RBMO

ELSEVIER

ARTICLE

Evaluation of an artificial intelligence-facilitated sperm detection tool in azoospermic samples for use in ICSI

BIOGRAPHY

Dale Goss is a PhD student at the University of Technology Sydney and a graduate of Stellenbosch University and Monash University. He is a clinical embryologist at IVFAustralia and as a scientific advisor for NeoGenix Biosciences. His research focuses on human embryology, male infertility, and technology in assisted reproduction.

Dale M. Goss[1,2,3,†], Steven A. Vasilescu[1,2,†], Phillip A. Vasilescu[2], Simon Cooke[3], Shannon HK. Kim[3,4], Gavin P. Sacks[1,3,4], David K. Gardner[2,5], Majid E. Warkiani[1,2,6,*]

KEY MESSAGE

This proof-of-concept study shows that an artificial intelligence image analysis tool can drastically improve sperm search times on testicular tissue samples, thus reducing physical strain and fatigue on embryologists and possibly improving the chance of finding spermatozoa. This is a highly translatable clinical tool for the treatment of severe male factor infertility.

ABSTRACT

Research question: Can artificial intelligence (AI) improve the efficiency and efficacy of sperm searches in azoospermic samples?

Design: This two-phase proof-of-concept study began with a training phase using eight azoospermic patients (>10,000 sperm images) to provide a variety of surgically collected samples for sperm morphology and debris variation to train a convolutional neural network to identify spermatozoa. Second, side-by-side testing was undertaken on two cohorts of non-obstructive azoospermia patient samples: an embryologist versus the AI identifying all the spermatozoa in the still images (cohort 1, $n = 4$), and a side-by-side test with a simulated clinical deployment of the AI model with an intracytoplasmic sperm injection microscope and the embryologist performing a search with and without the aid of the AI (cohort 2, $n = 4$).

Results: In cohort 1, the AI model showed an improvement in the time taken to identify all the spermatozoa per field of view ($0.02 \pm 0.30 \times 10^{-5}$s versus 36.10 ± 1.18s, $P < 0.0001$) and improved recall ($91.95 \pm 0.81\%$ versus $86.52 \pm 1.34\%$, $P < 0.001$) compared with an embryologist. From a total of 2660 spermatozoa to find in all the samples combined, 1937 were found by an embryologist and 1997 were found by the AI in less than 1000th of the time. In cohort 2, the AI-aided embryologist took significantly less time per droplet (98.90 ± 3.19 s versus 168.7 ± 7.84 s, $P < 0.0001$) and found 1396 spermatozoa, while 1274 were found without AI, although no significant difference was observed.

Conclusions: AI-powered image analysis has the potential for seamless integration into laboratory workflows, to reduce the time to identify and isolate spermatozoa from surgical sperm samples from hours to minutes, thus increasing success rates from these treatments.

[1] School of Biomedical Engineering, University of Technology Sydney, Sydney, New South Wales, Australia.
[2] NeoGenix Biosciences Pty Ltd, Sydney, New South Wales, Australia.
[3] IVFAustralia, Sydney, New South Wales, Australia.
[4] University of New South Wales, Sydney, New South Wales, Australia.
[5] Melbourne IVF, Melbourne, Victoria, Australia.
[6] Institute for Biomedical Materials & Devices (IBMD), University of Technology Sydney, Sydney, New South Wales, Australia.
† The authors should be considered as joint first authors.

*Corresponding author. E-mail address: majid.warkiani@uts.edu.au (M.E. Warkiani).
https://doi.org/10.1016/j.rbmo.2024.103910 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

KEY WORDS

Azoospermia
Male infertility
Microdissection testicular sperm extraction
Spermatozoa
Surgical sperm collection

INTRODUCTION

Male infertility is increasing worldwide at an alarming rate, sperm counts having declined by 50% over the past 50 years (*Levine et al., 2023*). Around 30% of cases of human infertility are caused solely by male infertility and 50% of cases are attributed to having male infertility as a contributing factor (*Agarwal et al., 2015*). While assisted reproductive technology has proved to be effective in treating infertile couples, some forms of male infertility remain difficult to treat. Azoospermia, defined as the absence of spermatozoa in centrifuged semen on at least two occasions, is the most severe form of male infertility, affecting 10–20% of infertile men and 1% of the general male population (*Verheyen et al., 2017*; *Wosnitzer et al., 2014*).

Azoospermia can be classified as obstructive and/or non-obstructive. Obstructive azoospermia occurs due to obstruction of the reproductive tract and constitutes 40% of azoospermic cases, while non-obstructive azoospermia (NOA) results from primary, secondary or incomplete/ambiguous testicular failure, which compromises sperm production and constitutes 60% of cases of azoospermia (*Jarow et al., 1989*; *Wosnitzer, et al., 2014*). Patients with obstructive azoospermia can attempt reconstruction (vasovasostomy, vasoepididymostomy or transurethral resection of the ejaculatory duct) when possible, or surgical sperm collection can be performed from the testis via testicular sperm aspiration (TESA), testicular sperm extraction (TESE) or microdissection TESE (mTESE), or from the epididymis via microsurgical epididymal sperm aspiration or percutaneous epididymal sperm aspiration (*Flannigan et al., 2017*; *Schrepferman et al., 2001*). Patients with NOA require sperm extraction from the testis (TESA, TESE or mTESE), and the surgically collected spermatozoa are then used for intracytoplasmic sperm injection (ICSI).

