Original Article

Advanced Embryo Ploidy Classification Using Vision Transformers: Integration of Sequential Time-Lapse Imaging and Undersampling Techniques: A Retrospective Study

Muhammad Fauzan Avidiansyah[1], Nining Handayani[1], Tri Aprilliana[1], Szeifoul Afadlal[1,2], Arief Boediono[1,2,3,4], Arie A. Polim[1,2,5], Batara Sirait[1,2,6], Irham Suheimi[1,2], Muhammad Farhan Aditya[2], Ivan Sini[1,2,3]

[1]IRSI Research Centre, [2]Morula IVF Jakarta Clinic, [5]Department of Obstetrics and Gynecology, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, [6]Department of Obstetrics and Gynaecology, Faculty of Medicine, Universitas Kristen Indonesia, Jakarta, [3]Faculty of Medicine, IPB University, [4]Department of Anatomy, Physiology and Pharmacology, IPB University, Bogor, West Java, Indonesia

Abstract

Background: Reliable identification of embryo ploidy is essential for optimising outcomes in assisted reproductive technology (ART). Conventional deep learning models, however, are limited by class imbalance, particularly due to the underrepresentation of mosaic embryos. **Aim:** This study aimed to improve embryo ploidy classification by integrating Vision Transformers (ViTs) with sequential time-lapse imaging and applying random undersampling (RUS) to mitigate data imbalance. **Settings and Design:** A retrospective study using blastocyst-stage time-lapse imaging data from a fertility clinic. Customised deep learning models were developed to predict embryo ploidy status. **Materials and Methods:** A total of 1020 blastocyst videos with genetically confirmed ploidy were analysed, generating 99,324 sequential frames representing the final 10 h of development before biopsy. To address imbalance, RUS produced a balanced dataset of 17,000 images per class: Euploid, aneuploid and mosaic. Two ViT architectures (ViT-B/16 and ViT-B/32) were fine-tuned for binary and multiclass tasks. Model performance was evaluated using accuracy, precision, recall, and F1-score on both balanced and imbalanced datasets. **Statistical Analysis Used:** Model performance was evaluated using accuracy, precision, recall, and F1-score. A 5-fold cross-validation procedure was applied to ensure robustness and reduce variance across data splits. **Results:** The ViT-B/16 achieved 0.84 accuracy in binary and 0.67 in multiclass classification on the balanced dataset, whereas performance dropped to 0.49 on the imbalanced set. RUS improved the prediction of minority classes, particularly mosaic embryos. **Conclusion:** Combining ViTs with sequential time-lapse imaging and RUS provides a promising non-invasive approach for embryo ploidy classification, enhancing accuracy for mosaic embryos and supporting more informed embryo selection in ART.

Keywords: *Assisted reproductive technology, embryo ploidy classification, time-lapse imaging, vision transformers*

Introduction

The integration of artificial intelligence (AI) in embryo assessment has transformed reproductive medicine by improving the accuracy and efficiency of embryo viability predictions. Traditional embryo evaluation, relying on subjective morphological assessments, shows limitations in predicting implantation success consistently.[1-3] Machine learning and AI

Address for correspondence: Dr. Ivan Sini, Faculty of Medicine, IPB University, Rasamala Lingkar Kampus Street, IPB Dramaga, Bogor 16680, West Java, Indonesia.
E-mail: ivan.sini@apps.ipb.ac.id

Received: 05-06-2025 **Revised:** 03-11-2025
Accepted: 15-11-2025 **Published:** 26-12-2025

Access this article online

Quick Response Code:

Website: www.jhrsonline.org

DOI: 10.4103/jhrs.jhrs_111_25

How to cite this article: Avidiansyah MF, Handayani N, Aprilliana T, Afadlal S, Boediono A, Polim AA, *et al.* Advanced embryo ploidy classification using vision transformers: Integration of sequential time-lapse imaging and undersampling techniques: A retrospective study. J Hum Reprod Sci 2025;18:208-17.

algorithms now provide objective, standardised methods through image analysis and data-driven insights, improving selection processes.[4-6] For instance, studies have demonstrated that AI models can outperform human embryologists in identifying embryos with a higher likelihood of leading to successful pregnancies, thereby addressing the inherent variability and subjectivity of conventional methods.[7,8] Furthermore, the application of AI extends beyond mere classification; it encompasses the analysis of time-lapse imaging data and the integration of multicentric clinical data, which collectively enhance predictive accuracy and facilitate personalised reproductive strategies.[9,10] As the field continues to evolve, the potential for AI to standardise and optimise embryo selection not only promises to improve clinical outcomes but also underscores the necessity for rigorous validation and ethical considerations in its implementation.[11-13]

Day 5, when embryos reach the blastocyst stage, plays a critical role in optimising embryo viability and genetic integrity, as biopsies at this stage significantly reduce mosaicism risk compared to earlier stages.[14] Research also shows higher implantation success rates and overall better outcomes in preimplantation genetic testing for embryos biopsied at this stage.[15] The final 10 h before biopsy represent a critical window in embryo development, during which key cellular and morphological changes occur, optimising viability for implantation.[16] Recent AI advancements, such as the Life Whisperer model, have enhanced non-invasive assessments of embryo viability through optical microscopy, refining embryo selection criteria before biopsy and ultimately improving reproductive success rates.[3]

