



Open Access

ORIGINAL ARTICLE

Male Health

# A preliminary study of sperm identification in microdissection testicular sperm extraction samples with deep convolutional neural networks

Daniel J Wu<sup>1</sup>, Odgerel Badamjav<sup>2</sup>, Vikrant V Reddy<sup>3</sup>, Michael Eisenberg<sup>4</sup>, Barry Behr<sup>3</sup>

Sperm identification and selection is an essential task when processing human testicular samples for *in vitro* fertilization. Locating and identifying sperm cell(s) in human testicular biopsy samples is labor intensive and time consuming. We developed a new computer-aided sperm analysis (CASA) system, which utilizes deep learning for near human-level performance on testicular sperm extraction (TESE), trained on a custom dataset. The system automates the identification of sperm in testicular biopsy samples. A dataset of 702 de-identified images from testicular biopsy samples of 30 patients was collected. Each image was normalized and passed through glare filters and diffraction correction. The data were split 80%, 10%, and 10% into training, validation, and test sets, respectively. Then, a deep object detection network, composed of a feature extraction network and object detection network, was trained on this dataset. The model was benchmarked against embryologists' performance on the detection task. Our deep learning CASA system achieved a mean average precision (mAP) of 0.741, with an average recall (AR) of 0.376 on our dataset. Our proposed method can work in real time; its speed is effectively limited only by the imaging speed of the microscope. Our results indicate that deep learning-based technologies can improve the efficiency of finding sperm in testicular biopsy samples.

*Asian Journal of Andrology* (2021) 23, 135–139; doi: 10.4103/aja.aja\_66\_20; published online: 23 October 2020

**Keywords:** artificial intelligence; computer vision; male infertility; microdissection testicular sperm extraction; sperm

## INTRODUCTION

Approximately 1% of all men and 10% of infertile men have azoospermia, the absence of sperm in the ejaculate.<sup>1</sup> Of all azoospermic men, over half have nonobstructive azoospermia (NOA) due to spermatogenic failure. NOA is defined as the absence of spermatozoa in semen because of minimal or no sperm production in the testis.<sup>2</sup> In men with NOA, microdissection testicular sperm extraction (micro-TESE), a meticulous microsurgical exploration of the testicular parenchyma to search for seminiferous tubules housing intact spermatogenesis, is routinely performed.<sup>3–5</sup>

As part of this procedure, most centers perform intra- and postoperative microscopic assessment of the extracted testicular tissue. This manual process is highly labor intensive, time consuming, and shows significant variability;<sup>6</sup> however, this remains the gold standard for identification and collection of potential sperm recovered from NOA patients.<sup>7</sup>

There are no existing computer-aided sperm analysis (CASA) systems for analysis of testicular biopsies, the source of sperm for patients with azoospermia. Clinics treating patients with severe infertility primarily rely on manual image analysis for sperm identification in testicular tissue.<sup>7</sup> This process requires careful processing in the laboratory in order to separate and identify sperm within the biopsied tissue. Not only is manual microscopic testicular

specimen examination time consuming and tedious, but also it is highly dependent on skill of the personnel. Sperm can easily be overlooked due to several variables, including incomplete cell dissociation, elevated levels of other cell types, inexperience, exhaustion, and human error.<sup>8</sup> For patients who produce a small number of sperm to begin with, failing to identify even a couple of sperm could mean the difference between a successful or unsuccessful testicular sperm extraction.

The field of computer vision technologies powered by deep learning methods offers exciting opportunities, especially when considering applications in low-cell number or single-cell analysis. With the increasing efficiency and robustness of deep learning methods for computer vision tasks,<sup>9</sup> we are optimistic that a CASA system utilizing deep neural networks can achieve robust performance on tasks associated with testicular biopsy analysis. However, careful sample preparation and processing is necessary for building robust CASA systems.