The gold-standard for treating patients with NOA is mTESE, which has a high sperm retrieval rate of up to 64% in suitable patients (*Deruyver et al., 2014*; *Ramasamy et al., 2005*; *Schiff et al., 2005*). Although these rates seem promising, the current manual examination process to find spermatozoa within tissue recovered from mTESE operations is time-consuming and inefficient, typically taking anywhere between 1 and 6 h of laboratory time, and in some cases even up to 14 h (*Mangum et al., 2020*; *Ramasamy et al., 2011*). This extended time is due to the requirement for manual searching through prepared suspensions of testicular tissue with a microscope, before using isolated sperm for ICSI.

The outcome of such searching is heavily dependent upon the complexity and contamination of the suspension provided to the embryologists by the surgeon. Viable sperm are easily overlooked due to variables such as collateral cell density, resulting in a process that is prone to human error, combined with inexperience and fatigue of laboratory staff (*Ramasamy, et al., 2011*). For patients with NOA, overlooking spermatozoa because of human error could wrongly indicate absolute infertility (*Samuel et al., 2016*). Similarly, for extended sperm searches in semen as a diagnostic test or as a last check of the ejaculate before surgery, failure to identify any spermatozoa present could unnecessarily direct patients into surgery. Furthermore, prolonged sample examination procedures can have adverse effects on the viability of the spermatozoa, consequently affecting their potential for fertilization and thus undermining the efforts of sperm searches and the considerable cost and physical strain caused to patients during these procedures (*Ouitrakul et al., 2018*). For individuals with NOA, a more efficient and higher throughput method capable of locating and isolating spermatozoa from the suspension would therefore greatly benefit the clinical workflow of assisting severe forms of male infertility.

Panning through surgically collected sperm samples under a microscope is a form of manual image analysis which machine learning and artificial intelligence (AI) have the potential to automate and improve. Therefore, with preliminary works showing promising results (*Goss et al., 2023*), this study aims to comprehensively assess the use of an assistive convolutional neural network (CNN) AI that was developed and trained to identify spermatozoa in complex tissue suspensions in real time (**FIGURE 1**). Using a YOLOv8 model (Ultralytics, USA), an open-source, high-speed, high-accuracy object detection and image segmentation model, this software works in tandem with an embryologist to instantly identify and alert embryologists to spermatozoa of interest for their assessment from the camera feed mounted into their microscope. The objective of this study was, for cohort 1, to compare the AI with embryologists working without the aid of the AI aid in terms of time, recall and number of spermatozoa found using still images, and then for cohort 2, to run a simulated sperm search with the AI integrated into an ICSI microscope kit, to demonstrate its potential for clinical implementation.

MATERIALS AND METHODS

Ethical approval

Ethical approval for healthy sperm samples was received from the University of Technology Sydney ethics review board (ETH19-3677, approved 12 December 2019), and for the use of discarded testicular tissue samples was received from the IVFAustralia Human Research Ethics Committee (DG01192, approved 2 August 2022) and University of Technology Sydney ethics review board (ETH22-7189, approved 30 September 2022).

Preparation of specifically prepared samples

To generate images for the training dataset prior to access to clinical testicular tissue samples, specifically prepared samples were used for the initial training of the AI model (**FIGURE 1**A). These samples consisted of donor spermatozoa, fingerprick blood and cells from epithelial cell culture lines.

Human semen samples were obtained through ejaculation after 2–5 days of sexual abstinence (*WHO, 2021*). Raw semen samples were left at room temperature for 20 min to allow for liquefaction. Samples were centrifuged for 8 min at 500*g* to separate the sperm pellet from the seminal plasma. Red blood cells (RBC) were obtained from whole-blood specimens within 3 days of collection. The collected blood samples were also resuspended in G-MOPS Plus (Vitrolife, Sweden) medium. Mixed cell suspensions were created to simulate testicular tissues samples containing spermatozoa, RBC, white blood cells (WBC), epithelial cells and C2C12 and THP-1 cells (Sigma-Aldrich, USA). All the cells were mixed in warmed G-MOPS Plus (37°C). Raw semen samples were diluted down to between 1×10^7 and 1×10^8 spermatozoa/ml, the RBC concentration was in the range

