The illusion of vision

The illusion of vision: limits of blastocyst morphology integration in AI-driven IVF pregnancy prediction

L. Nadirashvili[1], S. Sergeev[1,2]; V. Kazarinov[1]; I. Diakova[3]

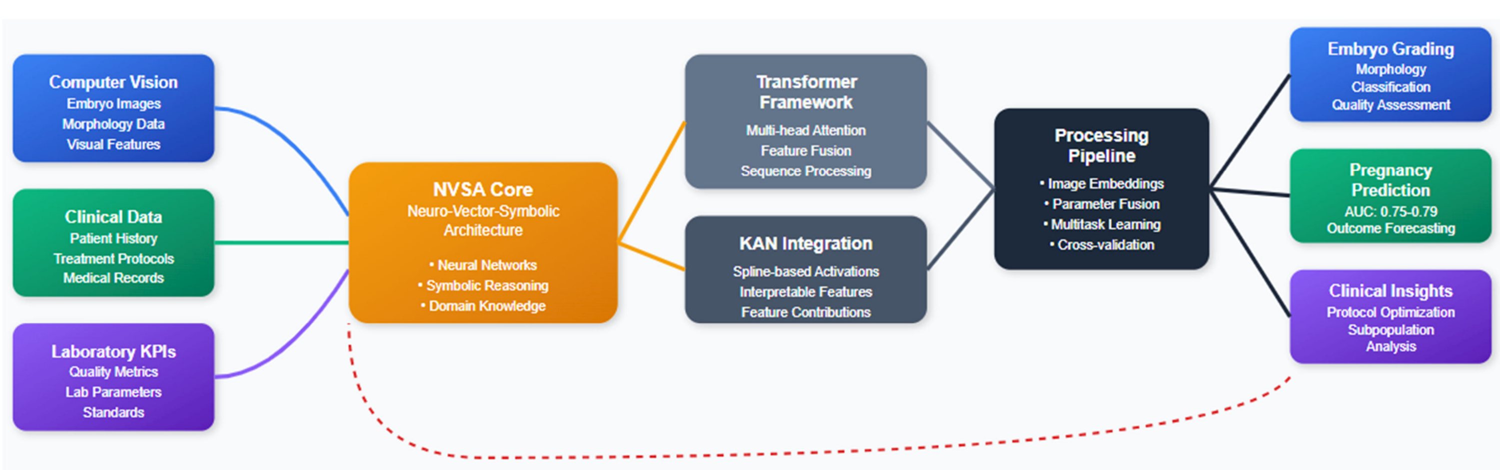
[1]Georgian-German Reproductive Center, Tbilisi, Georgia, [2]IVF and Genetic Center, 105043 Moscow, Russian Federation
[3]ART-IVF Clinic, 119435 Moscow, Russian Federation

