Transformers to Predict Embryo Quality Using Images and External Factors

Adnane Soulaimani[1,†], Carina Schwaiger[2,†], Reza Khoshkangini[3], Magnus Johnsson[4] and Thomas Ebner[5]

[1]Malmö University, Malmö, Sweden
[2]Malmö University, Malmö, Sweden
[3]Malmö University, Malmö, Sweden
[4]Kristianstad University, Kristianstad, Sweden
[5]Kepler University Clinic, Linz, Austria

Abstract

This study aims to integrate embryo images and environmental laboratory factors to predict embryo quality using a complex machine learning model. A challenge was data misalignment, which was solved by using a Random Forest Regressor to synthesise data for a complete dataset. The fine-tuned Inception V3 model with the added attention mechanisms inherent to transformers was used for multitask learning to predict three different scores for embryo quality: cell expansion (EXP), inner cell mass (ICP) and trophectoderm (TE). The model achieved an accuracy and F1 Score of 92.76%, 92.74% for EXP, 72.63%, 59.52% for TE and 63.69%, 10.77% for ICM prediction. These results indicate a great performance for 2 of the three scores and build a basis for a reliable model for prediction of embryo quality.

Keywords

Transformers, Attention Mechanism, Multitask Learning, Transfer Learning, Inception V3, Embryo Quality

1. Introduction

Choosing embryos with a high quality is an important progress for In-vitro fertilisation (IVF), a life-changing procedure for individuals and couples trying to have a baby. However, IVF can be expensive and has a limited amount of attempts, so it is important to investigate different factors that can influence embryo quality [1, 2]. The idea of this study was initially introduced by Khoshkangini et al. (2024) [3] and the focus is on external environmental factors as well as embryo images to mitigate negative influences in the future, making IVF more reliable and affordable.

Assessing the quality of embryos in their blastocyst stage can be done by using the Gardner score. This score was created and evaluated by Gardner et al. in 2000 and has been commonly used since then [4].

SAIS2025: Swedish AI Society Workshop 2025, 16-17 June 2025, Halmstad, Sweden.
†These authors contributed equally.
adnanesoulaimani92@gmail.com (A. Soulaimani); carina.schwaiger98@gmail.com (C. Schwaiger); reza.khoshkangini@mau.se (R. Khoshkangini); magnus.johnsson@hkr.se (M. Johnsson); Thomas.Ebner@kepleruniklinikum.at (T. Ebner)

This research will work with a published embryo image dataset by Kromp et al. (2023) [5, 6] together with external factors from Linz, Austria, which is the city where all images were taken, and Malmö, Sweden.

The overall approach consists of two main modules: a preprocessing pipeline for data preparation and a deep learning model for prediction. In the preprocessing stage, laboratory-specific environmental factors are synthesized using a Random Forest Regressor trained on paired external and internal environmental measurements from a reference site. This model enables the generation of temperature, humidity, and air pressure values for locations where only external data is available. These synthesized features are then combined with embryo images and Gardner scores as input to the predictive model. The model architecture itself is composed of three stages, illustrated in figure 2. In the first stage, tabular features, comprising environmental factors and Gardner scores, are processed through a feed-forward network. In the second stage, visual features are extracted from embryo images using a convolutional backbone. In the final stage, tabular and visual features are fused and passed through a classification head to predict embryo quality across three binary outcomes: TE, ICM, and EXP.

This study addresses the following research question:
RQ 1) To what extent can transformer (or attention mechanism) predict embryo quality utilizing embryo images?
This research question explores whether a machine learning model using the attention mechanism can accurately assess embryo quality based on images. The aim is to evaluate how effectively the model can learn from visual and tabular data and make reliable predictions about embryo quality.

This study gives a novel approach to predicting embryo quality using the transformers' attention mechanism and utilising both embryo images and external factors. The results from this study can be applied to make IVF more efficient and accessible. This will help couples and individuals who want to have IVF pregnancies, as well as help the medical personal doing these procedures. Showing that transfer learning performs well in a medical scenario with limited data is very important. It can be difficult to gather a large amount of data in the medical field due to its specificity and sensitivity inherent to medical data, so this approach can be applied to various other medical studies in the future.

