Original Article

Deep Learning-Based Robust Automated System for Predicting Human Sperm DNA Fragmentation Index

Roopini Sathiasai Kumar, Swapnil Sharma[1], Arunima Halder, Vipin Gupta[1]

Manipal Hospitals Pvt. Ltd, Bengaluru, Karnataka, [1]Linkedin Technology Information Pvt. Ltd., Mumbai, Maharashtra, India

Abstract

Background: Determining the DNA fragmentation index (DFI) by the sperm chromatin dispersion (SCD) test involves manual counting of stained sperms with halo and no halo. **Aims:** The aim of this study is to build a robust artificial intelligence-based solution to predict the DFI. **Settings and Design:** This is a retrospective experimental study conducted in a secondary *in vitro* fertilisation setup. **Materials and Methods:** We obtained 24,415 images from 30 patients after the SCD test using a phase-contrast microscope. We classified the dataset into two, binary (halo/no halo) and multiclass (big/medium/small halo/degraded (DEG)/dust). Our approach consists of a training and prediction phase. The 30 patients' images were divided into training (24) and prediction (6) sets. A pre-processing method *M* was developed to automatically segment the images to detect sperm-like regions and was annotated by three embryologists. **Statistical Analysis Used:** To interpret the findings, the precision-recall curve and F1 score were utilised. **Results:** Binary and multiclass datasets containing 8887 and 15,528 cropped sperm image regions showed an accuracy of 80.15% versus 75.25%. A precision-recall curve was determined and the binary and multiclass datasets obtained an F1 score of 0.81 versus 0.72. A confusion matrix was applied for predicted and actuals for the multiclass approach where small halo and medium halo confusion were found to be highest. **Conclusion:** Our proposed machine learning model can standardise and aid in arriving at accurate results without using expensive software. It provides accurate information about healthy and DEG sperms in a given sample, thereby attaining better clinical outcomes. The binary approach performed better with our model than the multiclass approach. However, the multiclass approach can highlight the distribution of fragmented and non-fragmented sperms.

Keywords: *Clinical outcome, DNA fragmentation index, machine learning*

Introduction

Over the past 10 years, the percentage of infertile couples with a male component has grown to almost 40%. Unexplained infertility is detected in about 30% of these couples. Semen analysis is the primary diagnostic test to evaluate the male partner. However, most often it is not sufficient to determine the reason for infertility in case of unexplained infertility.[1] In a study conducted by Oleszczuk *et al*., it is stated that in conventional diagnostic techniques, a sizable portion of males identified as having unexplained infertility have extraordinarily high levels of fragmented sperm DNA.[2] High DNA fragmentation index (DFI) is also associated with increased miscarriage and lower live birth rate.[3,4] In such cases, performing sperm function tests specifically, DNA fragmentation can aid in understanding the reason for infertility in the couple.

Received: 05-01-2023 **Revised:** 14-03-2023
Accepted: 15-03-2023 **Published:** ***

Access this article online

Quick Response Code:

Website: www.jhrsonline.org

DOI: 10.4103/jhrs.jhrs_4_23

Address for correspondence: Ms. Roopini Sathiasai Kumar, Manipal Hospitals, ITPL Main Rd, KIADB Export Promotion Industrial Area, Krishnarajapura, Bengaluru - 560 066, Karnataka, India.
E-mail: roopinissk@gmail.com

How to cite this article: Kumar RS, Sharma S, Halder A, Gupta V. Deep learning-based robust automated system for predicting human sperm DNA fragmentation index. J Hum Reprod Sci 2023;XX:XX-XX.