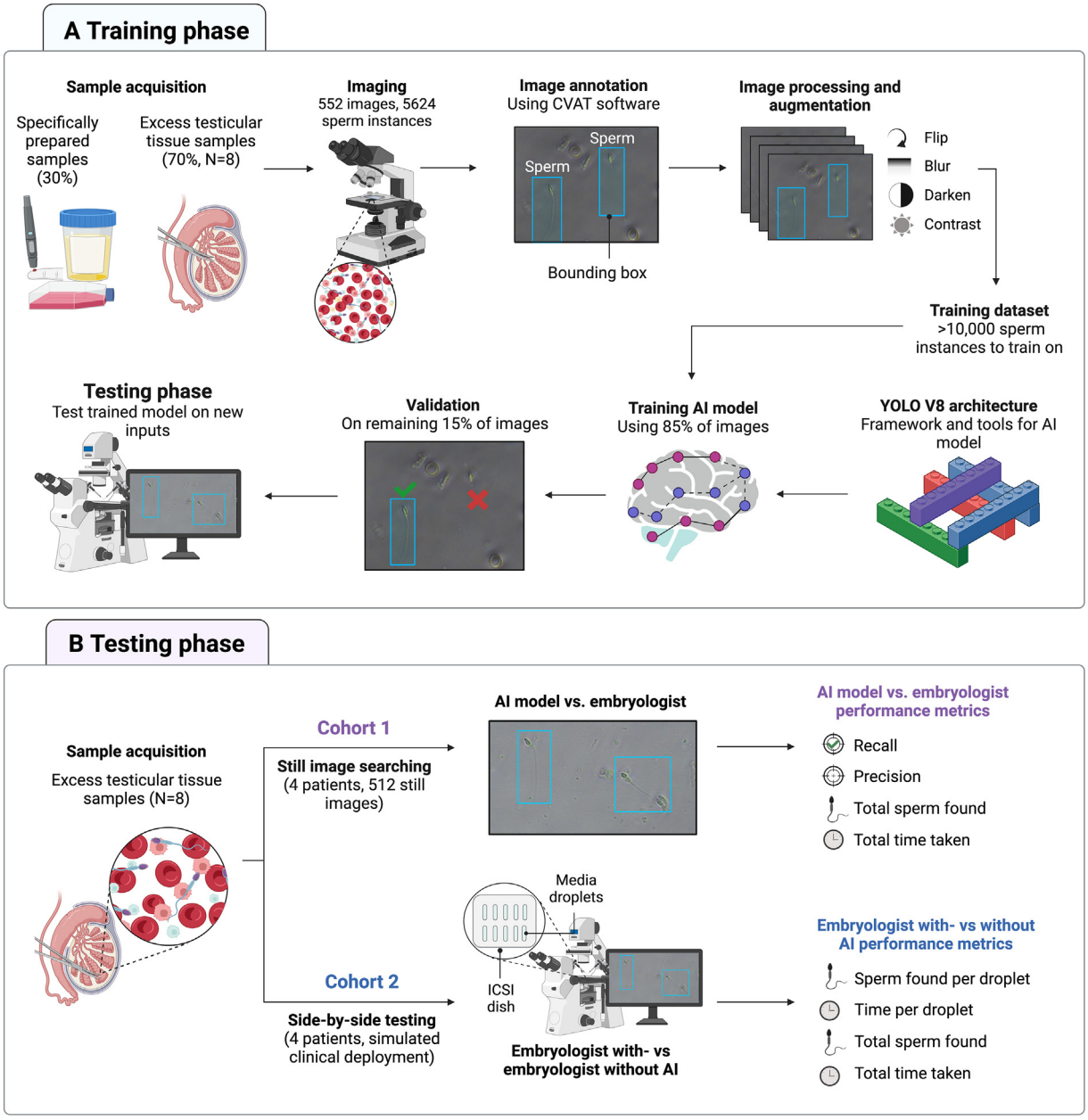

FIGURE 1 Overview of the study phases. (A) The training phase begins with sample acquisition from 30% specifically prepared samples and 70% testicular tissue samples, which are plated in dishes and imaged at a magnification of 200 ×. The images are then annotated using the Computer Vision Annotation Tool (CVAT), creating bounding boxes around all the spermatozoa in each image. Each image is processed and augmented to create a training dataset, which is then used to train a model created using the YOLOv8 architecture and tools. Then 85% of these images are used to train the model and 15% are used to validate the model performance before testing. (B) The testing phase begins with sample acquisition of excess testicular tissue from patients with non-obstructive azoospermia (NOA) and testing of the model's performance on still images (cohort 1, n = 4) versus an embryologist, and side-by-side testing of an embryologist with and without the aid of the artificial intelligence (AI) in a simulated real-world sperm search using an intracytoplasmic sperm injection (ICSI) microscope (cohort 2, n = 4).

2–15 × 10^6 cells/ml (approximated ranges for an mTESE sample), WBC (10 × 10^6 cells/ml; purchased from IQ Biosciences USA) were diluted to a concentration between 5 × 10^5 and 1 × 10^6 cells/ml, and epithelial cells were diluted to a concentration of between 7 × 10^5 and 1 × 10^6 cells/ml.

To add extra complexity, background cells from sperm donors were isolated from donors with high concentrations of background cell populations and cryopreserved until needed. These cells helped to simulate the conditions of poor-quality samples with high levels of collateral cell contamination from surgery and for infertile semen samples with high levels of contamination in the ejaculate.

Testicular biopsy retrieval and processing

Surgical sperm collection was performed in accordance with the routine workflow for each method of sperm collection

(mTESE and TESA) from azoospermic patients scheduled for surgical sperm collection for obstructive azoospermia or NOA. Surgical sperm collections were performed under general anaesthesia, and the samples were immediately placed in a sterile conical tube containing 1 ml of G-MOPS Plus (37°C) and transported to the IVF laboratory.

During mTESE, embryologists search through seminiferous tubules handed to them by the surgeon, with simultaneous further searching by the surgeon for dilated seminiferous tubules. Further samples are then sent to the IVF laboratory for a further search before being placed in 1–2 ml of G-MOPS Plus in a sterile Petri dish under a stereo-microscope to wash off excess blood from the tissue, and then moved to a new Petri dish with 300 μl of G-MOPS Plus. The tissue was gently teased apart using sterile syringes to release potential spermatozoa from the tubules into the surrounding G-MOPS Plus medium. The macerated tissue and large pieces were then removed and placed into a separate tube, and the remaining suspension was used for the sperm search and treatment. In cases whereby imaging and/or testing was not possible on the same day or the following day, the samples were fixed with 4% formalin to preserve their morphological integrity and prevent any microbial growth until use in the study.

To prepare samples for comparison between an AI-enabled sperm search and a sperm search by an embryologist in cohort 1, samples that were recorded having no spermatozoa found in clinical searches were spiked with low concentrations of spermatozoa from semen donors (prepared as described in 'Preparation of Specifically Prepared Samples'). To help create a master count of total spermatozoa in the plated samples, spiked spermatozoa were stained with propidium iodide and washed to remove excess stain before spiking. This was done to help identify the total number of spermatozoa to be found in each sample for comparison with the AI and embryologist performance groups. Samples that had spermatozoa present in the clinics were not spiked with donor semen and were preserved in their clinical state for processing.