2. Related Works

In-vitro fertilization, IVF, is a process for medically assisted reproduction where previously collected cumulus-oocyte-complexes are fertilised in a laboratory in a short-time embryo culture in vitro for up to 6 days. Then, the embryo with the best prognosis will be selected for intrauterine transfer. Extra embryos can be cryopreserved for subsequent embryo transfers [6].

For a successful IVF process, choosing an embryo with high quality is important. This is done by using the Gardner score to assess the quality of embryos in their blastocyst stage. This score was created and evaluated by Gardner et al. in 2000 and has been commonly used since then [4]. This score consists of three different parts: the blastocyst expansion, inner cell mass and trophectoderm. The blastocyst expansion is on a scale from 1-6 from blastocyst development and stage status to hatched out of shell. The inner cell mass and trophectoderm

are both assessed by how many cells the embryo has from A-C or not definable [7, 6].

The Gardner score focuses on the expansion and development of the embryo, but many other factors can have an impact on embryo development. Gardner and Kelly (2017) mention oxygen level, temperature, pH levels, and whether the embryo is alone in the culture or not as such factors [8].

There is various research analysing the influence of different environment factors, but the potential of modern machine learning models have not been harnessed in this field yet. Transformer-based models are state-of-the-art and excel at identifying long-term dependencies in data. A study by Zhao et al. (2023) created a TransFM model based on a CNN and Transformer hybrid framework. This combination of the two machine learning techniques used both the attention mechanism from transformers and the CNN's capabilities at handling images as input to achieve great results [9]. Parvaiz et al. (2023) claim that the usage of the global attention mechanism inherent to transformer models resolve long-range dependencies and can be used to decipher information in images as well [10].

Recent research has also explored how different transformer architectures can be combined or adjusted to improve performance in specific contexts. For example, models like Swin Transformers and DeiT have been tailored for medical imaging tasks where data might be limited but highly structured. The flexibility of transformers to work with sequences, images, or even mixed data types makes them very useful in this type of study. This study builds on these insights by designing and evaluating transformer-based models specifically suited for embryo quality prediction and environmental impact assessment.

A study done by Kromp et al. (2023) gathered a dataset with images of embryos on the 5th day in the IVF cycle during the blastocyst stage together with their Gardner scores to have a basis for research in IVF using machine learning. They also trained their own adapted transformer models on this dataset. The models were XCeption, Deit transformer and Swin transformer [6]. A study by Mazroa et al. (2024) created a Computer Vision-Aided Swin transformer model with Boosted-Dipper-Throated Optimisation to detect embryo development [11]. A study by Kim et al. (2024) analysed images from time-lapse videos of embryos and electronic health records of the parents to predict the embryo variability. To do so, they created their own multimodal transformer model [12].

Another approach to model training is transfer learning which can produce great results with only a short training period. This approach uses a pre-trained model and fine-tunes it for a similar problem. Even with a small amount of data, good performance can be achieved this way [13]. One study by Yusuf Abas et al. (2023) compared a CNN model and the fine-tuned VGG-16 model for classifying the quality of embryos based on their images. They showed that the model using transfer learning performed better than the CNN model. Due to the small amount of training data, they performed data augmentation to prevent overfitting and balanced the dataset for better performance [14].

To summarise, there has been various research analysing the influence of different factors on embryos, as well as using transformer models to classify the embryo quality. There has also been a study using transfer learning in the prediction of embryo quality. But there has not yet been a focus on the environmental factors for the quality using transformer's attention mechanism and transfer learning. This study will address that gap by combining usage of the transformer mechanism, transfer learning and the focus on the environmental factors for

embryo quality.

3. Data Representation

In this section, the datasets used in this study will be further explained. The first dataset was created by Kromp et al. (2023) and contains images of embryos in their blastocyst stage together with the Gardner scores. The images were taken from 2018 to 2021 in Linz, Austria. The dataset is already split into a training and testing dataset containing 2044 and 300 images respectively. The Gardner score consists of three categories: cell expansion, quality of inner cell mass, and trophectoderm [6] and is used to assess the quality of the embryo. To be able to match the images with the laboratory factors later on, the date when the image was taken was also added. This dataset is publicly available.

The second dataset consists of time-series environmental measurements recorded by two internal sensor modules in a laboratory in Malmö. These modules track factors such as temperature, humidity, CO_2 levels, air pressure, light intensity (lux), motion (PIR), and total volatile organic compounds (TVOC) across multiple years. The sensor modules are referred to using general labels for clarity.

The third data source is from the Swedish Meteorological and Hydrological Institute (SMHI) and has the data from Malmö city regarding the temperature, humidity and air pressure from 2018 - 2021 [15]. Data published by SMHI was chosen because it is easily accessible, well documented and from an official source, so of a high quality and reliable. Similar to the data taken from Malmö, the fourth data source is from Linz and also contains information about the cities temperature, humidity and air pressure in the same time frame as Malmö. This data is published by GeoSphere Austria, the Federal Institute for Geology, Geophysics, Climatology, and Meteorology in Austria, so also a reliable data source with good documentation [16]. For both Malmö and Linz' environmental factors (temperature, humidity and air pressure), it is important to only use trusted sources to ensure high data accuracy and quality.

In all datasets, if multiple sensors contained information, the mean was used to synthesise the multiple data entries. The mean was also used to resample the data to one daily value to ensure the same timespans across all datasets used in this study.

4. Proposed Approach

Laboratory-specific environmental factors are first synthesized from city-level environmental data using a Random Forest Regressor trained on paired measurements from a reference site. These synthesized features are then used as part of the input to the model. The proposed approach consists of three main stages, illustrated as modules in figure 2. In the first stage, the environmental tabular features are processed through a feed-forward network. In the second stage, visual features are extracted from embryo images using a convolutional backbone. In the third stage, the extracted tabular and image features are fused to perform binary classification of embryo quality, targeting the TE, ICM, and EXP scores. Each module of the architecture is detailed in the following sections.

4.1. Synthetic Data Generation

To generate the laboratory environmental factors for Linz, a machine learning approach was used because there was no access to real lab sensor data from that location. Instead, this study relied on data from Malmö, where both weather conditions and corresponding IVF laboratory factors were available. A Random Forest Regressor, a powerful model that can learn complex relationships, was trained to understand how Malmö's weather (temperature, humidity, and air pressure) influenced laboratory conditions. After training, this model was applied to the weather data from Linz to predict what the laboratory environment would likely have been under similar circumstances. The model performed well during validation, achieving a low mean squared error (MSE) of approximately 1.89, which indicated that it was making accurate predictions. Because of this strong performance, it can confidently be used to generate realistic and reliable synthetic laboratory factor data for Linz, allowing the project to move forward despite the lack of direct lab measurements. In figure 1, the comparison between the laboratory factors and the city environment factors can be seen for both Malmö and Linz.

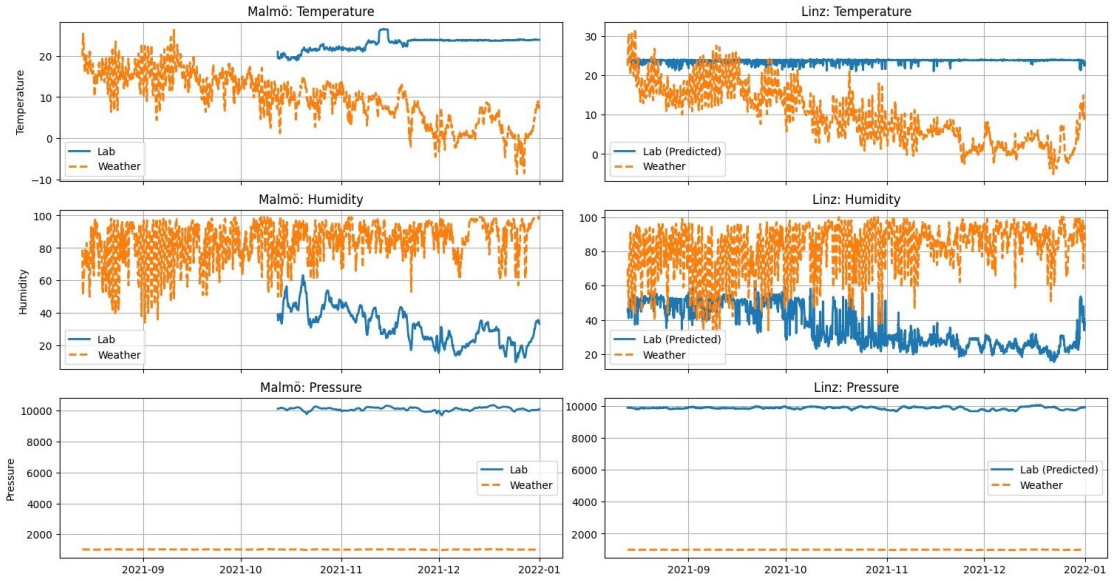

Figure 1: Laboratory and city external factors for Malmö and Linz (with synthesised Linz laboratory factors)

4.2. Model Architecture

The environmental features, including temperature, humidity, and air pressure, are first processed through a feed-forward neural network designed to encode tabular data. This stage, referred to as Module 1, applies two fully connected layers, each followed by batch normalization, ReLU activation, and dropout, to transform the raw environmental inputs into a 64-dimensional latent representation. This encoding facilitates effective integration with visual features in later stages, as illustrated in figure 2.

In parallel, embryo images are processed through a convolutional backbone based on a simplified InceptionV3 architecture. This component, forming the core of Module 2, captures multi-scale representations using a combination of 1×1, 3×3, and 5×5 convolutional filters alongside max pooling. The extracted feature maps are then refined through a feedforward block composed of fully connected layers, batch normalization, ReLU activation, and dropout, yielding high-level visual embeddings.

In Module 3, the environmental and visual embeddings are concatenated and passed through a fusion layer followed by a self-attention mechanism. The attention block models dependencies across modalities using a residual structure with a learnable scaling factor. The fused representation is then processed by a shared multi-layer perceptron to generate binary classification outputs for EXP, ICM, and TE scores, each using a sigmoid activation function. The overall fusion and prediction process is shown in figure 2.

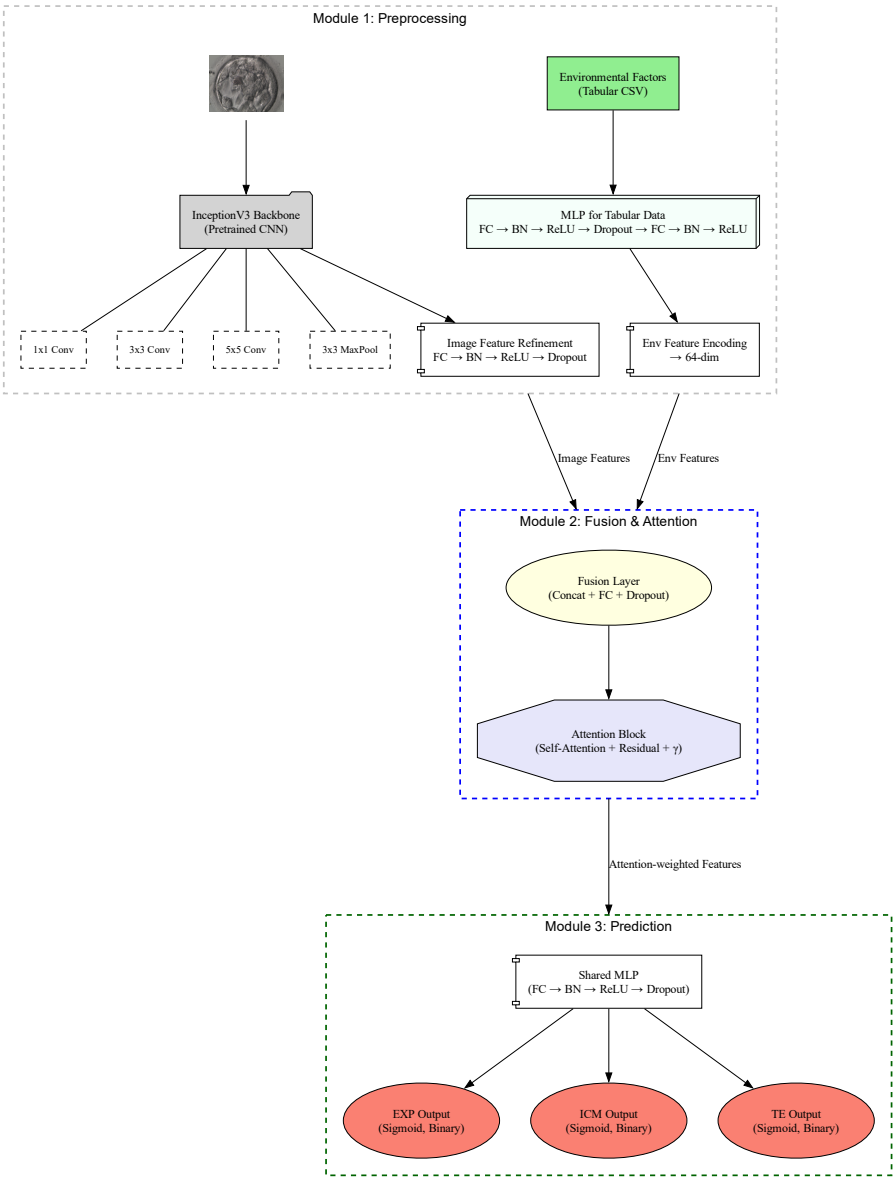

Figure 2: Embryo Quality Predictor architecture with InceptionV3; tabular features via MLP; fused with attention mechanism and multiple output heads

5. Results

This section will show and explain the results of the fine-tuned Inception V3 model that was trained for prediction on the cell expansion (EXP), inner cell mass (ICM) and trophectoderm (TE) into binary classes. The evaluation results can be seen in table 1.

The high values across all metrics for the cell expansion show that the model performs

Table 1
Results for EXP, ICM and TE predictions with attention block

Prediction	Accuracy	Precision	Recall	F1-Score	AUC
EXP	92.76%	0.95	98.22%	92.74%	0.98
ICM	63.69%	0.18	50.00%	10.77%	0.68
TE	72.63%	0.67	76.92%	59.52%	0.80

excellent for this factor with a high reliability. The trophectoderm prediction is moderate with a potential for improvement, especially in the precision of the model, so correctly identifying the positive instances. While EXP and TE are acceptable predictions, the model struggles to predict the ICM factor for the embryo quality. This can be clearly seen by an accuracy not much better than guessing and an especially poor performance regarding the precision. This means not a lot of positive cases, in this case embryos with high quality, were predicted correctly which could be explained by an imbalance in the data.

To put these results into context, the same model architecture was used, but this time without the attention-block. The performance results can be seen in table 2.

Table 2
Results for EXP, ICM and TE predictions without attention block

Prediction	Accuracy	Precision	Recall	F1-Score	AUC
EXP	95.79%	0.98	96.51%	97.36%	0.99
ICM	68.22%	0.91	39.05%	54.67%	0.71
TE	67.29%	0.93	46.34%	61.96%	0.79

While the cell expansion and inner cell mass have better accuracy without the attention block, the trophectoderm prediction worsens without the attention block. Looking at the other metrics, it is interesting to see that using the attention block improves the recall of the model, especially for TE and ICM predictions, but it worsens the precision notably for ICM prediction.

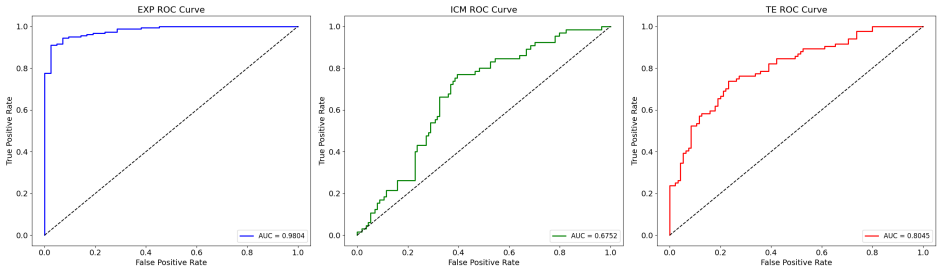

Figure 3: ROC curves for EXP, ICM and TE Prediction (with attention block)

Analysing the performance of the proposed model with the attention block, ROC curves are considered. The ROC curves can be seen in figure 3. For the EXP model, the ROC curve demonstrates exceptional discriminative ability with an AUC of 0.98. The curve rises steeply at

low false positive rates, indicating the model achieves high true positive rates even with strict classification thresholds. The TE model's ROC curve demonstrates good discriminative ability with an AUC of 0.80. The curve rises at a moderate pace and maintains reasonable distance from the diagonal reference line. Again, the performance is adequate but can be improved on. Analysing the ICM model's performance, the ROC curve shows limited discriminative ability with an AUC of 0.68. The curve rises gradually and stays relatively close to the diagonal reference line that represents random guessing.

Further delving into the ICM prediction with a visualisation of the prediction's distribution in figure 4, it becomes clear that the model gives similar scores to both good and poor ICMs. Most scores cluster around the middle (0.4), showing the model is uncertain.

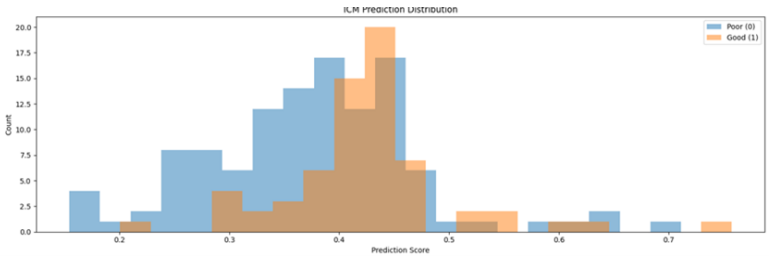

Figure 4: ICM Prediction Distribution (with attention block)

6. Conclusion

With the rising advancement of complex machine learning techniques, models can help identify embryo quality. This study proposed a complex model architecture based on the CNN Inception V3 model and an MLP neural network in combination that use the attention mechanism to predict three different scores using a multitask learning approach. This model was used to predict the quality of embryos in three scores EXP, ICM and TE based on their images and the environmental factors (temperature, humidity and air pressure) in the laboratory. The preliminary results show that the model achieved an accuracy of 92.76% for EXP, 72.63% for TE and 63.69% for ICM. This shows that the model performed excellent for classifying EXP scores, acceptable for TE score and leaves much room for improvement for ICM scoring. Comparing the model with and without the attention block shows that utilising the attention mechanism leads to better recall of the model, indicating a better ability to correctly identify embryos with a high quality. The reason for the models poor ICM performance should be further analysed in future work and possible data imbalance problems mitigated to improve performance. This will help build a reliable model for all three parts of the Gardner score. Replacing the Inception V3 model with another pre-trained model and comparing the results would also be interesting, just like utilizing a transformers model trained only for these tasks to widen the attention mechanism through the whole architecture.

Acknowledgments

This ongoing research is conducted as part of the EIVF-AI project funded by Vinnova, the Swedish Governmental Agency for Innovation Systems (Grant No.2024-00088).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (GPT-4) to assist with grammar and spelling checks and to improve clarity of phrasing in some sections. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Gardner, W. Schoolcraft, In-vitro culture of human blastocysts, in: R. Jansen, D. Mortimer (Eds.), Towards Reproductive Certainty: Infertility and Genetics Beyond, Parthenon Publishing, 1999, pp. 378–388.

[2] S. Vashevnik, B. Dall, R. Frydman, Influence of environmental factors on ivf success rates: A comprehensive review, Reproductive Biology and Endocrinology 19 (2021) 20.