Vision Transformers (ViTs) outperform Convolutional Neural Networks (CNNs) in image classification by modelling long-range dependencies and global context, addressing the limitations of CNNs, which primarily focus on local features.[17] Studies have demonstrated that ViTs achieve state-of-the-art results on benchmark datasets such as ImageNet, outperforming CNNs by leveraging their ability to capture intricate relationships across the entire image.[18,19] Furthermore, the flexibility of ViTs in adapting to various data distributions and their enhanced capacity for transfer learning have made them particularly effective in specialised domains, including medical imaging, where they have shown remarkable accuracy in classifying complex patterns.[20] As a result, the transition from CNNs to ViTs marks a significant shift in the landscape of image classification, underscoring the potential of transformer-based models to redefine performance standards in visual recognition tasks.[21]

Time-lapse imaging combined with ViTs represents a significant advancement in analysing dynamic environments, as ViTs model long-range dependencies and global context effectively.[22] Recent studies have demonstrated that ViTs can outperform traditional CNNs in various vision tasks, including image classification and object detection, by leveraging their self-attention mechanisms to align and process visual data more efficiently.[23,24] Moreover, the hierarchical structures of advanced ViTs, such as the HiViT and Multiscale ViTs, enhance their capability to manage multi-scale features, making them particularly suitable for the complex data generated in time-lapse sequences.[25,26] As the challenges of data association and temporal coherence in time-lapse imaging in embryo development continue to evolve, the application of ViTs offers promising solutions that can improve the accuracy and efficiency of visual data interpretation.[22] This integration not only enhances the analytical power of time-lapse imaging in embryo development but also opens new avenues for research in dynamic scene understanding and environmental monitoring.

Class imbalance is a pervasive issue in machine learning that can significantly hinder the performance of classification models, particularly when the majority class vastly outnumbers the minority class. In this study, we focus on random undersampling (RUS) as a straightforward yet effective technique to address this imbalance by selectively reducing the number of instances in the majority class. RUS has been widely recognised for its ability to enhance model performance by allowing classifiers to better learn from the minority class, which is often underrepresented in the training data.[27,28] While RUS is simple to implement, it is essential to acknowledge that it may lead to the loss of potentially informative majority class instances, which can distort the underlying data distribution.[29] Despite this drawback, RUS remains a popular choice due to its computational efficiency and ease of integration into various machine learning workflows.[30] Furthermore, studies have demonstrated that RUS can yield improved classification accuracy in high-dimensional datasets, making it particularly suitable for applications where the minority class is of critical importance.[30,31] Overall, the application of RUS in our experiment aims to create a more balanced dataset, thereby facilitating enhanced learning and improved predictive performance in the context of class imbalance.[28]

The utilisation of sequential time-lapse images from the final 10 h of embryo development has demonstrated a significant improvement in model accuracy for ploidy classification, as evidenced in recent studies.[32] This critical period, immediately preceding the blastocyst

biopsy, is marked by dynamic morphological and cellular changes that provide valuable predictive information on embryo viability and genetic integrity. Time-lapse imaging captures subtle features missed by static images. Integrating these sequential images into a ViT improves its ability to identify complex developmental patterns, enhancing ploidy classification accuracy. This approach underscores the potential of time-lapse imaging combined with ViTs to refine non-invasive embryo assessment methods, offering a promising tool for improving reproductive outcomes.

Mosaic embryos, which contain both euploid and aneuploid cell lines, pose a major clinical challenge in assisted reproductive technology (ART). Despite exhibiting chromosomal abnormalities, some mosaic embryos are capable of resulting in healthy live births, creating uncertainty in embryo selection. Accurate identification of mosaic embryos is, therefore, crucial for optimising transfer decisions, minimising unnecessary biopsies and reducing the risk of discarding potentially viable embryos. However, due to their subtle morphological features and underrepresentation in datasets, mosaic embryos remain difficult to classify accurately using both conventional and AI-based approaches. This study aims to address these challenges by leveraging ViTs and undersampling techniques to enhance non-invasive mosaic classification, thereby supporting more precise and ethical clinical decision-making in ART.

Materials and Methods

Data extraction and data preparation

This study utilised a sample of 425 individuals and 1,020 blastocyst videos with known ploidy status, collected from Morula IVF Jakarta Clinic. Each video captured the last 10 h of embryo development prior to biopsy. This time window was chosen because it represents the late blastocyst stage, during which critical morphological events such as blastocoel expansion, zona pellucida thinning and initiation of hatching occur. These dynamic processes are closely linked with embryo competence and chromosomal normality, making this period biologically informative for assessing ploidy-related features through time-lapse imaging. Time-lapse imaging data were extracted from these videos to assess critical morphological changes and cellular activity. The videos were pre-processed by segmenting each into frames at regular intervals to capture key developmental stages. Sequential frames were then resized and normalised for consistency across the dataset. Manual annotations were made by embryologists to label the ploidy status of each embryo, categorising them into euploid, aneuploid, or mosaic. These labelled images were further divided into training, validation and test sets, ensuring a balanced representation of each ploidy class. For addressing class imbalance, RUS of the majority class was applied during model training to enhance classification performance.