Deep neural networks have been used in a variety of domains to learn the underlying relationships between raw data and complex, higher-order tasks. Traditional methods often require researchers to specify an explicit methodology for solving a task – often a difficult, if not impossible, requirement. Deep neural networks, however, only require properly labeled datasets, and can learn how to solve the task from the data. Training deep neural networks involves defining a loss

<sup>1</sup>Department of Computer Science, Stanford University, Stanford, CA 94305, USA; <sup>2</sup>Department of Urology, University of Utah Health, Salt Lake City, UT 84108, USA; <sup>3</sup>Department of Obstetrics and Gynecology, Stanford Children's Health, Stanford, CA 94305, USA; <sup>4</sup>Department of Urology, Stanford University School of Medicine, Stanford, CA 94305, USA.

Correspondence: DJ Wu (danjwu@stanford.edu) or O Badamjav (odgerel.badamjav@hsc.utah.edu)

Received: 24 February 2020; Accepted: 03 August 2020

function – a heuristic measure of how well the network is accomplishing its task – and then using nonconvex optimization techniques to adjust the parameters of the network in order to minimize that loss function and thereby increase the performance of the neural network.

One task well suited for deep neural networks is object recognition – the parsing of an image, represented by the raw RGB values of each pixel, in order to determine the presence and location of objects in that image. Here, we translate basic research on deep learning for object recognition into an applied CASA tool for the automated identification of spermatozoa.

## MATERIALS AND METHODS

### Dataset

Deep learning methods require a large amount of training data. To the best of our knowledge, there are no existing large-scale image datasets of testicular biopsies. After institutional review board (IRB) approval at Stanford University, Stanford, CA, USA (Approval No. 41652), we collected a novel dataset of 702 de-identified images from testicular biopsy samples of 30 patients. Consent from all participants was received at the time of semen collection (Figure 1).

### Sample processing

The samples we obtained for this study were limited to the residual fraction of the prepared TESE samples for cryopreservation following micro-TESE procedures. The images were collected from these samples immediately after the procedure(s). In order to effectively obtain this fraction, we added 1.0 ml of Quinn's Advantage Medium with HEPES (SAGE, Trumbull, CT, USA) with 20% Human Serum Albumin (SAGE), analyzed under the inverted microscope (IX73; Olympus, Tokyo, Japan) and captured images of sperm found in these samples. Hence, these TESE samples were more diluted and less complex compared to original TESE samples.

### Characteristics

The data are all composite images of testicular biopsies from 30 distinct patients. These images vary in sperm phenotype, cellular clutter, tissue superstructure, imaging modality, size, and resolution. The dataset is comprised of two types of images: dense testicular tissue, collected from an inverted microscope (IX73; Olympus), at  $1680 \times 1050$  resolution, or diffuse tissue, collected at  $640 \times 400$  resolution.

### Data preprocessing

At the time of collection, each image was normalized, passed through glare filters and diffraction correction (Hamilton Thorne, HT video and image capture software version 3, Hamilton Thorne, Inc., Beverly, MA, USA), and had microscopy artifacts removed.

Images were then anonymized and aggregated. A single

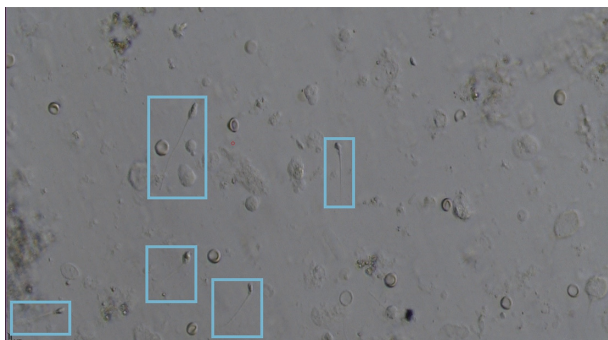


Figure 1: Sample testicular biopsy image with embryologist-labeled bounding boxes.

embryologist annotated each image with bounding boxes around each identified spermatozoon. The data were split 80%, 10%, and 10% into training, validation, and test sets, respectively, and kept separate.

The dataset, with corresponding annotations, was parsed into the same format as Microsoft's Common Objects in Context (COCO) image dataset, a standard convention for object detection datasets. This raw dataset was then augmented at training time by the methods described below.

