Explainable Deep Neural Architectures for Morphological Classification of Human Sperm Cells in Automating Reproductive Health Applications

Rincy T.A., PhD
Associate Professor
Prajyoti Niketan College, Pudukad,
Thrissur, Kerala, India

ABSTRACT
Accurate morphological classification of human sperm cells is critical in diagnosing male infertility and automating assisted reproductive technologies (ART). Deep learning-based solutions have demonstrated remarkable performance in this domain. However, the opaque nature of these models poses challenges in clinical acceptance. This study presents an Explainable AI (XAI) framework for the analysis of sperm morphology in reproductive health, using CNN architectures to classify and interpret morphological features of sperm cells. The study employs Grad-CAM (Gradient-weighted Class Activation Mapping) to provide visual explanations for the model's decision-making process, enabling enhanced interpretability of deep learning models in the biomedical domain. The framework focuses on key sperm morphological components, including the head, vacuole, tail, and acrosome, assessing the performance of VGG16, ResNet34 and DenseNet-121 across these categories. Through a comparative evaluation, this study demonstrates the performance of various CNN architectures and the effectiveness of Grad-CAM in highlighting important regions within sperm images, thus providing transparency into the classification process and ensuring model trustworthiness.

General Terms
Deep Learning, Explainable AI.

Keywords
Male Infertility, Human Sperm Morphology.

1. INTRODUCTION
Sperm morphology analysis represents a critical diagnostic factor in male infertility. Infertility is a prevalent global health issue affecting millions of individuals and exerting significant impacts on families and communities. Epidemiological data suggest that nearly one-sixth of the reproductive-age population experiences infertility over their lifetime. Male infertility is predominantly associated with abnormalities in semen ejaculation, sperm concentration, morphology, and motility. Disruptions in sperm shape or movement are particularly detrimental to male fertility outcomes.

The majority of countries still have issues with the availability, quality, and accessibility of infertility treatments [1]. Public health rarely funds the diagnosis and treatment of infertility, and national population, development policies and reproductive health strategies usually do not prioritize this issue. Furthermore, major challenges include a lack of qualified personnel, the necessary infrastructure and equipment, and the high cost of treatment drugs, even for countries that are actively addressing the needs of people with infertility. Manual assessment of male infertility is inherently subjective, time-consuming, and prone to differences between evaluators. Although the relationship between sperm morphology and the outcomes of Assisted Reproductive Technologies (ART) has been extensively researched, changes in how sperm morphology is assessed and ongoing advancements in ART methods have made direct comparisons between studies from different periods challenging. Although in vitro fertilization (IVF) and other assisted reproductive technologies (ART) have helped create millions of children worldwide for over thirty years, they remain largely unavailable and hard to access. They are also very expensive in many parts of the world, especially in low and middle-income countries (LMIC) [2].

Since the first WHO manual in 1980, sperm morphology criteria have evolved. In the 1st and 2nd editions, sperm displaying at least one distinct abnormality, as defined by Macleod and Gold, were classified as morphologically abnormal. [3]. Previous studies by Menkveld et al. have defined the criteria for normal sperm morphology [4]. The sperm cell is a microscope and flagellated motile cell with a unique structure that is made up of a head, a mid-piece and a tail. A mature spermatozoa is covered by a larger plasma membrane, measures approximately 60-65 mm in length. A normal spermatozoon has a head with a smooth, regular, oval contour in which the acrosomal region occupies 40%–70% of the total head area, with no large vacuoles and no more than two small vacuoles. Abnormalities of the head include an acrosome that is either smaller than 40% or larger than 70% of the head area, a length-to-width ratio less than 1.5 or greater than 2, amorphous or asymmetrical shapes, non-oval forms, double heads, or vacuoles occupying more than one-fifth of the head area. The midpiece of a normal sperm is slender, about the same length as the head, and aligned along the same axis, whereas abnormalities include irregular shape or thickness, asymmetry, angulation, or bending. The tail is normally of uniform caliber, about ten times the length of the head, and free of sharp angulations, while abnormalities may present as sharp bends, hairpin turns, coiling, shortened length and width, or multiple tails. Cytoplasmic droplets are acceptable only if smaller than one-third of the head size, while residual cytoplasm exceeding this proportion is considered abnormal [5].

With the rise of computer-aided diagnostic systems, deep learning has emerged as a powerful tool in automating sperm morphology analysis. Despite their high accuracy, most deep models operate as "black boxes", limiting their applicability in clinical environments that demand interpretability. CNNs help detect subtle defects and classify sperm cells efficiently, accurately, and automatically — which is crucial in fertility research, sperm banking, and medical diagnosis. Convolutional neural network (CNN) is a neural network that has

outperformed computer vision problems [6]. Convolutional neural networks (CNNs) have been applied since the late 1980s. The first multilayered CNN, ConvNet, introduced by LeCun et al., employed supervised training with backpropagation, establishing the foundation for modern 2D CNNs [7]. ConvNet achieved notable success in handwritten digit and ZIP code recognition. Its refinement as LeNet-5 in 1989 enabled character classification in document recognition [8], and by the early 1990s, CNNs had proven highly effective in fingerprint recognition. Subsequent research over the past decades has further advanced CNN architectures and applications.