Following segmentation and normalisation of the frames, the dataset was split into training and validation sets with an 80:20 ratio. To ensure model robustness and minimise overfitting, a 5-fold cross-validation procedure was employed during the hyperparameter tuning phase. Each fold served as a validation set once, while the remaining folds were used for training, enabling comprehensive performance evaluation across all data subsets. This approach provided a more reliable estimation of model generalisability compared to a single split validation. The training set was utilised to fine-tune the ViT models (ViT-B/16 and ViT-B/32), whereas the validation set was reserved for performance evaluation.

To address class imbalance, the RUS technique was applied to the training set, resulting in a balanced dataset with 17,000 images per class (euploid, aneuploid and mosaic). The number of 17,000 images was determined based on the smallest available class (mosaic embryos) to achieve class balance without data duplication. The majority of classes (euploid and aneuploid) were randomly undersampled to match this number. This data-driven approach ensured balanced representation while minimising potential sampling bias.

To ensure that the model captures relevant temporal and morphological features from the time-lapse sequences, feature extraction was applied to the image sequences. Developmental events such as blastocyst expansion, zona pellucida thinning and inner cell mass (ICM) formation were not manually encoded but were implicitly represented through pixel intensity patterns and spatial-temporal changes across sequential frames. Each frame was treated as a token input, allowing the ViTs to learn dynamic morphological variations through patch embeddings and self-attention mechanisms. Temporal continuity was maintained using frame indexing to numerically represent developmental progression within the sequence. This step provided the ViT model with a comprehensive understanding of both spatial resolution and the temporal progression of embryo development, aiding in the ploidy classification task.

A total of 99,324 images were extracted from 1020 blastocyst time-lapse videos, each covering the final 10 h of embryo development. Of these, 17,772 images corresponded to euploid blastocysts, 43,637 images to aneuploid blastocysts and 37,915 images to mosaic blastocysts. Each image was labelled based on the known ploidy status of the embryos, enabling precise classification during model training.

The significant variation in image counts across the three ploidy classes reflects the natural distribution within the dataset. To mitigate the class imbalance, RUS was applied to the euploid, aneuploid and mosaic classes, ensuring a more balanced training set. Specifically, the number of images in each class was rounded down to 17,000, aligning all classes for uniformity. This step was crucial to prevent the model from being biased toward the majority class and to improve the overall classification accuracy for minority classes, particularly euploid and mosaic embryos. Figure 1 illustrates the class target distribution after the application of the undersampling method.

The implementation of the ploidy classification model leveraged several key Python libraries. For building the model, torchvision was utilised due to its extensive support for variation ViT and efficient handling of image data, particularly in the ViT architecture. Data pre-processing and manipulation of tabular data, including the ploidy labels and image metadata, were performed using pandas and numpy, which provided powerful data

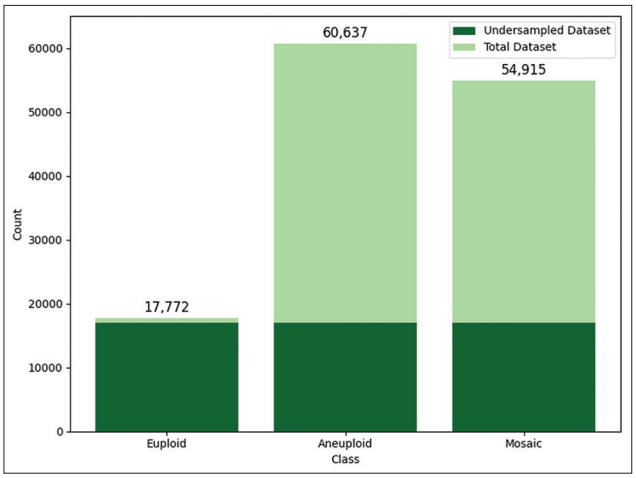

Figure 1: Class Target Distribution After Undersampling Method

structures and mathematical operations. To enhance the robustness of the model and prevent overfitting, image augmentation techniques were applied using OpenCV, which allowed for real-time adjustments such as rotations, flips and contrast changes, thereby increasing the diversity of the training data. These combined tools enabled a streamlined workflow, from data extraction to model training and evaluation.

Ethical approval and data governance

This study was conducted in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki Declaration and its later amendments. The study protocol was approved by the Ethical Committee of Universitas Indonesia (KET-469/UN2.F1/ETIK/PPM.00.02/2024). The data used in this study were fully anonymised embryo images obtained from the IVF laboratory database. As only de-identified data were analysed, no direct patient consent was required.