### Data augmentation

In order to increase the adequate size of our dataset, we applied various data augmentation techniques. Images were normalized (the RGB values of images were scaled to be between 0 and 1), randomly flipped horizontally, randomly cropped, padded, and jittered. Applying these augmentation techniques has been shown to increase the robustness of deep neural networks.<sup>10</sup> Images were then linearly resampled to match the input size of  $640 \times 640$ .

### Hard example mining

Our labelers were instructed to draw bounding boxes around all the spermatozoa they identified in each image. This means that our dataset only contains bounding boxes of "positive" examples and does not contain "negative" examples, *i.e.*, bounding boxes for the "background" class. However, the neural network must also learn to recognize which parts of the image do not contain sperm, *i.e.*, correspond to the "background" class. As such, we dynamically generate "negative" bounding boxes in each image to train the model on.

In more technical terms, a bounding box is considered a negative example if it has less than a 0.5 intersection-over-union ratio (IOU, defined in our Metrics section, below) with any labeled spermatozoa in an image. These boxes are dynamically selected during training. As there are significantly more possible boxes containing "background" rather than sperm, we decided to limit the number of negative boxes to three per positive box. In each image with  $k$  spermatozoa, we select the  $3k$  negative examples with the largest loss (*i.e.*, the images the model made the greatest error on) in that image. These negative examples are temporarily added to the dataset for that iteration of model training. This procedure is known as hard example mining and is a method by which we pick the most informative negative examples for training the network while preventing a class imbalance between negative and positive examples.

### Model

Sperm identification is an object detection task – object detection models take in an image and annotate it with bounding boxes. Each bounding box has an associated label (spermatozoon or background) and a confidence rating.

This deep object detection network is composed of two parts: a feature extraction network, which takes in raw images, and outputs feature maps, and an object detection network, which interprets those feature maps, and outputs predictions for the location and presence of items.

In this work, MobileNetV2<sup>11</sup> was used as a feature extraction network, and single-shot detector (SSD)<sup>12</sup> was used as an object detection network. While we present summaries of these network architectures below, we encourage readers to see the original papers for more information.

Medical images require specialized expertise to annotate, and, as a result, medical image datasets are often much smaller than general image datasets. Transfer learning offers an effective way to mitigate the challenges of small training sets. Related work done by

the medical imaging group at the National Institutes of Health (NIH) explored transfer learning of CifarNet, AlexNet, VGGNet, Overfeat, and GoogLeNet for image classification on lymph node detection and interstitial lung disease diagnosis<sup>13</sup> and found that models pretrained on ImageNet, an extensive database of miscellaneous images,<sup>14</sup> outperformed models trained from scratch, and achieved expert-level performance. Thus, we pretrained our feature extractor, MobileNetV2, on ImageNet.

### Architectures

We use two architectures in our CASA system – MobileNetV2 and SSD.

The original MobileNet is a feature extraction network optimized for fast predictions on mobile devices.<sup>11</sup> MobileNetV2 is an updated version composed of a fully convolutional layer with thirty-two  $3 \times 3$  filters, and then 19 residual bottleneck layers, taking advantage of depth-wise convolutions for speed, and linear bottlenecks for memory efficiency. This results in a lightweight, small, and fast network while maintaining an acceptable standard of accuracy. By selecting this network for our feature extraction network, we optimize for speed over accuracy.

The SSD is an object detection network which makes the use of the convolutional layers of VGG16, appending additional convolutional layers, and extracting feature maps at each layer for prediction. These feature maps are a variety of sizes, which allow for predictions at a variety of scales. This utilizes the Multibox algorithm,<sup>12</sup> which generates 8732 default box proposals, at predetermined distributed locations and sizes. The model then classifies each proposal as either a spermatozoon or background and regresses the bounding box shape to match identified component objects.