Image acquisition and processing

To train the model, specifically prepared samples containing mixtures of spermatozoa, RBC, WBC and epithelial cells from cell culture media were prepared and plated in a similar manner to a clinical sperm search, using 10 long drops of G-MOPS Plus of 2–3 mm in length under OVOIL (Vitrolife, Sweden) in an ICSI dish (Vitrolife, Sweden), and imaged at 200 × magnification using cellSens Imaging Software (Olympus Life Science, Japan; FIGURE 1A). This approach was chosen to initiate training, and once the model's ability to identify spermatozoa was confirmed, clinically obtained testicular tissue samples from eight azoospermic patients (six with NOA and two with obstructive azoospermia) were then used to train the model with more representative backgrounds.

The training dataset comprised 540 images (152 from specifically prepared samples and 388 from testicular tissue samples), containing 5624 unique sperm instances, duplicated and augmented generating at least one augmented copy per image, which resulted in over 10,000 spermatozoa to train the identification function (FIGURE 1A). Synthetic data (duplication) during the model training was used to create more unique images for the model to learn from and is commonly performed to improve dataset fidelity (*Chavez-Badiola et al., 2020*; *Cubuk et al., 2018*; *Trembley et al., 2018*). By creating these flipped and augmented duplicate images, these images can be used in the training process as they may be considered functionally unique to their original copy (Supplementary Figure 1).

Images were annotated using the Computer Vision Annotation Tool (CVAT; Intel, USA) which is open-source software with a web-based interface designed for image and video annotation for computer vision tasks. This software was chosen to create the annotated dataset of images whereby the spermatozoa in these images were annotated with simple bounding boxes (Supplementary Figure 2) that enclose the entire visible spermatozoa including the head and tail. If the spermatozoon is partially occluded it is still bound by a single bounding box encompassing all the visible areas. CVAT was chosen for this purpose due to collaborative annotation from multiple users (including the AI model) as well as the user-friendly interface.

Dataset preparation

The training images were 2456 × 1842px JPG images with 95% compression. Images were saved in JPG format to better reflect real-world environments where images may be sent over a network and require rapid real-time feedback. These were resized to 1664 × 1664px with a black fill. A total of 85% of the images were used for training and 15% reserved for validation of the model's performance after training. Augmentations were applied to all images including duplicates from both the specifically prepared samples and the excess testicular tissue samples to inflate the dataset and make the trained model more robust to variations in microscope camera images, such as compression artefacts, changing focal length or lighting and colour variations.

A vertical flip was applied to each duplicate image, ensuring it was uniquely different from its source, and then with various probabilities a series of augmentation techniques were employed using the Python-based Albumentations library (*Buslaev et al., 2020*). Initially, a blurring effect with a kernel size of 2 × 2 pixels was applied to each image to simulate the effect of slight defocusing. Thereafter, JPG compression was implemented, adjusting the compression quality to a range between 60% and 80%, to mimic the common lossy compression artefacts (features identifiable with the human eye) found in digital imaging. An example of these augmentations is shown in Supplementary Figure 1.

Training of the AI model

Once the dataset of images had been compiled for training, an open-source machine-learning model architecture, YOLOv8 (Ultralytics, USA), was chosen, which provided the framework and tools to train the authors' own model. YOLOv8 was selected as it is a highly performant architecture for real-time object detection tasks, which suits the application of identifying spermatozoa in highly complex tissue samples. YOLOv8 was used with the 'small' size architecture configuration with 225 layers and 1,1166,560 parameters to prioritize minimal inference time (i.e. speed of identifying potential spermatozoa during searching) over potentially greater recall from more parameters (*Jocher et al., 2023*). Further image augmentations were applied by YOLOv8 during the training process, including horizontal flipping, scaling, translation and augmentations to hue, saturation and value. The training setup was restricted to a modest video random access memory of less than 8 GB, which limits the size of the model and training image resolution. Thus,

to maintain a high image resolution required to differentiate fine detail and the desired model size on this set-up, the study used a small batch size of four images being trained in parallel. The model was trained for 300 training iterations or epochs with a learning rate of 0.01. The stochastic gradient descent optimizer was used with 0.937 momentum and 0.005 weight decay.

The trained model was then used to make inferences on unseen, unlabelled images from the 15% of images allocated for validation (FIGURE 1A). The performance of the model was validated on images with a ground truth sperm number showing 85% precision and 78% accuracy after 300 epochs; the model was then considered ready for side-by-side testing against an embryologist as the training dataset is purposely compiled to validate performance on edge-cases and relatively difficult to identify spermatozoa. This approach has been proven to produce robust and unbiased image detection models (*Vabalas et al., 2019*).

Comparison of the AI model versus embryologist performance

Side-by-side testing was split into two cohorts both using immotile spermatozoa for their ability to standardize sperm spatial detection. The first cohort used fixed samples at University of Technology Sydney research laboratories and consisted of comparing the time, recall and precision of sperm detection on still images between the AI model and an embryologist (FIGURE 1B). The AI model was loaded onto a desktop computer (Intel Core i5-10600K CPU @ 4.10 GHz [6 cores] (Intel, USA), RTX 3070 graphics card Zotac Gaming, Hong Kong) and annotated spermatozoa in still images of plated discarded testicular tissue samples (n = 4 NOA patients, 512 images acquired with a total of 2660 spermatozoa to be found) in droplets at 200 × magnification. The embryologist used CVAT to annotate the location of spermatozoa independently in the same images while being timed. Sperm annotations from both the AI model and the embryologist (using the annotation software, CVAT) were then compared with a ground truth of verified sperm labels for each image to attain comparable metrics, i.e. precision, recall, time per field of view (FOV) and total spermatozoa found. Consensus for the ground truth annotation for each image was performed by two scientists independent of the embryologist used to test against the AI model.