Vision transformer model

The ViT represents a significant advancement in image classification by leveraging self-attention mechanisms to model global dependencies within images, contrasting with the localised feature extraction used in CNNs. ViT divides an image into fixed-size patches, which are treated as input tokens, enabling the model to capture both fine-grained details and long-range spatial relationships. This characteristic is particularly advantageous for tasks involving complex morphological changes over time, such as embryo development in time-lapse imaging.

Figure 2 illustrates the ViT architecture utilised in this study, detailing its key components, including the multi-head self-attention mechanism, the positional embedding and the transformer encoder layers. This architecture was selected due to its capability to capture global

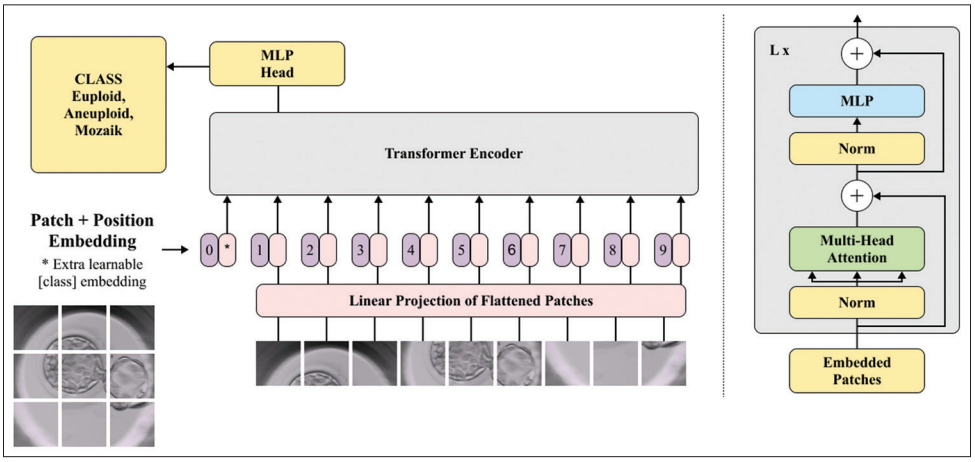

Figure 2: Diagram illustrating the Vision Transformer (ViT) Architecture

dependencies and its proven performance in image classification tasks. The ViT model processes sequences of image patches, enabling efficient representation learning for embryo ploidy classification.

In this study, we employed two variation ViT models: ViT-B/16 and ViT-B/32, which differ in the size of image patches used for tokenisation. ViT-B/16 utilises 16 × 16 pixel patches, offering higher granularity and potentially improving feature extraction for subtle morphological variations. In contrast, ViT-B/32 uses larger 32 × 32 pixel patches, providing computational efficiency but potentially sacrificing some fine detail. Both models were fine-tuned on our embryo ploidy classification dataset, with input images resized to 224 × 224 pixels. Positional encodings were applied to maintain the spatial relationships between patches, which is crucial for analysing the sequential developmental stages of the embryos.

The choice of patch size plays a critical role in balancing the trade-off between computational efficiency and model performance. ViT-B/16, with its smaller patches, is hypothesised to capture more detailed spatial features, which could be particularly beneficial in distinguishing between euploid, aneuploid and mosaic embryos, where subtle morphological differences are key. The models were trained using cross-entropy loss, with hyperparameters optimised through grid search. To enhance model generalisation, data augmentation techniques were applied, and early stopping was employed to prevent overfitting. Evaluation metrics such as accuracy, precision, recall and F1-score were used to assess the performance of each model, with particular attention given to the minority classes.

We employed the AdamW optimiser[33] due to its superior performance compared to both the Adamax and RMSprop optimisers. A fixed learning rate of 1×10^{-3} was used throughout training. For both the ViT-B16 and ViT-B32 models, the batch size was set to 256, and the models were trained for 300 epochs. Embryo images were uniformly resized to 224 × 224 pixels. These hyperparameters were determined experimentally to achieve optimal results.

Table 1: Hyperparameter configuration

Configuration	Value
Optimizer	AdamW
Epoch	300
Batch size	256 for VIT-B16 and 256 for VIT-B32
Learning rate	1×10^{-3}
Loss function	Categorical cross entropy

VIT=Vision transformers

Table 1 summarises the hyperparameter configurations used for training the ViT models. Key parameters include the learning rate, batch size, optimiser type and number of epochs, which were fine-tuned to optimise model performance. These configurations were determined through extensive experimentation and cross-validation to ensure robust classification outcomes.

For evaluation metrics, we used accuracy, precision, recall and the F1 score as our metrics. The pre-processing, model development and evaluation were carried out using PyTorch on an Nvidia RTX 2060 GPU.

Evaluation metrics

In evaluating the performance of our deep learning-based embryo ploidy classification model, we utilised several standard metrics: Accuracy, precision, recall and F1 score. These metrics provide a comprehensive overview of how well the model performs,[34] particularly given the challenges of class imbalance in the dataset.

a) True Positive (TP): The data are expected to be positive and are predicted correctly as positive
b) False Negative (FN): The data are expected to be negative and are predicted correctly as positive
c) True Negative (TN): The data are expected to be negative and are predicted correctly as negative
d) False Positive (FP): The data are expected to be positive and are predicted correctly as negative.