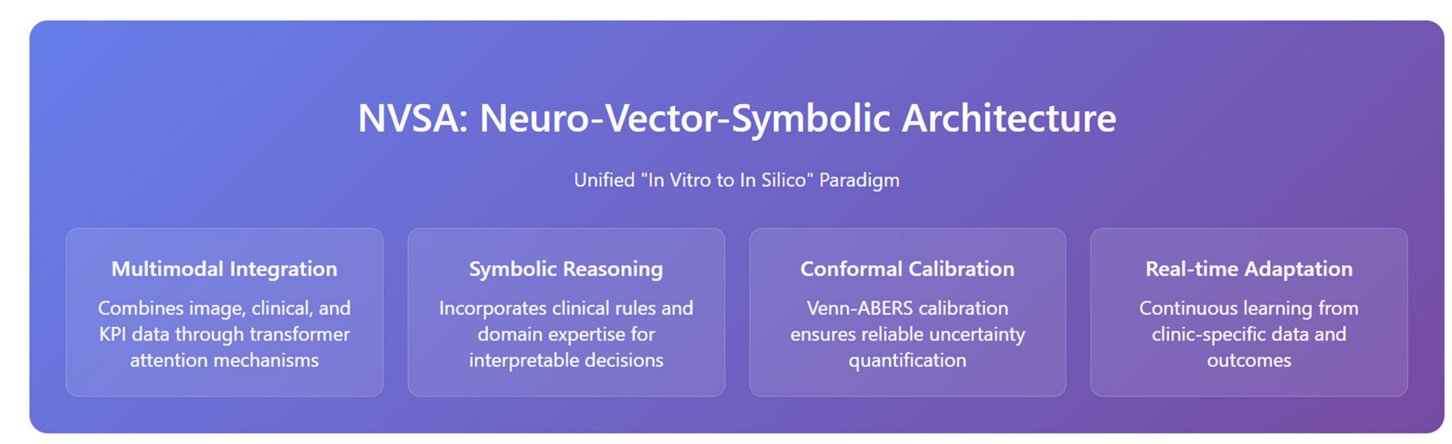
INTRODUCTION

While artificial intelligence has shown promise in IVF, current approaches often overemphasize embryo image analysis, potentially introducing unnecessary complexity and analytical bias into predictive models. The subjective nature of embryo assessment and the push toward image-based AI solutions have created a paradigm where visual data dominates. Deep learning models applied to embryo images frequently fail to achieve significant improvements over established embryologist-led selection methods with laboratory KPIs, raising questions about the clinical utility of complex imaging pipelines. This study challenges the prevailing emphasis on image-centric AI by explicitly demonstrating the limitations of excessive image integration in pregnancy prediction models. We propose a strategic delineation: while clinical and laboratory data should drive pregnancy outcome predictions, embryo imaging retains its value specifically for individual embryo implantation assessment. Our multimodal ensemble framework combining Kolmogorov-Arnold