Precision is a measure of how many sperm detections are correct, i.e. the ratio of the correctly predicted positive observations to the total number of predictions made, and recall (sensitivity or true positive rate) is a measure of how many of the spermatozoa in an FOV the model finds, i.e. the ratio of correctly predicted positive observations to the total of actual spermatozoa in the FOV. Precision and recall are defined by:

$$Precision = \frac{\sum TP}{\sum (TP + FP)}$$

$$Recall = \frac{\sum TP}{\sum (TP + FN)}$$

where TP is true positives, FP is false positives and FN is false negatives. Potential sperm detections (bounding boxes) with significant overlap (>40% intersection of union) with confirmed spermatozoa were counted as positive detections and those without as negatives. Spermatozoa bordering the edge of an image are often cut off and lack enough information to distinguish them as either positive or negative, so any potential spermatozoa within 2px of the edge of the image were omitted.

For the second cohort, to better simulate real-time clinical deployment, a side-by-side test of the AI comparing the performance of an embryologist with and without the AI was performed. Dishes were plated and prepared testicular tissue samples were added to the dishes in a similar manner to a clinical sperm search, with 10 long drops of G-MOPS Plus of 2–3 mm in length under OVOIL (Vitrolife, Sweden) in an ICSI dish (Vitrolife, Sweden) per patient sample. The embryologist recorded and compared the number of spermatozoa found per droplet for each tissue sample (n = 4 NOA patients) that they processed with (Supplementary Videos 1 and 2) and without AI, as well as the time taken to complete their assessment (FIGURE 1B). No ground truth total sperm number was acquired for each drop or dish and therefore a direct comparison of spermatozoa found per unit time in each drop of medium was compared between using the AI and not using the AI. The embryologist was blinded to the dishes and these were reordered to prevent any memory of the sperm location by the embryologist when performing each search. Confidence of sperm identification functionality was added to the AI whereby the confidence range was indicated by a green (>0.75) and orange (between 0.75 and 0.4) scale (Supplementary Video 1) or a red (>0.75) and blue (between 0.75 and 0.4) scale (Supplementary Videos 2 and 3).

Statistical analysis

All statistical analyses were performed using GraphPad Prism 9.0 (GraphPad Software). Normal distribution was assessed using the Shapiro–Wilk test. The statistical significance of the differences between the groups were tested using the Mann–Whitney U-test as the data were not normally distributed. Two-way analysis of variance was performed to assess the effects of the counting method and group. A value of $P < 0.05$ was considered statistically significant, and the means are expressed with the standard error of the mean (SEM) as a measure of the sample mean estimates.

RESULTS

In the first cohort of this study (n = 4 NOA patients), when assessing the performance of sperm identification from the still images, the AI model showed a dramatic improvement in the time taken to identify the spermatozoa in each FOV, improved recall in identifying spermatozoa and provided a high level of precision (TABLE 1). The AI was able to identify all the spermatozoa within each FOV in significantly less time compared with the trained embryologist, with durations of $0.02 \pm 0.3 \times 10^{-5}$ s versus 36.10 ± 1.18 s, respectively ($P < 0.0001$; TABLE 1). This represents an approximate 99.95% reduction in time per FOV. The AI model demonstrated a significant difference in recall compared with the trained embryologist (91.95 ± 0.81% versus 86.52 ± 1.34%, $P < 0.001$; TABLE 1). The model exhibited a precision of 89.58 ± 0.87%, considering the correct identification of spermatozoa and false positives relative to the control count (TABLE 1). In contrast, the embryologist had a precision of 98.18 ± 0.38% ($P < 0.0001$). Out of a total of 2660 spermatozoa, the embryologist identified 1937, while the AI model detected 1997 (TABLE 1).

In the second cohort of this study (n = 4 NOA patients), a simulated deployment of the AI was performed in a research laboratory whereby the AI was used as an assistive tool to guide the embryologists to identify spermatozoa on an ICSI microscope kit (see Supplementary Videos

TABLE 1 COMPARISON OF AI AND EMBRYOLOGIST SPERM SEARCH PERFORMANCE METRICS

Parameter	Embryologist	AI	*P*-value
Cohort 1 (still images)			
Time per FOV (s)	36.10 ± 1.18	$0.02 \pm 0.3 \times 10^{-5}$	<0.0001[a]
Recall (%)	86.52 ± 1.34	91.95 ± 0.81	0.0006[a]
Precision (%)	98.18 ± 0.38	89.58 ± 0.87	<0.0001[a]
No. of sperm found (from 2660)	1937	1997	N/A
Cohort 2 (side-by-side deployment)			
Time taken per drop (s)	168.7 ± 7.84	98.9 ± 3.19	<0.0001[b]
Total time taken (s)	6749.71	3955.89	N/A
Sperm found per drop	31.85 ± 3.09	34.9 ± 3.43	0.3843[b]
Total no. of sperm found	1274	1396	N/A

Data are presented as the mean ± SEM or total. Between-group differences were tested using a Mann–Whitney *U*-test[a], and variance effects between groups were assessed using two-way analysis of variance.[b]

AI, artificial intelligence; FOV, field of view; N/A, not applicable.

1 and 2). As for cohort 1, the AI-assisted embryologist outperformed the individual assessment of an embryologist across all four samples. The embryologist using the AI took significantly less time to find all the spermatozoa per droplet (98.9 ± 3.19s versus 168.7 ± 7.84s, $P < 0.0001$) and found a total of 1396 spermatozoa, while they found 1274 without the use of the AI (TABLE 1). There was no significant difference in the number of spermatozoa found per droplet for the embryologist using AI versus not using AI although a slight trend of consistently more spermatozoa found was observed (34.9 ± 3.43 versus 31.85 ± 3.09 spermatozoa, respectively).