### Loss

The loss function is a numerical measure of the performance of the model and is minimized through backpropagation. SSD uses a loss function composed of the sum of a regularization loss and a weighted average of a classification loss and a localization loss (Equation 1, **Supplementary Information**). The classification loss represents how accurate the network is at identifying a given object in an image as belonging to the correct class (*e.g.*, sperm or background) and is given by a softmax loss between the predicted and actual classes for each bounding box (Equation 2, **Supplementary Information**). The localization loss represents how precise the network is determining the coordinates of the bounding box and is given by the sum of smoothed distances between the center XY coordinates, width, and height of each predicted bounding box with the corresponding coordinates of the actual bounding box (Equation 3, **Supplementary Information**). We also add a regularization loss  $L_r$  which is given by the square of the L2 norm of the weights in the network, multiplied by a regularization strength hyperparameter (Equation 4, **Supplementary Information**). This regularization loss serves the pragmatic function of penalizing overly large values for model parameters and mitigates model overfitting.

### Training and tuning

The model was trained on our dataset on Google Cloud in 8 h on an NVIDIA K80 GPU, using minibatch gradient descent. The RMSProp optimizer, which accumulates a moving average for the first-order momentum across batches, was used to update our weights (Equation 5, **Supplementary Information**). This smoothens training updates compared to standard stochastic gradient descent optimization while preventing slow training.

We then used model performance on the validation data to tune the hyperparameters of the model. The main goal of the model was

to automate manual sperm identification. Thus, the model has to make rapid predictions on quickly changing microscopy images. Furthermore, for patients with severe male factor infertility where no sperm is found in the semen, it must err on the side of high sensitivity.

### Nonmaximal suppression

Due to the design of SSD, the output of the model on any given image is 8732 predictions, the vast majority of which correspond to the “background” class. Because these boxes have significant overlap with one another, we applied nonmaximal suppression,<sup>12</sup> a technique which keeps only the most confident prediction of each set of overlapping boxes, to declutter the resulting image.

### Metrics

Following the standard of Microsoft’s Common Objects in Context (COCO) challenge, the following metrics were used to assess model performance:

The IOU, also known as the Jaccard Index (**Figure 2**), is a measure of how well a given bounding box aligns with another bounding box. It is equal to the percentage of the overlapping area of the two boxes. The IOU was used to determine if a given model prediction was correct; any bounding box with an IOU of greater than some set threshold, generally 0.5 (*i.e.*, an overlap of 50%), is considered a correct bounding box.

Mean average precision (mAP) is a measure of the precision-recall trade-off of a model. We used the COCO mAP metric, which averages the precision over 101 evenly distributed recall values, from 0 to 1, to calculate the mAP.

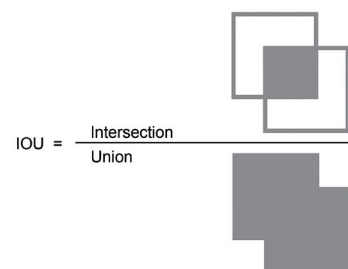
Average recall (AR) is a measure of the sensitivity of a model at a given number of detections per image. In this paper, we used 100 detections per image, *i.e.*, after nonmaximal suppression, only the 100 most confident detections out of all 8732 detections were assessed and all other detections were removed. We used the COCO AR metric, which is defined as the maximum recall given this fixed number of detections per image, averaged over categories and IOUs. In keeping with the COCO standard, 10 IOU thresholds were used (50%, 55%, 60%, 65%, ..., 95%).<sup>15</sup>

The F1 score is an overall measure of the accuracy of our classifier and is defined as the harmonic mean of precision (mAP) and recall (AR).

## RESULTS

### Benchmarks

While our training dataset was labeled by a single embryologist, we wanted to compare the performance of our model against human error rates. Thus, three embryologists independently labeled, with bounding boxes, our test dataset of 110 images, containing 111 sperm, with the VIA Image Annotator.<sup>16</sup> These annotations were then parsed to form a



**Figure 2:** The visual calculation of the intersection-over-union ratio (IOU).



ground-truth set of labels, where, if in a given image, at least two out of three embryologists drew bounding boxes with at least 0.5 IOU, then we took the ground truth to be a bounding box with the average coordinates of the embryologists' labels (**Figure 3**). We then calculated IOUs from each embryologist's prediction for the corresponding bounding box in our ground truth and then calculated mAP and AR for human performance.

### Model assessment

We calculated mAP and AR on our test set, with our tuned hyperparameters. We also report the F1 score, which is the harmonic mean of the precision and recall. The performance of our model is given in **Table 1**.

