used because it is an open-source dataset which contains 85 sperm videos and biological analysis data from different patients. The video samples went through pre-processing stage to obtain the suitable modalities for training and validation. The developed system has been proven to be capable of improving performance which was as proposed, after the results had been compared to other similar research works. Average mean absolute error (MAE) for sperm motility was observed with high accuracy up to 8.05, and competent performance for sperm concentration with Pearson's correlation coefficient (R_P) of 0.853.

ABSTRAK

Analisis semen yang konvensional menggunakan kaedah manual untuk menilai ketidaksuburan lelaki, termasuk pemeriksaan yang dijalankan oleh juruteknik makmal pada parameter motoliti sperma dan kepekatan sperma. Pemeriksaan manual tersebut mudah terdedah kepada kesilapan manusia dan menyumbang kepada isu ketepatan. Tujuan kajian ini adalah untuk mengguna teknik pembelajaran dalam penglihatan mesin dan pembelajaran multimodal dalam analisis parameter sperma menggunakan input berdasarkan video dan imej. Rangkaian saraf perlingkaran (CNN) telah memanfaatkan industri teknologi sejak kebelakangan ini, dan ia juga digunakan secara secara meluas dalam penyelidikan yang melibatkan penglihatan mesin. Kebanyakan input model yang mantap telah direka dan dilatih terlebih dahulu berasaskan imej input, manakala maklumat yang berdasarkan video input mungkin tidak dapat diekstrak dengan betul menggunakan model tersebut. CNN tiga-dimensi (3DCNN) akan menjadi model alternatif kerana ia direka untuk mengekstrak ciri pergerakan dan temporal, yang amat penting untuk analisis motoliti sperma. Untuk kepekatan sperma, memandangkan CNN dua-dimensi sudah cekap dalam mengestrak dan mengenali ciri ruangan, model ResNet digunapakai untuk penjangkaan kepekatan sperma dengan pengubahsuaian yang minimum pada model asal. Selain itu, pembelajaran multimodal ialah teknik untuk mengagregatkan ciri-ciri yang dipelajari daripada model dan modaliti yang berbeza, serta memberikan model pembelajaran dalam dengan cerapan yang lebih baik untuk melaksanakan tugasnya. Oleh itu, pembelajaran multimodal telah diperkenalkan dalam kajian ini, di mana model pembelajaran dalam yang dimuktamadkan akan menerima kedua-dua input berasaskan imej (imej yang diekstrak dari sampel video) dan video (imej yang disusun sebelum diproses dari sampel video) yang mewakili ciri ruangan dan temporal untuk ramalan parameter sperma. Dalam penyelidikan ini, dataset VISEM digunakan kerana ia adalah sumber terbuka dan mengandungi 85 video sperma dan data analisis biologi daripada pesakit yang berbeza. Sampel video telah melalui peringkat pra-pemprosesan untuk mendapatkan modaliti yang sesuai untuk tujuan latihan dan pengesahan model. Sistem yang diperkenalkan telah dibuktikan bahawa ia dapat meningkatkan prestasi seperti yang dicadangkan selepas perbandingan dengan kerja penyelidikan lain yang serupa. Purata Min Ralat Mutlak (MAE) untuk motoliti sperma mencapai ketepatan tinggi yang tidak kurang daripada 8.05, manakala prestasi yang kompeten juga dicapai untuk kepekatan sperma dengan 0.843 dalam korelasi Pearson (R_P).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CNN	-	Convolutional Neural Network
3DCNN		Three-dimensional Convolutional Neural Network
2DCNN	-	Two-dimensional Convolutional Neural Network
MAE	-	Mean Absolute Error
WHO	-	World Health Organization
CASA	-	Computer Aided Sperm Analyzer
SQA	-	Sperm Quality Analyzer
IQA	-	Internal Quality Assessment
EQA	-	External Quality Assessment
CI	-	Confidence Interval
SVM	-	Support Vector Machine
ANN	-	Artificial Neural Network
ReLU	-	Rectified Linear Units
GPU	-	Graphics Processing Unit
CPU	-	Central Processing Unit
TPU	-	Tensor Processing Unit
ILSVRC	-	ImageNet Large Scale Visual Recognition Challenge
ResNet	-	Residual Network
SOTA	-	State-of-the-Art
MRI	-	Magnetic Resonance Imaging
MLP	-	Multilayer Perceptron
IDE	-	Integrated Development Environment
RMSE	-	Root Mean Squared Error
MSE	-	Mean Squared Error
PR	-	Progressive
NPR	-	Non-Progressive
IM	-	Immotile