Since the early 2010s, CNNs—including pre-trained architectures like VGG16, ResNet, and DenseNet—have been used extensively for analyzing sperm morphology. Around 2020, in particular, researchers began using these deep learning models to automate and standardize the analysis of semen. Studies have demonstrated that retraining these models on sperm datasets can result in high accuracy, matching or surpassing human experts in the classification of sperm morphology. These models are effective because they can learn intricate features from images. CNNs are essential for increasing the accuracy and efficiency of automated sperm analysis, as evidenced by the history's progression from traditional image processing to ever-more-advanced deep learning methods.

This study investigates the morphological classification of human spermatozoa through a comparative analysis of established explainable deep neural network architectures applied to sperm morphology classification. I employ Grad-CAM to visualise the decision-making process, thus enabling transparent evaluation of model predictions. This research aims to bridge the gap between high-performance automation and explainability in reproductive health applications.

2. PREVIOUS WORK

The use of deep learning techniques has greatly advanced research into the automatic selection and classification of sperm cells. Below is a curated list of notable studies, detailing their methodologies and findings:

The paper by Violeta Chang et al. (2017), titled ″Automatic classification of human sperm head morphology," [9] presents a computational approach for classifying sperm head shapes using image analysis techniques. The authors developed and evaluated machine learning models that automatically classify sperm heads into different morphological categories based on WHO guidelines. Their method uses a set of shape descriptors and texture features extracted from microscopic images, aiming to reduce subjectivity and improve consistency in sperm morphology assessment. The study demonstrated promising accuracy in classifying normal and abnormal sperm head shapes, showing potential for enhancing computer-aided diagnosis systems in male fertility evaluations

The researchers in [10] developed a public gold-standard dataset of human sperm head images designed to provide a benchmark for evaluating and comparing morphological classification techniques. Testing revealed significant inter-expert variability and showed that current standard descriptors and learning methods are inadequate, achieving a peak accuracy of only 49%. Consequently, the study highlights an urgent need for specialized shape-based descriptors and classification models to better identify the high variability found in abnormal sperm cells.

Deep learning methods offer the ability to automatically learn both discriminative features and accurate classification rules, making them an effective solution to the limitations of traditional approaches. Among these methods, convolutional neural networks (CNNs) have shown strong performance in image classification tasks [11]. The application of deep learning techniques to sperm morphology analysis is relatively recent. Initial studies primarily addressed binary classification tasks, distinguishing normal sperm from abnormal samples [12, 13, 14]. In 2019, one of the earliest implementations of a pre-trained deep learning model was introduced for sperm health assessment. In this work, researchers combined a smartphone-based microscope with a deep convolutional neural network to classify sperm images as normal or abnormal. They demonstrated that this integrated system could potentially enable at-home evaluation of human sperm quality [13].

In 2019, Javadi and Mirroshandel [15] introduced a deep learning–based framework for automated human sperm morphology assessment, aiming to overcome the subjectivity, time demands, and error susceptibility of manual evaluation. They also released a publicly available dataset, termed the Modified Human Sperm Morphology Analysis (MHSMA). Their approach combined an effective preprocessing strategy, incorporating data augmentation and class imbalance mitigation, with the training of a convolutional neural network (CNN) to perform morphology classification.

The study titled "Deep learning for the classification of human sperm" [16] presents a method using a retrained VGG16 CNN model to classify human sperm heads into WHO-defined morphological categories. Trained on HuSHeM and SCIAN datasets, the model achieved higher true positive rates than conventional methods like CE-SVM and comparable results to APDL, while avoiding heavy computational costs. It demonstrates the potential of transfer learning for standardizing and automating sperm morphology analysis.

The paper by Ilhan, Yuzkat, and Aydin [17] introduces a low-cost, smartphone-based system for sperm motility analysis using recursive Kalman filters. Users capture sperm videos via a smartphone microscope, which are uploaded to a server for automated processing. The system stabilizes video, tracks sperm using Kalman filtering, and classifies motility into four categories based on kinematic features. Tested on data from six subjects, the system showed results comparable to a commercial CASA device, with minimal differences. This approach offers an accessible, portable, and reliable alternative to traditional sperm analysis methods.

In the 2020 study by Ilhan, Serbes, and Aydin [18], titled ″Automated sperm morphology analysis approach using a directional masking technique," the authors present a novel method for evaluating sperm morphology using computer vision techniques. The core of their approach lies in the use of a directional masking technique**,** which enhances the accuracy of morphological classification by focusing on specific directional features of sperm cell images. The study demonstrates that this technique significantly improves the performance of automated sperm analysis systems, offering a reliable and efficient alternative to manual assessment, which is often time-consuming and subject to inter-observer variability. Overall, the paper contributes to advancing computer-aided diagnosis in male fertility assessment by introducing a precise and effective image analysis framework.