DISCUSSION

AI image analysis can identify spermatozoa faster and with better recall than an embryologist in still images and significantly faster in a simulated sperm search scenario when integrated into an ICSI microscope. This is the first known application of machine learning AI for surgical sperm searches for the clinical treatment of azoospermia and results in a streamlining of a historically laborious process.

Machine learning is an algorithmic method of data analysis whereby a predictive model is trained to recognize patterns and associations from the input data (*Bannach-Brown et al., 2019*). Supervised machine learning models can be trained on labelled images and/or videos to understand how to predict the labels of unseen data. CNN algorithms are a type of deep-learning model that attempts, through iterative training, to transform input data into the desired output labels. There have been a considerable number of studies on the utility of machine learning and AI-based image analysis on the selection of embryos for the prediction of euploidy status, implantation potential and incidence of miscarriage (*Barnes et al., 2023*; *Diakiw et al., 2022*; *Duval et al., 2023*; *Hariharan et al., 2019*; *Tran et al., 2018*; *VerMilyea et al., 2020*). Studies have also proven the application of machine learning in the selection and assessment of spermatozoa for use in ICSI by tracking spermatozoa correlated with better quality blastocysts (*Joshi et al., 2023*; *Mendizabal-Ruiz et al., 2022*). Furthermore, studies have used images of spermatozoa that have been labelled as normal or abnormally shaped by a professional or stained for DNA integrity; given a sufficient volume and variety of these labelled images, machine learning models have been trained to label the morphology of predicted DNA fragmentation of new, unseen, images of spermatozoa (*McCallum et al., 2019*; *Wang et al., 2019*). Whereas CNN, commonly referred to as AI, have largely looked at spermatozoa in a clear environment, the current authors applied a CNN to complex, processed tissues from testicular sperm retrieval procedures and implemented it in a live video feed for the real-time identification of spermatozoa for use in ICSI.

The application of a computer vision-based machine learning model to identify spermatozoa in real time during sperm searches outperforms embryologists' manual searching in simulated searches using still images in terms of the time taken, recall and sperm count. The biggest noticeable difference is the time reduction, where image analysis is almost instant (0.02 s per FOV) but does not consider clinical tasks such as dish set-up, panning and magnification change, and the collection of identified spermatozoa using a micromanipulator needle. Recall and precision were measured as metrics of both the AI and the embryologists' performance against a ground truth number of spermatozoa per image. The significantly lower time taken to identify spermatozoa per FOV, higher recall and increase in the total number of spermatozoa found show the clear superiority of AI image analysis compared with the eyes and focus of trained embryologists (TABLE 1). Although the AI had a lower precision value than the embryologists in the first cohort, it is worth noting that this is a result of the annotation approach taken when training the AI, and precision values are particularly relevant in applications when the cost of false-positive results is high. For the application of this AI model in sperm searching, the cost of false negatives is much higher, whereby a potential spermatozoon suitable for ICSI could be missed, as opposed to an extra 2 s of an embryologist's attention potentially being wasted in the case of a false-positive result. Recall is, however, essential when the cost of false negatives is high, as is in sperm searches of samples from individuals with NOA.

In the second cohort, testicular tissue samples with supplemented spermatozoa (for better quantification of efficacy) were searched by an embryologist in plated ICSI dishes to better simulate a clinical sperm search on an ICSI kit with and without the aid of the AI (see Supplementary Videos 1 and 2). It was determined that the AI reduced the time taken to identify all the spermatozoa in the droplet by around 50% (TABLE 1), with no drop in the number of spermatozoa identified per drop and a higher total number of spermatozoa identified in total (TABLE 1).

Using an exhaustively trained image analysis model to identify spermatozoa based on tens of thousands of sperm images has clinical utility in directing an embryologist's attention to what the AI deems may be of interest and can thus drastically reduce the time taken or number of manual extended sperm searches when integrated with a micromanipulator microscope. The model

trained in this study is designed to cater for multiple clinics that may have different microscopes, light environments, filters and cameras. These environmental and equipment factors may affect the performance of the AI and have thus been catered for. The image augmentations such as blur, colour variations, focus changes, image saturation and colour balance changes and flipping of images used to train the AI model follow a common strategy in computer vision image analysis whereby these augmentations artificially replicate variant circumstances that may appear in images that were not necessarily widely represented in the training data that comes from a relative few, largely homogenous samples (*Chavez-Badiola et al., 2020*; *Cubuk et al., 2018*; *Trembley et al., 2018*). It is common for microscope images to be slightly blurry or have different lighting conditions and this is replicated in the training data through the authors' choice of augmentations, such that the model is resilient to these conditions. This is another area that with further tuning could improve model performance in the future. The model was also trained using both epididymal and testicular spermatozoa to broaden the sample dataset empowering the AI to broaden target sperm prompting. Importantly, the model can also identify spermatozoa with a broad range of motility, from immotile to hyperactivated, and adjusts and adapts to magnification change and panning in real time (Supplementary Video 3).

The role of this model is not to replace an embryologist, but to be a guide towards spermatozoa of interest, leaving the embryologist to make the final determination on the suitability of a spermatozoon for ICSI. AI can negate the biological limits of human error and observation as well as the effects of fatigue, which have long been a limiting factor to extended sperm searches of heterogenous samples obtained via surgical sperm collection. It is important to remember, however, that the AI is limited to detection within the manually directed FOV, and thus if the embryologist has overlooked an area in the sample, the AI will not be able to detect a spermatozoon without having it within view.