[3] R. Khoshkangini, E. Mangrio, M. Johnsson, Enhancing in vitro fertilization with environment optimization utilizing artificial intelligence (eivf-ai), in: Proceedings of EAI Pervasive Health 2024, EAI, Heraklion, Crete, Greece, 2024.

[4] D. K. Gardner, M. Lane, J. Stevens, T. Schlenker, W. B. Schoolcraft, Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer, Fertil. Steril. 73 (2000) 1155–1158.

[5] Kaglle, Human blastocyst dataset for ivf, https://www.kaggle.com/datasets/iamshahzaibkhan/human-blastocyst-dataset-for-ivf, 2024.

[6] F. Kromp, R. Wagner, B. Balaban, V. Cottin, I. Cuevas-Saiz, C. Schachner, P. Fancsovits, M. Fawzy, L. Fischer, N. Findikli, B. Kovačič, D. Ljiljak, I. Martínez-Rodero, L. Parmegiani, O. Shebl, X. Min, T. Ebner, An annotated human blastocyst dataset to benchmark deep learning architectures for in vitro fertilization, Sci. Data 10 (2023) 271.

[7] N. Nasiri, P. Eftekhari-Yazdi, An overview of the available methods for morphological scoring of pre-implantation embryos in in vitro fertilization, Cell J. 16 (2015) 392–405.

[8] D. K. Gardner, R. L. Kelley, Impact of the ivf laboratory environment on human preimplantation embryo phenotype, Journal of Developmental Origins of Health and Disease 8 (2017) 418–435. doi:10.1017/S2040174417000368.

[9] L. Zhao, G. Tan, B. Pu, Q. Wu, H. Ren, K. Li, TransFSM: Fetal anatomy segmentation and biometric measurement in ultrasound images using a hybrid transformer, IEEE J. Biomed. Health Inform. PP (2023) 1–12.

[10] A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, M. M. Fraz, Vision transformers in medical computer vision—a contemplative retrospection, Engineering Applications of Artificial Intelligence 122 (2023) 106126. URL: https://www.sciencedirect.com/science/article/pii/S095219762300310X. doi:https://doi.org/10.1016/j.engappai.2023.106126.

[11] A. A. Mazroa, M. Maashi, Y. Said, M. Maray, A. A. Alzahrani, A. Alkharashi, A. M. Al-Sharafi, Anomaly detection in embryo development and morphology using medical computer vision-aided swin transformer with boosted dipper-throated optimization algorithm, Bioengineering 11 (2024). URL: https://www.mdpi.com/2306-5354/11/10/1044. doi:10.3390/bioengineering11101044.

[12] J. Kim, Z. Shi, D. Jeong, J. Knittel, H. Y. Yang, Y. Song, W. Li, Y. Li, D. Ben-Yosef, D. Needleman, H. Pfister, Multimodal learning for embryo viability prediction in clinical ivf, in: M. G. Linguraru, Q. Dou, A. Feragen, S. Giannarou, B. Glocker, K. Lekadir, J. A. Schnabel (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, Springer Nature Switzerland, Cham, 2024, pp. 542–552.

[13] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, K. Meenakshi, K. Swaraja, P. Plawiak, U. Rajendra Acharya, Transfer learning techniques for medical image analysis: A review, Biocybernetics and Biomedical Engineering 42 (2022) 79–107.

URL: https://www.sciencedirect.com/science/article/pii/S0208521621001297. doi:https://doi.org/10.1016/j.bbe.2021.11.004.

[14] Y. A. Mohamed, U. K. Yusof, I. S. Isa, M. M. Zain, An automated blastocyst grading system using convolutional neural network and transfer learning, in: 2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE), 2023, pp. 202–207. doi:10.1109/ICCSCE58721.2023.10237105.

[15] SMHI, Data och analyser för väder samt Sveriges klimat och miljö| SMHI — smhi.se, https://www.smhi.se/data, ???? [Accessed 22-02-2025].

[16] G. Austria, GeoSphere Austria Data Hub — data.hub.geosphere.at, https://data.hub.geosphere.at/dataset/klima-v2-1m, 2024. [Accessed 22-02-2025].