

Simultaneous Morphology, Motility, and Fragmentation Analysis of Live Individual Sperm Cells for Male Fertility Evaluation

Keren Ben-Yehuda, Simcha K. Mirsky, Mattan Levi, Itay Barnea, Inbal Meshulach, Sapir Kontente, Daniel Benvaish, Rachel Cur-Cycowicz, Yoav N. Nygate, and Natan T. Shaked*

A new technique for sperm analysis is presented, measuring DNA fragmentation, morphology with virtual staining, and motility, all three criteria on the same individual unstained live cell. The method relies on quantitative stain-free interferometric imaging, providing unique topographic structural and content maps of the cell, becoming available for the first time for clinical use, together with deep-learning frameworks and least-squares linear approximation. In the common clinical practice, only motility evaluation can be carried out on live human cells, while full morphological evaluation and DNA fragmentation assays require different staining protocols, and therefore cannot be performed on the same cell, resulting in inconsistencies in fertility evaluation. A clinic-ready interferometric module is used to acquire dynamic sperm cells without chemical staining, together with deep learning to evaluate all three scores per cell with accuracy of 93.1%, 88%, and 90% for morphology, motility, and DNA fragmentation, respectively. It is shown that the expected number of cells that pass all three criteria based on the current evaluations performed separately does not correspond with the number of cells that pass all criteria, demonstrating the importance of the suggested method. The proposed stain-free evaluation method is expected to decrease uncertainty in infertility diagnosis, increasing treatment success rates.


1. Introduction

Infertility is a common condition, as roughly one in six couples face some degree of infertility, defined as failing to naturally conceive over the course of 1 year. Male infertility factors contribute to approximately half of all cases.^[1] In vitro fertilization (IVF) has made parenthood possible for many people that could not conceive otherwise. IVF enables fertilization of a female egg by a male sperm outside the female body. The selection of sperm cells is crucial especially for intracytoplasmic sperm injection (ICSI), a common type of IVF, in which the clinician chooses a single sperm cell using a micropipette and injects it into the female egg in a dish. Approximately 10 eggs are obtained in each IVF cycle through a surgery subsequent to the woman's hormonal treatment, and it is not uncommon that no fertilized egg develops into a good-quality embryo.^[2] Unfortunately, much less efforts are invested in analyzing sperm cells rather than eggs for IVF sperm selection. The

World Health Organization (WHO) provides criteria for male fertility evaluation based on imaging of the sperm sample.^[3] These criteria mainly consist of evaluating the sperm morphology, motility, and DNA fragmentation status. While motility assessment is done on live sperm cells swimming in a dish, internal morphological assessment and DNA fragmentation assays are typically performed on fixed and stained sperm cells, and on different portions of the sample; hence, different evaluations are done on different cells. Certain sperm cells may possess WHO2010-normal morphology, but not motility, or may possess both WHO2010-normal morphology and motility, but fragmented DNA, causing defective embryo development or delivery.^[4] Current tests cannot distinguish between these scenarios leading to a lack of consistency in sperm evaluation and selection performed by different clinicians, as well as a large margin of error, even when automated analysis is performed.

Deep convolutional neural networks (CNNs) have recently proven to be an efficient tool for image analysis and

K. Ben-Yehuda, S. K. Mirsky, M. Levi, I. Barnea, I. Meshulach, S. Kontente, D. Benvaish, R. Cur-Cycowicz, Y. N. Nygate, N. T. Shaked
Department of Biomedical Engineering
Faculty of Engineering
Tel Aviv University
Tel Aviv 69978, Israel
E-mail: nshaked@tau.ac.il

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202100200>.

© 2021 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202100200

classification.^[5–9] The model is generally built from sequential layers, each providing a nonlinear mapping of the previous layer output to the following layer. Recently, generative adversarial neural networks (GANs)^[10] have been successfully used for virtual staining of microscopic images.^[11–15] These neural networks include a generator model and a discriminator model, where the generator takes random noise and maps it to an image, and the discriminator classifies images as real or generated.^[16] Deep learning automatic classification of sperm cells could be the next gold standard. However, it is still difficult to obtain reliable automatic sperm classifiers or virtual staining using only qualitative 2D images as an input. The biological mechanisms that connect sperm movement, morphology, and contents to fertilization potential and normal pregnancy is not fully understood yet.^[17,18]

In the absence of staining, sperm cells are nearly transparent under bright-field microscopy, as their optical properties differ only slightly from those of their surroundings, resulting in a weak image contrast. An internal contrast mechanism that can be used when imaging sperm cells without staining is their refractive index.^[19–21] Phase imaging creates stain-free quantitative image contrast based on the optical path delay (OPD) induced in the light beam as it interacts with the sample, which can be recorded by interferometry. Conventional phase contrast imaging methods for sperm cells, such as Zernike phase contrast microscopy, differential interference contrast (DIC) microscopy, and Hoffman modulation contrast microscopy, are not quantitative, thus they do not enable interpretation of the resulting phase images in terms of the quantitative optical thickness of the sperm cell. In addition, these techniques present significant imaging artifacts, especially near the cell edges. Quantitative phase imaging records the full sample complex wave front including the optical thickness map of the cell, which is equal to the integral of the refractive-index values across the cell thickness. This map is proportional to the cell dry mass surface density, thereby providing cellular parameters that have not been available to clinicians.^[22] Until recently, quantitative phase imaging implementations were limited to optics labs, due to the optical setup bulkiness, difficulty of alignment, and sensitivity to mechanical vibrations. In recent years, we made significant efforts and succeeded to make these wave front sensors applicable and affordable for direct clinical use.^[21,23–25]

DNA fragmentation is a critical biomarker in sperm cells. Studies have shown that even in normal semen samples, $\approx 20\%$ of the sperm cells have fragmented DNA, which becomes worse with age.^[26] Sperm DNA fragmentation has been associated with reduced fertilization rates, reduced embryo quality, reduced pregnancy rates, and increased miscarriage rates. Thus, sperm cells with fragmented DNA should not be selected for IVF. DNA fragmentation is not well correlated with morphology and motility^[27] and cannot be imaged at the moment without staining. Detecting DNA fragmentation requires molecular staining, which cannot be carried out during IVF.

In this work, we present a new approach for measuring the DNA fragmentation status of live and unstained dynamic sperm cells, at the individual-cell level, in parallel to measuring the cell motility and morphology as though they have been stained. This is achieved by using a clinic-ready interferometric module to acquire dynamic sperm cells and record their quantitative phase profiles, followed by using an algorithmic architecture that

includes CNNs to virtually stain the cells and classify them based on their morphology, motility, and DNA fragmentation status. Each of the scores for morphology, motility, and fragmentation is mapped to a number between 0 and 1, where 0 means the lowest possible quality and 1 means the highest possible quality. All information is then mapped to a 3D scatter plot for each cell.

In contrast to previous attempts of finding correlations between pairs of stain-free bright-field (low-contrast) images and DNA-fragmentation-stained images, as well as for motility or morphology, using machine learning,^[28–34] here we use rapid interferometric imaging, providing not only the amplitude low-contrast image, but also the quantitative phase profile of the cell, including internal cellular organelle contrast and content-related cellular texture, attributed to the DNA fragmentation level of the cell. As the phase profile is quantitative, having meaningful and highly informative structural (topographic) and content (refractive-index)-related values on all the cell points, the gap between the images in the input pair is smaller compared to the previous methods that used bright-field imaging; thus, the presented approach allows individual-cell DNA fragmentation classification, in addition to internal-morphological-structure virtual staining of each cell (which is not possible by bright-field imaging), enabling more accurate fertility grading, based on the triple generalized score (WHO morphology, motility, and DNA fragmentation).

We show that the number of cells that pass all three criteria cannot be accurately determined by the number of cells that pass each criterion separately, which necessitates a different fertility grading procedure than is used today.