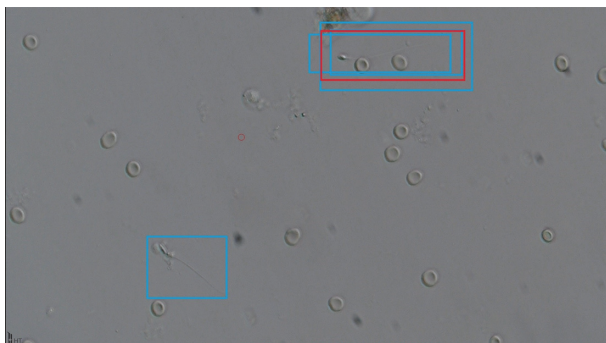
### Qualitative results and limitations

The sample output from our model is shown in **Figure 4**. The

**Table 1: Comparison of performance of architecture on test set**

Feature extractor	Object detector	mAP	AR	F1 Score
Human eye (embryologists)	Occipital cortex	0.925 <sup>a</sup>	0.642	0.758
MobileNet V2 (CNN)	SSD	0.741	0.376	0.499

Reported mAP is at 0.5 IOU, and AR is at 100 detections per image. <sup>a</sup>We report the raw precision here. Under the assumption that humans predict with 100% confidence, raw precision is equivalent to mAP. mAP: mean average precision; AR: average recall; SSD: single-shot detector; IOU: intersection-over-union ratio



**Figure 3:** A (faux) visual example of the methodology behind how the ground-truth labels were built. Top: if at least 2 embryologists labeled the same area (blue), then a ground-truth bounding box is created from the average of embryologist labels (red). Bottom: if only 1 embryologist labeled an area, it is not added to the ground-truth labels.

model has a difficult time identifying distorted sperm, particularly those with bent tails, deformities, or those occluded by microscopy artifacts. Detection rates and localization performance on sperm with normal and subnormal morphology, our primary use case, align with embryologist-level performance.

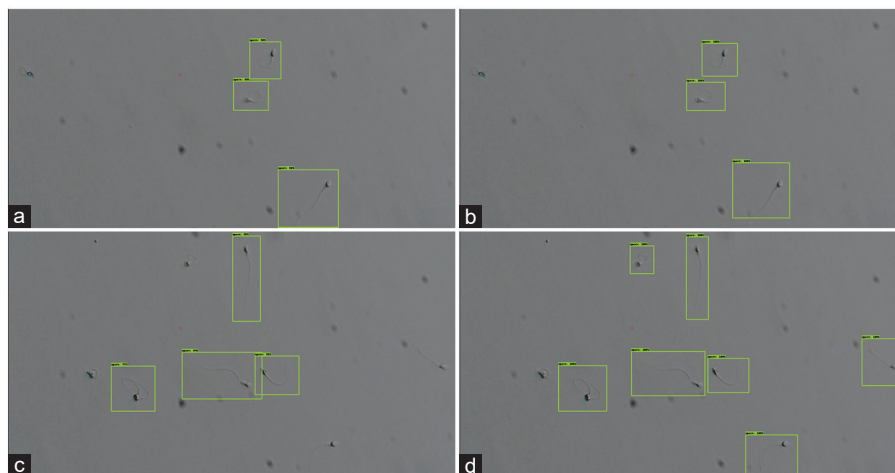
Furthermore, the determinations were made on random static images. The accuracy of the deep neural network is lower than that of the embryologists; however, the model has the potential to be significantly more time efficient. On our GPU hardware, the model makes a prediction on all 702 images in the dataset in approximately 25 s, whereas it took three embryologists total of approximately 129 600 s (approximately 36 h) in comparison to manually identify and label all 702 images. In a clinical setting, without the need to label the images, it would take embryologists several (approximately 2–3) hours to visually identify sperm in these images, several orders of magnitude longer than the speed of the model.