This study was performed solely on immotile spermatozoa for the most accurate quantification for spatially identifying and locating spermatozoa, although the AI identifies motile spermatozoa very well (see Supplementary Video 3), and a true clinical deployment will better prove the clinical utility of the model. This proof-of-concept study demonstrates the potential for AI-assisted sperm searches, both in semen for extended sperm searches and in testicular tissue. While the results of this study are promising, continuing to improve the core dataset and image variety will make the model more robust and adoptable for clinics with significantly different microscope arrangements, as well as achieving a higher level of recall.

The limitation of a simulated sperm search using an ICSI workstation with and without the use of the AI, using samples spiked with spermatozoa, is that it does not consider the time spent confirming the locations of the spermatozoa in the FOV during panning (so as not to re-count or miss spermatozoa). This is a disadvantage of the testing method and might be contributing to the lower difference in time taken per method in cohort 2. Therefore, a robust clinical deployment study has been planned, involving consenting in-treatment patients, whereby embryologists will be able to perform sperm searches with the aid of the AI model.

Furthermore, there is potential for the expansion of this AI to include motility and morphological assessments of identified spermatozoa to help in the choice of spermatozoa for insemination when the spermatozoa outnumber the number of oocytes suitable for injection. Another useful addition to the AI would be a sensitive measure of spermatozoan 'twitching' in these cases. 'Twitching' sperm movement in cases of severe NOA confirms the vitality of the spermatozoa without the need for other interventions to prove sperm vitality such as the hyperosmotic swelling test, which also reduces the time taken when selecting the spermatozoa found.

In conclusion, azoospermia affects 10% of infertile men, with NOA, the most severe form, constituting 60% of these cases (*Verheyen et al., 2017*). The current approaches to recover spermatozoa from men who undergo surgery from this condition are antiquated and potentially detrimental to the quality of the spermatozoa found. This study has successfully demonstrated a proof-of-concept application of an AI image analysis model to drastically reduce the sperm search time in testicular tissue samples in simulated clinical sperm searches. When applying the AI to a simulated real-time search workflow, a 50% reduction in time taken to identify the spermatozoa has been demonstrated. This presents the potential to avoid or at least reduce the negative effect of the extended exposure of spermatozoa to biopsied testicular tissue containing a host of molecules capable of reducing sperm viability. By applying this approach with further development and ergonomic optimization, the authors believe it could result in a standardized and more efficient workflow, greatly improving the current processing procedure of all surgically retrieved samples and azoospermic ejaculates by increasing access to treatment for azoospermia and reducing staff time required, as well as increasing sample coverage to ultimately increase chances of finding spermatozoa.

DATA AVAILABILITY

The data that has been used is confidential.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support and facilitation provided by IVFAustralia Eastern Suburbs doctors Dr Jeffrey Persson and Dr Shadi Khashaba, in providing access to patients undergoing treatment for severe male factor infertility, and the technical support and input from the embryology team. M.E.W. would like to acknowledge the support of the Cancer Institute New South Wales through the Career Development Fellowship (2021/CDF1148).

AUTHOR CONTRIBUTIONS

D.M.G., S.A.V., P.A.V., S.C., S.H.K.K., D.K.G. and M.E.W. designed and conceptualized the study. D.M.G., S.A.V. and P.A.V. were responsible for the data acquisition. P.V. and S.A.V. designed and trained the AI model. D.M.G., S.H.K.K. and S.A.V. facilitated clinical sample acquisition. D.M.G., S.A.V. and P.A.V. drafted the manuscript and all the authors critically revised and finally approved it. All authors have agreed to be accountable for the academic integrity and accuracy of the research.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.rbmo.2024.103910.

REFERENCES

Agarwal, A., Mulgund, A., Hamada, A., Chyatte, M.R., 2015. A unique view on male infertility around the globe. Reproductive biology and endocrinology 13, 1–9.

Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A.S., Ananiadou, S., Liao, J., Macleod, M.R., 2019. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Systematic reviews 8, 1–12.

Barnes, J., Brendel, M., Gao, V.R., Rajendran, S., Kim, J., Li, Q., Malmsten, J.E., Sierra, J.T., Zisimopoulos, P., Sigaras, A., 2023. A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: A retrospective model development and validation study. The Lancet Digital Health 5, e28–e40.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: fast and flexible image augmentations. Information 11, 125.

Chavez-Badiola, A., Flores-Saiffe-Farías, A., Mendizabal-Ruiz, G., Drakeley, A.J., Cohen, J., 2020. Embryo Ranking Intelligent Classification Algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. Reproductive BioMedicine Online 41, 585–593.

Cubuk E.D., Zoph B., Mane D., Vasudevan V., Le Q.V., 2018. Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501.

Deruyver, Y., Vanderschueren, D., Van der Aa, F., 2014. Outcome of microdissection TESE compared with conventional TESE in non-obstructive azoospermia: a systematic review. Androl 2, 20–24.

Diakiw, S., Hall, J., VerMilyea, M., Amin, J., Aizpurua, J., Giardini, L., Briones, Y., Lim, A., Dakka, M., Nguyen, T., 2022. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. Human Reproduction 37, 1746–1759.

Duval, A., Nogueira, D., Dissler, N., Maskani Filali, M., Delestro Matos, F., Chansel-Debordeaux, L., Ferrer-Buitrago, M., Ferrer, E., Antequera, V., Ruiz-Jorro, M., 2023. A hybrid artificial intelligence model leverages multi-centric clinical data to improve fetal heart rate pregnancy prediction across time-lapse systems. Human Reproduction 38, 596–608.