Based on these values, metrics such as accuracy, recall, precision and F1 score were calculated. The formulations were as follows:

Accuracy

Accuracy represents the proportion of correctly classified embryos (euploid, aneuploid or mosaic) out of the total number of embryos. However, in deep learning models dealing with imbalanced data, accuracy may not always reflect the model's true performance, as the majority class can skew the results.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Precision focuses on the model's ability to avoid false positives. It measures the proportion of embryos classified as belonging to a particular class (e.g. euploid) that are actually in that class. High precision in deep learning ensures the model minimises incorrect classifications.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

Recall is essential in deep learning for ensuring that the model identifies all relevant instances. It measures the model's ability to detect embryos that belong to a specific class, such as euploid embryos and minimises false negatives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score

The F1 score balances precision and recall, making it especially useful for imbalanced datasets in deep learning. As the harmonic mean of precision and recall, it provides a single, unified measure of the model's performance.

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}}$$

To ensure reliable performance evaluation, all reported metrics represent the mean values obtained across five cross-validation folds. This approach minimised random variation and provided a more stable estimate of the model's generalisation performance.

Results

Extraction of sequential TL blastocyst images over the last 10 h before the biopsy procedure yielded approximately 80–100 unique images per TL video. From a pool of 1020 blastocyst TL videos, a total of 99,324 images were captured within the 10-h period preceding biopsy. Specifically, 17,772 images were obtained from euploid blastocysts, 43,637 from aneuploid blastocysts and 37,915 images from mosaic blastocysts. To address class imbalance, a RUS method was applied, reducing each class to 17,000 images. The classification results will be compared between the undersampled dataset and the total dataset. Using ViTs, multiclass classification involving euploid, aneuploid and mosaic will be compared to binary classification distinguishing euploid and aneuploid. The slight disparities between the sequential embryo images increase model complexity, indicating that using multiple time-point images could yield more accurate classification outcomes than single time-point images.

In this study, we employed ViTs to classify embryo ploidy status, including binary (euploid and aneuploid) and multiclass (euploid, aneuploid and mosaic) classification tasks. Table 2 summarises the performance metrics of the classification models on the undersampled dataset. The ViT-B/16 model achieved an accuracy of 0.84 in binary classification, with precision, recall and F1-scores showing consistent results for both euploid and aneuploid categories. The multiclass classification task was more complex, with ViT-B/16 achieving 0.67 accuracy, whereas the ViT-B/32 scored 0.66, suggesting that the addition of mosaic classification reduces the overall performance.

In contrast, Table 3 presents the classification results for the total dataset, where a notable decline in accuracy is observed for both binary and multiclass tasks. The ViT-B/16 model achieved only 0.49 accuracy in the binary classification task, whereas the multiclass task resulted in a lower accuracy of 0.27. These results indicate that the RUS method mitigates the class imbalance issue, improving the model's ability to classify minority classes accurately. The confusion

Table 2: Experiment model with undersampled dataset

Classification class	Variation VIT	Classification report									
		Accuracy	Recall			Precision			F1-score		
			Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic
Biner classification	VIT-B/16	0.84	0.83	0.85	-	0.85	0.83	-	0.84	0.84	-
Biner classification	VIT-B/32	0.83	0.83	0.82	-	0.83	0.82	-	0.82	0.83	-
Multi-class classification	VIT-B/16	0.67	0.72	0.72	0.59	0.73	0.65	0.65	0.72	0.68	0.61
Multi-class classification	VIT-B/32	0.66	0.73	0.71	0.55	0.71	0.62	0.67	0.72	0.66	0.6

VIT=Vision transformers

Table 3: Experiment model with total dataset

Classification class	Variation VIT	Classification report									
		Accuracy	Recall			Precision			F1-score		
			Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic
Biner classification	VIT-B/16	0.49	0.17	0.86	-	0.59	0.47	-	0.26	0.6	-
Biner classification	VIT-B/32	0.44	0.17	0.76	-	0.45	0.44	-	0.24	0.56	-
Multi-class classification	VIT-B/16	0.27	0.09	0.63	0.21	0.27	0.2	0.49	0.14	0.3	0.29
Multi-class classification	VIT-B/32	0.25	0.1	0.53	0.21	0.25	0.17	0.48	0.14	0.26	0.3

VIT=Vision transformers

matrices and ROC curves for these models further confirm that the undersampling approach leads to better performance across all classification tasks.

The results show that the undersampled dataset outperformed the total dataset in both binary and multiclass classification tasks. As shown in Tables 2 and 3, models trained on the undersampled dataset consistently achieved higher accuracy, recall, precision and F1 scores compared to those trained on the total dataset. Specifically, the ViT-B/16 model achieved an accuracy of 0.84 in binary classification with the undersampled dataset, whereas its performance dropped to 0.49 with the total dataset. A similar trend was observed in the multiclass classification, where the undersampling ViT-B/16 model achieved 0.67 accuracy, compared to only 0.27 when using the total dataset. These results underscore the effectiveness of RUS in balancing the class distribution and improving classification performance, particularly for the minority class (mosaic), which was significantly underrepresented in the total dataset.

Although the primary focus of this study was quantitative evaluation through accuracy, precision, recall, and F1-score, future extensions may include visual analyses such as confusion matrices or ROC curves to enhance interpretability.

Discussion

The performance comparison with recent methods for embryo ploidy classification is shown in Table 4. In our binary classification experiment, the undersampled dataset with the ViT (ViT-B/16) model outperformed most previous studies utilising deep learning techniques for embryo image classification. For instance, Handayani *et al*.[32] reported an accuracy of 0.67 using 99,324 embryo images in their study 'Improving Deep Learning-Based Algorithm for Ploidy Status Prediction Through Combined U-NET Blastocyst Segmentation and Sequential Time-Lapse Blastocyst Images'. Our model achieved the same accuracy with fewer images (51,000), demonstrating that our undersampling method effectively balances the dataset while maintaining performance.

Similarly, our ViT models also outperform the results of the study by Danardono *et al*.,[35] which achieved a 0.74 accuracy using only 527 embryo images. Although that study used a relatively small dataset, our binary classification model achieved a significantly higher accuracy of 0.84, indicating that ViTs can deliver more robust performance, even when dealing with more complex datasets and classification tasks. This improvement highlights the model's capability to effectively distinguish among euploid, aneuploid and mosaic categories.

In addition, the Embryo Ranking Intelligent Classification Algorithm, developed by Chavez-Badiola *et al*.,[36] reported an accuracy of 0.70 using 1231 images. In contrast, our two-class classification model achieved a binary classification accuracy of 0.84, demonstrating that our ViT-B/16 model, in combination with the undersampling technique, outperforms previous CNN-based methods in binary classification tasks. This improvement underscores the model's ability to effectively distinguish among euploid, aneuploid and mosaic embryo categories.

Moreover, Barnes *et al*.,[37] in their study, 'A Non-Invasive Artificial Intelligence Approach for the Prediction of Human Blastocyst Ploidy' reported an accuracy of 0.693 using 10,378 images in a binary classification task. In contrast, our model, utilising a larger dataset of 17,000 images per class, achieved a significantly higher accuracy of 0.84. This demonstrates that the undersampling strategy enhances the model's generalisation across embryo ploidy categories, even with a substantially larger dataset.

These findings suggest that our ViT-B/16 model, fine-tuned with undersampled data, offers better generalisation performance compared to the existing CNN-based models, especially in binary classification tasks. The inclusion of a larger dataset combined with undersampling has significantly enhanced the robustness of our model, ensuring consistent performance across various ploidy classification scenarios. To further

Table 4: Performance comparison with state-of-the-art methods using ploidy images for ploidy prediction

Reference	Classification	Embryo images	Classification report									
			Accuracy	Recall			Precision			F1-score		
				Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic	Euploid	Aneuploid	Mosaic
[32]	Multi-class	99,324	0.67	-	-	-	-	-	-	-	-	-
[35]	Biner-class	527	0.74	0.33	0.84	-	0.35	0.83	-	0.34	0.83	-
	Multi-class	865	0.43	0.07	0.78	0.46	-	-	-	-	-	
[36]	Biner-class	1231	0.7	-	-	-	-	-	-	-	-	-
[37]	Biner-class	10,378	0.693	-	-	-	-	-	-	-	-	-

Some studies included in Table 4 reported only accuracy without providing recall, precision, or F1-score metrics. Therefore, only the available performance indicators from each reference are presented to ensure consistency with the original publications

minimise model bias and overfitting, we employed an internal 5-fold cross-validation strategy during training. This approach strengthened the reliability of the reported results, though future studies should incorporate external validation using multicentre datasets to confirm generalisability.

Although the present study primarily focused on predictive performance, incorporating visual interpretability techniques such as Grad-CAM in future work could help elucidate which morphological regions or developmental events (e.g. ICM formation and blastocoel expansion) are most influential in determining ploidy classification. This would not only enhance the biological interpretability of ViT decisions but also provide valuable insights for embryologists regarding the morphological correlates of euploid and aneuploid development.

One limitation of this study is that the data were sourced from a single clinic, which may limit the generalisability of the findings. To improve the robustness and external validity of the results, future research should consider using multicentre datasets. In addition, correlating the model's predictions with actual clinical outcomes, such as implantation and pregnancy rates, would further strengthen the clinical relevance of this approach. Future studies incorporating these outcomes are warranted to validate the model's applicability in real-world IVF settings. Clinically, the integration of ViT-based models for non-invasive ploidy prediction holds promise for reducing dependence on invasive genetic testing, thereby potentially improving the efficiency and accessibility of embryo selection procedures. Beyond improving prediction accuracy, this approach could support embryologists in prioritising embryos with higher predicted euploid probabilities before biopsy or transfer, allowing for more informed and less invasive clinical decision-making. Integrating AI-based assessments into existing IVF workflows may help optimise embryo selection strategies, reduce the number of unnecessary biopsies and ultimately enhance clinical outcomes for patients undergoing assisted reproductive treatment.