Networks (KANS) with transformer architectures (KAT) for clinical data analysis, alongside selective vision-based approaches aims to optimize the balance between predictive accuracy and clinical interpretability while avoiding the analytical pitfalls of image-heavy methodologies.

METHOD

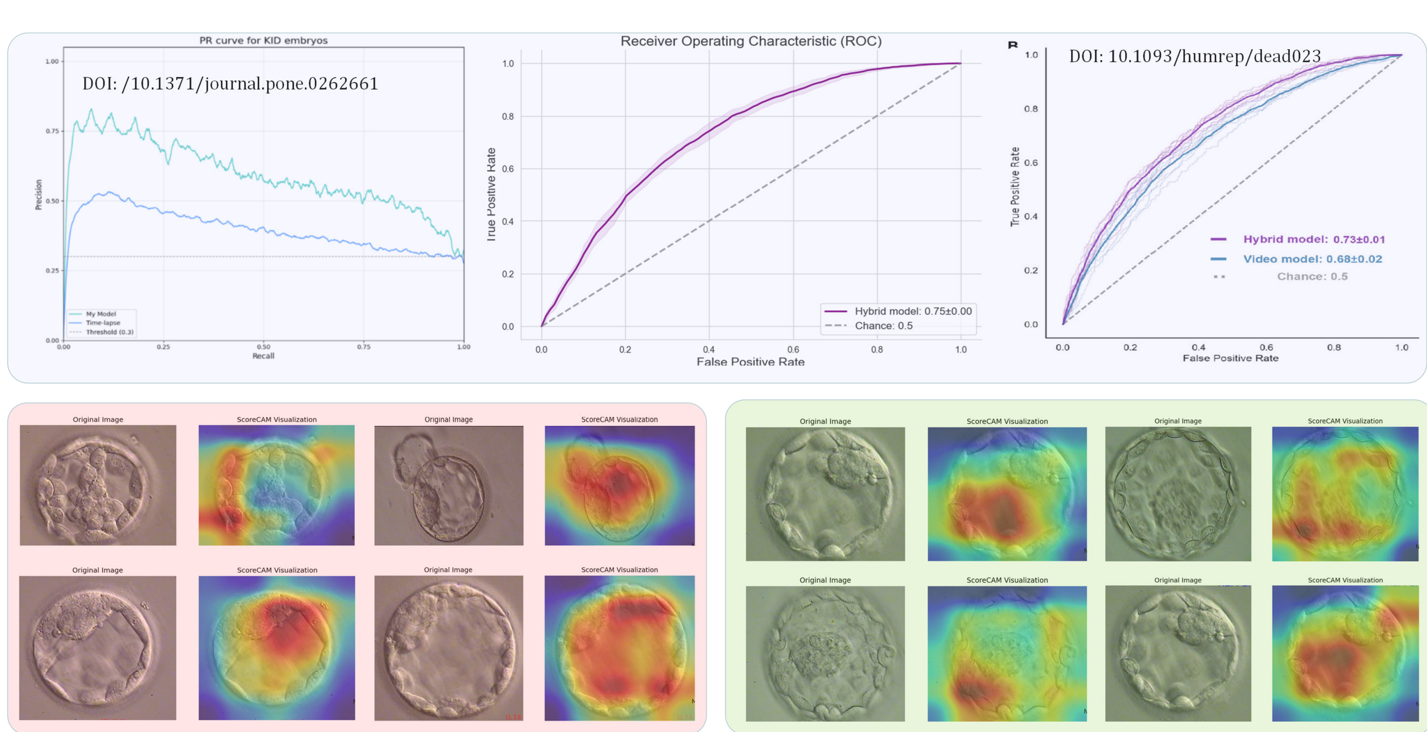
We developed parallel AI architectures using Kolmogorov-Arnold Networks combined with transformer mechanisms - "from in vitro to in silico pipeline", comparing models trained exclusively on clinical and laboratory KPI data (n=15,779 IVF cycles across six countries) versus hybrid models integrating blastocyst image analysis (n=2,344 training images, 300 test images). Image processing utilized ResNet34 convolutional networks (CNN) and Vision Transformers with DINO architecture (ViT). Model performance was assessed using area under curve (AUC), expected calibration error (ECE), and mean squared error (MSE) metrics. Cross-architectural explainability analysis was performed using Structural Similarity Index (SSIM) and Intersection over Union measurements (IoU).





RESULTS

Laboratory KPIs data-only models achieved robust predictive performance (AUC=0.75, 95% CI: 0.69-0.77) with superior calibration (ECE=0.06). Image-enhanced variants demonstrated inconsistent improvements: CNN-integrated models showed decreased performance (AUC=0.67, EGE=0.07), while ViT integration achieved higher AUC (0.78) at the cost of substantially degraded calibration (ECE=0.10, MSE=0.22). Cross-architectural analysis revealed disagreement in morphological feature prioritization, with extremely low spatial overlap (IoU=0.078) between attention regions and weak correlation (Spearman's p=-0.071) in pixel importance rankings CNN and ViT and between activation and attention maps (SSIM=0.388 and 0.059), indicating inconsistent visual reasoning strategies in image recognition models. Analysis of a test group images of single euploid blastocyst transfers, verified with commercially available AI solution ERICA from IVF2.0, revealed a 0.27 error rate in clinical pregnancy prediction (AUC = 0.67, 95% CI = 0.62-0.75) with image enhanced models.



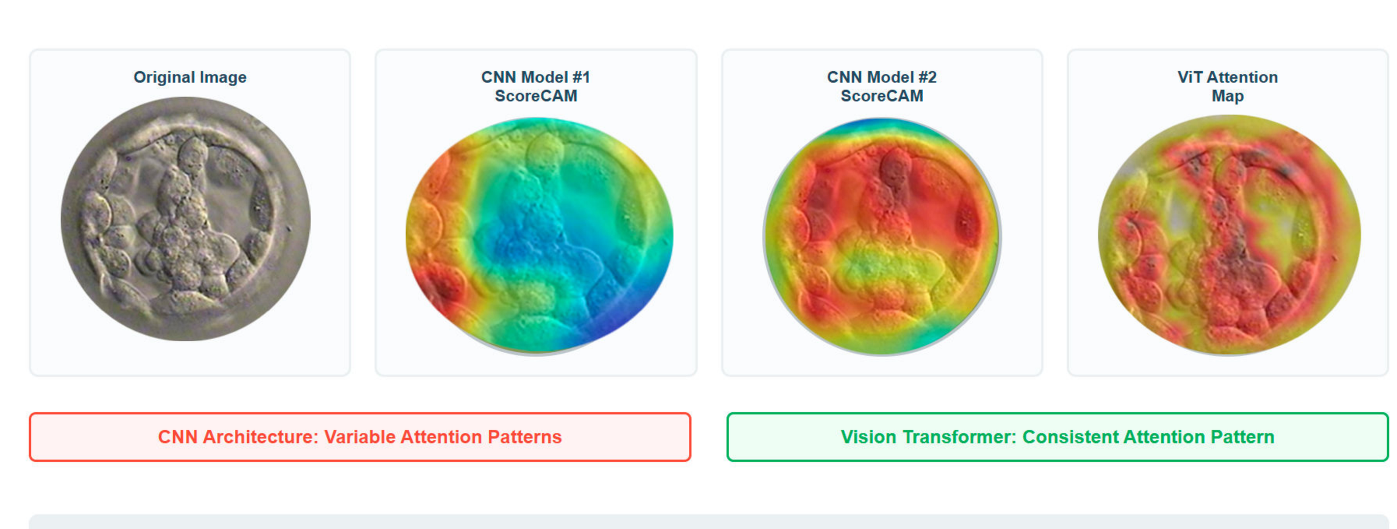
Performance metrics comparison of time-lapse CNN models and KAT predictive pipeline: same (p>0,05) ROC-AUC curves with existing TL hybrid models with enhanced blastocyst morphology patterns but with significantly higher (p<0,001) PRC curve for tabular based data. The proposed deep learning framework combines visual analysis of blastocyst morphology with clinical and laboratory metadata through a dual-branch neural architecture with NVSA adjustment for blastocyst grading for non-implanted (red) and implanted (green) patterns through attention mapping SCORE-CAM visualization.

MODEL	PRETRAINING / STRATEGY	ARCHITECTURE	AUC IMAGE RECOGNITION	AUC (95% CI) CLASSIFICATION	KEY CHARACTERISTICS
ViT (DINO)	ImageNet *Pseudo-labeling with consistency loss*	ViT-DINOv2	0.85	*N/A*	Strong attention-based performance, fast inference, robust on small datasets
ViT (MAE)	*Masked Autoencoder*	ViT-B/16	0.61	*N/A*	Reconstructive learning of spatial structure, lower performance
ViT-KAT	ImageNet *Clinical data integration*	ViT-DINOv2 + Tabular Fusion	0.78	0.73 *(0.69–0.79)*	High AUC with clinical fusion, but higher prediction error (MSE = 0.22)
Swin-S	*Visual-Temporal Contrastive Learning*	Swin-S Transformer	0.85	0.81 *(0.77-0.84)*	Best AUC, comparable to ViT, but slower inference (60 vs 45 msec)
CNN-ERICA	*ImageNet*	Convolutional ResNet Backbone	0.69	0.67 *(0.62-0.75)*	Baseline CNN
CNN-KAT	*ImageNet*	Convolutional ResNet Backbone	0.74	0.77 *(0.74-0.79)*	CNN with laboratory KPI data had high AUC score but prediction variance in attention patterns

Transformer-based Models — CNN Models

CONCLUSION

Integration of embryo imaging into AI frameworks fails to provide consistent or meaningful enhancement in pregnancy prediction accuracy. KPIs-parameter-only models demonstrate superior reliability and calibration properties. Our results emphasize the need to go beyond performance scores when validating AI systems for IVF and to include interpretability and robustness as core criteria in model selection.



REFERENCES

Sergeev S, Diakova I, Nadirashvilil L Neural networks pipeline for quality management in IVF laboratory. JIFww. 2024;2(4):1-9. doi:10.46989/001c.124947

CONTACT

Embryology lab Director, Neural network development for IVF data an...

Mail to:embryossa@gmail.com

https://rm2030.cme-congresses.com/