### DISCUSSION

Object detection via deep convolutional neural networks is an active and robust field of research. Robust models for object detection have been trained for a variety of image modalities – magnetic resonance imaging (MRI), computed tomography (CT), X-ray, and mammography – for a variety of tasks – detection, diagnosis, and assistive labeling.<sup>17</sup>

Computer vision tasks appear in a variety of medical imaging domains. Since medical images often contain nuanced and high-level features in localizable parts of the image, and the classification (assessment, diagnosis, and monitoring) task is often abstract, deep convolutional neural networks are the method of choice.<sup>18–21</sup>

As such, the confluence of deep convolutional neural networks and existing object detection architectures are well suited for sperm identification and morphological classification. Currently, TESE success rates vary between centers.<sup>8</sup> While patient demographics play a role, we know that the laboratory also plays a significant role in determining the success of a sperm extraction.<sup>22</sup> Our results indicate that deep learning-based technologies can improve the efficiency of finding sperm in testicular biopsy samples. Deep learning algorithms can help improve the efficiency and accuracy of sperm identification by reducing the tedium of the procedure. However, for cell identification deep learning-based vision technologies to be successful for micro-TESE, these systems must possess some specific characteristics:



**Figure 4:** Example predictions from SSD MobileNetV2 on two spermatozoon-rich images. (a) Model detections on sample 1. (b) Embryologist labels on sample 1. (c) Model detections on sample 2. (d) Embryologist labels on sample 2. SSD: single-shot detector.

1. Since the number of sperm present in the sample is typically low, the system should err on the side of false positives
2. The system should be able to account for variation in the samples in terms of the constituents of the sample
3. The final clinical utilization of the system must be simple, cost-effective, and require low maintenance.

For this study, we trained a deep learning architecture on a novel dataset of testicular biopsy images, and achieved a mAP at IOU 0.5 of 0.741, with an AR at 100 detections per image of 0.376. This is likely because deformed sperm are not well represented in our dataset and exist as a small subclass within the sperm class and thereby suffer from the effects of class imbalance. Given that MobileNetV2 is optimized for speed rather than accuracy, this is a robust result. For example, on the MS COCO dataset, a dataset used to benchmark object detection models, MobileNetV2 achieved a mAP of only 0.22.<sup>23</sup> Our model achieves subhuman performance but at extremely high throughput, thus we suspect our model will be a useful “prefiltering” step to assist embryologists in quickly finding areas of interest in testicular biopsy samples.

We emphasize that our aim is to create a practical tool for real-time sperm identification. A properly functioning network might not be accurate enough to outperform an embryologist on the number of spermatozoa found in each microscopy image but must be fast and accurate enough to outperform an embryologist in the amount of time it takes to find spermatozoa in a sample. We expect that future work will focus on locating and identifying sperm in TESE samples in tandem with embryologists.

We note that MobileNetV2 is optimized for fast inference; on a Google Pixel I smartphone, the model processes an image every 75 ms,<sup>11</sup> and on dedicated hardware, the model can make much faster predictions – we speculate that the practical limitation will only be based on the imaging speed of the microscope. These deep learning architectures can be applied to the micro-TESE procedure without a great deal of difficulty by integrating with existing imaging software. Using automated technologies, the potential exists to minimize the need for skilled tissue processing personnel, while at the same time increasing the efficiency and consistency of the process, leading to increased sperm recovery rates.

After initial testing of the model, it will be deployed in an academic *in vitro* fertilization center as a research tool for further testing. We intend to deploy the model as a real-time video classification pipeline, which automatically identifies sperm in testicular tissue. Our novel tool will improve the efficiency of searching for sperm in testicular biopsy samples, which is currently a labor-intensive process dependent on the skill of the embryologist.

#### AUTHOR CONTRIBUTIONS

DJW preprocessed the data, trained the model, analyzed results, and prepared the manuscript. OB collected and prepared the dataset, analyzed results, and prepared the manuscript for publication. VR assisted in labeling the dataset and read and reviewed the manuscript for publication. ME, who is the senior operating surgeon for all cases described in the paper, read and reviewed the manuscript. BB read and reviewed the manuscript. All authors read and approved the final manuscript.

#### COMPETING INTERESTS

ME declares potential personal conflicts of interest in the companies Sandstone Diagnostics, Dadi, and Underdog which may hold an intellectual property license related to the technology described in this paper.