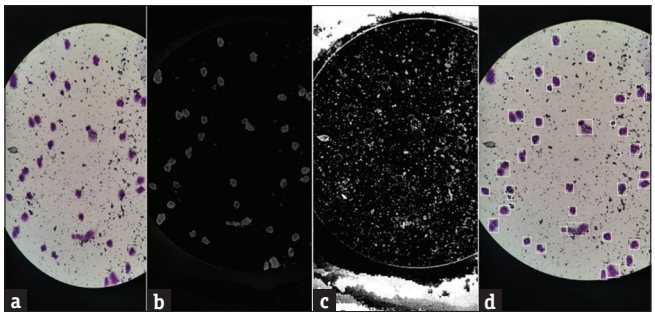

Figure 1: Image pre-processing. Full image of a single field as seen under the microscope. (a) Raw image of field seen under the microscope, (b) Masking applied, (c) After removing the mask the field is exposed showing dust and other artifacts in the image, (d) Segmentation applied to separate the sperm and non-sperms cells

Assessing the gametes and embryos is solely based on morphological evaluation and results show high inter- and intrapersonnel variability which also lacks standardisation. Computer-assisted/aided semen analysis (CASA) offers testing with reduced variability, however, is expensive and limits its use in clinics. The automated interpretation of semen parameters produced by computed-aided systems such as CASA has shown to decrease operator subjectivity when compared to manual testing.[5] Although the precision of results is being maintained, according to Talarczyk-Desole *et al*., the technician's personal interpretation could skew the computer-assisted study. The technicians need not be experts in semen analysis, however, basic knowledge of seminology is required and to get accurate findings, it is crucial to strictly follow CASA's producer settings and to evaluate human sperm in Leja chambers.[6]

The aim and objective of the current investigation are to build a robust artificial intelligence (AI)-based solution to predict the DFI of samples subjected to sperm chromatin dispersion (SCD) test by leveraging human-annotated stained images to produce unbiased results inexpensively.

Determination of sperm DNA fragmentation can be done using multiple tests. The most popular ones include Sperm Chromatin Structure Assay (SCSA), terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL), COMET and SCD. The SCSA analyses the sperm DNA's sensitivity to denaturation when subjected to heat or acids. TUNEL assay uses fluorescent nucleotides to identify DNA 'nicks' or free ends. The comet assay employs neutral and alkaline electrophoresis to assess the many patterns of DNA damage per cell. SCSA and TUNEL require a flow cytometry and/or fluorescence microscope which reduces their appeal to clinical andrology laboratories due to the expense of the necessary equipment. COMET assay is not appropriate for rapid diagnosis and requires

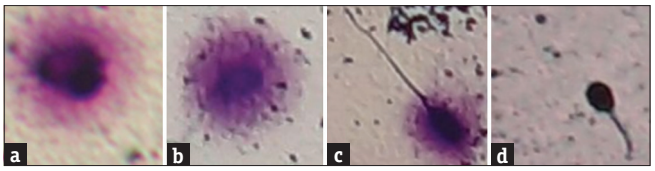

Figure 2: Images of sperms after segmentation. (a) BH, (b) MH, (c) SH, (d) DEG. BH = Big halo, MH = Medium halo, SH = Small halo, DEG = Degraded

highly specialised staff to interpret the results. Based on the idea that fragmented sperm do not create the distinctive halo of dispersed DNA loops that are seen in sperm with non-fragmented DNA after acid denaturation and removal of nuclear proteins, the SCD, also known as the halo test, is used to identify sperm. The method is straightforward and does not call any sophisticated equipment. Although a competent assay for DFI quantification, the classification of the halos may have some interobserver subjectivity.[7,8]

McCallum *et al*. developed a deep-learning model to predict the DFI of sperm based on sperm morphology. The DNA fragmentation test was performed by SCSA and acridine orange (AO) and results were obtained. The images of sperms that underwent DNA fragmentation test were mapped with the morphology of the same cohort of sperms subjected to the test. These images were then fed into the deep-learning model. The study, however, was conducted only on 6 donors and 1064 images were extracted which offers very little scope for accuracy when compared to a large dataset.[9] A recent study comparing sperm morphology with DFI found that 8.23% of sperm in their dataset with normal morphology were apoptotic.[10] Avendano *et al*. have reported that in infertile men, spermatozoon with apparently normal morphology may have DNA fragmentation, and the presence of an increased proportion of normal spermatozoon with damaged DNA was negatively associated with embryo quality and pregnancy outcome after intracytoplasmic sperm injection (ICSI). Therefore, morphology cannot be solely depended upon for testing the DFI of sperms.[11]

Our research is an experimental attempt to create a machine learning (ML) model that can calculate the DFI of sperms that have undergone an SCD test to produce more accurate findings without the use of costly equipment.