Flannigan, R., Bach, P.V., Schlegel, P.N., 2017. Microdissection testicular sperm extraction. Translational Androl. and Urology. 6, 745.

Goss, D., Vasilescu, S., Vasilescu, P., Sacks, G., Gardner, D., Warkiani, M., 2023. O-136 Artificial intelligence to assist in surgical sperm detection and isolation. Hum. Reproduction. 38, dead093. 163.

Hariharan, R., He, P., Meseguer, M., Toschi, M., Rocha, J.C., Zaninovic, N., Malmsten, J., Zhan, Q., Hickman, C., 2019. Artificial intelligence assessment of time-lapse images can predict with 77% accuracy whether a human embryo capable of achieving a pregnancy will miscarry. Fertility and Steril. 112, e38–e39.

Jarow, J.P., Espeland, M.A., Lipshultz, L.I., 1989. Evaluation of the azoospermic patient. The J. of Urology. 142, 62–65.

Jocher G., Chaurasia A., Qiu J. YOLO by Ultralytics. 2023. Ultralytics, GitHub.

Joshi, K., Simbulan, R.K., Rajah, A.M., Burd, G., Gupta, S., Behr, B., Guarnaccia, M., Singh, G., 2023. A proof-of-concept prospective study of applying artificial intelligence for sperm selection in the IVF laboratory. Reproductive BioMedicine Online, 103329.

Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Jolles, M., Pinotti, R., Swan, S.H., 2023. Temporal trends in sperm count: a systematic review and meta-regression analysis of samples collected globally in the 20th and 21st centuries. Hum. Reproduction Update 29, 157–176.

Mangum, C.L., Patel, D.P., Jafek, A.R., Samuel, R., Jenkins, T.G., Aston, K.I., Gale, B.K., Hotaling, J.M., 2020. Towards a better testicular sperm extraction: novel sperm sorting technologies for non-motile sperm extracted by microdissection TESE. Translational Androl. and Urology. 9, S206.

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, 250.

Mendizabal-Ruiz, G., Chavez-Badiola, A., Figueroa, I.A., Nuño, V.M., Farias, A.F-S., Valencia-Murilloa, R., Drakeley, A., Garcia-Sandoval, J.P., Cohen, J., 2022. Computer software (SiD) assisted real-time single sperm selection associated with fertilization and blastocyst formation. Reproductive BioMedicine Online 45, 703–711.

Ouitrakul, S., Sukprasert, M., Treetampinich, C., Choktanasiri, W., Vallibhakara, S.A-O., Satirapod, C., 2018. The Effect of Different Timing after Ejaculation on Sperm Motility and Viability in Semen Analysis at Room Temperature. J. of the Méd Association of Thail. 101.

Ramasamy, R., Reifsnyder, J.E., Bryson, C., Zaninovic, N., Liotta, D., Cook, C-A., Hariprashad, J., Weiss, D., Neri, Q., Palermo, G.D., 2011. Role of tissue digestion and extensive sperm search after microdissection testicular sperm extraction. Fertil. and Steril. 96, 299–302.

Ramasamy, R., Yagan, N., Schlegel, P.N., 2005. Structural and functional changes to the testis after conventional versus microdissection testicular sperm extraction. Urology 65, 1190–1194.

Samuel, R., Badamjav, O., Murphy, K.E., Patel, D.P., Son, J., Gale, B.K., Carrell, D.T., Hotaling, J.M., 2016. Microfluidics: The future of microdissection TESE? Systems Biology in Reproductive Medicine 62, 161–170.

Schiff, J.D., Palermo, G.D., Veeck, L.L., Goldstein, M., Rosenwaks, Z., Schlegel, P.N., 2005. Success of testicular sperm injection and intracytoplasmic sperm injection in men with Klinefelter syndrome. The J. of Clinical Endocrinology & Metabolism. 90, 6263–6267.

Schrepferman, C.G., Carson, M.R., Sparks, A.E., Sandlow, J.I., 2001. Need for sperm retrieval and cryopreservation at vasectomy reversal. The J. of Urology 166, 1787–1789.

Tran, A., Cooke, S., Illingworth, P., Gardner, D., 2018. Artificial intelligence as a novel approach for embryo selection. Fertil. and Steril. 110, e430.

Tremblay J., Prakash A., Acuna D., Brophy M., Jampani V., Anil C., To T., Cameracci E., Boochoon S., Birchfield S. Training deep networks with synthetic data: Bridging the reality gap by domain randomization Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018, pp. 969-977.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PloS one 14, e0224365.

Verheyen, G., Popovic-Todorovic, B., Tournaye, H., 2017. Processing and selection of surgically-retrieved sperm for ICSI: a review. Basic and Clinical Androl. 27, 1–10.

VerMilyea, M., Hall, J., Diakiw, S., Johnston, A., Nguyen, T., Perugini, D., Miller, A., Picou, A., Murphy, A., Perugini, M., 2020. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. Hum. Reproduction. 35, 770–784.

Wang, Y., Riordon, J., Kong, T., Xu, Y., Nguyen, B., Zhong, J., You, J.B., Lagunov, A., Hannam, T.G., Jarvi, K., 2019. Prediction of DNA integrity from morphological parameters using a single-sperm DNA fragmentation index assay. Advanced Science 6, 1900712.

WHO, 2021. Laboratory manual for the examination and processing of human semen, Sixth edn World Health Organization.

Wosnitzer, M., Goldstein, M., Hardy, M.P., 2014. Review of azoospermia. Spermatogenesis. 4, e28218.

Received 7 September 2023; received in revised form 31 January 2024; accepted 9 February 2024.