2. Results

The architecture of the overall system for attaining a generalized fertility evaluation per patient is shown in **Figure 1**. The proposed technique enables simultaneously obtaining, for each of the dynamic sperm cells imaged in a dish, the full stain-like morphology, DNA fragmentation, and motility status, without the need for chemical staining. As shown in Figure 1, the analysis includes four deep neural networks (DNNs) and least-squares linear approximation: DNN1 performs morphological virtual staining, DNN2 performs morphological scoring, DNN3 performs DNA fragmentation scoring, DNN4 performs DNA virtual staining, and motility scoring is performed using least-squares linear approximation. To generate quantitative stain-free imaging data, we implemented a clinic-ready holographic setup (Figure S1, Supporting Information) to acquire stain-free quantitative phase maps of the cells. This setup is composed of an inverted microscope and a custom-built common-path, compact interferometric module, connected to the microscope camera port (see Experimental Section). The acquired off-axis interferograms were processed into the optical thickness maps of the sperm cells, taking into consideration both their physical thickness and their refractive index contents. As shown in Figure S2 and in Video S1, Supporting Information, these maps can be visualized in HEMA-staining-like colors, thereby distinguishing between different subcellular structures. We designed a custom-built tracking algorithm and tracked all cells dynamically, resulting in a space–time array per cell.

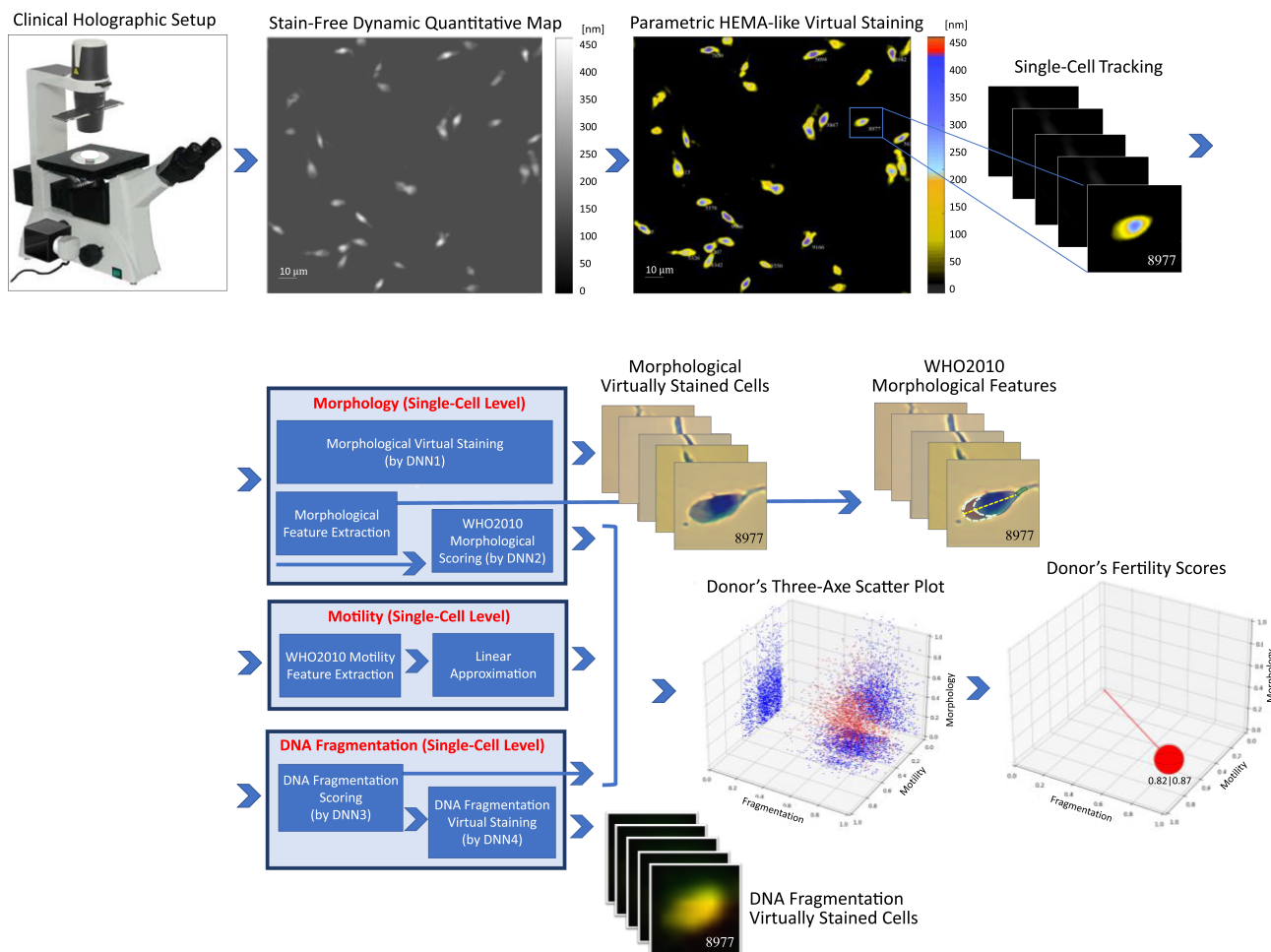


Figure 1. System schematics and working modes. Architecture for analyzing the full sperm morphology, motility, and DNA fragmentation status of large numbers of dynamic live sperm cells at the single-cell level, using stain-free quantitative imaging and deep learning, yielding more accurate donor fertility scores.

Using the networks trained on well-established ground-truth stain-based labels for sperm morphology and DNA fragmentation, as well as quantitative motility analysis using least-squares linear approximation, we can obtain for each live and dynamic cell the intracellular morphology and associated parameters, motility parameters, and the DNA fragmentation status using only the quantitative stain-free imaging data. We then map each cell to one point in a 3D space, with the axes being morphology, motility, and fragmentation values, thereby displaying the complete cellular status of each cell, for live cells without using cell staining. The combined cell points per patient in the 3D space are depicted as a sphere, with the sphere volume representing the patient's generalized fertility score. Regardless of the triple scoring of each individual cell, DNN1 and DNN4 use GAN models for virtual staining for visualization purposes, which is complementary to the automatic scoring capability.

Using this scheme, we quantitatively imaged 5101 dynamic human sperm cells from eight donors, resulting in 51 809 images. The collection of all cells measured per donor and their representation on the 3D scatter plot uniquely characterizes the

fertility status of the donor. Instead of having three different criteria measured on three different populations of the sample, where there is the risk of having cells that pass one criterion and fail one or more of the other criteria, we are now able to generate a triple-criteria score per cell.

2.1. Full Morphology Evaluation and Virtual Staining

As shown in Figure 1 and Video S1, Supporting Information, the morphological evaluation of each cell includes virtual staining of the cell, extraction of the cell morphological features, and the final cell classification, as recommended by the 2010 World Health Organization^[3] (WHO2010) for stained cells, although no chemical staining is used and the cells are not fixated but rather dynamically swim in a dish. For both the cell virtual staining and the WHO2010-based morphology scoring tasks, we used DNNs. The first DNN in Figure 1, DNN1, performs virtual staining of individual sperm cells.^[14] It is based on a conventional GAN architecture that is designed to generate

new high-quality images. DNN1 was trained with the stain-free optical thickness maps of sperm cells and their chemically stained counterparts as labels, for the exact same sperm cell. Then, the trained network could transform the stain-free optical thickness maps into virtually stained images, making them look as though they were chemically stained, without actual staining (Figure S3, Supporting Information), thereby providing the information necessary for gold-standard evaluation. The network was tested by classification of a trained embryologist, after randomizing the data order, and yielded very similar results to those obtained by chemical staining, as presented below. In the current study, virtual staining is implemented on live and highly dynamic sperm cells for the first time, in contrast to our previous research that demonstrated virtual staining on fixed sperm cells.^[14] In addition, we analyzed the optical thickness maps of the cells using classical image processing techniques to automatically extract six morphological features, including nucleus area, acrosome area, total head area, mean posterior–anterior difference, dry mass, and the variance of the optical thickness values quantifying the texture of the cell. These six features were calculated as explained in the study by Mirsky et. al^[35] and were used as input to DNN3 for DNA fragmentation scoring, as well as to give the clinician further quantitative morphological parameters, which are yet to be standard for use in the future. Note that virtual staining is demonstrated as an additional annotation guide to the embryologist, and not as the input or ground-truth label to DNN2 for morphological scoring or DNN3 for DNA fragmentation scoring presented below.