Subjects and Methods

This is an experimental retrospective study conducted in a secondary *in vitro* fertilisation (IVF) setup. The patient data used in the study maintain anonymity, the institution's ethics committee has granted ethical

clearance and waived off consent as it involves retrospective data. Thirty patients' samples were taken for this study. Inclusion criteria were patients who underwent the SCD test for determining the cause of infertility. No patients were recruited specifically for the purpose of this study.

We obtained 24,415 images from 30 patients after the SCD test using a phase-contrast microscope. Different approaches were explored to classify sperm appearance into binary (halo/no halo) and multiclass (big/medium/small halo/degraded (DEG)/dust) problems. Our approach consists of (1) training and (2) the prediction phase.

The SCD test was performed using the sperm chroma kit (Cryotec, SAR HEALTHLINE). The sperms were classified into four types based on the dispersion of stain, referred to as halo, big halo (BH), medium halo (MH), small halo (SH) and DEG. Sperm with big and MH were considered to have intact DNA, sperms with SH and those that were DEG were considered to have poor DNA integrity. The sperms were counted as per the manufacturer's instruction manual.

Image pre-processing included hue separation and morphological operation which were used to reduce the noise from the images. Sperm cells take up the purple stain when subjected to the test and therefore, were isolated by hue and they were segmented according to their respective dimensions [Figure 1]. The sperm cells were separated from dust based on their size. The sperm cells were then separately cropped with the help of a segmentation technique known as connected-component analysis [Figure 2]. These images of single sperms were then augmented using rotation, saturation and Gaussian blur/noise. These images were then annotated by three experienced embryologists.

The dataset was classified into two:

1. Multiclass: Having distinct classifications between the sperms, big, medium small halo and DEG
2. Binary: Having two classes: Fragmented (Small-Deg (SD): SH and DEG) and unfragmented (Big-Medium (BM): BH and MH) sperm DNA.

For training both the multiclass and binary approaches, we took 110 full images for training and 33 full images for testing. The images were randomly assigned to the training and testing groups to avoid investigator bias and to test the efficiency of the model. In this study, an image classifier was used to train a unique classifier to differentiate between various sperm cell types by utilising custom vision. Until the loss was as small as possible and the value plateaued, this process was repeated thousands of times. Azure's custom vision is trained on various models including deep network, clustering algorithms and selected best-trained model for testing. Custom vision is part of the family of cognitive services. It enables to quickly customise state-of-the-art computer vision models for a particular scenario, with a small set of labelled images, and image processing algorithms that provide image classification, captioning, optical character recognition and content moderation. Computer vision application programming interface (API) provides information about the objects that are found in an image. The APIs return the tags that are most pertinent to the image and a caption that describes it after the image has been evaluated. The APIs identify the text in the image and deliver the recognised characters as a JSON payload. Custom vision trains a specific model for a particular scenario using transfer learning and data augmentation methods.[12] In our study, the training set was rerandomised, and training was repeated once all batches in the training dataset had been evaluated (i.e. after 1 epoch), covering the full training set. To avoid overtraining, the model was run on a predefined subset of photos put aside for use in the training process. The overall precision and recall measures are provided following training. In addition, each tag's (i.e. label's or class) performance was displayed.

Several epochs of the train–validate cycle were completed before a sufficiently stable model with a low loss function was produced. The best models were integrated into a final ensemble model at the end of the series of train–validate cycles. The same dataset was used to train a variety of models, including convolutional neural networks, clustering algorithms and random forests [Figure 4]. The best model was chosen based on the parameters of precision, recall and average precision. Models received training for about 8 h. For both models, the probability threshold employed in the results was set at 50%.

Results

110 and 33 full images were taken for training and testing. A total of 24,415 images from 30 patients were obtained after a series of image pre-processing methods and augmentation. The multiclass dataset included 15,528 images (BH - 2763, MH - 3265, SH - 3254, DEG - 3208 and dust - 3038). To prevent the model from overtraining, the images were augmented to avoid having an unequal number of images in each. Multiclass dataset showed an accuracy of 75.25% [Figure 5]; an F1 score of 0.72. A confusion matrix [Figure 6] was applied to the multiclass dataset with class accuracy measures. The confusion matrix assessed true positives,

false negatives, false positives and true negatives as the basis for model classification and misinterpretation. From Figure 5, it is observed that maximum confusion occurred between classifying SH and MH. The binary class data set included 8887 images (BM - 2814, SD - 2871 and dust - 3202). Dust was included only for training and not for testing, therefore, testing was done only on two variables. The binary class dataset showed an accuracy of 80.15%; an F1 score of 0.80.

Discussion

Different AI algorithms are being developed to identify the various sperm parameters as ML in sperm parameter analysis is attracting increasing attention. The purpose of our study is to determine sperm DNA fragmentation using custom vision employing deep learning algorithms. The absence of standardisation is the primary problem with determining results manually or using automated machines. However, ML models can lessen this variability and deliver more precise outcomes. Our deep learning model, which categorises sperms, was trained using two different ways. Knowing only the DFI is insufficient, therefore, it is also vital to understand how many sperms in the sample have different degrees of DNA fragmentation which our study provides. So far, ML models have been developed for determining DFI preliminarily using tests such as SCSA, AO, single sperm DFI assay as well as SCD.[5,9,13,14]

The outcome of our investigation contrasts the manual counting of DFI calculations using our deep-learning model. To calculate the DFI of sperms subjected to single-cell sperm DNA fragmentation experiment, Wang *et al*. developed an ML model. Although the sample size was limited (1056 sperm from six donors), their test had an accuracy of 0.827. However, without making any apparent distinctions between the physical characteristics, they divided the sample set into excellent and bad sperm. The small sample size and unclear distinction between good and bad sperms leave this study to be implored further.[13] In a recent study conducted by Yang *et al*., the rates of pregnancy, early abortion, oocyte fertilisation and high-quality embryos from IVF and ICSI cycles were compared between the low DFI (DFI 15%) and medium DFI (15 DFI groups. In the high, medium and low sperm, DFI groups after IUI, the clinical pregnancy rates were not statistically significant. However, these groups had statistically significant early abortion rates [P = 0.02]). They also found that sperm DFI was negatively associated with sperm density, vitality and normal morphology; It was positively correlated with age, abstinence time and unhealthy lifestyles.[15] Therefore, this does not support the study conducted by McCallum *et al*. which determines the DFI of sperms using brightfield images of the raw sample.[9]

In our study, the multiclass dataset's confusion matrix is shown in Figure 3. With respect to a recent study which claims to be the first to assess the intraclass agreement amongst testers of the SCD assay to evaluate DFI in men shows >80% agreement with normal and high DFI categories, 60% intermediate DFI categories.[16] Our model shows agreement of 90% of normal sperms and 81% of DEG sperms, 65 and 66% of intermediate DFI, seen in the confusion matrix. The setback of the study is that only 400 sperms were assessed from each recipient. Meanwhile, our model can assess million sperms and maintain the same accuracy.

On the other hand, our model misidentifies SH as MH 31% of the time and MH as BH 29% of the time, reducing model accuracy. Although the multiclass data performed less well than expected when compared to the binary class dataset, we still regard it as a superior classification approach because it makes a clear distinction between the degree of sperm fragmentation, which can help us understand the DNA damage that is present in sperms.

The drawbacks of our study are that it did not assess the model's accuracy using expert calculations and that it does not offer any information on the developmental status of the embryo that leads to a miscarriage or clinical pregnancy. As of now, there are no imaging or testing techniques to identify the sperm DNA fragmentation of live samples which will help in the selection of sperms with low DNA fragmentation during IVF or ICSI. The stained sperm sample used for the SCD test cannot be used for egg insemination. Our model only provides results for the samples that were tested and cannot be used to identify live sperm cells with minimal DNA fragmentation.

DFI plays a vital role in predicting clinical pregnancy. Our AI model offers insights to determine the best

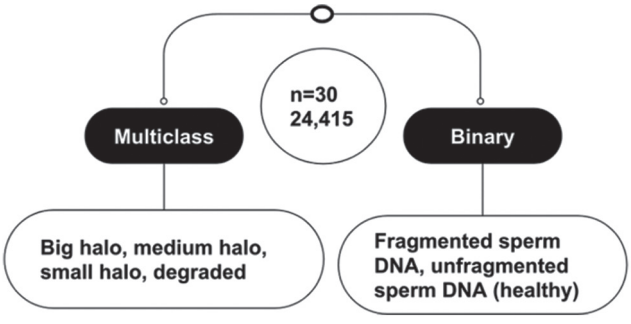

Figure 3: Classification of the image dataset

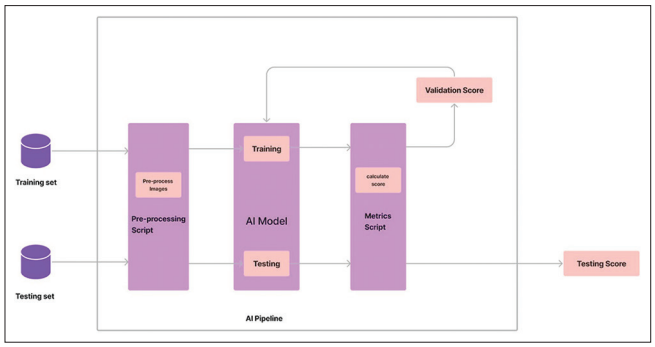

Figure 4: AI pipeline. AI = Artificial intelligence

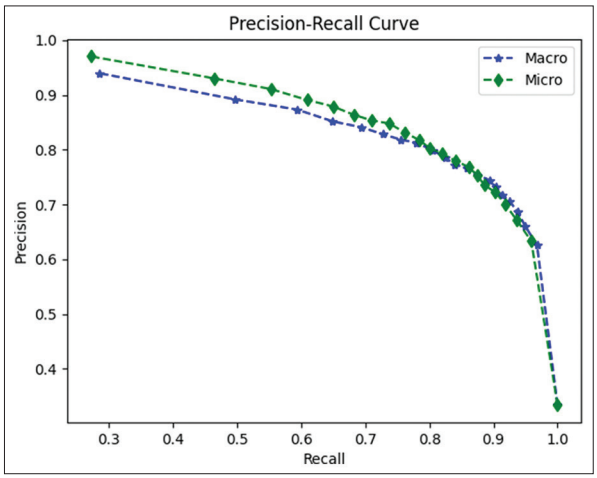

Figure 5: Plot of micro and macro precision-recall values

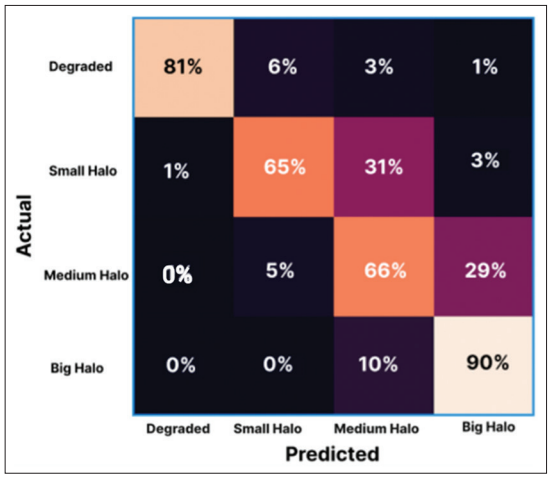

Figure 6: Confusion matrix applied to the multiclass dataset

sperm selection process by giving accurate information of healthy and DEG sperms in a given sample, thereby attaining better clinical outcomes. A confusion matrix clarified that the binary experimental data set performed better with our model. However, the multiclass approach can highlight the distribution of fragmented and non-fragmented sperms. Our proposed ML model can aid in arriving at accurate results without the use of expensive software and reduce intra- and interobserver variability.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Data availability

The information will be accessible through the corresponding author and can be provided upon request.

References

1. Zandieh Z, Vatannejad A, Doosti M, Zabihzadeh S, Haddadi M, Bajelan L, *et al.* Comparing reactive oxygen species and DNA fragmentation in semen samples of unexplained infertile and healthy fertile men. Ir J Med Sci 2018;187:657-62.
2. Oleszczuk K, Augustinsson L, Bayat N, Giwercman A, Bungum M. Prevalence of high DNA fragmentation index in male partners of unexplained infertile couples. Andrology 2013;1:357-60.
3. Leach M, Aitken RJ, Sacks G. Sperm DNA fragmentation abnormalities in men from couples with a history of recurrent miscarriage. Aust N Z J Obstet Gynaecol 2015;55:379-83.
4. Malić Vončina S, Stenqvist A, Bungum M, Schyman T, Giwercman A. Sperm DNA fragmentation index and cumulative live birth rate in a cohort of 2,713 couples undergoing assisted reproduction treatment. Fertil Steril 2021;116:1483-90.
5. Dearing C, Jayasena C, Lindsay K. Can the sperm class analyser (SCA) CASA-Mot system for human sperm motility analysis reduce imprecision and operator subjectivity and improve semen analysis? Hum Fertil (Camb) 2021;24:208-18.
6. Talarczyk-Desole J, Berger A, Taszarek-Hauke G, Hauke J, Pawelczyk L, Jedrzejczak P. Manual versus. Computer-assisted sperm analysis: Can CASA replace manual assessment of human semen in clinical practice? Ginekol Pol 2017;88:56-60.
7. Agarwal A, Majzoub A, Esteves SC, Ko E, Ramasamy R, Zini A. Clinical utility of sperm DNA fragmentation testing: Practice recommendations based on clinical scenarios. Transl Androl Urol 2016;5:935-50.
8. Evenson DP. The sperm chromatin structure assay (SCSA (®)) and other sperm DNA fragmentation tests for evaluation of sperm nuclear DNA integrity as related to fertility. Anim Reprod Sci 2016;169:56-75.
9. McCallum C, Riordon J, Wang Y, Kong T, You JB, Sanner S, *et al.* Deep learning-based selection of human sperm with high DNA integrity. Commun Biol 2019;2:250.
10. Ferrigno A, Ruvolo G, Capra G, Serra N, Bosco L. Correlation between the DNA fragmentation index (DFI) and sperm morphology of infertile patients. J Assist Reprod Genet 2021;38:979-86.
11. Avendañ C, Oehninger S. DNA fragmentation in morphologically normal spermatozoa: How much should we be concerned in the ICSI era? In Journal of Andrology 2011;32:356-63. https://doi.org/10.2164/jandrol.110.012005.
12. Salvaris M, Dean D, Tok WH. Cognitive services and custom vision. In: Deep Learning with Azure. Apress; 2018. p. 99-128.
13. Wang Y, Riordon J, Kong T, Xu Y, Nguyen B, Zhong J, *et al.* Prediction of DNA integrity from morphological parameters using a single-sperm DNA fragmentation index assay. Adv

Sci (Weinh) 2019;6:1900712.
14. Mendizabal-Ruiz G, Chavez-Badiola A, Aguilar Figueroa I, Martinez Nuño V, Flores-Saiffe Farias A, Valencia-Murilloa R, *et al.* Computer software (SiD) assisted real-time single sperm selection associated with fertilization and blastocyst formation. Reprod Biomed Online 2022;45:703-11.
15. Yang H, Li G, Jin H, Guo Y, Sun Y. The effect of sperm DNA fragmentation index on assisted reproductive technology outcomes and its relationship with semen parameters and lifestyle. Transl Androl Urol 2019;8:356-65.
16. Esteves SC, López-Fernández C, Martínez MG, Silva EA, Gosálvez J. Reliability of the sperm chromatin dispersion assay to evaluate sperm deoxyribonucleic acid damage in men with infertility. Fertil Steril 2022;117:64-73.