LIST OF SYMBOLS

R_P	-	Pearson's correlation coefficient
R_S	-	Spearman's rank correlation

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Infertility, is a medical condition where sexually active and non-contracepting couples are unable to successfully achieve clinical pregnancy, as defined by World Health Organization (WHO) [1]. Infertility could happened in both men and women, however more than half of the failure in childbearing was contributed by infertile men [2]. Male infertility is normally due to genetic issue, unhealth environment and lifestyle. To identify the male factor infertility, the initial evaluation should at least include one properly performed semen analysis and physical examination performed by experienced examiner, along with a detailed reproductive history [3]. If abnormal values or condition has been observed after initial screening evaluation, only a full evaluation by urologist or other specialist be carried out [3]. Hence, semen parameter analysis is one of the primary and important analysis required to study the probability of male infertility issue of an infertile couple, then only treatment planning options are available for conception. Most of the semen analysis are not open-sourced. Nevertheless, the dataset used in this research study is VISEM dataset, which is an online multimodal dataset that contains videos and biological analysis data from 85 anonymized participants. Semen parameter analysis is mostly carried out by manual approach, which the guidelines have been provided clearly in semen analysis manual given by WHO, however it is often susceptible to human related errors. This stimulated the development of automated semen analyser such as CASA (Computer Aided Semen Analyzer) few decades ago, and a recent model SQA (Sperm Quality Analyzer). However, it received several criticisms regarding its inconsistent handling methods for routine clinical analysis, as well as credibility on accuracy and precision [4]–[12]. Therefore, these limitations allow a room for improvement to be explored by other approach such as deep learning. In this era, where artificial intelligence concept is vastly used in image processing, recognition, and classification, it is expected that

technology advancement would slowly improve and overcome the criticisms received 30 years back then.

1.2 Problem Statement

Most common semen parameter analysis is done manually by experienced laboratory technician to evaluate the sperm conditions, which required intensive training and regular participation in quality assurance programs. Even though it is the most common practice for semen analysis, it has been revealed with limitations such as lack of standardization in methodologies and tools used in lab, that caused precision and accuracy are being compromised [4], [13]–[14]. Besides the procedures that are causing inaccuracies in semen analysis, some sperm parameters evaluations are merely based on human subjective judgments such as sperm morphology classification. Even though guidelines are provided by WHO in the semen analysis manual, several studies have revealed that time spent for technician training will also produce inconsistent results [15]. The parameters which would normally be included in the manual semen analysis are sperm concentration, total sperm count, sperm motility, sperm morphology, semen volume, semen viscosity, pH values of semen sample and sperm vitality.

The inconsistencies in human evaluation method induced the development of automated semen analyzer, which started about 30 years ago with CASA and a recent model named as SQA [16]. From several comparison studies of both models using manual evaluation as gold standard, generally SQA series presented closer results with manual evaluation than CASA. However there were also contradict findings from different research groups [4]–[7] and both devices still required minimal training to operate the system despite being identified as automated system. Some of the laboratory personnel were reported not running standard operating procedures and laboratory environments as suggested by the manufacturing companies. Hence, these downsides of automated semen analysers are WHO's concerns to not encourage them in routine clinical analysis, but it could be the chance to explore other better approaches for semen analysis.