DNN2 in Figure 1 performs sperm scoring based on the WHO2010 morphology guidelines, which include five criteria: head shape, acrosome size, number of vacuoles, midpiece shape and orientation, and cytoplasmic droplets. Regardless of our experiment that examined all cell population without bias, these

WHO2010 guidelines for sperm morphology were previously formulated by examining sperm cells retrieved from the human female cervical secretion following sexual intercourse, where it is assumed that sperm cells that reached there have higher fertilization potential. During our training process, the network received the optical thickness map of the cell along with the ground-truth WHO2010 classification of a trained embryologist as labels, which is based on the chemically stained image. Once the model is trained, the network predicts the morphology score without using the chemical staining-based label. The network outputs six parameters; the five WHO2010 criteria per cell (head shape, acrosome size, number of vacuoles, midpiece shape and orientation, and cytoplasmic droplets) together with a sixth parameter: the direct overall prediction of whether the cell passes all five criteria (independent of the first five outputs) according to the embryologist. In addition, a combined overall prediction checking if the model predicted a “pass” for all five qualifications presented by the first five outputs was calculated, for comparison with the direct overall prediction (see Experimental Section). **Figure 2** presents the architecture of DNN2, as well as the receiver operating characteristic curve (ROC) and precision recall curve (PRC). This network was built as a standard CNN because the entire information necessary for cell classification based on the WHO2010 guidelines is available in the image. We started with very few layers and slowly increased the number of layers till generalized learning was obtained, as indicated by the training curves as well as the test success metrics.

The ROCs for the five qualifications attained areas under the curves (AUCs) of 95.7%, 96.3%, 100.0%, 98.7%, and 86.7%, respectively. The PRCs for the five qualifications attained AUCs of 93.6%, 95.5%, 100.0%, 99.7%, and 97.4%, respectively. The direct overall parameter, representing the direct classification, attained AUC of 96.2% for the ROC and 93.5% for the

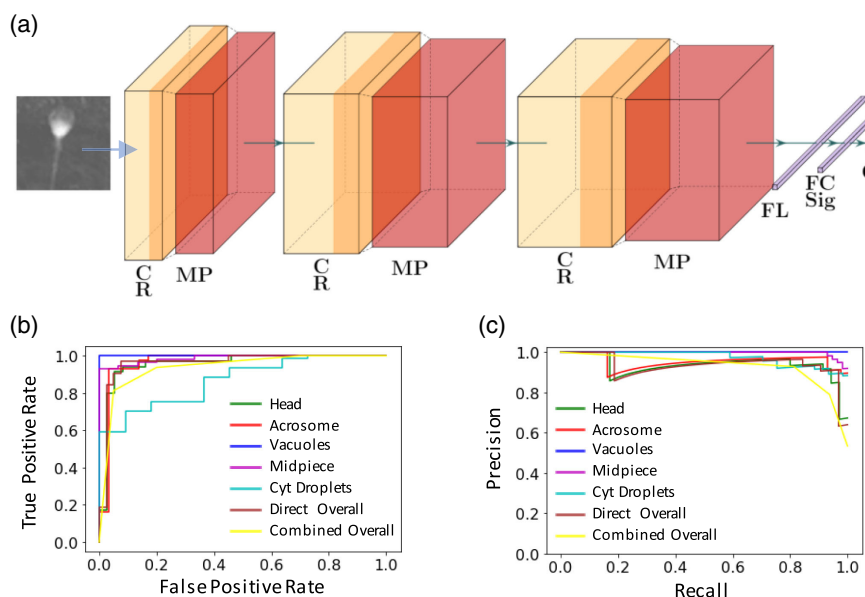


Figure 2. DNN1 for morphological sperm classification. a) CNN architecture predicting morphological features per stain-free image. b) ROC curves for the seven outputs of the network. The direct overall score is a prediction whether the cell passes all five criteria (independent of the first five outputs), and the combined overall score indicates whether the first five output are positive. c) Precision-recall curves for the seven outputs of the network. C, convolutional layer; R, ReLU activation; MP, max pooling layer; FL, flattening layer; FC, fully connected layer; Sig, sigmoid activation function; O, output.

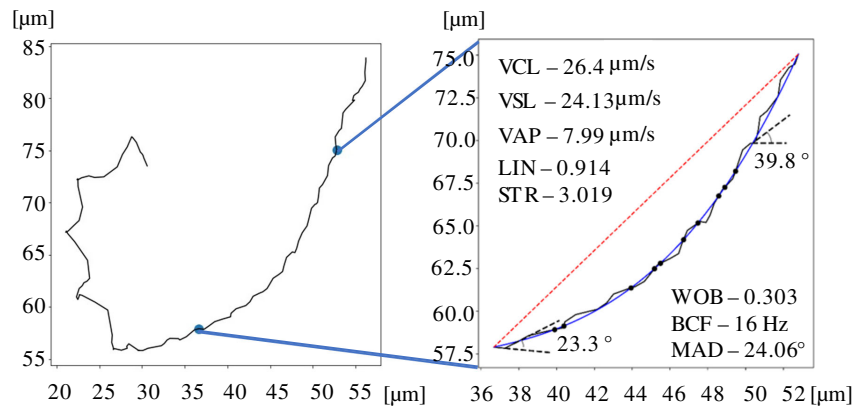


Figure 3. Automatic motility feature extraction. Example of a sperm cell tracked over 4.9 s (left), enlargement of a 1-sec segment and automatic calculation motility parameters (right).

PRC, precision of 90.9%, and accuracy of 93.1%. AUCs are used here as an indicator of network performance. Precision is expected to be a dominant predictor for a successful sperm classifier, as a person's potential of fertilization is mostly dependent on his best cells. These cells reach the egg in natural insemination and are the only cells that should be selected for IVF. Once the network is trained and tested, it can classify the cell without having the ground-truth label from the trained embryologist. The direct overall output was set as the coordinate value of the morphological axis in the 3D scatter plot for the cell examined.

2.2. Motility Evaluation

To assess the sperm motility on an individual-cell basis, we used the space-time arrays extracted from the cell tracking algorithm, composed of the locations of each cell in each frame. The trajectory for one representative cell is shown in **Figure 3**. We divided the trajectory into one-second windows with a half window stride, as the qualification of swimming linearly in progressively motile cells is only expected over short time intervals. One of these windows is shown on the right of **Figure 3**. We then calculated eight motility parameters suggested by the WHO2010. These include curvilinear velocity (VCL), straight-line velocity (VSL), average path velocity (VAP), linearity (LIN), wobble (WOB), straightness (STR), beat-cross frequency (BCF), and mean angular displacement (MAD). The median values of these parameters over all windows per cell were taken as the final motility parameters per cell. For each donor, another subsample was tested by an experienced embryologist according to the qualitative WHO2010 protocol, classifying each sperm cell into one of three motility classes: immotile, nonprogressively motile, and progressively motile. To compare the qualitative and quantitative motility tests, we conducted two comparative experiments. First, we calculated the correlation between the qualitative and quantitative motility test results over all eight donors. To do this, we chose four quantitative motility parameters (VCL, VSL, VAP, and $VSL \times LIN$), which most resembled the qualitative assessment, resulting in high significant correlations. The other parameters yielded small or no correlation. To cancel the effect of sampling errors, a video containing 87 sperm cells was processed

both quantitatively by our automatic algorithm and qualitatively by the embryologist. This increased the correlations (from 0.49 to 0.75) and their statistical significance (p -values decrease from 0.15 to 10^{-16}). We then used a least-square approximation to define a linear equation that maps all automatically extracted quantitative motility parameters to the three qualitative classes defined above, to ensure that no previously acquired fertility score is overlooked by our protocol. The normalized function value of the linear equation was used as the coordinate of the motility axis in the 3D scatter plot.

2.3. DNA Fragmentation Evaluation and Virtual Staining

We used DNN3 to grade each live sperm cell according to its DNA fragmentation level. In contrast to Barnea et al.^[36] that shows only statistical difference for specific parameters in large populations of sperm cells measured by interferometry, with population overlaps, here we provide an individual-cell DNA measurement without staining. For training, we used pairs of images: the stain-free quantitative thickness map of the cell and the image of the same cell after it was stained by acridine orange, a DNA fragmentation indicator emitting green fluorescence for double-stranded DNA (nonfragmented) and red fluorescence for single-stranded DNA or RNA (fragmented) as labels. Results are shown in **Figure S4**, Supporting Information. The automatically extracted morphological parameters were also inserted into DNN3, creating a bimodal neural network. **Figure 4a** presents the network architecture (see also Experimental Section). The ROC AUC and PRC AUC were all above 0.98. After training and testing, the network can take a stain-free quantitative optical thickness map of the cell and output the cell fragmentation level without the need for chemical staining. DNN3 used a bimodal neural network architecture. It receives both the optical thickness map and the six morphological features calculated based on this map, in addition to the chemically stained cell image label, because these features were found to be correlative to the DNA fragmentation level of the cell,^[36] practically guiding the network. When training without these six features, we obtain inferior results: accuracy of 0.90, ROC AUC of 0.78, and PRC AUC of 0.98.