#### ACKNOWLEDGMENTS

The authors would like to thank the subjects that took part and the embryologists at Stanford Reproductive Clinic who participated in the collection and annotation of images.

Supplementary Information is linked to the online version of the paper on the *Asian Journal of Andrology* website.

#### REFERENCES

- 1 Gudeloglu A, Parekattil SJ. Update in the evaluation of the azoospermic male. *Clinics* 2013; 68: 27–34.
- 2 Devroey P, Liu J, Nagy Z, Goossens A, Tournaye H, *et al*. Pregnancies after testicular sperm extraction and intracytoplasmic sperm injection in non-obstructive azoospermia. *Hum Reprod* 1995; 10: 1457–60.
- 3 Schlegel PN. Testicular sperm extraction: microdissection improves sperm yield with minimal tissue excision. *Hum Reprod* 1999; 14: 131–5.
- 4 Corona G, Minhas S, Giwercman A, Bettocchi C, Dinkelman-Smit M, *et al*. Sperm recovery and ICSI outcomes in men with non-obstructive azoospermia: a systematic review and meta-analysis. *Hum Reprod Update* 2019; 25: 733–57.
- 5 Donoso P, Tournaye H, Devroey P. Which is the best sperm retrieval technique for non-obstructive azoospermia? A systematic review. *Hum Reprod Update* 2007; 13: 539–49.
- 6 Popal W, Nagy ZP. Laboratory processing and intracytoplasmic sperm injection using epididymal and testicular spermatozoa: what can be done to improve outcomes? *Clinics* 2013; 68: 125–30.
- 7 Auger J, Eustache F, Ducot B, Blandin T, Daudin M, *et al*. Intra- and inter-individual variability in human sperm concentration, motility and vitality assessment during a workshop involving ten laboratories. *Hum Reprod* 2000; 15: 2360–8.
- 8 Anderson RE, Hotaling JM. Inherent difficulties of meta-analysis for surgical techniques in male infertility: an argument for standardizing reporting and outcomes. *Fertil Steril* 2015; 104: 1127–8.
- 9 Agarwal S, Terrail JO, Jurie F. Recent advances in object detection in the age of deep convolutional neural networks. *arXiv preprint* 2018; arXiv: 1809.03193.
- 10 Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. Deep neural networks-based recognition of plant diseases by leaf image classification. *Comput Intell Neurosci* 2016; 2016: 3289801.
- 11 Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p4510–20.
- 12 Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, *et al*. SSD: Single Shot Multibox Detector. In European Conference on Computer Vision; 2016. p21–37.
- 13 Shin HC, Roth HR, Gao M, Lu L, Xu Z, *et al*. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016; 35: 1285–98.
- 14 Deng J, Dong W, Socher R, Li LJ, Li K, *et al*. Imagenet: A Large-Scale Hierarchical Image Database. In IEEE Conference on Computer Vision and Pattern Recognition; 2009. p248–55.
- 15 Lin TY, Maire M, Belongie S, Hays J, Perona P, *et al*. Microsoft COCO: Common Objects in Context. In European Conference on Computer Vision; 2014. p740–55.
- 16 Dutta A, Zisserman A. The VIA annotation software for images, audio and video. *arXiv preprint* 2019; arXiv: 1904.10699.
- 17 Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 2016; 35: 1153–9.
- 18 Brosch T, Tang LY, Yoo Y, Li DK, Traboulssee A, *et al*. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans Med Imaging* 2016; 35: 1229–39.
- 19 Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016; 35: 1207–16.
- 20 Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging* 2016; 35: 1240–51.
- 21 Dou Q, Chen H, Yu L, Zhao L, Qin J, *et al*. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans Med Imaging* 2016; 35: 1182–95.
- 22 Go KJ. ‘By the work, one knows the workman’: the practice and profession of the embryologist and its translation to quality in the embryology laboratory. *Reprod Biomed Online* 2015; 31: 449–58.
- 23 Huang J, Rathod V, Sun C, Zhu M, Korattikara A, *et al*. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p7310–1.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

©The Author(s)(2020)

