Abou Bekr Belkaid University
Tlemcen Algeria
جامعة أبي بكر بلقايد
تلمسان الجزائر
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
الجمهورية الجزائرية الديمقراطية الشعبية
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
وزارة التعليم العالي و البحث العلمي
FACULTY OF TECHNOLOGY
DEPARTMENT OF BIOMEDICAL ENGINEERING

A Thesis

Presented for obtaining the degree of:

MASTER

In: Biomedical Engineering
Specialty: Biomedical and Hospital Informatics

By:

Aissa BenFettoume Souda

Theme

Development of an intelligent system for the automatic classification of embryo in In Vitro Fertilization

Thesis defended on June 25, 2025 at Tlemcen in Front of the Jury Composed of:

Mr Tarik Taleb	MCB	University of Tlemcen	President
Mme Souaad Hamza-Cherif	MCA	University of Tlemcen	Supervisor
Mme Nesma Settouti	MCA	LabISEN - Yncréa Ouest	Co-Supervisor
Mr Gaouar Adil	MAA	University of Tlemcen	Examiner

Academic Year 2024/2025

"Whoever follows a path in pursuit of knowledge, Allah will make a path to Paradise easy for him."

– Prophet Muhammad (PBUH)

Dedication

First and foremost, I would like to extend my sincere gratitude to **Allah** for granting me the strength, patience, and guidance necessary to complete this work.

I extend my deepest gratitude to **my parents**, who have supported me in every aspect of my life, always stood by me, and gave me everything they could. Their endless love and sacrifices are what have brought me to where I am today.

To **my sister**, I owe a huge thank you for her constant encouragement, moral support, and faith in me, which has been invaluable throughout this journey.

To **my friends**, thank you for your wonderful company through thick and thin. Your friendship, shared laughter, and unforgettable memories are treasures I will always cherish.

I also extend my sincere gratitude to **my professors**, who generously dedicated their time and effort to teaching me not only academic knowledge but also valuable life lessons. Their guidance has shaped me into a more conscientious and responsible person.

This humble work would not have been possible without the guidance and support of **Mme. Hamza-Cherif Souaad**, **Mme. Settouti Nesma**, and an **Anonymous** individual whose contributions have greatly contributed to my progress. They have not only carried this humble work with me, but have also guided, taught, and illuminated my path. Their impact on me is deep, and words cannot describe my gratitude. Their kindness is a beautiful burden on me, one I hope to repay one day.

– Aissa BenFettoume Souda

Acknowledgements

First, and foremost, I would like to express my heartfelt thanks to **Mme. Settouti Nesma**. for her continuous support, guidance, and supervision throughout this work. Her expertise, patience, and encouragement were incredibly valuable from the very beginning, and her dedication helped me stay focused and motivated during every step of this thesis. I am especially grateful for the trust she placed in me and for inspiring my interest in this fascinating topic.

I would also like to sincerely thank **Mme. Hamza-Cherif Souaad** for her valuable academic input, thoughtful feedback, and kind encouragement. Her insightful comments helped me improve the quality and direction of this work, and I truly appreciated her clear explanations and professional advice throughout the process.

I would like to give a special thanks to an **anonymous contributor** who provided important technical help, especially with the coding part of the project. Their input, suggestions, and shared resources made a real difference in reaching the final outcome.

My gratitude also goes to **LabISEN - Yncréa Ouest** for offering a productive and supportive research environment. The resources and atmosphere there played an important role in the completion of this work.

Finally, I would like to thank the members of the jury **Mr. Taleb Tarik**, **Mr. Gaouar Adil** as well as my supervisors **Mme. Settouti Nesma** and **Mme. Hamza-Cherif Souaad**, for taking the time to read, review, and evaluate my work. Your contributions are deeply appreciated.