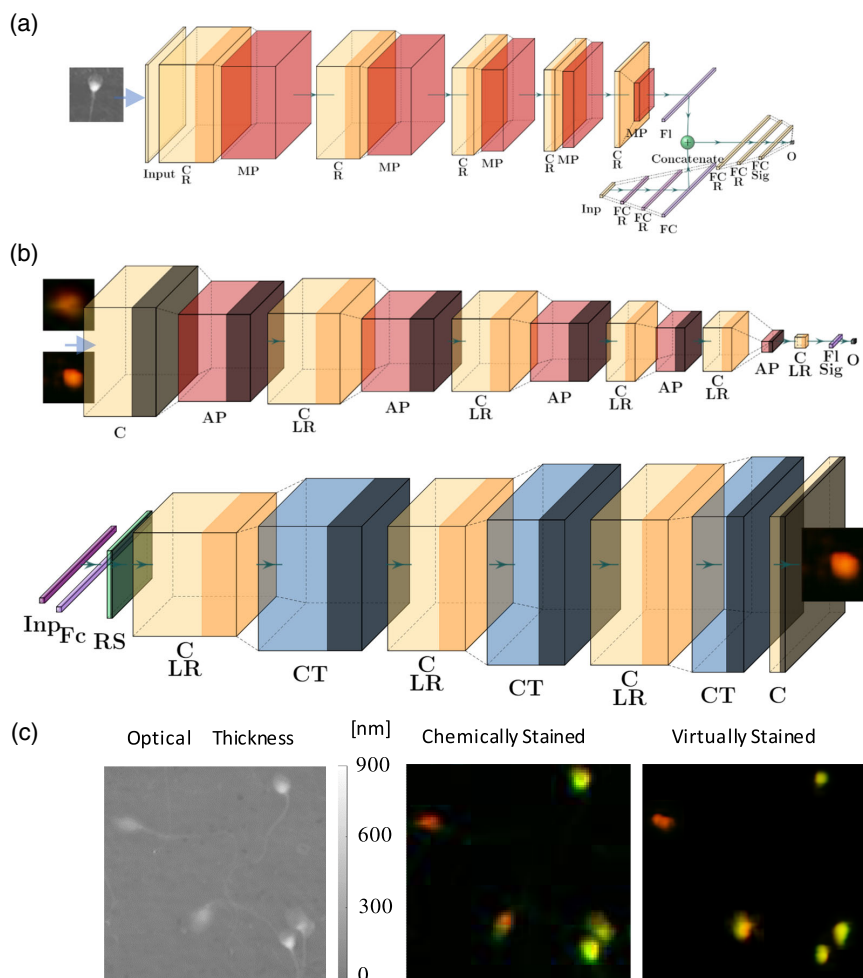


Figure 4. Stain-free DNA fragmentation classification using DNN3 and virtual staining using DNN4. a) Architecture of the bimodal neural network, predicting the fragmentation level for each stain-free optical thickness map. b) Architecture for the GAN model used for DNA fragmentation virtual staining with the discriminator network (top) and generator network (bottom). c) Comparison between unstained optical thickness map (left) with the actual chemically stained cell image (center) and the generated virtually stained cell image (right). C, convolutional layer; R, ReLU activation; MP, max pooling layer; FI, flattening layer; FC, fully connected layer; Sig, sigmoid activation function; O, output; AP, average pooling layer; LR, leaky ReLU activation; RS, resampling layer; CT, convolution transpose layer.

Another DNN (DNN4) used a GAN architecture to generate virtually stained DNA fragmentation images. It was trained to virtually stain the quantitative optical thickness map, creating a semblance of its acridine-orange-stained counterpart. The network architecture is shown in Figure 4b and its virtual staining operation is demonstrated in Figure 4c, yielding a virtually stained image that is very similar to the chemically stained sperm cell image shown in the center. Note that this virtual staining was not used to score the cells presented in Figure 5 and 6 according to their DNA fragmentation status. This scoring was done by DNN3. Only then, we used DNN4 with the fragmentation score from DNN3 as label to virtually stain the optical thickness map and make it look as if the cell was chemically stained with acridine orange, as a visual tool to the embryologist. The verification of DNN4 virtual staining results was done by a trained embryologist (see Section 4.7).

2.4. Patient's Fertility Scoring

To evaluate the patient's fertility, we calculated each sperm-cell morphology, motility, and DNA fragmentation as explained above, for live cells and without staining, and then mapped the cell into a single point in a 3D space with axes representing these three criteria. Our evaluation also consists of virtual morphological and acridine-orange staining per cell. As shown in Figure 5, the presented technique enables gathering all single-cell triple-criteria information regarding the population of sperm cells. The left column in Figure 5 shows the resulting 3D scatter plots for the eight donors measured while presenting both the 2-D projection of each cell on each of the three planes and also the location of that cell in the 3D space. The right column in Figure 5 shows the intersections of the three criteria and the left column shows the Venn diagrams, where each of the three circles represents a specific criterion. The number of cells that