Several research works have been explored by researchers to solve above mentioned issues using machine or deep learning approach. The difference between machine and deep learning is that machine learning is a "shallow" classifier, much smaller and simpler architecture than deep learning architecture [17]. However, it is not suitable to learn non-linear input such as image and speech recognition without designing specifically tailored feature extractor, which is harder to adapt same architecture/method on other application [17]. Thus, deep learning convolutional neural network (CNN) is often being explored by other researchers to tackle computer vision related problems, due to its ability to compute non-linear input and deliver non-linear output. In the research works focusing on sperm parameters prediction, most of the previous studies adopted transfer learning techniques from state-of-the-art (SOTA) 2DCNN or self-customized 3DCNN architectures, with image-based or video-based modality extracted from video samples to predict sperm motility, concentration, and morphology [18]–[23]. By looking at sperm motility prediction, no significant improvements were observed which indicates either the architectures or modalities chosen by the researchers were less effective in learning the temporal information [18]–[23]. Currently, there were no research studies on sperm concentration prediction using computer vision deep learning approach. Nevertheless, there was a similar research work demonstrated concentration prediction using artificial neural networks (ANN) but the results were not comparable as the accuracy calculation were inaccurate [24].

Multimodal learning approach has been introduced in medical and assistive technology fields to unravel more information from different data sources. For instance, knee angle estimation and sensor-based human activity recognition adopted several different types of sensors data, mental illness studies and medical image segmentation that comprised of multimodalities which are structural and functional magnetic resonance imaging (MRI) data, and etc [25]–[30]. However, similar research studies on semen parameters prediction mostly adopted unimodal learning where the proposed architectures used only one type of modality, either image-based or video-based modality [19]–[23]. There were other attempts made by researchers to incorporate multimodal learning concepts into motility and morphological predictions, where image-based and tabular data served as the input types for 2DCNN architecture [18]. However, results showed that not only the model's performance using

multimodal learning approach (image-based and tabular data) did not outperform the unimodal learning approach but degrades instead. Here it might indicate that the choice of tabular data as additional information did not provide any advantages. All in all, multimodal learning approach could probably achieve better performance than unimodal learning, but the choice of modalities and architecture selection shall be considered wisely.

1.3 Objectives

Based on the aforementioned problem statement, this research study aims to achieve objectives as stated below:

(a) To formulate multimodal deep learning methods in sperm parameters prediction by adopting image-based and video-based input.

(b) To integrate different deep learning architectures comprised of 3DCNN and 2DCNN for sperm parameters prediction focusing on sperm motility and concentration prediction. To predict/develop by combining architectures

(c) To compare and validate the results obtained from the proposed method with related research works.

1.4 Project Scopes

Sperm motility and sperm concentration have been observed as better fertility predictors than other sperm parameters, hence these two parameters were chosen as the focus of this study. The aim was to predict sperm motility and concentration by using combined deep learning architectures that employ multimodal learning methods. The desired framework was a combination of 2DCNN and 3DCNN that accepted multiple modalities as input, which were image-based and video-based modalities. It was expected to harness spatial and temporal data better than a model that only takes

in 2D data, which was an image-based input in this context. The modalities (image-based and video-based) were prepared by pre-processing the extracted frames of video samples from an online and open-source multimodal database, Simula VISEM [31]. It is a multimodal dataset containing 85 videos of semen samples and participants' anonymized data [31]. As the purpose of this study was to perform prediction using computer vision deep learning approach, therefore only video samples from this dataset were used for modalities generation, and the manual biological data analysis was used as reference for ground truth. The model architectures were developed using PYTHON language with PyTorch as the framework and Google Colaboratory as the development environment. Data analysis and visualization were done by using Microsoft Excel to compare the performance. The summary of project scopes was visualized as shown in Figure 1.1.

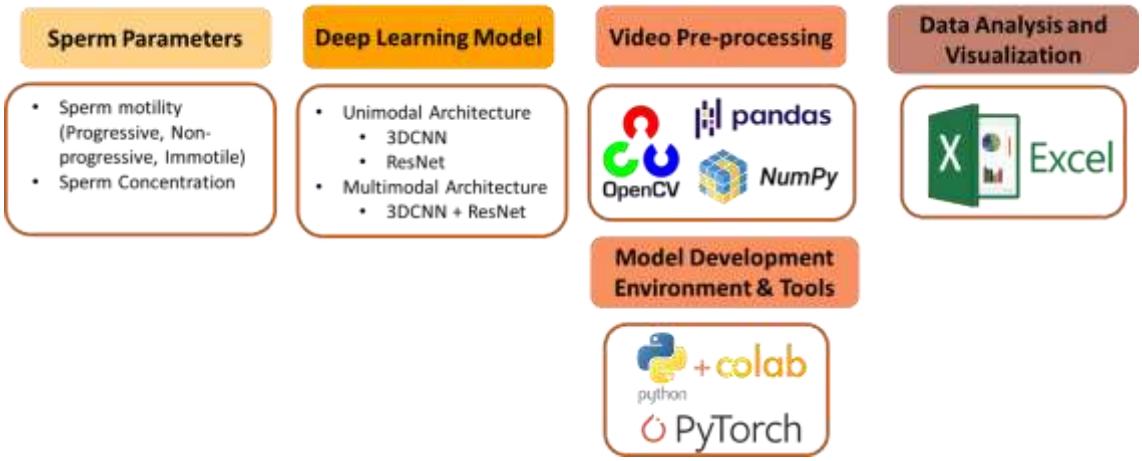

Figure 1.1 Scope of Studies

1.5 Significance of Study

This research study focused on delivering a multimodal deep learning architecture that can predict sperm motility and sperm concentration, which are one of the primary parameters to indicate the fertility condition of a male with comparable accuracy to existing studies. Manual evaluation and automated semen analyzer required a certain level of human judgments and intervention during semen analysis or minimal human operations on the semen analyzer. By achieving the objectives of

the study, it allows sperm analysis procedures to be less dependent on humans compared with the current clinical approach that includes complicated procedures. Besides, it introduced an automated multiple semen parameters prediction system using multimodal deep learning approach which has yet to be introduced by previous similar research works.

REFERENCES

[1] World Health Organization, “Infertility,” 13-Sep-2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/infertility. [Accessed: 31-Jan-2023].

[2] M. C. Inhorn and P. Patrizio, “Infertility around the globe: New thinking on gender, reproductive technologies and global movements in the 21st century,” *Hum. Reprod. Update*, vol. 21, no. 4, pp. 411–426, 2014.

[3] C. L. R. Barratt *et al.*, “The diagnosis of male infertility: An analysis of the evidence to support the development of global WHO guidance-challenges and future research opportunities,” *Hum. Reprod. Update*, vol. 23, no. 6, pp. 660–680, 2017.

[4] J. Lammers, S. Chtourou, A. Reignier, S. Loubersac, P. Barrière, and T. Fréour, “Comparison of two automated sperm analyzers using 2 different detection methods versus manual semen assessment,” *J. Gynecol. Obstet. Hum. Reprod.*, vol. 50, no. 8, 2021.

[5] J. Lammers, C. Splingart, P. Barrière, M. Jean, and T. Fréour, “Double-blind prospective study comparing two automated sperm analyzers versus manual semen assessment,” *J. Assist. Reprod. Genet.*, vol. 31, no. 1, pp. 35–43, 2014.

[6] K. M. Engel, S. Grunewald, J. Schiller, and U. Paasch, “Automated semen analysis by SQA Vision ® versus the manual approach—A prospective double-blind study,” *Andrologia*, vol. 51, no. 1, pp. 1–10, 2019.

[7] O. M. Yis, “Comparison of fully automatic analyzer and manual measurement methods in sperm analysis and clinical affect,” *Exp. Biomed. Res.*, vol. 3, no. 4, pp. 224–230, 2020.

[8] J. Talarczyk-Desole, A. Berger, G. Taszarek-Hauke, J. Hauke, L. Pawelczyk, and P. Jedrzejczak, “Manual vs. computer-assisted sperm analysis: Can CASA replace manual assessment of human semen in clinical practice?,” *Ginekol. Pol.*, vol. 88, no. 2, pp. 56–60, 2017.

[9] W. M. Medical Electronic Systems, “SQA-Vision Automated sperm quality analyzer.”

[10] Source Medical, “SQA-V Gold.” [Online]. Available:

http://sourcemedi.com/product/sqa-v-gold/1. [Accessed: 31-Jan-2023].

[11] T. G. Cooper and C. H. Yeung, “Computer-aided evaluation of assessment of ‘grade a’ spermatozoa by experienced technicians,” *Fertil. Steril.*, vol. 85, no. 1, pp. 220–224, 2006.

[12] S. T. Mortimer, G. Van Der Horst, and D. Mortimer, “The future of computer-aided sperm analysis,” *Asian J. Androl.*, vol. 17, no. 4, pp. 545–553, 2015.

[13] J. Auger *et al.*, “Intra- and inter-individual variability in human sperm concentration, motility and vitality assessment during a workshop involving ten laboratories,” *Hum. Reprod.*, vol. 15, no. 11, pp. 2360–2368, 2000.

[14] M. J. Tomlinson, “Uncertainty of measurement and clinical value of semen analysis: has standardisation through professional guidelines helped or hindered progress?,” *Andrology*, vol. 4, no. 5, pp. 763–770, 2016.

[15] U. Punjabi, C. Wyns, A. Mahmoud, K. Vernelen, B. China, and G. Verheyen, “Fifteen years of Belgian experience with external quality assessment of semen analysis,” *Andrology*, vol. 4, no. 6, pp. 1084–1093, 2016.

[16] A. Agarwal and R. K. Sharma, “Automation is the key to standardized semen analysis using the automated SQA-V sperm quality analyzer,” *Fertil. Steril.*, vol. 87, no. 1, pp. 156–162, 2007.

[17] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] S. A. Hicks *et al.*, “Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019.

[19] V. Thambawita, P. Halvorsen, H. Hammer, M. Riegler, and T. B. Haugen, “Extracting temporal features into a spatial domain using autoencoders for sperm video analysis,” *arXiv*, vol. 2670, pp. 3–5, 2019.

[20] V. Thambawita, P. Halvorsen, H. Hammer, M. Riegler, and T. B. Haugen, “Stacked dense optical flows and dropout layers to predict sperm motility and morphology,” *arXiv*, vol. 10, pp. 9–11, 2019.

[21] S. Hicks, T. B. Haugen, P. Halvorsen, and M. Riegler, “Using deep learning to predict motility and morphology of human sperm,” *CEUR Workshop Proc.*, vol. 2670, pp. 5–7, 2019.

[22] J. M. Rosenblad, S. Hicks, H. K. Stensland, T. B. Haugen, P. Halvorsen, and M. Riegler, “Using 2D and 3D convolutional neural networks to predict semen

quality," *CEUR Workshop Proc.*, vol. 2670, no. October, pp. 27–29, 2019.

[23] Priyansi, B. Bhattacharjee, and J. H. Rahim, "Predicting Semen Motility using three-dimensional Convolutional Neural Networks," pp. 1–8, 2021.

[24] A. Lesani *et al.*, "Quantification of human sperm concentration using machine learning-based spectrophotometry," *Comput. Biol. Med.*, vol. 127, no. August, p. 104061, 2020.

[25] H. Wu, Q. Huang, D. Wang, and L. Gao, "A CNN-SVM combined regression model for continuous knee angle estimation using mechanomyography signals," *Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019*, no. Itnec, pp. 124–131, 2019.

[26] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, "CNN-based sensor fusion techniques for multimodal human activity recognition," *Proc. - Int. Symp. Wearable Comput. ISWC*, vol. Part F1305, pp. 158–165, 2017.

[27] V. Calhoun and S. Plis, "205. Deep Learning Approaches to Unimodal and Multimodal Analysis of Brain Imaging Data With Applications to Mental Illness," *Biol. Psychiatry*, vol. 83, no. 9, pp. S82–S83, 2018.

[28] M. Tang, P. Kumar, H. Chen, and A. Shrivastava, "Deep multimodal learning for the diagnosis of autism spectrum disorder," *J. Imaging*, vol. 6, no. 6, 2020.

[29] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 162–169, 2019.

[30] R. Balakrishnan and R. Priya, "Multimodal Medical Image Fusion based on Deep Learning Neural Network for Clinical Treatment Analysis," *Int. J. ChemTech Res.*, vol. 11, no. 06, pp. 160–176, 2018.

[31] T. B. Haugen *et al.*, "VISEM: A multimodal video dataset of human spermatozoa," *Proc. 10th ACM Multimed. Syst. Conf. MMSys 2019*, pp. 261–266, Jun. 2019.

[32] D. S. Cram, M. K. O'Bryan, and D. M. De Kretser, "Male infertility genetics - The future," *J. Androl.*, vol. 22, no. 5, pp. 738–746, 2001.

[33] M. N. Mascarenhas, S. R. Flaxman, T. Boerma, S. Vanderpoel, and G. A. Stevens, "National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys," *PLoS Med.*, vol. 9, no. 12, pp. 1–12, 2012.

[34] J. Cunningham, “Infertility: A primer for primary care providers,” *J. Am. Acad. Physician Assist.*, vol. 30, no. 9, pp. 19–25, 2017.

[35] K. P. Nallella, R. K. Sharma, N. Aziz, and A. Agarwal, “Significance of sperm characteristics in the evaluation of male infertility,” *Fertil. Steril.*, vol. 85, no. 3, pp. 629–634, 2006.

[36] World Health, “Examination and processing of human semen,” *World Health*, vol. Edition, V, no. 10, p. 286, 2010.

[37] B. A. Keel, P. Quinn, C. F. Schmidt, N. T. Serafy, N. T. Serafy, and T. K.schalue, “Results of the American Association of Bioanalysts national proficiency testing programme in andrology,” *Hum. Reprod.*, vol. 15, no. 3, pp. 680–686, 2000.

[38] H. . Ahrendt and K.J. Buhling, “Reproduktionsmedizin und Endokrinologie,” *J. Reprod. Med. Endocrinol.*, vol. 1, no. 3, pp. 194–201, 2006.

[39] R. Walczak-Jedrzejowska, K. Marchlewska, E. Oszukowska, E. Filipiak, L. Bergier, and J. Slowikowska-Hilczer, “Semen analysis standardization: Is there any problem in Polish laboratories?,” *Asian J. Androl.*, vol. 15, no. 5, pp. 616–621, 2013.

[40] C. Mallidis, T. G. Cooper, B. Hellenkemper, M. Lablans, F. Ückert, and E. Nieschlag, “Ten years’ experience with an external quality control program for semen analysis,” *Fertil. Steril.*, vol. 98, no. 3, 2012.

[41] L. Björndahl, C. L. R. Barratt, D. Mortimer, and P. Jouannet, “‘How to count sperm properly’: Checklist for acceptability of studies based on human semen analysis,” *Hum. Reprod.*, vol. 31, no. 2, pp. 227–232, 2016.

[42] N. Kumar and A. Singh, “Trends of male factor infertility, an important cause of infertility: A review of literature,” *J. Hum. Reprod. Sci.*, vol. 8, no. 4, pp. 191–196, 2015.

[43] T. G. Cooper *et al.*, “World Health Organization reference values for human semen characteristics,” *Hum. Reprod. Update*, vol. 16, no. 3, pp. 231–245, 2009.

[44] H. O. Ilhan and N. Aydin, “Smartphone based sperm counting - an alternative way to the visual assessment technique in sperm concentration analysis,” *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 6409–6435, 2020.

[45] M. reza Mohammadi, M. Rahimzadeh, and A. Attar, “Sperm Detection and Tracking in Phase-Contrast Microscopy Image Sequences using Deep

Learning and Modified CSR-DCF," 2020.

[46] B. DUCOT, A. SPIRA, D. FENEUX, and P. JOUANNET, "Male factors and the likelihood of pregnancy in infertile couples. 11. Study of clinical characteristics — practical consequences," *Int. J. Androl.*, vol. 11, no. 5, pp. 395–404, 1988.

[47] S. Ottl, M. Gerczuk, S. Amiriparian, and B. Schuller, "A Machine Learning Framework for Automatic Prediction of Human Semen Motility," 2021.

[48] J. F. Moruzzi, A. J. Wyrobek, B. H. Mayall, and B. L. Gledhill, "Quantification and classification of human sperm morphology by computer-assisted image analysis," *Fertil. Steril.*, vol. 50, no. 1, pp. 142–152, 1988.

[49] T. Akashi, I. Mizuno, A. Okumura, and H. Fuse, "Usefulness of sperm quality analyzer-V (SQA-V) for the assessment of sperm quality in infertile men," *Arch. Androl.*, vol. 51, no. 6, pp. 437–442, 2005.

[50] S. T. Mortimer, G. Van Der Horst, and D. Mortimer, "The future of computer - aided sperm analysis," no. April, 2015.

[51] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," pp. 1–11, 2015.

[52] J. Leitner, "Towards adaptive and autonomous humanoid robots: From Vision to Actions," no. September, 2014.

[53] D. Jakhar and I. Kaur, "Artificial intelligence, machine learning and deep learning: definitions and differences," *Clin. Exp. Dermatol.*, vol. 45, no. 1, pp. 131–132, 2020.

[54] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "GauGAN," no. July, pp. 1–1, 2019.

[55] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[56] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2242–2251, 2017.

[57] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2868–2876, 2017.

[58] "Understanding of Convolutional Neural Network (CNN) — Deep Learning | by Prabhu | Medium." [Online]. Available:

https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148. [Accessed: 25-Apr-2021].

[59] "Convolutional Neural Networks from the ground up | by Alejandro Escontrela | Towards Data Science." [Online]. Available: https://towardsdatascience.com/convolutional-neural-networks-from-the-ground-up-c67bb41454e1. [Accessed: 24-Apr-2021].

[60] "Introduction to Different Activation Functions for Deep Learning | by Shruti Jadon | Medium." [Online]. Available: https://medium.com/@shrutijadon10104776/survey-on-activation-functions-for-deep-learning-9689331ba092. [Accessed: 25-Apr-2021].

[61] "The best explanation of Convolutional Neural Networks on the Internet! | by Harsh Pokharna | TechnologyMadeEasy | Medium." [Online]. Available: https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8.

[62] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Journal. Pract.*, vol. 10, no. 6, pp. 730–743, Feb. 2015.

[63] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[64] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[66] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.

[67] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.

[68] "Understanding 1D and 3D Convolution Neural Network | Keras | by Shiva Verma | Towards Data Science." [Online]. Available:

https://towardsdatascience.com/understanding-1d-and-3d-convolution-neural-network-keras-9d8f76e29610. [Accessed: 25-Apr-2021].

[69] J. Summaira, X. Li, A. M. Shoib, S. Li, and J. Abdul, *Recent Advances and Trends in Multimodal Deep Learning: A Review*. 2021.

[70] K. Liu, Y. Li, N. Xu, and P. Natarajan, "Learn to Combine Modalities in Multimodal Deep Learning," 2018.

[71] Y. Li *et al.*, "CR-Net: A Deep Classification-Regression Network for Multimodal Apparent Personality Analysis," *Int. J. Comput. Vis.*, vol. 128, no. 12, pp. 2763–2780, 2020.

[72] Y. Mroueh, E. Marcheret, and V. Goel, "Deep multimodal learning for Audio-Visual Speech Recognition," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2015-Augus, pp. 2130–2134, 2015.

[73] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800–1807, 2017.

[74] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 4278–4284, 2017.

[75] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016.

[76] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2017-Decem, pp. 319–323, 2018.

[77] S. Lessmann and S. Voß, "Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy," *Int. J. Forecast.*, vol. 33, no. 4, pp. 864–877, 2017.

[78] K. Wang, M. Bansal, and J. M. Frahm, "Retweet wars: Tweet popularity prediction via dynamic multimodal regression," *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 1842–1851, 2018.

[79] M. J. Tomlinson *et al.*, "Validation of a novel computer-assisted sperm analysis (CASA) system using multitarget-tracking algorithms," *Fertil. Steril.*, vol. 93, no. 6, pp. 1911–1920, 2010.

[80] P. Sedgwick, “Pearson’s correlation coefficient,” *BMJ*, vol. 345, no. 7864, pp. 1–2, 2012.

[81] “Correlation - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Correlation. [Accessed: 25-Jun-2022].

[82] “Monotonic function - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Monotonic_function. [Accessed: 25-Jun-2022].