In a 2021 study, Abbasi, Miahi, and Mirroshandel [19] proposed two deep learning–based approaches for automated sperm morphology classification in the context of male

infertility assessment. The first approach applied deep transfer learning, while the second, termed Deep Multi-Task Transfer Learning (DMTL), integrated transfer learning with multi-task learning to simultaneously classify multiple morphological components—namely the head, acrosome, and vacuole—within a single framework. This work represents the first use of multi-task learning for sperm morphology analysis. Evaluated on the publicly available MHSMA dataset, both methods demonstrated state-of-the-art performance.

In a 2025 study, Prabaharan and Saravanan [20] presented a three-stage CNN-based framework for automated sperm abnormality detection. The method begins with image reprocessing using a Wiener filter to reduce noise and enhance image quality. Next, individual sperm cells are segmented from the background using a modified Havrda–Charvat entropy-based thresholding technique. In the final stage, a convolutional neural network is used to classify sperm morphology as normal or abnormal. Experiments conducted on a benchmark sperm image dataset showed that the proposed approach achieved a classification accuracy of 98.99%, highlighting its superior performance compared to existing methods.

Prior studies have leveraged CNNs for sperm morphology classification with promising results. Many works applied deep models for head classification and motility assessment. However, few have focused on multi-class, multi-label outputs corresponding to biologically relevant structures. Moreover, explainability remains largely unexplored, with only limited efforts incorporating attention mechanisms or Grad-CAM visualization.

3. DATASET

This study utilizes the publicly available Modified Human Sperm Morphology Analysis (MHSMA) dataset [21], obtained from Kaggle. The dataset consists of grayscale sperm images collected from 235 patients diagnosed with male factor infertility. Each image is annotated by domain experts to indicate normal or abnormal conditions across four key morphological components: head, acrosome, vacuole, and tail. MHSMA is an enhanced version of the original Human Sperm Morphology Analysis Dataset (HSMA-DS) [22]. The dataset is organized into training, validation, and test subsets, with all images resized to a uniform resolution of 128 × 128 pixels.

Table 1 shows the number of positive and negative examples in the dataset.

Table 1. Distribution of samples in MHSMA

Set	Label	# +ve	# -ve	% +ve
Whole Dataset	Acrosome	1086	454	70.52
	Head	1122	418	72.86
	Vacuole	1301	239	84.48
	Tail	1471	69	95.52
Training Set	Acrosome	699	301	69.90
	Head	727	273	72.70
	Vacuole	830	170	83.00
	Tail	954	46	95.40
Validation Set	Acrosome	174	66	72.50
	Head	176	64	73.33
	Vacuole	209	31	87.08
	Tail	233	7	97.08
Test Set	Acrosome	213	87	71.00
	Head	219	81	73.00
	Vacuole	262	38	87.33
	Tail	284	16	94.67

4. METHODOLOGY

This study adopts a methodology focused on the development and evaluation of explainable deep neural network architectures for the morphological classification of human sperm cells, with the objective of supporting automated applications in reproductive health assessment. The overall workflow consists of five major components: (i) dataset acquisition (ii) comparison of deep neural architectures for sperm cell morphology classification, (iii) incorporation of explainable artificial intelligence (XAI) techniques to enhance interpretability, (iv) evaluation based on classification accuracy and explainability metrics. The steps used in the deep neural architectures for sperm cell morphology classification is described in detail in the following subsection.

VGG16 [23]: Deep learning model for sperm morphology meta-analysis, classifying images based on four categories: head, vacuole, acrosome, and tail is built. First, a custom SpermDataset class is defined to load grayscale sperm images and their respective labels from .npy files, applying data augmentation transformations like random flips and rotations during training. The model architecture is based on a modified VGG16 network: its first convolutional layer is adapted for single-channel (grayscale) images, and four separate fully connected layers are added to independently predict each morphology label. Data is loaded into PyTorch DataLoaders for batching and shuffling. The model is trained using the AdamW optimizer and CrossEntropyLoss, with a learning rate scheduler that reduces learning rate upon validation loss plateaus. During training, the model minimizes the sum of the four classification losses, and validation is performed after each epoch. After training, the model's performance is evaluated on the test set by calculating accuracy for each sperm structure separately as well as an overall accuracy, giving a complete performance measure of the multi-output model. This setup allows simultaneous and efficient prediction of multiple important features in sperm analysis.

Resnet34 [24, 25]: The proposed methodology employs a multi-task deep learning framework based on a modified ResNet34 architecture to enable the explainable morphological classification of human sperm cells. Grayscale sperm cell images of size 128×128128 times 128128×128 pixels were organized into NumPy arrays and paired with labels corresponding to four morphological attributes, namely head, vacuole, acrosome, and tail. A custom dataset loader was implemented in PyTorch to manage image–label mapping, while preprocessing involved intensity normalization and, in the case of training data, augmentations such as horizontal flipping and random rotations to improve generalization. The ResNet34 backbone, pretrained on ImageNet, was adapted by modifying its first convolutional layer to accept single-channel input, and its final classification layer was replaced with four independent fully connected layers, each dedicated to one morphological attribute. Training was performed using the AdamW optimizer with a learning rate of 1×10−4, while a ReduceLROnPlateau scheduler dynamically adjusted the rate based on validation loss. Cross-entropy loss was applied to each output branch, and the total loss was defined as the sum of all four components. Model performance was validated after each epoch and subsequently evaluated on the test set by computing individual accuracies for head, vacuole, acrosome, and tail classifications, alongside an overall accuracy metric

averaged across all outputs. This multi-output formulation enables the simultaneous learning of diverse morphological features, thereby supporting reliable automation in reproductive health applications.

DenseNet-121[26]: In this study, I implemented a deep learning pipeline for multi-label sperm morphology analysis using PyTorch and a DenseNet-121 backbone. First grayscale sperm images are loaded and their corresponding labels for four morphological traits: head, vacuole, acrosome, and tail. Images are loaded from .npy files and preprocessed with data augmentations such as horizontal flipping and random rotations for training, while validation and testing images are simply normalized. A modified version of DenseNet-121 is defined, where the first convolutional layer is adjusted to accept single-channel (grayscale) images. After feature extraction, the network uses a shared dense block and four separate fully connected heads — one for each label — to perform independent binary classifications.

The training process involves optimizing the sum of four cross-entropy losses (one per label) using the AdamW optimizer and a learning rate scheduler that reduces the learning rate if validation loss plateaus. The model is trained over multiple epochs from 1 to 10, tracking both training and validation losses. Finally, the model's performance is evaluated on the test set, calculating and printing the classification accuracy for each morphological trait individually as well as an overall average accuracy across all traits. This approach enables efficient, multi-output prediction from a single CNN, making it well-suited for automated, high-throughput sperm meta-analysis tasks.

The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of a binary classification model across varying decision thresholds. It illustrates the trade-off between the true positive rate and false positive rate, thereby indicating the model's ability to discriminate between the two classes. The ROC curve shows

X-axis (False Positive Rate, FPR) → how often your model incorrectly predicts the positive class when it's actually negative.

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

Y-axis (True Positive Rate, TPR / Recall / Sensitivity) → how often your model correctly predicts the positive class when it's actually positive.

$$TPR = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Each point on the ROC curve represents the performance of the classifier at a specific decision threshold, reflecting the corresponding trade-off between true positive and false positive rates.

All the above model's interpretability was enhanced through explainable artificial intelligence (XAI) techniques. Gradient-weighted Class Activation Mapping (Grad-CAM) [27] were applied to visualize the region's most influential in model predictions, thereby providing insights into the morphological features driving classification. This integration of deep learning and XAI not only improved predictive accuracy but also enhanced the transparency and clinical trustworthiness of the models in the context of automating reproductive health applications.

5. RESULTS AND DISCUSSIONS

In this section, the performance of the selected convolutional neural network architectures—VGG16, ResNet34, and DenseNet121—is systematically evaluated. The comparative study is presented through three distinct outputs: training and validation loss curves, which provide insights into convergence behavior and generalization capability; Receiver Operating Characteristic (ROC) curves, which quantify classification performance across varying thresholds; and Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations, which enhance interpretability by localizing the image regions most influential to the models' predictions. These complementary perspectives enable both quantitative and qualitative assessment of the models, thereby facilitating a comprehensive comparison of their effectiveness for the given task.

VGG16: The proposed multi-output VGG16-based architecture demonstrated strong performance in simultaneous classification of sperm head, vacuole, acrosome, and tail morphologies. Over the course of 10 training epochs, the model exhibited a consistent reduction in both training and validation losses, with the validation loss decreasing from 1.7199 in the first epoch to 1.4891 in the final epoch, indicating improved generalization capability. The training loss decreased steadily from 1.9442 in the first epoch to 1.6068 in the final epoch, indicating continuous improvement in fitting the training data. The relatively close alignment of training and validation losses, without significant divergence, further confirms that the network avoided severe overfitting and achieved stable convergence. On the independent test set, the model achieved high classification accuracies across all morphological components, with 74.00% for head detection, 88.33% for vacuole detection, 72.00% for acrosome detection, and 94.67% for tail detection. The aggregated overall accuracy across all four tasks was 82.25%, highlighting the effectiveness of the shared feature extraction backbone combined with task-specific classification heads. These trends validate the robustness of the proposed architecture in learning discriminative features for multi-output sperm morphology classification.

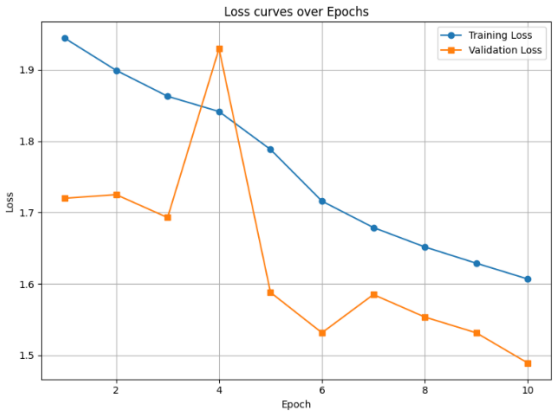

Fig 1: VGG 16 - Loss curves over epochs

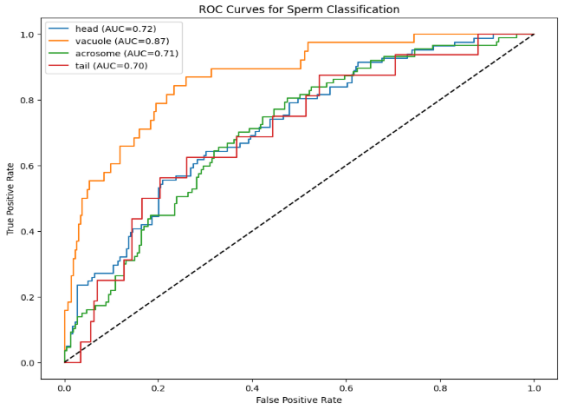

Fig 2: VGG 16 – ROC curves for sperm classification

Grad-CAM visualizations using a single sample for VGG16 show that the model predominantly focuses on the sperm head, vacuole and acrosome regions, indicating effective feature localization in these areas. However, the attention maps for the tail appear more dispersed, suggesting limited sensitivity to finer morphological details. These results confirm the suitability of the proposed deep learning framework for accurate, multi-aspect morphological assessment in automated sperm analysis systems.

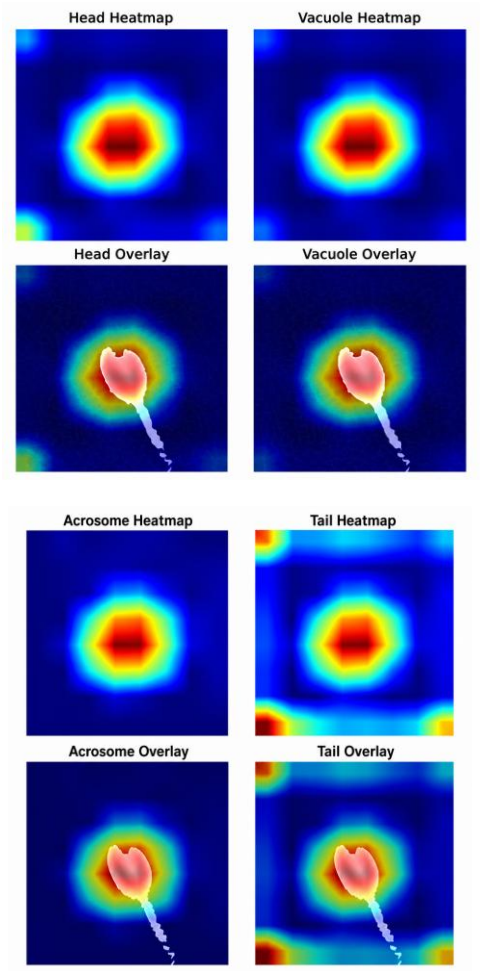

Fig 3: VGG 16 – Heatmaps and overlays

Resnet34: During training, the model demonstrated a consistent decrease in training loss from 2.2699 in Epoch 1 to 1.1128 in Epoch 10, indicating effective fitting to the training data. The validation loss exhibited fluctuations, starting at 1.7679 in Epoch 1, reaching a minimum of 1.4525 in Epoch 6, and slightly increasing thereafter, suggesting a mild onset of overfitting beyond this point.

On the held-out test set, the model achieved high classification performance across all four morphological categories: Head (74.67% accuracy), Vacuole (91.33%), Acrosome (73.67%), and Tail (94.67%). The overall aggregated accuracy across all categories was 83.58%, demonstrating the model's robustness in simultaneously predicting multiple sperm morphological features. Notably, the highest performance was obtained for the vacuole and tail categories, suggesting that these features are more visually distinctive in the dataset, whereas head and acrosome classification posed greater challenges due to their subtler morphological differences. These results indicate that a multi-output deep convolutional network can effectively model the complex morphological variations present in sperm cell images, achieving competitive performance across multiple tasks without the need for separate models.

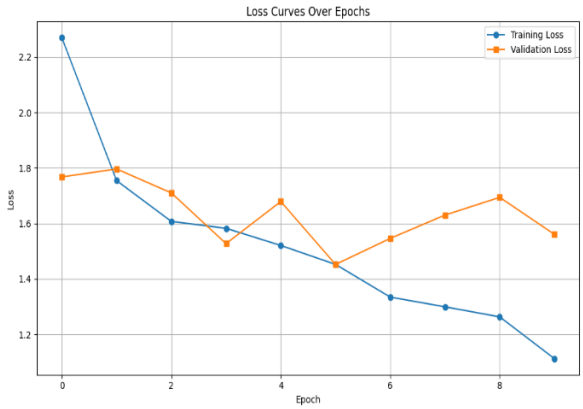

Fig 4: Resnet34 - Loss curves over epochs

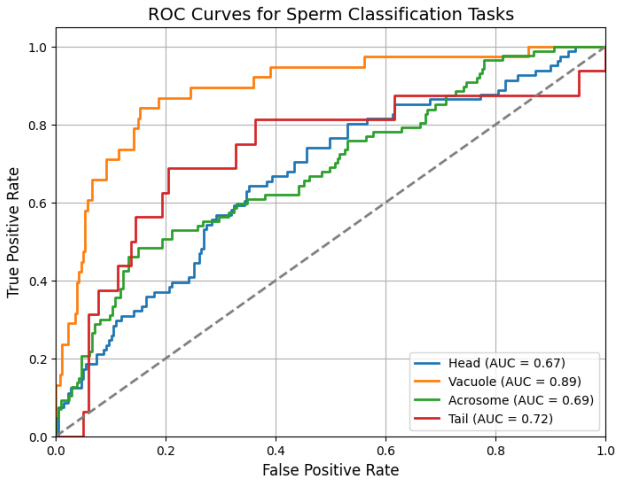

Fig 5: Resnet34 – ROC curves for sperm classification

The plot presents ROC curves for each sperm classification task, namely head, vacuole, acrosome, and tail, with the corresponding Area Under the Curve (AUC) values reported in the legend. The Area Under the ROC Curve (AUC) is a single scalar value that summarizes the overall performance of the classifier. A higher AUC indicates better performance for that

specific task. For example, the "Vacuole" task has the highest AUC (0.89), suggesting your model is best at classifying vacuoles. It has a good balance between True Positive Rate and False Positive Rate across various thresholds.

Grad-CAM visualizations of the same sample for ResNet34 reveal that the model successfully highlights the sperm head and acrosome regions, but the attention is relatively scattered across the background. This indicates that while the model captures key morphological features, its focus is less precise compared to VGG16.

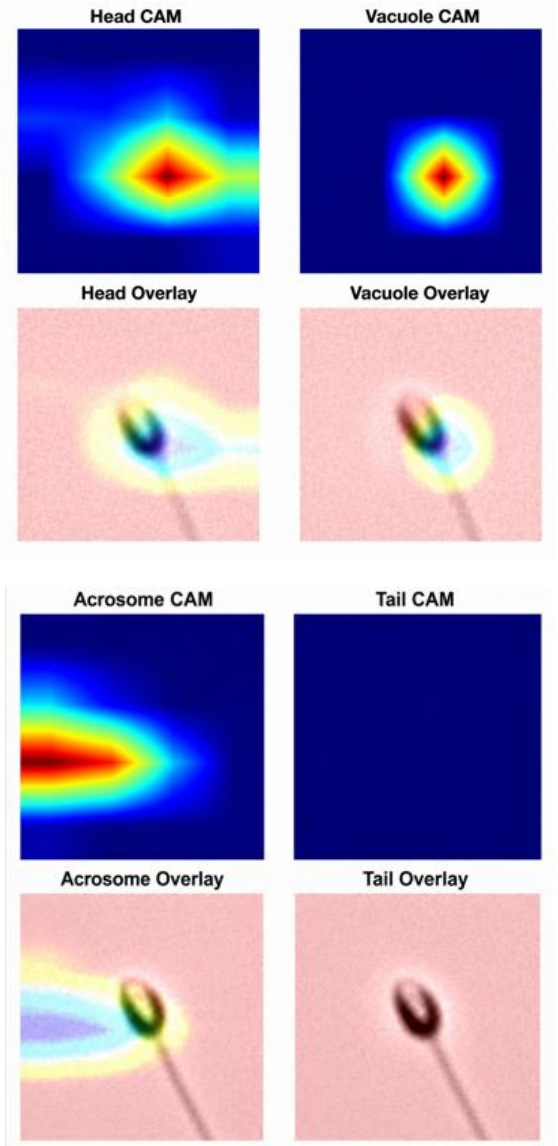

Fig 6: Resnet34 – Heatmaps and overlays

DenseNet-121: The multi-output convolutional neural network model based on a modified DenseNet-121 architecture was trained to classify human sperm morphology across four categories—head, vacuole, acrosome, and tail—using grayscale microscopy images. Over the course of 10 training epochs, the model demonstrated a consistent decline in both training and validation loss indicating effective learning and generalization. Upon evaluation on the test set, the model achieved high classification accuracies for vacuole (91.00%) and tail (94.00%) abnormalities, and reasonable performance for head (76.00%) and acrosome (68.33%) features. The overall mean classification accuracy across all four morphological traits was 82.33%, suggesting the model's strong potential for assisting in automated, multi-label sperm morphology assessment in clinical or research settings.

The first plot shows the training and validation loss over 10 epochs. The Training Loss (blue line) decreases steadily throughout the epochs. This indicates that the model is learning from the training data. The Validation Loss (orange line) decreases for the first few epochs but then starts to fluctuate and slightly increase towards the end. This suggests that the model might be starting to overfit to the training data after epoch 7, as the performance on unseen validation data is no longer improving. The second plot shows the Receiver Operating Characteristic (ROC) curves for each of the four classification tasks (Head, Vacuole, Acrosome, and Tail). The Tail classification has the highest AUC (0.76), indicating the best performance among the four tasks.

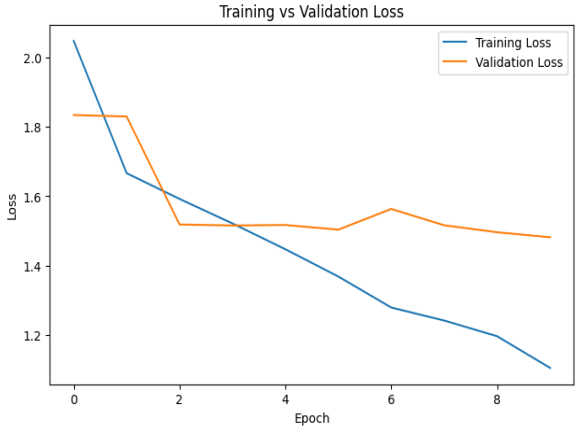

Fig 7: DenseNet121 - Loss curves over epochs

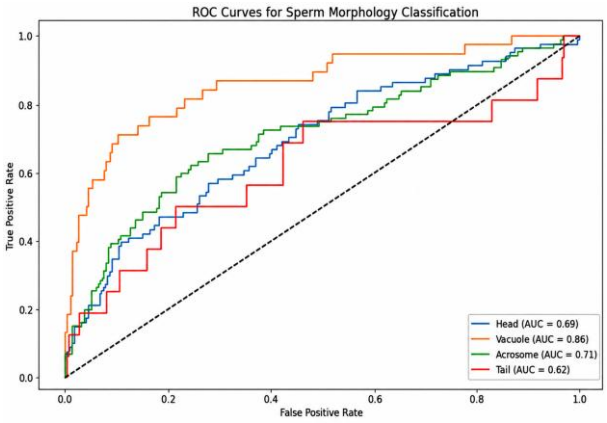

Fig 8: DenseNet121–ROC curves for sperm classification

In the Grad-CAM visualisation process, the original grayscale image undergoes a perceptual transformation due to the overlay of a colour-coded heatmap, which is designed to highlight the spatial regions most influential to the model's prediction. The heatmap highlights the important regions the model focuses on while making predictions. To show this visually, I overlaid the heatmap onto the RGB-converted version of the grayscale image using a blending technique. In DenseNet121, the Grad-CAM heatmaps demonstrate concentrated focus on the head, acrosome, and tail regions with minimal background interference. This suggests that DenseNet121 provides more refined feature localization than VGG16 and Resnet34, making it particularly effective in distinguishing subtle morphological variations.

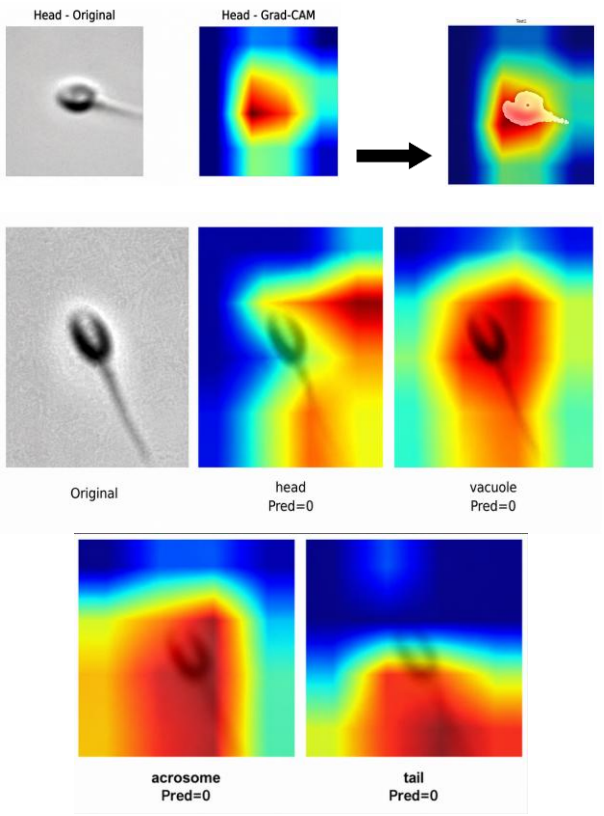

Fig 6: DenseNet121 – Heatmaps and overlays

6. CONCLUSION

The results in this study demonstrate the comparative performance of VGG16, ResNet34, and DenseNet121 in sperm morphology classification. The loss curves indicate effective convergence for all three models, with DenseNet121 achieving faster stabilization and lower overall loss. ROC curve analysis highlights strong discriminative ability across all morphological classes, with DenseNet121 showing the highest AUC values. Grad-CAM visualizations further reveal that DenseNet121 provides the most precise feature localization, particularly in the head and acrosome regions, while VGG16 and ResNet34 show more dispersed attention. Overall, DenseNet121 outperforms the other models in both classification accuracy and interpretability. Future work can focus on using larger and more diverse datasets from multiple laboratories to improve the generalization and reliability of the models. Advanced deep learning architectures such as Vision Transformers, EfficientNet, and hybrid CNN-transformer models can be explored to further enhance classification performance. The research can also be extended toward multi-task learning frameworks that simultaneously analyze sperm morphology, motility, and concentration for comprehensive fertility assessment.

7. REFERENCES

[1] Segal, T.R., and Giudice, L.C., 2019. Before the beginning: environmental exposures and reproductive and obstetrical outcomes. Fertility and Sterility; 112(4), 613-21.

[2] Zegers-Hochschild, F., and Dickens, B.M., and Dughman-Manzur, S., 2013. Human rights to in vitro fertilization. International Journal of Gynecology & Obstetrics; 123(1), 86-89.

[3] MacLeod, J., and Gold, R.Z., 1951. The male factor in fertility and infertility. IV. Sperm morphology in fertile and infertile marriage. Fertil Steril; 2(5):394–414.

[4] Menkveld, R., Holleboom, C.A., and Rhemrev, J. P., 2011. Measurement and significance of sperm morphology, Asian J. Androl. 13, 59.

[5] WHO. 2021. Laboratory manual for the examination and processing of human semen. 6th ed. Geneva: World Health Organization.

[6] Yim, J., Ju, J., Jung, H., and Kim, J., 2015. Image classification using convolutional neural networks with a multi-stage feature. Proceedings of the 3rd International Conference on Robot Intelligence Technology and Applications, November, Beijing, China, Springer. 587–594.

[7] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Computation. 1, no. 4, 541–551.

[8] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Muller, U. A., Sackinger, E., Simard, P., and Vapnik, V., 1995. Learning algorithms for classification: a comparison on handwritten digit recognition. Neural networks: The statistical mechanics perspective. 261.

[9] Violeta Chang, Laurent Heutte, Caroline Petitjean, Steffen Härtel, and Nancy Hitschfeld, 2017. Automatic classification of human sperm head morphology. Computers in Biology and Medicine, Vol 84, 205-216, ISSN 0010-4825.

[10] Chang, V., Garcia, A., Hitschfeld, N., and Härtel, S. 2017. Gold-standard for computer-assisted morphological sperm analysis. Comput. Biol. Med. 83, 143–150.

[11] Goodfellow, I., Bengio, Y., and Courville, Y. 2016. Deep Learning. online: https://mitpress.mit.edu/9780262035613/

[12] Thirumalaraju, P., Bormann, C.L., Kanakasabapathy, M., Doshi, F., Souter, I., Dimitriadis, I., and Shafiee, H. 2018. Automated sperm morpshology testing using artificial intelligence. Fertil. Steril. 110, e432.

[13] TP Kanakasabapathy, M.K., Bormann, C.L., Kandula, H., Sai Pavan, S.K., Yarravarapu, D., and Shafiee, H. 2019. Human sperm morphology analysis using smartphone microscopy and deep learning. Fertil. Steril. 112, e41.

[14] Ilhan, H.O., Serbes, G., and Aydin, N. 2020. Automated sperm morphology analysis approach using a directional masking technique. Comput. Biol. Med. 122, 103845.

[15] Javadi, S., and Mirroshandel, S.A., 2019. A novel deep learning method for automatic assessment of human sperm images. Comput. Biol. Med. 109, 182–194.

[16] Riordon, J., McCallum, C., and Sinton, D. 2019. Deep learning for the classification of human sperm. Comput Biol Med. doi: 10.1016/j.compbiomed.2019.103342. Epub: 2019 Jun 25. PMID: 31279166.

[17] Ilhan, H.O., Yuzkat, M., and Aydin, N. 2021. Sperm Motility Analysis by using Recursive Kalman Filters with the smartphone based data acquisition and reporting approach. Expert Syst. Appl. 186, 115774.

[18] Ilhan, H.O., Serbes, G., and Aydin, N. 2020. Automated sperm morphology analysis approach using a directional masking technique. Comput. Biol. Med. 122, 103845.

[19] Abbasi, A., Miahi, E., and Mirroshandel, S.A. 2021.

Effect of deep transfer and multi-task learning on sperm abnormality detection. Comput. Biol. Med. 128, 104121.

[20] Prabaharan, L., and Saravanan, N. 2025. A three stage framework for abnormality detection in sperm cell images using CNN, Biomedical Signal Processing and Control, Vol 99, ISSN 1746-8094.

[21] Ghasemian, F., Javadi, S., and Mirroshandel, S. A., 2020. MHSMA: The Modified Human Sperm Morphology Analysis Dataset. Mendeley Data, V2.

doi:10.17632/hjygss6sw2.2

[22] Ghasemian, F., Mirroshandel, S.A., Monji-Azad, S., Azarnia, M., and Zahiri, Z. 2015. An efficient method for automatic morphological abnormality detection from human sperm images. Comput. Methods Progr. Biomed. 122, 409–420.

[23] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint. https://arxiv.org/abs/1409.1556.

[24] He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[25] Wu, S., Zhong, S., and Liu, Y. 2018. Deep residual learning for image steganalysis. Multimedia Tools and Applications. 77, no. 9, 10437–10453, https://doi.org/10.1007/ s11042-017-4440-4

[26] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[27] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 618–626, doi: 10.1109/ICCV.2017.74.