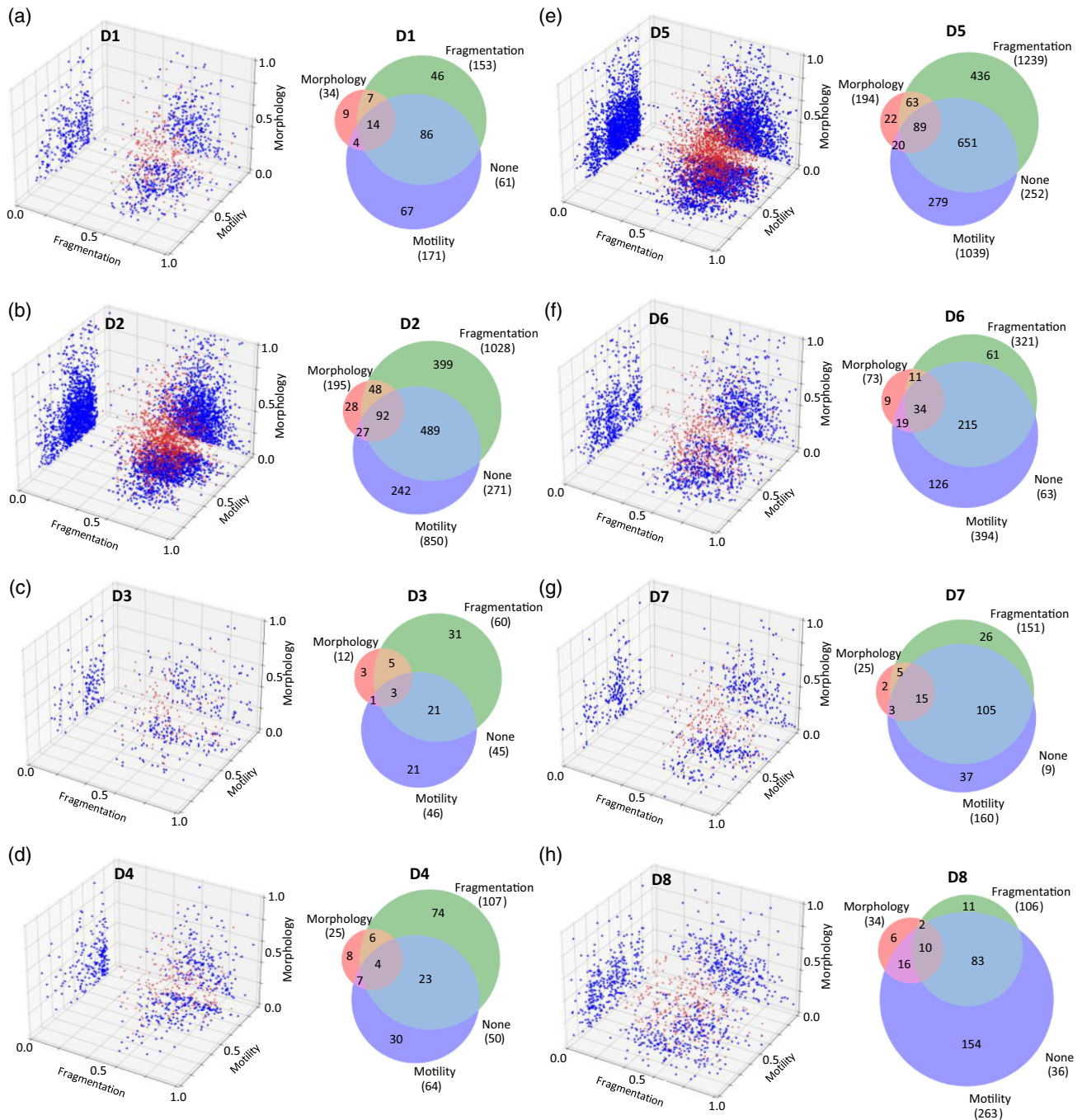


Figure 5. Triple-criteria sperm evaluation at the single-cell level for eight donors. (Left) 3D scatter plot, where each cell is represented by a red dot, and its three projections onto the 2D planes are represented by blue dots. (Right) Venn diagrams for the donors with each of the three fertility criteria and their intersections (morphology above a threshold of 0.5, motility above a threshold of 0.29, and fragmentation above a threshold of 0.61). a) Donor 1. b) Donor 2. c) Donor 3. d) Donor 4. e) Donor 5. f) Donor 6. g) Donor 7. h) Donor 8.

pass each criterion, together with the number of cells that pass each set of two criteria and all criteria, is displayed, where the passing thresholds are set to 0.5 for morphology, 0.29 for motility, and 0.61 for fragmentation. This figure demonstrates the great variability between donors. In relation to the other criteria, some donors have more cells that pass the fragmentation tests

(donors 2–5), while others have more cells that pass the motility test (donors 1, 6–8). Some donors have more cells that are both nonfragmented and morphologically intact (donors 1–3, 5, 7), while others have more cells that are both morphologically intact and motile (donors 4, 6, 8). The percentage of the overall passing cells differs among the donors: donor 5 (Figure 5e) has

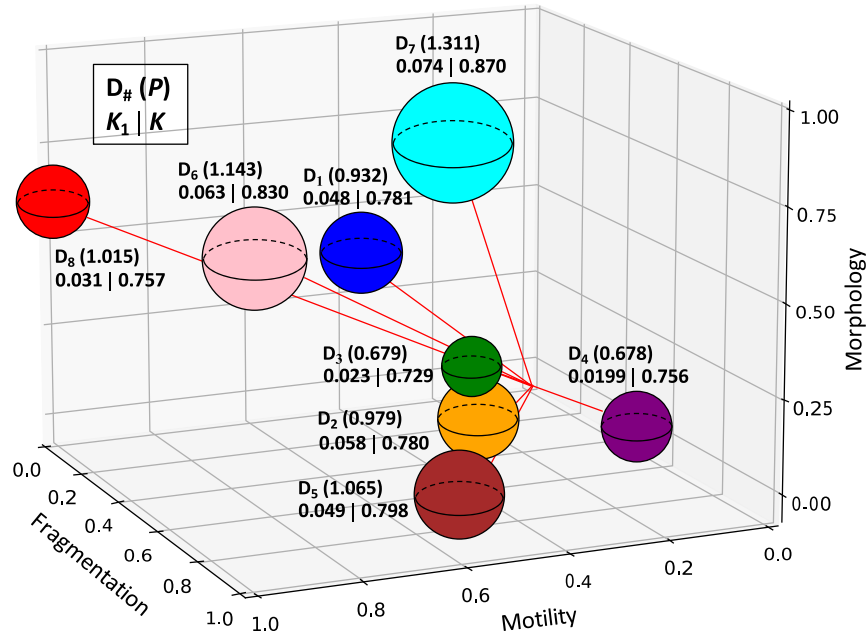


Figure 6. Donor fertility scoring. 3D plot representing all donors by their fertility score based on the current-practice score (in brackets, after the donor number) and the proposed scores $K_1 | K$, based on having the ability to measure all criteria together at the individual-cell level. The distance of each donor from the origin is correlative to the current-practice fertility score, whereas the size of each sphere is correlative to the proposed score.

significantly more cells than donor 2 (Figure 5b), yet donor 2 has three more sperm cells that pass all three criteria. The thresholds may be adjusted according to the patient's specific needs, as shown in Figure S5 and Video S2, Supporting Information.

To check whether the tests dependencies were consistent between the donors, we checked the statistical independence between the criteria per each donor. This was implemented by multiplying the three percentages of cells passing each criterion and comparing the result to the percentage of the cells passing all three criteria. For example, for donor 3 (Figure 5c), the number of cells that actually pass all three criteria is 1.53 times of the number of cells that are expected under the independent-criteria assumption, whereas for donor 4 (Figure 5d), the same ratio is only 0.95. This shows that the criteria are not completely independent, as the expected percentage of passing cells under the independent-criteria assumption is different from the end-result and that the criteria dependencies differ between the donors.

We now define the fertility score based on the current practice and two new fertility scores based on the presented triple-criteria capability. Due to the inability to measure morphology, motility, and DNA fragmentation all together on the same cell, fertility scoring today is carried out by checking each criterion separately, on different parts of the sample, and either assuming that the sample is homogenous or repeating the examination to check for consistency. The current fertility assessment relies on averaging the cells tested per each criterion separately, resulting in three different scores, yet missing the intersections between the criteria. To define the current-practice score, we first calculated the average results of each fertility assessment (morphology, motility, and DNA fragmentation) separately, averaging over all cells per donor. Then, we normalized each assessment score over all donors and found the distance of the resulting (x, y, z)

point from the origin of the 3D space. We define P as the distance from the origin in the 3D space, where each dimension corresponds to one of the three normalized coordinates (morphology $[x]$, motility $[y]$, and DNA fragmentation $[z]$)

$$P = \sqrt{\langle x \rangle^2 + \langle y \rangle^2 + \langle z \rangle^2} \quad (1)$$

where $\langle x \rangle = \sum_{i=1}^N x_i / N$ is the normalized morphology value averaged for all N cells measured per donor, $\langle y \rangle = \sum_{i=1}^N y_i / N$ is the normalized motility value averaged for all N cells measured per donor, and $\langle z \rangle = \sum_{i=1}^N z_i / N$ is the normalized DNA fragmentation value averaged for all N measured cells per donor. This normalized overall score P represents the fertility scores obtained by the current practice, as the averaging was done on each criterion separately, with the underlying assumption of sample homogeneity.

On the other hand, based on the presented unique capability to measure triple criteria per cell, we can define a new score, K , as follows

$$K = \frac{1}{N} \sum_{i=1}^N \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (2)$$

This parameter takes the distance from the origin for each cell in the normalized 3D space, and averages all these distances per donor.

We also define K_1 as the percentage of cells per donor that pass all three criteria, where \wedge is the logical "and" operator. The thresholds required to pass each criterion can be chosen per diagnosis goal. For example, in case of Figure 5, it is defined as follows

$$K_1 = \frac{1}{N} \sum_{i=1}^N (x_i > 0.5) \wedge (y_i > 0.29) \wedge (z_i > 0.61) \quad (3)$$

These new scores, K and K_1 , depend on how well the cells stand in all three criteria, whereas a donor might have no sperm cells that pass all three criteria together and still have a high previous fertility score P .

We now show that these new scores result in different donor grading, implying that erroneous donor grading is done today, even under the homogenous sample assumption. Figure 6 shows the previous fertility scores together with the new scores obtained for the eight donors. In this figure, each sphere represents a donor. The distance of the center of the sphere from the origin of the 3D space represents the donor's status as per the current practice, P , and can be seen as the first value next donor number. On the other hand, the sphere diameter is correlative to the donor's first new fertility score, K , which can be seen as on the right in the second row next to each donor's sphere. The second new score, K_1 , can be seen on the left in the second row next to each donor's sphere. As seen, the donors' rankings according to the old and new scores differ from each other. Donor 3, for instance, takes the last place according to the K score, as this donor's average sperm quality is the lowest, although donor 4 would be labeled as the worst donor according to the previous fertility grading, P . Donor 1 takes the fourth place according to the K score, while he takes fifth place according to the K_1 score and sixth place according to current score, P . On the other hand, the two best donors, donors 7 and 6, are consistent among all three scores, while donor 5 is the third best donor for both the K and the P scores but not for the K_1 score, implying that the new scores may be especially useful in discriminating situations of infertility.

3. Conclusion

In this work, we present a new capability to simultaneously measure individual sperm morphology, motility, and fragmentation, as well as virtually stain the sperm cells. Today, a clinician is unable to gather this necessary information per sperm cell, and instead the fertility examination provides independent percentages regarding each evaluation separately, on different parts of the cell population, where statistical changes may occur between subsamples. Thus, currently, there is no information regarding how many cells would have passed two or all three tests. In addition, the harsh fixation and staining protocols required for performing the morphology and DNA fragmentation assays might damage the reliability of the assays if not performed correctly, and create different results when performed by different labs. Our virtual classifiers solve these issues as they do not require staining and perform all three assays on the same cells, while they are alive and dynamic. We show that our virtual assays correlate with the chemical gold standards, and when applied together on each cell, give us the capability to consider the previously unavailable intersections between criteria when counting the normal and abnormal cells per patient. We show that patients may differ in the statistical dependencies of the different tests (Figure 5) and that consequently a patient with overall lower independent scores may very well have a higher score per

cell or even a higher percentage of qualified sperm cells (Figure 6), emphasizing the need for the new approach and the insufficiency of the current examination.

This approach is expected to give rise to personalized sperm quality evaluations by adjusting the thresholds to suit the current need. Figure S5 and Video S2, Supporting Information, show how different thresholds result in different sperm groups. Without any specialized needs, one may want to use the original thresholds, and obtain a result similar to that which is shown in Figure S5a, Supporting Information. However, another would possibly want to increase the fragmentation threshold, as shown in Figure S5b, Supporting Information, securing a low percent of fragmented cells from a potential sperm donation. In another case, one may personalize these thresholds for selecting sperm cells for fertilization and may want to raise all thresholds in order to select the most promising sperm cells, creating scatter plots similar to the one in Figure S5c, Supporting Information. Other situations may result in decreasing or increasing different thresholds, depending on which patient is being evaluated, what types of infertility problems he possesses and for what reason this evaluation is being held.

Here, we demonstrated the usefulness of the method for male fertility evaluation prior to IVF. In the future, the integration of a portable clinic-ready interferometric module that can be attached to the standard microscopic systems used today in the clinic, together with our virtual staining and classification triple-criteria method at the single-cell level, can enable computer-assisted sperm selection during IVF in real time, allowing clinicians to choose the most qualified sperm cells for egg injection, with the best morphology, motility, and DNA fragmentation scores. As the scoring is automatic, rather than manual, in the future, the system may be integrated with a robotic selector and independently choose the sperm cells with the highest probability of fertilization. Moreover, incorporating these systems will give rise to advanced research linking types of chosen sperm cells with fertilization and pregnancy success, with the ability to personalize medicine for patients suffering from fertility problems.

To conclude, we have suggested a new approach that utilizes stain-free quantitative phase imaging and deep learning, in order to improve male fertility assessments through combined morphology, motility, and DNA fragmentation scores, allowing clinicians and researchers to obtain previously unattainable fertility evaluations at the single sperm-cell level.

4. Experimental Section

Sample Preparation: Upon obtaining the institutional review board approval of Tel Aviv University, semen samples were collected from eight randomly selected donors between the ages of 18 and 45, after they signed an informed written consent. The samples were left to liquefy for 30 min. Sperm cells were isolated from the semen fluid using a PureCeption bilayer kit (ORIGIO, Måløv, Denmark) in accordance with the manufacturer's instructions. Then, the sample was suspended in human tubal fluid (HTF, 3 mL) medium (Irvine Scientific, CA, USA) and placed on top of a 40% and 80% silicon bead gradient and centrifuged for 20 min at 1750 rpm. The supernatant was removed and cells in the pellet were resuspended in 5 mL of HTF before centrifugation at 1750 rpm for 5 min. The supernatant was discarded and the pellet containing the living sperm cells was resuspended in HTF (100 μ L). For the stain-based morphological assays, 5 μ L of the sample was mixed with 5 μ L of QuickStain

(Biological Industries, Beit Haemek, Israel). For the stain-based DNA fragmentation assays, the sample was washed with HTF (5 mL) before centrifugation at 1750 rpm for 5 min. The supernatant was discarded, and the pellet was vortexed gently. The cells were then fixed by slowly adding 5 mL of methanol–acetic acid mix (3:1). After 5 min of incubation at room temperature, the cells were centrifuged at 1750 rpm for another 5 min and the supernatant was discarded. The pellet was resuspended in the fixative solution (0.1 mL). Next, the solution was placed on the cover slip and stained with acridine orange (0.19 mg mL⁻¹) for several hours before imaging.^[36]

State-of-the-Art Fertility Examination: For each donor, 10 µL of the sample was used for stain-based morphological evaluation, 10 µL for concentration and motility evaluation, 10 µL for stain-based DNA fragmentation evaluation, and the rest was used for stain-free interferometric assays. All state-of-the-art examinations were carried out according to the WHO2010 guidelines.^[3] For stain-based morphological analysis, 200 QuickStain-labeled cells were assessed on a slide according to WHO2010 guidelines using an inverted microscope (CKX53, Olympus) with 100× oil objective and a 10× ocular. Sperm concentration and motility were examined in a Makler counting chamber (Sefi Medical Instruments, Haifa, Israel) under an inverted microscope (Primover, Zeiss) with a 20× oil objective and 10× ocular. For stain-based DNA fragmentation analysis, the acridine-orange-labeled cells were imaged on a slide by a confocal fluorescence microscope (Zeiss LSM 510-META) using a 25×, 1.4 numerical-aperture microscope objective. Excitation wavelengths were 477–488 nm, and emissions were at 572–668 and 505–550 nm.

The correlations between our results and those obtained by the embryologist were $R = 0.85$, $p = 0.07$ for morphology, $R = 0.9$, $p = 0.03$ for motility, and $R = 0.84$, $p = 0.07$ for fragmentation.

Interferometric Phase Imaging: The optical system architecture is shown Figure S1, Supporting Information. The sample was illuminated by a HeNe laser (Thorlabs Model HRP170) with a wavelength of 632.8 nm, and was imaged by an inverted microscope consisting of microscope objective MO (Olympus PlanApo N, 60×/1.42 oil) and tube lens TL (focal length: 150 mm). The sample magnified image entered a clinic-ready portable interferometric module,^[24] where it was split by beam splitter BS. One beam was spatially filtered by lenses L1 and L2 ($f = 100$ mm) and a 15 µm pinhole, turning it into a reference beam, and then projected toward the camera by a corner retroreflector mirror, and the other beam was projected toward the camera at a small angle using a laterally shifted corner retroreflector mirror. The combined beams were projected through a 4f system composed of lenses L3 ($f = 50$ mm) and L4 ($f = 40$ mm), and the resulting off-axis image hologram was produced on the digital camera (iDS UI-3880CP Rev.2, 3088 × 2076 pixels, 2.4 × 2.4 µm each). In total, 231 one-minute interferometric videos were recorded at 40 frames per second on nonoverlapping field of views for the eight donors. Each off-axis hologram from each video frame was digitally Fourier transformed and one of the resulting cross-correlation terms was cropped and inverse Fourier transformed, thereby reconstructing the complex wave front of the sample. The phase argument underwent a phase unwrapping,^[37] and the result was divided by 2π and multiplied by the illumination wavelength, resulting in the optical path delay (OPD) or optical thickness map of the sample.

Cell Tracking Algorithm: The resulting OPD videos were processed using a specially designed cell-tracking algorithm, resulting in a space–time array per cell. This array consisted of the location of the cell in each frame and was used for the automatic motility, stain-free fragmentation, and stain-free morphological assessments. Connected component labeling is used to find the centroid of each cell. Detected cells in different frames that were close in time and space were grouped as the same cell; cells were included if recognized in less than five consecutive frames; cells containing less than 15 frames were not included.

Databases and Network Training: DNN1 for virtual staining was trained and tested (0.8:0.2) on 1328 pairs of cell images, each of them contained the stain-free OPD profile of a cell and the coinciding morphologically stained bright-field image of the same cell. DNN2 for morphological scoring was trained, tested, and validated (0.8:0.1:0.1) on 720 cells from this

database. DNN3 and DNN4 for DNA fragmentation scoring and virtual staining were trained, tested, and validated (0.8:0.1:0.1) on 595 pairs of cell images, each of them contained the stain-free OPD profile of a cell and the coinciding DNA-fragmentation-stained bright-field image of the same cell. The database of the swimming cells contained 5101 dynamic human sperm cells with trajectories tracking, originated from eight donors, resulting in 51 809 images. This database was used for motility scoring and for correlation of the results to the gold-standard morphology, motility, and DNA fragmentation evaluations of the donors performed by the embryologist according to the WHO2010 recommended protocols. In the network training, we have initialized the hyperparameters based on the default values in Tensorflow. Then, the Adam optimizer tuned the learning rate during training.

Automatic Stain-Free Morphological Evaluation: For morphological evaluation, the OPD maps were filtered so that only frames that resembled in-focus cells viewed from an axis perpendicular to the major axis of the cell remained. Cells with flat head orientation were chosen. In the clinic, the cells analyzed according to the WHO2010 morphological evaluation are stained and stuck to the floor of the dish in a flat-head orientation. To best train a network to perform similarly to the embryologists and also to enable a comparison with the WHO2010 guidelines, our filtering algorithm automatically used only images of cells that align with that criterion. Some of the necessary features for a sperm to be considered morphologically intact are barely seen if the sperm is in different orientations and therefore such images would be insufficiently informative for the model. We then extracted the total head area, nucleus area, acrosome area, mean posterior–anterior difference, mean OPD, and variance of the OPD. Ground-truth annotation for DNN2 training was performed by an experienced embryologist. The data were randomized several times and assessed by the embryologist in coded manner. The embryologist classified each cell according to the five following qualifications for the sperm head and provided a pass-or-fail result^[3,38]: a) The head should be smooth with an oval-like shape with a width-to-length ratio of roughly 3:5. b) The acrosome should be between 40% and 70% of the head area and should be clearly visible. c) There should be no more than two small vacuoles, and they should occupy less than 20% of the sperm head and exist outside the postacrosomal region. d) The midpiece should be aligned with the head's major axis, should be about the length of the sperm head, and should be regularly shaped and slender. e) Residual cytoplasm should not exceed one-third of the sperm's head size. The sperm head parameters were assessed directly in our analysis. The WHO2010 guidelines regarding the sperm tail state the following: a) The principal midpiece should be uniform along its length. b) The tail should be thinner than the midpiece. c) The tail should be ≈ 45 µm long. d) There is no sharp angle in the tail, indicative of a flagellar break. These tail defects are indirectly incorporated in our overall score of the sperm cell. The first and second tail defects are indirectly included in our analysis, as the clinician assessed the entire sperm cells, and the neural network received slightly more than just the midpiece. The third requirement is quantitative and normal gold-standard assessments typically do not incorporate it in practice. The fourth tail defect is also indirectly incorporated into our analysis by using the motility algorithm, as severely defected tails hinder the motility. The manual analysis was done by a single highly experienced clinical embryologist, head of an IVF lab in a major hospital, with more than 15 years of experience. To assess the embryologist subjectivity, we wanted to see that the difference between him and the model prediction is similar to the typical difference between different embryologists, which is estimated to be up to 5%.^[39] We provided the embryologist with a dataset given to him five times, each time in a random order, and gathered his annotation on the data. We then computed the correlation between his annotation and the prediction DNN2, and found a correlation of 96% with significance in the order of 10e-57. In addition, the correlation between his own annotations was 95% with a similar significance.

We used a previous sperm database^[35] and built here a new CNN classifier, DNN2. The input to the model was the preprocessed OPD image of the cell (size 200 × 200 pixels), as they were vertically centered and rotated to the same direction, with the major axis of the cell being horizontal.^[35] The images were 2× augmented with vertical flipping. The output was a

binary vector of length 6 predicting if the cell passed each of the five criteria (a–e), and a direct overall prediction (stating whether the cell passed all five criteria together). The error of each WHO2010 feature was used in the loss function, forcing the network to account for these features. Calculating these features also enabled checking whether the direct overall prediction of the model is better than the analysis based on the first five outputs.

The ground-truth labels of the cells are binary; 0 (negative) being abnormal according to the WHO2010 guidelines and 1 (positive) being normal. The model predicts if a cell is normal or abnormal by outputting a value between 0 and 1, where the closer it is to 0 the more the model is confident that cell is abnormal and vice versa. True positive (TP) is defined as a cell that is classified by the model as positive (above 0.5) and it is indeed positive (1) according to WHO2010. True negative (TN) is defined as a cell that is classified by the model as negative (below 0.5) and it is indeed negative (0) according to WHO2010. False positive (FP) is defined as a cell that is classified by the model as positive (above 0.5) but it is negative (0) according to WHO2010. False negative (FN) is defined as a cell that is classified by the model as negative (below 0.5) but it is positive according to WHO2010. Based on TP, TR, FP, and FN, the confusion matrices are defined (see Figure S6, Supporting Information), as well as the true positive rate (recall), precision, and accuracy. The ROC and PRC curves are obtained by changing the default threshold of 0.5 to a range of values between 0 and 1.

The architecture of the model is shown in Figure 2. It consisted of three CNN layers with 32, 64, and 128 kernels of size 3×3 , ReLU activation functions and padding set to 'same', with max pooling layers of size 2×2 , dropout layers with probability of 0.5 and 2 fully connected (FC) layers of size 1500 and five with the same dropouts and activations, with an L2 kernel regularizer consisting of a penalty of 0.001 and a final sigmoid activation function. The Adam optimizer was used with a learning rate of 0.001 and the compilation consisted of a custom binary cross entropy weighted loss, computing a weighted sum of all the individual feature's binary cross entropy losses^[40,41] (see equations in the Supporting Information). The model was trained with a batch size of 100 and 6000 epochs with a training–validation–test split of 80%–10%–10%. The code was written in Python in Google Collaboratory using Tensorflow 2.3.0 with Tensorflow Keras 2.4.0.

Automatic Motility Evaluation: To automatically assess sperm motility, we created an algorithm that takes the space–time arrays of the cell tracking algorithm as the inputs, and outputs each donor's motility scores. A progressive sperm cell swims fast and in a linear projection at a velocity of at least $25 \mu\text{m s}^{-1}$.^[3] As linearity is expected only for short time-periods and short distances, one-second windows in increments of half a second were selected to assess progressivity^[38] and the median values were chosen for each motility score per cell. The motility features extracted were: a) VCL—time-averaged velocity of a sperm head along its actual curvilinear path. b) VSL—time-averaged velocity of a sperm head along its linear estimated path—a straight line connecting its first and last spatial points. c) VAP—time-averaged velocity of a sperm head along its average estimated path. The average path was calculated as a third-degree polynomial approximation of the original path. d) LIN—the linearity of the curvilinear path calculated as the ratio of the VSL to the VCL. e) WOB—oscillation of the actual path about the average path, calculated as the ratio of the VAP to the VCL. f) STR—linearity of the average path, calculated as the ratio of the VSL to the VAP. g) BCF—the average rate at which the curvilinear path crosses the average path. h) MAD—the time-averaged absolute values of the instantaneous turning angle of the sperm head along its curvilinear trajectory. i) Progressiveness (PROG)—a cell is considered progressive if its VSL is at least $25 \mu\text{m s}^{-1}$ and its LIN is at least 0.6. To compare our results to the previous qualitative motility results according to the WHO2010, we created a least-squares-based linear transformation that uses all nine features (a–i) and maps them to a prediction in the original scale of 1–3 (immotile to progressively motile) that were given to the cell by the embryologist. After all cells were given a motility score based on this three-class assessment, the average was taken as the final motility score per donor.

Stain-Free DNA Fragmentation Evaluation: We created neural networks for DNA fragmentation classification and virtual staining. For training, we used the database described in our previous work, which contains pairs of OPD maps (size 224×224 pixels) and the same cell labeled in acridine orange and imaged under a confocal fluorescence microscope (Zeiss LSM 510-META).^[36] Each fluorescent cell was evaluated in a random order by the experimentalist in a color scale of 1 (red) to 5 (green). This number was used as the gold standard while training a five-class CNN classifier. The inputs to the network were the stain-free OPD map, and the cell dry mass, mean anterior–posterior difference, head, acrosome and nucleus areas and head OPD mean and variance. The CNN classifier architecture is shown in Figure 4a. The OPD image went through five CNN layers, while the features went through three FC layers. All the layers in the first two parts had the ReLU activation function. The results of both the subnetworks were concatenated and inserted into the third and final part consisting of an additional three FC layers and the output, while the activation function of the hidden layers was the leaky ReLU with a slope coefficient (α) of 0.3, and the final activation was a sigmoid with the loss being weighted binary cross-entropy. The ground-truth of each cell was between zero and one, while class 1 was given a value of 0.2, class 2 a value of 0.4 and so on, with class 5 being given a value of 1. The predicted class of each cell was the one whose value was closest to the model prediction. The classifier had a train–validation–test data split of 80%–10%–10%. The model was trained on five Monte-Carlo cross-validation sets. The weight for each class was defined as the ratio between the number of samples from the major class and the number of samples from the current class,^[42] in a similar way to DNN2. To avoid overfitting, a dedicated train set was used during training, and the model weights finally selected were those that resulted in the best validation accuracy. The final model's success metrics were based on the test set that does not overlap with the validation and training sets. See the resulting test-set confusion matrices in Figure S6, Supporting Information.

Another network (DNN4) was used for individual-cell virtual staining. It mapped the predicted fragmentation status from the classifier to the expected acridine orange color and visualized the cells as though they have been stained by acridine orange, producing virtually stained images (size 112×112 pixels), but without using chemical staining. First, we took all the fluorescent images of each cell and split the cell bounding box into nine three by three patches. For each patch, we calculated the ratio between the average green and red channel values. Then, for each cell, we created a nine-feature vector of the results sorted in an ascending order. We used a *k*-means algorithm ($k=9$) to cluster the images, and then grouped the clusters into five groups of colors. Then, we created a GAN, with architecture shown in Figure 4b, which was trained separately for each class, as the input to the discriminator would switch from a resized version of the original image belonging to that class, to the generated image. This DNA fragmentation virtual staining model was evaluated in a similar manner to the morphological virtual staining model. An experienced embryologist received 100 virtually stained cells and classified them to the five classes, 4 times in a row, where in each run the cells were shown to him in random order. The correlation between the ground-truth classes and the embryologist came out to be 0.967 with p -value < 0.0005 .

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Applied Sciences and Engineering grant from the Ministry of Science and Technology of Israel.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon a reasonable request.

Keywords

cell imaging, deep learning, digital holographic microscopy, fertility, quantitative phase microscopy, sperm fragmentation

Received: October 5, 2021

Revised: November 27, 2021

Published online: December 26, 2021

-
- [1] A. Agarwal, A. Mulgund, A. Hamada, M. R. Chyatte, *Reprod. Biol. Endocrinol.* **2015**, 5, 1.
- [2] N. Gleicher, V. A. Kushnir, D. H. Barad, *Human Reproduction Open* **2019**, Human Reproduction Open **2019**, p. hoz017.
- [3] World Health Organization, *WHO Laboratory Manual for Examination and Processing of Human Semen*, 5th ed., WHO Press, Geneva **2010**.
- [4] D. P. Evenson, Z. Darzynkiewicz, M. R. Melamed, *Science* **1980**, 210, 1131.
- [5] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, 521, 436.
- [6] J. Schmidhuber, *Neural Netw.* **2015**, 61, 85.
- [7] L. Deng, D. Yu, *Found. Trends Signal Process.* **2014**, 7, 1.
- [8] H. Greenspan, B. Van Ginneken, R. M. Summers, *IEEE Trans. Med. Imag.* **2016**, 35, 1153.
- [9] M. T. McCann, K. H. Jin, M. Unser, *IEEE Signal Process. Mag.* **2017**, 34, 85.
- [10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Adv. Neural Inf. Process. Syst.* **2014**, 27, 2111.
- [11] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, *Cell* **2018**, 173, 792.
- [12] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, G. R. Johnson, *Nat. Methods* **2018**, 15, 917.
- [13] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, A. Ozcan, *Light: Sci. Applic.* **2019**, 8, 1.
- [14] Y. N. Nygate, M. Levi, S. K. Mirsky, N. A. Turko, M. Rubin, I. Barnea, G. Dardikman-Yoffe, M. Haifler, A. Shalev, N. T. Shaked, *Proc. Natl. Acad. Sci.* **2020**, 117, 9223.
- [15] M. E. Kandel, Y. R. He, Y. J. Lee, T. H. Y. Chen, K. M. Sullivan, O. Aydin, M. T. A. Saif, H. Kong, N. Sobh, G. Popescu, *Nat. Commun.* **2020**, 11, 6256.
- [16] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2017**, p. 1125.
- [17] Z. Li, A. Y. Wang, M. Bowman, K. Hammarberg, C. Farquhar, L. Johnson, N. Safi, E. A. Sullivan, *Hum. Reprod.* **2018**, 33, 1322.
- [18] L. Simon, A. Zini, A. Dyachenko, A. Ciampi, D. T. Carrell, *As. J. Androl.* **2017**, 19, 80.
- [19] M. Haifler, P. Girshovitz, G. Band, G. Dardikman, I. Madjar, N. T. Shaked, *Fertility Sterility* **2015**, 104, 43.
- [20] G. Di Caprio, M. A. Ferrara, L. Miccio, F. Merola, P. Memmolo, P. Ferraro, G. Coppola, *J. Biophotonics* **2015**, 8, 779.
- [21] M. U. Daloglu, W. Luo, F. Shabbir, F. Lin, K. Kim, I. Lee, J. Q. Jiang, Cai, V. Ramesh, M. Y. Yu, A. Ozcan, *Light: Sci. Applic.* **2018**, 7, 17121.
- [22] P. Girshovitz, N. T. Shaked, *Biomed. Opt. Express* **2012**, 3, 1757.
- [23] P. Girshovitz, N. T. Shaked, *Light: Sci. Applic.* **2014**, 3, 151.
- [24] P. Girshovitz, N. T. Shaked, *Opt. Express* **2013**, 21, 5701.
- [25] N. T. Shaked, *Opt. Lett.* **2012**, 37, 2016.
- [26] N. P. Singh, C. H. Muller, R. E. Berger, *Fertility Sterility* **2003**, 80, 1420.
- [27] P. Cohen-Bacrie, S. Belloc, Y. J. Ménézo, P. Clement, J. Hamidi, M. Benkhalifa, *Fertility Sterility* **2009**, 91, 1801.
- [28] Y. Wang, J. Riordon, T. Kong, Y. Xu, B. Nguyen, J. Zhong, J. B. You, A. Lagunov, T. G. Hannam, K. Jarvi, D. Sinton, *Adv. Sci.* **2019**, 6, 1900712.
- [29] C. McCallum, J. Riordon, Y. Wang, T. Kong, J. B. You, S. Sanner, A. Lagunov, T. G. Hannam, K. Jarvi, D. Sinton, *Commun. Biol.* **2019**, 2, 250.
- [30] J. Riordon, C. McCallum, D. Sinton, *Com. Biol. Med.* **2019**, 111, 103342.
- [31] S. A. Hicks, J. M. Andersen, O. Witczak, V. Thambawita, P. Halvorsen, H. L. Hammer, T. B. Haugen, M. A. Riegler, *Sci. Rep.* **2019**, 9, 16770.
- [32] S. Javadi, S. Abolghasem Mirroshande, *Com. Biol. Med.* **2019**, 109, 182.
- [33] I. Iqbal, G. Mustafa, J. Ma, *Diagnostics* **2020**, 10, 325.
- [34] J. Bem You, C. McCallum, Y. Wang, J. Riordon, R. Nosrati, D. Sinton, *Nat. Rev. Urol.* **2021**, 18, 387.
- [35] S. K. Mirsky, I. Barnea, M. Levi, H. Greenspan, N. T. Shaked, *Cytometry A* **2017**, 91, 893.
- [36] I. Barnea, L. Karako, S. K. Mirsky, M. Levi, M. Balberg, N. T. Shaked, *J. Biophotonics* **2018**, 11, 20180013.
- [37] D. C. Chiglia, M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, Wiley-Interscience, Hoboken, NJ **1998**.
- [38] T. B. Hargreave, *Male Infertility*, Springer, New York **1994**.
- [39] A. P. Brady, *Insights Imaging* **2017**, 8, 171.
- [40] G. King, L. Zeng, *Political Analysis* **2001**, 9, 137.
- [41] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA **2016**.
- [42] W. Dubitzky, M. Granzow, D. Berrar, *Fundamentals of Data Mining in Genomics and Proteomics*, Springer, New York **2009**, p. 178.