– Aissa BenFettoume Souda

Abstract

This Master thesis addresses the challenge of embryo selection in IVF, aiming to improve pregnancy outcomes through more objective methods. It proposes an intelligent system for automatically classifying human embryo developmental stages using time-lapse imaging. The study compares 2D and 3D CNNs to assess spatial information and explores temporal models like TimeSformer to capture embryo dynamics. A hierarchical classification strategy is also introduced to handle 15 developmental phases. Results show that while 2D CNNs provide a solid baseline, temporal models significantly outperform them by leveraging morphokinetic features, highlighting the potential of AI to enhance accuracy and consistency in embryo selection.

keywords: In Vitro Fertilization (IVF), Embryo Classification, Deep Learning, Time-Lapse Imaging, Convolutional Neural Networks (CNN), Vision Transformer, TimeSformer, Temporal Modeling, Hierarchical Classification

Résumé

Ce travail de master aborde le défi de la sélection d'embryons en FIV, visant à améliorer les résultats de grossesse grâce à des méthodes plus objectives. Elle propose un système intelligent de classification automatique des stades de développement de l'embryon humain par imagerie accélérée. L'étude compare les réseaux neuronaux convolutif 2D et 3D pour évaluer les informations spatiales et explore des modèles temporels comme TimeSformer pour capturer la dynamique embryonnaire. Une stratégie de classification hiérarchique est également introduite pour gérer 15 phases de développement. Les résultats montrent que si les réseaux neuronaux conjoncturels 2D constituent une base solide, les modèles temporels les surpassent nettement en exploitant les caractéristiques morphocinétiques, soulignant ainsi le potentiel de l'IA pour améliorer la précision et la cohérence de la sélection d'embryons.

mots-clés: Fécondation in vitro (FIV), classification des embryons, apprentissage profond, imagerie accélérée, réseaux de neurones convolutifs (CNN), transformateur de vision, TimeSformer, modélisation temporelle, classification hiérarchique

مُلخّص

تتناول مذكرة الماستر هذه تحدي اختيار الأجنة في التلقيح الاصطناعي (IVF) ، بهدف تحسين نتائج الحمل من خلال أساليب أكثر موضوعية ودقة. تقترح المذكرة نظامًا ذكيًا للتصنيف الآلي لمراحل تطور الجنين البشري باستخدام التصوير بالفيديو بالفاصل الزمني (time-lapse) . تقارن الدراسة بين الشبكات العصبية الالتفافية ثنائية الأبعاد (2D CNN) وثلاثية الأبعاد (3D CNN) لتقييم أهمية المعلومات المكانية في تحليل مراحل التطور، كما تستكشف النماذج الزمنية مثل نموذج (TimeSformer) لالتقاط الديناميكيات التطورية للجنين عبر الزمن. بالإضافة إلى ذلك، تم تقديم استراتيجية تصنيف هرمية لمعالجة ١٥ مرحلة تطورية مختلفة للجنين. تُظهر النتائج أنه بينما توفر شبكات (2D CNN) خط أساس قويًا، تتفوق النماذج الزمنية بشكل ملحوظ من خلال استغلال الخصائص المورفوكينية والزمنية، مما يؤكد إمكانات الذكاء الاصطناعي في تحسين دقة واتساق عملية اختيار الأجنة.

الكلمات المفتاحية: الإخصاب في المختبر (IVF) ، تصنيف الأجنة، التعلم العميق، التصوير بالفاصل الزمني، الشبكات العصبية الالتفافية (CNN) ، المحول البصري (Vision Transformer) ، (TimeSformer) ، النمذجة الزمنية، التصنيف الهرمي.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

- **2D**: Two-Dimensional
- **3D**: Three-Dimensional
- **AI**: Artificial Intelligence
- **ART**: Assisted Reproductive Technology
- **AUC**: Area Under the Curve
- **B**: Blastocyst
- **C**: Cleavage
- **CNN**: Convolutional Neural Network
- **DenseNet**: Densely Connected Convolutional Network
- **DeiT**: Data-efficient Image Transformer
- **F0**: Central Focal Plane
- **GRU**: Gated Recurrent Unit
- **GNN**: Graph Neural Network
- **ICM**: Inner Cell Mass
- **ICSI**: Intracytoplasmic Sperm Injection
- **IVF**: In Vitro Fertilization
- **LSTM**: Long Short-Term Memory
- **LightGBM**: Light Gradient Boosting Machine
- **M**: Morula
- **OHSS**: Ovarian Hyperstimulation Syndrome
- **PGT-A**: Preimplantation Genetic Testing for Aneuploidy
- **R(2+1)D**: Residual (2+1)D Convolutional Network
- **ResNet**: Residual Network
- **RNN**: Recurrent Neural Network
- **STORK**: Inception-V1 CNN for Blastocyst Classification

- **SVM**: Support Vector Machine
- **TLI**: Time-Lapse Imaging
- **TLM**: Time-Lapse Monitoring
- **TE**: Trophectoderm
- **TimeSformer**: Time-Attention Transformer for Video Understanding
- **TransUNet**: Transformer-based U-Net for Medical Image Segmentation
- **UNet**: U-shaped Convolutional Neural Network for Biomedical Image Segmentation
- **ViT**: Vision Transformer
- **XCeption**: Extreme Inception Model
- **Z**: Pronuclei

Introduction

In Vitro Fertilization (IVF) has significantly changed the field of Assisted Reproductive Technology (ART), offering renewed hope to countless couples striving to achieve parenthood. IVF is a way that effectively help human eggs and sperm meet outside the body to aid in achieving a successful pregnancy. Yet, despite its promise, one of the most critical and unresolved challenges in IVF remains the selection of the most viable embryo for transfer.

Identifying which embryo has the highest potential for implantation and healthy development is a complex task. Traditionally, embryologists have relied on Humain observation and subjective morphological assessments practices, despite years of refinement, remain limited in accuracy and consistency. The continued reliance on such subjective evaluations has led to high rates of false positives and false negatives, ultimately contributing to missed opportunities and suboptimal outcomes in many clinics. Even in well-intentioned and rigorously managed centers, embryo selection often suffers from imprecision, leading to failed cycles that cannot always be ethically or clinically justified.

This master thesis addresses these limitations by proposing the development of an intelligent system designed to automatically classify embryos based on their implantation potential. To achieve this goal, we rely on deep neural networks, computer vision algorithms, and established techniques from the field of Artificial Intelligence (AI). These tools are leveraged to create a robust and

interpretable model capable of assisting clinicians in making more informed and objective decisions during IVF cycles.

This study is focused on three main goals:

- Enhance the prediction of embryo viability using deep learning models applied to time-lapse imaging data.
- Improve the accuracy of embryonic phases classification by exploring and comparing various deep learning methods.
- Leverage the temporal information embedded in time-lapse sequences by investigating transformer-based models to enhance the precision of embryo phase detection.

This work is part of a broader research project conducted in collaboration with **Brest University Hospital**, and **LabISEN - Yncréa Ouest**. The goal is to create a fully dependable and practical system for choosing embryos using advanced deep learning techniques that can minimize diagnostic mistakes and enhance patient results.

Throughout this work, each section aims to provide a clear and structured presentation of the key concepts and techniques that form the foundation of our system, highlighting their role within the overall analytical and decision-making pipeline. The structure of this research is organized as follows: **Chapter 1** gives an overview of the clinical setting, explaining the IVF process, past techniques for choosing embryos, and the difficulties involved in evaluating embryo quality. **Chapter 2** provides a detailed overview, examining current AI-based methods in embryology, the datasets that are accessible to the public and a description of the dataset we used. This chapter wraps up by pointing out important research gaps and laying out our proposal to tackle them. In **Chapter 3**, we go over

our experimental methods and the findings from our exploration of various deep learning architectures, which are central to the technical contributions of this study. In **conclusion**, this section wraps up the main findings, discusses the goals that were met, and suggests possible areas for future research.

Chapter I: Foundations of In Vitro Fertilization

"The way to get started is to quit talking and begin doing."

– Walt Disney

1 Introduction

IVF is a medical procedure that has enabled many couples facing infertility to fulfill their desire to have children. One of the most critical steps in the IVF process is the selection of the embryo to be transferred, as this decision significantly influences the chances of a successful pregnancy.

This chapter provides an overview of the key stages involved in the IVF procedure. It also traces the evolution of embryo selection practices, from early approaches based solely on visual assessment to more recent techniques incorporating advanced imaging and genetic testing. