



## OPEN Sperm motility assessed by deep convolutional neural networks into WHO categories

Trine B. Haugen<sup>1✉</sup>, Oliwia Witzczak<sup>1</sup>, Steven A. Hicks<sup>2</sup>, Lars Björndahl<sup>3</sup>,  
Jorunn M. Andersen<sup>1,4</sup> & Michael A. Riegler<sup>2,4</sup>

Semen analysis is central in infertility investigation. Manual assessment of sperm motility according to the WHO recommendations is the golden standard, and extensive training is a requirement for accurate and reproducible results. Deep convolutional neural networks (DCNN) are especially suitable for image classification. In this study, we evaluated the performance of the DCNN ResNet-50 in predicting the proportion of sperm in the WHO motility categories. Two models were evaluated using tenfold cross-validation with 65 video recordings of wet semen preparations from an external quality assessment programme for semen analysis. The corresponding manually assessed data was obtained from several of the reference laboratories, and the mean values were used for training of the DCNN models. One model was trained to predict the three categories progressive motility, non-progressive motility, and immotile spermatozoa. Another model was used in predicting four categories, where progressive motility was differentiated into rapid and slow. The resulting average mean absolute error (MAE) was 0.05 and 0.07, and the average ZeroR baseline was 0.09 and 0.10 for the three-category and the four-category model, respectively. Manual and DCNN-predicted motility was compared by Pearson's correlation coefficient and by difference plots. The strongest correlation between the mean manually assessed values and DCNN-predicted motility was observed for % progressively motile spermatozoa (Pearson's  $r = 0.88$ ,  $p < 0.001$ ) and % immotile spermatozoa ( $r = 0.89$ ,  $p < 0.001$ ). For rapid progressive motility, the correlation was moderate (Pearson's  $r = 0.673$ ,  $p < 0.001$ ). The median difference between manual and predicted progressive motility was 0 and 2 for immotile spermatozoa. The largest bias was observed at high and low percentages of progressive and immotile spermatozoa. The DCNN-predicted value was within the range of the interlaboratory variation of the results for most of the samples. In conclusion, DCNN models were able to predict the proportion of spermatozoa into the WHO motility categories with significantly lower error than the baseline. The best correlation between the manual and the DCNN-predicted motility values was found for the categories progressive and immotile. Of note, there was considerable variation between the mean motility values obtained for each category by the reference laboratories, especially for rapid progressive motility, which impacts the training of the DCNN models.

A basic semen analysis is central in infertility investigation and implies examinations of sperm number, sperm motility, and sperm morphology<sup>1</sup>. Manual assessment of these variables according to WHO recommendations is regarded as the gold standard<sup>1,2</sup>. This standardization is essential in clinical work, for comparability across laboratories and over time, and for the establishment of reference limits for semen parameters<sup>1,3,4</sup>. For sperm motility, it includes to estimate the proportion of spermatozoa in different categories of motility, or as immotile spermatozoa. However, the manual motility assessment is time-consuming and requires comprehensive training, which is a prerequisite to obtain accurate and reproducible results.

In recent years, computer-aided sperm analysis (CASA) systems have been improved, however, these systems still have limitations in analyzing undiluted human ejaculates<sup>1,5,6</sup>. Moreover, the various CASA systems differ in technology, and the result of the analyses will be instrument-dependent. CASA may be valuable in identifying

<sup>1</sup>Department of Life Sciences and Health, OsloMet – Oslo Metropolitan University, Oslo, Norway. <sup>2</sup>Simula Metropolitan Center for Digital Engineering, Oslo, Norway. <sup>3</sup>ANOVA, Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>These authors contributed equally: Jorunn M. Andersen and Michael Riegler. ✉email: tribha@oslomet.no

characteristics of spermatozoa prepared for medically assisted reproduction, but is not recommended in routine assessment<sup>1</sup>.

The potential application of artificial intelligence in reproductive medicine has resulted in increased research activity to develop algorithms for use in semen analysis<sup>7,8</sup>. Most of the work has focused on sperm morphology to improve the accuracy of the assessment and make the classification automated. There have also been several approaches to developing machine learning models for tracking sperm movement, however, few studies have investigated how algorithms are able to predict sperm motility<sup>9–14</sup>.

Within the field of machine learning, deep learning is the current state of the art. Deep convolutional neural networks (DCNNs) are especially suitable for image classification and may therefore be applicable in assessing sperm motility. DCNNs learn the important features automatically from the data. These features might not make sense to humans, but they can contain important information to solve the task at hand. This can also allow us to discover new knowledge through model explanations that describe what information the DCNN uses to make a specific prediction.

However, the algorithms need training data and can only perform as well as the quality of the provided data permits. To develop DCNN models with high predictive power for sperm motility categorization, the dataset used should be based on accurate and reproducible methods and of a quality as good as possible in terms of image resolution and ground truth data.

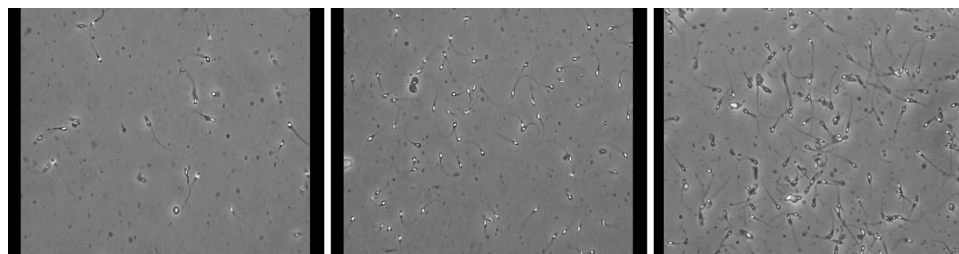
We have previously investigated the performance of various machine learning models in analyzing sperm motility from videos of fresh semen samples<sup>14</sup>. The data in that study was based on manual assessment of three categories of motility as described in WHO 2010 Manual<sup>15</sup>, performed by one person. This dataset has also been used in two recent studies<sup>12,13</sup>. According to the WHO recommendations from 2010, the assessment of sperm motility is based on a system that distinguishes motile spermatozoa from immotile and progressive motility from non-progressive. We found that the DCNN model ResNet-50 overall performed best. In the most recent recommendations from WHO<sup>1</sup>, the differentiation of progressive motility into rapid and slow was again recommended, as in the WHO 1999 manual<sup>16</sup>.

The aim of the present study was to test ResNet-50 architecture on sperm motility videos obtained from an external quality assessment programme for semen analysis, using the corresponding manual data provided by several reference laboratories participating in the programme. The assessment did also include the differentiation of progressive motility into rapid and slow. The performance of the DCNN models in predicting the proportion of spermatozoa into the motility categories was compared with the manual assessments. The evaluation of each sample was performed on the same clips of the videos of wet semen preparations.

## Materials and methods

**Dataset.** Videos of wet preparations (400× magnification) of 65 fresh semen samples were obtained from the ESHRE-SIGA EQA Programme for Semen Analysis between 2006 and 2018 and used in the development of the DCNN model. All participating laboratories analyzed the same videos and sent their results back to the EQA programme manager after analysis. The video clips were assessed by personnel from four to ten laboratories. Data was stored and shared for comparison between labs. The dataset used for training the DCNN models contains these videos and mean values of motility results provided by the reference laboratories. The dataset used to train the models in this study is openly available and can be downloaded at<sup>17</sup>.

The videos were recorded at the Andrology Unit, Karolinska University Hospital, and Karolinska Institutet, and distributed among the participating laboratories in the programme. The ejaculates were incubated at 37 °C, and the videos were made 30–60 min after collection. The temperature during the recordings was kept at 37 °C by use of pre-heated slides placed on microscope trays with a controlled temperature of 37 °C. Randomly chosen fields were recorded for 5–10 s, depending on the number of visible spermatozoa. The number of fields was chosen to allow the assessment of at least 200 spermatozoa. Some of the investigators used a mask with a 25 × 25 µm reticle to facilitate the assessment of sperm motility. Micrometer scale recorded at the beginning of each DVD, with the same magnification as the actual sample recording, allowed each investigator to calibrate their own screen used. The manual assessment of the videos was performed according to WHO 1999 criteria<sup>16</sup>. Sperm motility was categorized into rapid progressive (grade a), slow progressive (grade b), non-progressive (grade c), and immotile (grade d). In addition, we calculated the proportion of total progressive motility (a + b), which is a part of the categorization system in the WHO 2010 manual<sup>15</sup>. Examples of frames from videos are shown in Fig. 1.



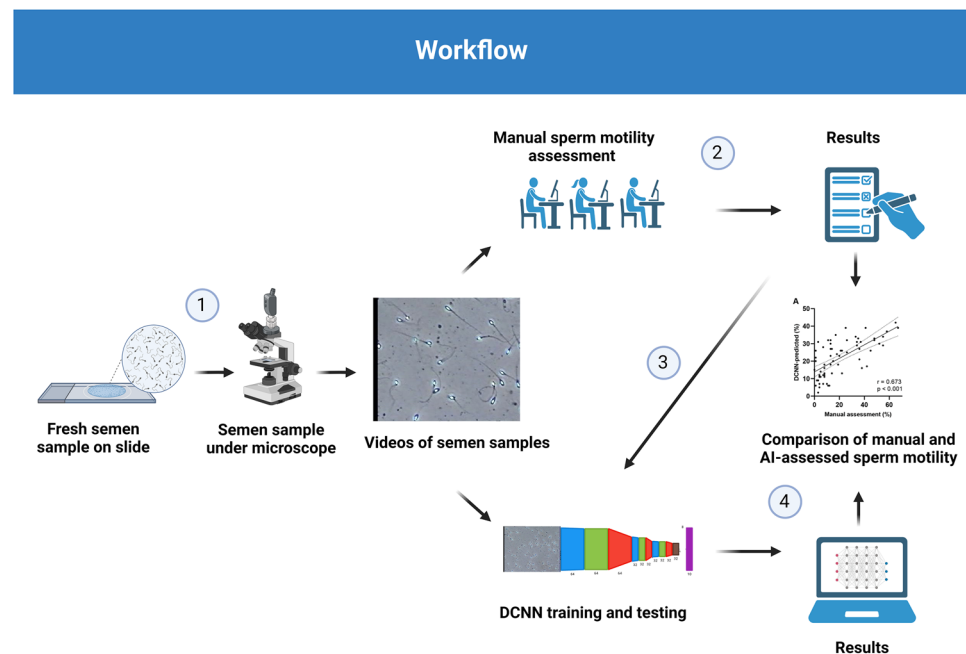
**Figure 1.** Example frames from three different videos from the presented dataset.

**Model development.** Lucas–Kanade optical flow was estimated for every second of the video and then visualized as an image, which was then used as input to the DCNN. In essence, this compresses the temporal information about the movements of the spermatozoa into a single image that can more easily be interpreted by the DCNN model. The videos have a framerate of 30 frames per second, making each optical flow-generated image represent motion calculated across 30 frames. The DCNN architecture is based on ResNet-50, which has previously shown promising results when analyzing videos of live spermatozoa<sup>14</sup>. The ResNet-50 is built with a Global Average Pooling layer at the end and trained on optical flow-based images generated by Lucas-Kanade. For training, we used the optimizer Adam with a learning rate of 0.0004 and calculated loss using mean absolute error (MAE). Otherwise, the default parameters included in the Keras Python package were used. For the baseline, ZeroR (pseudo regression) was performed. Results are reported as MAE, where the closer MAE is to 0, the more accurate the model is. Ten-fold cross-validation was used to compensate for the relatively small dataset, where per fold, one-tenth of the data was used as an independent validation dataset and left out from training. The model was trained for a maximum of 1,000 epochs (rounds of training the model), where we stopped training if the model did not improve on the validation dataset for the last 15 epochs. Using the DCNN architecture, we trained two models, one for predicting the percentage of spermatozoa in the motility grades a, b, c, and d, the four-category model, and one for predicting the percentage of spermatozoa in the motility grades a + b, c, and d, the three-category model. An overview of the work is presented in Fig. 2.

**Statistical analysis.** Analyze-it for Excel version 5.30.1 and GraphPad Prism version 9.0 were used for statistical analysis and for the generation of plots. Scatter plots were used to visualize the correlations between manual and DCNN-predicted motility. Associations were assessed by Pearson's correlation coefficient. P-values < 0.01 were considered significant. Difference plots were used to compare manual and DCNN-predicted motility. The bias was estimated as the median of the differences and is shown as a horizontal line, and 95% limits of agreement (LoA) were added in dashed lines to display outliers. Line plots were used to visualize the difference in results between the manual assessment and the results predicted with DCNN. Mean values from the manual assessment were sorted from lowest to highest for the respective categories. Minimum and maximum values were added to indicate the interlaboratory variation for each sample. Minimum and maximum values were available for 43 samples.

## Results

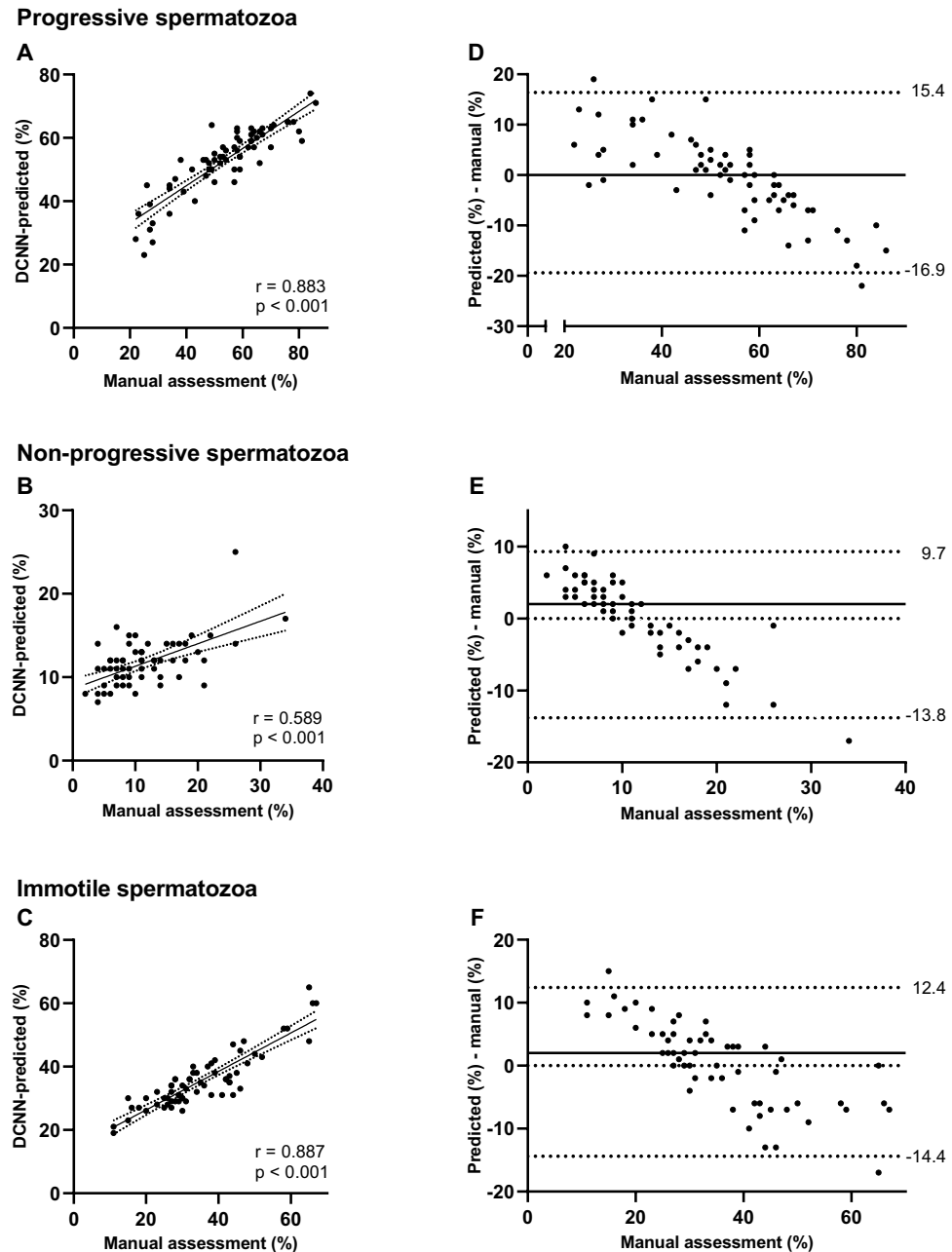
Both DCNN models showed high predictive power. For the three-category model, the MAE was 0.06 for progressive, 0.04 for non-progressive, and 0.05 for immotile spermatozoa, resulting in an average MAE of 0.05. The average three-category ZeroR baseline was 0.09. For the four-category model, the MAE was 0.11 for rapid



**Figure 2.** Chart of workflow. (1) Semen samples from 65 individuals were placed on slides, and sperm motility was recorded by a microscope-mounted video camera. (2) Videos were distributed to reference laboratories for manual sperm motility assessment according to WHO recommendations. (3) Videos ( $n = 65$ ) and mean values ( $n = 65$ ) from the manual sperm motility assessment were used to train and test the three-category and four-category DCNN models. Optical flow generated from batches of 30 frames collected across the entire length of each video, was used to develop the models. (4) DCNN-predicted values were compared to manual results. The figure was created by the authors with Biorender.com.

progressive, 0.09 for slow progressive, 0.04 for non-progressive, and 0.05 for immotile spermatozoa. The average MAE was 0.07, and the average four-category ZeroR baseline was 0.10.

We compared the mean values from the manual assessment with the DCNN-predicted for the three motility categories recommended in the WHO manual from 2010<sup>15</sup>; progressive, non-progressive, and immotile spermatozoa. The correlation between the mean values based on the reference laboratories and DCNN-predicted motility for the various categories is shown in Fig. 3. Strongest correlation was observed for % progressively motile spermatozoa (Pearson's  $r = 0.88$ ,  $p < 0.001$ ) (Fig. 3A) and % immotile spermatozoa ( $r = 0.89$ ,  $p < 0.001$ )



**Figure 3.** Comparison of manually and DCNN-predicted sperm motility in a three-category model. Correlation graphs (A–C) and difference plots (D,E) between sperm motility assessed manually by reference laboratories and motility values predicted for three categories by deep convolutional neural network (DCNN); (A) and (D) progressive spermatozoa, (B) and (E) non-progressive spermatozoa, (C) and (F) immotile spermatozoa. In the correlation graphs, solid lines represent the trendlines while dashed lines the 95% confidence interval. Pearson's correlation coefficient ( $r$ ) and  $p$ -value ( $p$ ) between results are also depicted. In the difference plots, the bias estimated as the median of the differences between predicted and manually assessed motility is shown as a solid line, while the 95% limits of agreement (LoA) are represented in the figure and shown in dashed lines to display outliers.

(Fig. 3C). At low proportion of progressively motile spermatozoa, the discrepancy between the assessments was most pronounced.

We also included a comparison of the results for rapid progressive motility, which is based on the assessment of four categories; rapid progressive, slow progressive, non-progressive, and immotile spermatozoa<sup>1,16</sup> (Fig. 4 and Supplementary Fig. S1). The correlation between the manually assessed and the DCNN-predicted rapid progressive motility was moderate (Pearson's  $r = 0.673$ ,  $p < 0.001$ ) (Fig. 4A).

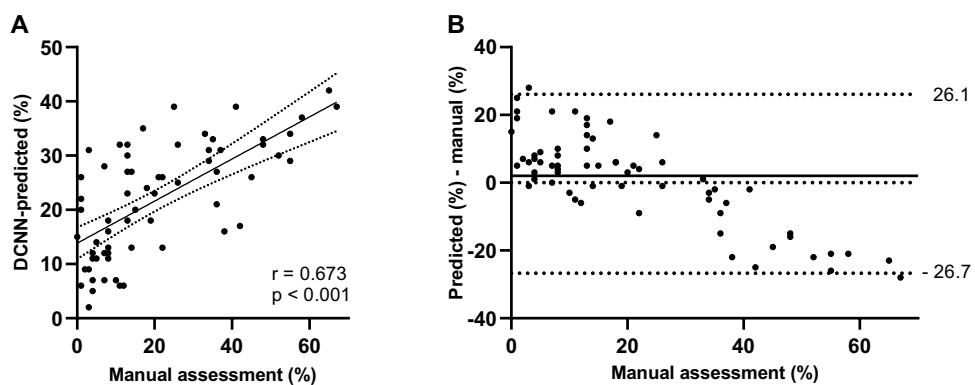
For the three-category model, the median difference (bias) when comparing manual against predicted progressive motility, non-progressive motility, and immotility was 0, 2, and 2, respectively (Fig. 3D–F). The 95% LoA was broader for the progressive category than for the immotile. Furthermore, the predicted results had a tendency to be overestimated in the lower percentage range and underestimated for high percentages, for all the motility categories. These trends were also observed for the four-category model (Supplementary Fig. S1).

Figure 5A–D depicts the manually assessed mean values of progressive, non-progressive, immotile, and rapid progressive spermatozoa for the individual values in increasing order, as well as the predicted values. The maximum and minimum values for the reference laboratories are also plotted in the figures, showing the discrepancy in the manual assessments. For most of the samples, the DCNN-predicted value was within the range of the interlaboratory variation of the results. The deviation in predicted value was largest at the lower proportion of rapid, non-progressive, and immotile spermatozoa, where an overestimation was observed.

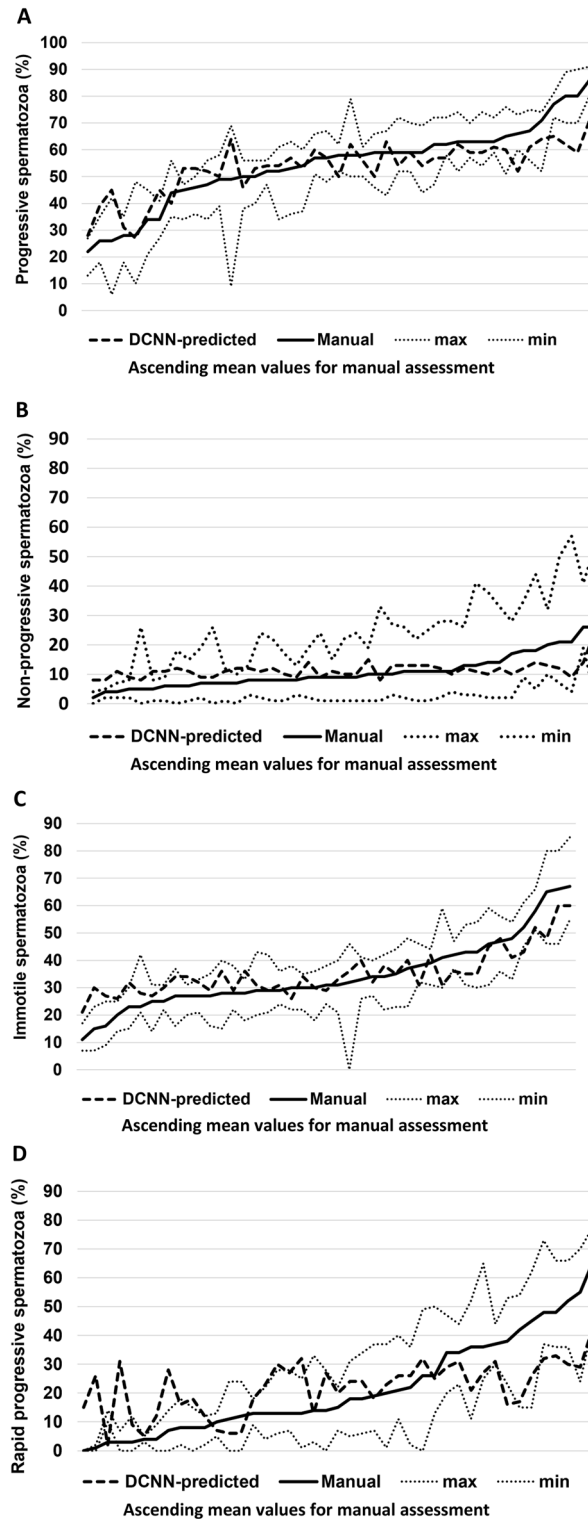
## Discussion

A more objective assessment of sperm motility has been one of the goals in the development of CASA systems, but this has been a challenge, especially with undiluted human semen samples<sup>1,5,6</sup>. Previous studies involving the use of AI for motility assessment have focused on comparing the performance of various models or exploring the sperm kinematics in prepared samples using for example traditional supervised machine learning<sup>9</sup> and in fresh semen samples applying a Region-Based Convolutional Neural Networks (R-CNN) architecture<sup>10</sup>. The current state-of-the-art method for image and video analysis is DCNNs, which can extract relevant features and utilize them for classification or segmentation. Thus, DCNNs have the potential to assess sperm motility from videos. We have previously demonstrated that deep learning models are able to predict sperm motility better than the classical machine learning algorithms in terms of mean MAE for each motility category in a three-category system<sup>14</sup>. Of the DCNN models used, the ResNet-50 model overall performed best. The manual assessment of sperm motility in that previous study was performed according to WHO recommendations, but by only one person. In the present work, the architecture from the former study was used to train both a three-category and a four-category model based on new data of higher quality. The DCNN models were trained by using the mean of the results from several well-trained personnel at different locations, leading to more reliable models. The earliest data in our study was from 2006, when the recommendations from WHO 1999 were followed<sup>16</sup>. The ESHRE-SIGA External Quality Assessment Programme continued with the four-category system for sperm motility assessment also after the WHO 2010 manual was published<sup>18</sup>. Categorization of progressive motility into rapid and slow was not a part of the recommendation of the WHO 2010 manual, mainly based on the argument that this distinction was difficult to obtain accurately<sup>15</sup>. However, the identification of rapid progressive spermatozoa has been found to be of clinical importance, and the four-category system is again recommended in the latest edition of the manual<sup>1,19–24</sup>.

### Rapid progressive spermatozoa



**Figure 4.** Comparison of manually and DCNN-predicted rapid progressive sperm motility. Correlation graph (A) and difference plot (B) between rapid progressive motility assessed manually by reference laboratories and motility values predicted for four categories by deep convolutional neural network (DCNN). In the correlation graph, solid line represents the trendlines while dashed lines the 95% confidence interval. Pearson's correlation coefficient ( $r$ ) and  $p$ -value ( $p$ ) between results are also depicted. In the difference plot the bias estimated as the median of the differences between predicted and manually assessed motility is shown as a solid line, while the 95% limits of agreement (LoA) are represented in the figure and shown in dashed lines to display outliers. For the results between all the four categories, manually and DCNN-predicted, see Supplementary Fig. S1.



**Figure 5.** Comparison of manually and DCNN-predicted sperm motility at increasing order of manual mean values for the individual samples. Line plots showing mean, maximum and minimum values ( $n = 43$ ) from manual assessment and DCNN-predicted values for rapid progressive (A), slow progressive (B), non-progressive motility (C), and immotility (D).

We found that DCNN worked well in predicting progressive motility and immotility. Among the motility categories, we found the lowest correlation between the manual and the DCNN-predicted values for slow

progressive motility. However, when comparing the manual and the predicted values for the individual samples (Fig. 5), it can be seen that there is considerable variation between the reference laboratories, especially for the rapid progressive motility. Thus, the challenge in obtaining accurate assessments for this category is also reflected in the DCNN-predicted results, which are within the range of variation between the reference laboratories for most samples. More data with high-quality and precise annotations will be needed to develop models that perform better. However, from a clinical point of view, the presence of rapid progressive spermatozoa ( $\geq 25$   $\mu\text{m/s}$ ) is important, although the exact velocities of the rapid spermatozoa are not determined. Determination of the proportion and the total number of rapid spermatozoa would therefore be of clinical interest.

The performance of the DCNN has two main directions for future improvements. Firstly, the quality in terms of visual details and resolution of the provided videos is mid to low level based on present technical standards. In the future, it is important to produce video datasets of higher quality, e.g., of high resolution like 4 k and no artifacts, in combination with assessments of the videos by well-trained staff. Such a dataset could also be used to develop more sophisticated methods. Secondly, for now, we are processing the videos in batches of 30 frames, disregarding any shifts in the video. In the future, a strategy would be to create ensemble networks that can recognize temporal and detailed information from the spermatozoa by providing, for example, segmentation masks that track individual sperm across the entire video. Multimodal input learning data could allow tackling the challenge with multi-instance learning. For example, one task could be to learn the number of spermatozoa and the second one the movement patterns, which in combination could lead to a better overall prediction. This would also be completely different from the current CASA.

The relatively small dataset is a limitation. To compensate for the size, ten-fold cross-validation was used. However, even though cross-validation is acceptable for testing model performance and comparing it to other models on the same dataset<sup>9,25</sup>, it does not test the generalizability of the results. An independent test set evaluation should be performed, preferentially across different clinics<sup>26</sup>.

Moreover, the dataset consists of videos of relatively homogenous samples. There are differences in sperm concentration and motility, but no samples had a high degree of debris, agglutinates, aggregates, or round cells. The consequence of this is that it is unknown how the model will perform with more heterogeneous samples, and more rigorous testing is needed in this regard. However, the results from our previous study indicated that the predictive power of the DCNN model was independent of concentration ranging from  $4 \times 10^6/\text{mL}$  to  $350 \times 10^6/\text{mL}$ <sup>14</sup>.

A strength of this study is that the same video clips were evaluated by the laboratory personnel and by the DCNN models. Furthermore, the data used for training the DCNN models was based on an assessment performed by several well-trained laboratory personnel. The interlaboratory variation between the reference laboratories, as shown in Fig. 5, reflects the challenge of obtaining reproducible results, even for highly trained personnel. In this respect, a DCNN model is promising for developing a more objective and reproducible method for motility assessment. This is important for the development of accurate and clinically useful techniques to evaluate the functions of the male reproductive organs and male infertility factors. An additional DCNN model could be trained in future work in tracking the spermatozoa to perform more fine-grained analysis (e.g., include distance moved and movement patterns). Although it is clear that AI will not be able to replace manual motility assessment in the near future, support from the AI model for manual evaluation can be beneficial. It is considerably less time-consuming and compared with CASA systems, might also be performed at a low cost. There would still be a need for a manual external and internal quality control programme, but these could be run more effectively with the support of automated systems. For example, one could do a baseline performance analysis using the AI model to quickly identify poor samples. Overall, this could encourage the establishment of a DCNN method in the clinics as the basis for improved diagnosis and management of male infertility, and for examining the semen sample of vasectomized men. A more objective and less resource-demanding tool will also be of value for performing multicentre studies. Finally, it is important to develop AI models for the evaluation of other sperm characteristics to obtain a more objective semen analysis.

## Conclusion

DCNN models performed well in predicting the proportion of spermatozoa in the WHO motility categories, especially for the progressive and immotile spermatozoa. For most samples, the DCNN-predicted value was within the range of the interlaboratory variation. For clinical purposes, a more reliable model for evaluating rapid progressive spermatozoa should be developed. DCNN models could provide a cost- and time-effective method for motility assessment, however, there is still a need for improvement. A prerequisite is larger manually assessed datasets of high quality for training the models. It is crucial to test the DCNN models in studies with data from different clinics and with varying degrees of quality of the samples.

## Data availability

The datasets generated during and/or analysed during the current study are available at <https://datasets.simula.no/visem-qc>.

Received: 29 June 2023; Accepted: 1 September 2023

Published online: 07 September 2023

## References

1. World Health Organization. *WHO laboratory manual for the examination and processing of human semen* 6th edn. (World Health Organization, 2021).
2. Bjorndahl, L. & Kirkman Brown, J. Other editorial board members of the WHO laboratory manual for the examination and processing of human semen. The sixth edition of the WHO laboratory manual for the examination and processing of human semen.

- Ensuring quality and standardization in basic examination of human ejaculates. *Fertil. Steril.* **117**, 246–251. <https://doi.org/10.1016/j.fertnstert.2021.12.012> (2022).
3. Cooper, T. G. *et al.* World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update* **16**, 231–245. <https://doi.org/10.1093/humupd/dmp048> (2010).
  4. Campbell, M. J. *et al.* Distribution of semen examination results 2020—A follow up of data collated for the WHO semen analysis manual 2010. *Andrology* **9**, 817–822. <https://doi.org/10.1111/andr.12983> (2021).
  5. Mortimer, S. T., van der Horst, G. & Mortimer, D. The future of computer-aided sperm analysis. *Asian J. Androl.* **17**, 545–553. <https://doi.org/10.4103/1008-682X.154312> (2015).
  6. Mortimer, D. & Mortimer, S. T. Routine application of CASA in human clinical andrology and ART laboratories. In *XIIIth International Symposium on Spermatology* (eds Björndahl, L. *et al.*) 183–197 (Springer International Publishing, 2021). [https://doi.org/10.1007/978-3-030-66292-9\\_26183-197](https://doi.org/10.1007/978-3-030-66292-9_26183-197).
  7. Chu, K. Y. *et al.* Artificial intelligence in reproductive urology. *Curr. Urol. Rep.* **20**, 52. <https://doi.org/10.1007/s11934-019-0914-4> (2019).
  8. Riegler, M. A. *et al.* Artificial intelligence in the fertility clinic: Status, pitfalls and possibilities. *Hum. Reprod.* **36**, 2429–2442. <https://doi.org/10.1093/humrep/deab168> (2021).
  9. Goodson, S. G. *et al.* CASAnova: A multiclass support vector machine model for the classification of human sperm motility patterns. *Biol. Reprod.* **97**, 698–708. <https://doi.org/10.1093/biolre/iox120> (2017).
  10. Valiuskaite, V., Raudonis, V., Maskeliunas, R., Damasevicius, R. & Krilavicius, T. Deep learning based evaluation of spermatozoid motility for artificial insemination. *Sensors (Basel)* **21**, 72. <https://doi.org/10.3390/s21010072> (2020).
  11. Somasundaram, D. & Nirmala, M. Faster region convolutional neural network and semen tracking algorithm for sperm analysis. *Comput. Methods Programs Biomed.* **200**, 105918. <https://doi.org/10.1016/j.cmpb.2020.105918> (2021).
  12. Alameri, M. *et al.* Multistage optimization using a modified gaussian mixture model in sperm motility tracking. *Comput. Math. Methods Med.* **2021**, 6953593. <https://doi.org/10.1155/2021/6953593> (2021).
  13. Ottl, S., Amiriparian, S., Gerczuk, M. & Schuller, B. W. motilitAI: A machine learning framework for automatic prediction of human sperm motility. *iScience* **25**, 104644. <https://doi.org/10.1016/j.isci.2022.104644> (2022).
  14. Hicks, S. A. *et al.* Machine learning-based analysis of sperm videos and participant data for male fertility prediction. *Sci. Rep.* **9**, 16770. <https://doi.org/10.1038/s41598-019-53217-y> (2019).
  15. World Health Organization. *WHO laboratory manual for the examination and processing of human semen* 5th edn. (World Health Organization, 2010).
  16. World Health Organization. *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interactions* 4th edn. (Cambridge University Press, 1999).
  17. <https://datasets.simula.no/visem-qc>.
  18. Barratt, C. L., Björndahl, L., Menkveld, R. & Mortimer, D. ESHRE special interest group for andrology basic semen analysis course: A continued focus on accuracy, quality, efficiency and clinical relevance. *Hum. Reprod.* **26**, 3207–3212. <https://doi.org/10.1093/humrep/der312> (2011).
  19. Aitken, R. J., Sutton, M., Warner, P. & Richardson, D. W. Relationship between the movement characteristics of human spermatozoa and their ability to penetrate cervical mucus and zona-free hamster oocytes. *J. Reprod. Fertil.* **73**, 441–449. <https://doi.org/10.1530/jrf.0.0730441> (1985).
  20. Mortimer, D., Pandya, I. J. & Sawers, R. S. Relationship between human sperm motility characteristics and sperm penetration into human cervical mucus in vitro. *J. Reprod. Fertil.* **78**, 93–102. <https://doi.org/10.1530/jrf.0.0780093> (1986).
  21. Comhaire, F. H., Vermeulen, L., Hinting, A. & Schoonjans, F. Accuracy of sperm characteristics in predicting the in vitro fertilizing capacity of semen. *J. In Vitro Fert. Embryo Transf.* **5**, 326–331. <https://doi.org/10.1007/bf01129567> (1988).
  22. Barratt, C. L., McLeod, I. D., Dunphy, B. C. & Cooke, I. D. Prognostic value of two putative sperm function tests: Hypo-osmotic swelling and bovine sperm mucus penetration test (Penetrak). *Hum. Reprod.* **7**, 1240–1244. <https://doi.org/10.1093/oxfordjournals.humrep.a137834> (1992).
  23. Bollendorf, A., Check, J. H. & Lurie, D. Evaluation of the effect of the absence of sperm with rapid and linear progressive motility on subsequent pregnancy rates following intrauterine insemination or in vitro fertilization. *J. Androl.* **17**, 550–557 (1996).
  24. Sifer, C. *et al.* World Health Organization grade “a” motility and zona-binding test accurately predict IVF outcome for mild male factor and unexplained infertilities. *Hum. Reprod.* **20**, 2769–2775. <https://doi.org/10.1093/humrep/dei118> (2005).
  25. Shaker, F., Monadjemi, S. A., Alirezaie, J. & Naghsh-Nilchi, A. R. A dictionary learning approach for human sperm heads classification. *Comput. Biol. Med.* **91**, 181–190. <https://doi.org/10.1016/j.compbiomed.2017.10.009> (2017).
  26. Abbasi, A., Miah, E. & Mirroshandel, S. A. Effect of deep transfer and multi-task learning on sperm abnormality detection. *Comput. Biol. Med.* **128**, 104121. <https://doi.org/10.1016/j.compbiomed.2020.104121> (2021).

## Acknowledgements

The authors acknowledge the ESHRE-SIGA External Quality Assessment Programme for Semen Analysis for the data and are grateful to Heidi Eilertsen for support with the difference plots.

## Author contributions

Conception of the study was initiated by T.B.H., O.W., J.M.A. and M.A.R. All authors contributed to the final study design and planning of analyses. L.B. contributed to the data collection. M.A.R. and S.A.H. developed and trained the DCNN models. O.W. and J.M.A. performed the statistical analysis and visualized the data. T.B.H. drafted the manuscript. All authors contributed to interpretation of the data, critical revision of content, and approval of the final manuscript. All authors agreed to be personally accountable for own contributions and ensure that questions related to the accuracy or integrity of any part of the work, are appropriately investigated and resolved.

## Funding

The work was funded by the Norwegian Research Council, project number 288727.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41871-2>.



**Correspondence** and requests for materials should be addressed to T.B.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023