Error analysis revealed that most misclassifications occurred between mosaic and euploid embryos. This overlap likely reflects the morphological and developmental similarities observed between these two classes during the late blastocyst stage, where subtle chromosomal variations may not always manifest as distinct visual differences. In contrast, aneuploid embryos exhibited more pronounced morphological deviations, resulting in fewer misclassifications. These findings suggest that while the ViT effectively distinguishes major ploidy categories, mosaic prediction remains challenging due to its intermediate phenotypic characteristics.

Despite the promising performance of the proposed model, predicting mosaic embryos remains challenging. Mosaic embryos often exhibit intermediate or subtle morphological patterns between euploid and aneuploid states, making them difficult to distinguish based solely on static visual features. These embryos may display normal morphological development while containing a mixture of chromosomally normal and abnormal cells, resulting in overlapping visual characteristics that limit model discrimination.

The incorporation of temporal information from time-lapse imaging offers a potential solution to this limitation. Sequential imaging enables the model to capture subtle temporal cues, such as variations in the timing of blastocoel expansion, ICM formation and hatching initiation, which may be associated with chromosomal mosaicism. By modelling these temporal dynamics, ViTs can learn to identify developmental irregularities that are not apparent in single-frame analyses, thus improving the biological interpretability and potential diagnostic accuracy for mosaic prediction.

Conclusion

This study demonstrates the efficacy of integrating ViTs with sequential time-lapse imaging and undersampling techniques to improve the classification of embryo ploidy status. By addressing class imbalance through RUS and leveraging the self-attention mechanisms of ViTs, we achieved a notable improvement in ploidy prediction, particularly in the minority classes. The ViT-B/16 model outperformed traditional CNN-based approaches in both binary and multiclass classification tasks, achieving 0.84 accuracy in binary classification with an undersampled dataset. In addition, time-lapse imaging during the final 10 h of embryo development provided critical morphological information that significantly enhanced the model's predictive performance.

Importantly, this non-invasive AI-based approach holds potential clinical value by reducing the dependence on invasive biopsy procedures and enabling earlier, safer assessment of embryo viability. Such integration of AI models into embryology workflows may improve embryo selection efficiency and patient outcomes in ART. Future work should focus on refining segmentation methods, validating the model using multicentric datasets and exploring multimodal data to further enhance predictive accuracy and clinical applicability.

Acknowledgements

Not applicable.

Author contribution

IS designed and coordinated the study. MFA and NH developed the AI model, performed data analysis and drafted the manuscript. TA, SA, AB, AAP and BS contributed to data collection and literature review. BS and IS supervised the project and provided critical revision of the manuscript. All authors contributed to the article and approved the final version of the manuscript.

Financial support and sponsorship

Nil.

Conflict of interest

There are no conflicts of interest.

Data availability statement

The datasets used and analysed during the current study are not publicly available due to patient privacy restrictions but are available from the corresponding author upon reasonable request.

References

1. Allahbadia GN, Allahbadia SG, Gupta A. In contemporary reproductive medicine human beings are not yet dispensable. J Obstet Gynaecol India 2023;73:295-300.
2. Allahbadia G. Embryo transfer is the last frontier for deep machine learning and artificial intelligence in medically assisted reproduction (MAR). J Reprod 2023;2:28-38.
3. VerMilyea M, Hall JM, Diakiw SM, Johnston A, Nguyen T, Perugini D, *et al.* Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. Hum Reprod 2020;35:770-84.
4. Chavez-Badiola A, Flores-Saiffe Farias A, Mendizabal-Ruiz G, Garcia-Sanchez R, Drakeley AJ, Garcia-Sandoval JP. Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. Sci Rep 2020;10:4394.
5. Bori L, Paya E, Alegre L, Viloria TA, Remohi JA, Naranjo V, *et al.* Novel and conventional embryo parameters as input data for artificial neural networks: An artificial intelligence model applied for prediction of the implantation potential. Fertil Steril 2020;114:1232-41.
6. Swain J, VerMilyea MT, Meseguer M, Ezcurra D, Fertility AI Forum Group. AI in the treatment of fertility: Key considerations. J Assist Reprod Genet 2020;37:2817-24.
7. Babayev E. Man versus machine in *in vitro* fertilization-can artificial intelligence replace physicians? Fertil Steril 2020;114:963.
8. Wang R, Pan W, Jin L, Li Y, Geng Y, Gao C, *et al.* Artificial intelligence in reproductive medicine. Reproduction 2019;158:R139-54.
9. Duval A, Nogueira D, Dissler N, Maskani Filali M, Delestro Matos F, Chansel-Debordeaux L, *et al*. A hybrid artificial intelligence model leverages multi-centric clinical data to improve fetal heart rate pregnancy prediction across time-lapse systems. Hum Reprod 2023;38:596-608.
10. Lundin K, Park H. Time-lapse technology for embryo culture and selection. Ups J Med Sci 2020;125:77-84.
11. Rolfes V, Bittner U, Gerhards H, Krüssel JS, Fehm T, Ranisch R, *et al.* Artificial intelligence in reproductive medicine – An ethical perspective. Geburtshilfe Frauenheilkd 2023;83:106-15.
12. Curchoe CL, Bormann CL. Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018. J Assist Reprod Genet 2019;36:591-600.
13. Trolice MP, Curchoe C, Quaas AM. Artificial intelligence-the future is now. J Assist Reprod Genet 2021;38:1607-12.
14. Huang L, Bogale B, Tang Y, Lu S, Xie XS, Racowsky C. Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. Proc Natl Acad Sci U S A 2019;116:14105-12.
15. Toporcerová S, Špaková I, Šoltys K, Klepcová Z, Kľoc M, Bohošová J, *et al*. Small non-coding RNAs as new biomarkers to evaluate the quality of the embryo in the IVF process. Biomolecules 2022;12:1687.
16. Chen TJ, Zheng WL, Liu CH, Huang I, Lai HH, Liu M. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. Fertil Reprod 2019;1:51-6.
17. Fayyaz M, Koohpayegani SA, Rezaei Jafari F, Sengupta S, Vaezi Joze HR, Sommerlade E, *et al*. Adaptive Token Sampling for Efficient Vision Transformers; 2021.
18. Gai L, Chen W, Gao R, Chen Y, Qiao X. Using Vision Transformers in 3-D Medical Image Classifications; 2022.
19. Niu C, Wang G. Unsupervised contrastive learning based transformer for lung nodule detection. Phys Med Biol 2022;67:204001.
20. Li Y, Hou Q, Jiang Z, Feng J, Yan S. VOLO: Vision outlooker for visual recognition. IEEE Trans Pattern Anal Mach Intell 2023;45:6575-86.
21. Melnychenko AB. Incorporating attention score to improve foresight pruning on transformer models. Visnyk Zaporizhzhya Natl Univ Phys Math Sci 2023:34–40. [doi: 10.26661/2786-6254-2023-2-03]
22. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: A survey. ACM Comput Surv 2022;54:200.
23. Morel Omer AA. Image classification based on vision transformer. J Comput Commun 2024;12:49-59. [doi: 10.4236/jcc.2024.124005].
24. Touvron H, Cord M. DeiT III: Revenge of the ViT. Proc Eur Conf Comput Vis 2022;13684:516-33. [doi: 10.1007/978-3-031-20053-3_30].
25. Zhang X, Tian Y, Huang W, Ye Q, Dai Q, Xie L, *et al*. HiViT: A Simpler and More Efficient Design of Hierarchical Vision Transformer. Int Conf Learn Represent 2023.
26. Fan H, Xiong B, Mangalam K, Li Y, Yan Z, Malik J, *et al*. Multiscale vision transformers. Proc IEEE/CVF Int Conf Comput Vis 2021:6824-35. [doi: 10.48550/arxiv.2104.11227].
27. Vuttipittayamongkol P, Elyan E. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. Inf Sci (N Y) 2020;509:47-70.
28. Devi D, Namasudra S, Kadry S. A boosting-aided adaptive cluster-based undersampling approach for treatment of class imbalance problem. Int J Data Warehousing Min 2020;16:60-86. [doi: 10.4018/ijdwm.2020070104].
29. Grina F, Elouedi Z, Lefèvre É. Learning From Imbalanced Data Using an Evidential Undersampling-Based Ensemble; 2022.
30. Guo H, Diao X, Liu H. Improving undersampling-based ensemble with rotation forest for imbalanced problem. Turk J Electr Eng Comput Sci 2019;27:1371-86.
31. Khairy M. The effect of rebalancing techniques on the classification performance in cyberbullying datasets. Neural

Comput Appl 2023;36:1049-65. [doi: 10.1007/s00521-023-09084-w].

32. Handayani N, Danardono GB, Boediono A, Wiweko B, Sini I, Sirait B, *et al*. Improving deep learning-based algorithm for ploidy status prediction through combined U-NET blastocyst segmentation and sequential time-lapse blastocysts images. J Reprod Infertil 2024;25:110-9.
33. Ankalaki S, Biradar VG, Kishor Kumar Naik P, Hukkeri GS. A deep learning approach for malnutrition detection. Int J Online Biomed Eng 2024;20:116-38.
34. Tan L, Lu J, Jiang H. Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods. AgriEngineering 2021;3:542-58.
35. Danardono GB, Handayani N, Louis CM, Polim AA, Sirait B, Periastiningrum G, Afadlal S, *et al*. Embryo ploidy status classification through computer-assisted morphology assessment. AJOG Glob Rep 2023;3:100209.
36. Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, Cohen J. Embryo ranking intelligent classification algorithm (ERICA): Artificial intelligence clinical assistant predicting embryo ploidy and implantation. Reprod Biomed Online 2020;41:585-93.
37. Barnes J, Brendel M, Gao VR, Rajendran S, Kim J, Li Q, *et al*. A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: A retrospective model development and validation study. Lancet Digit Health 2023;5:e28-40.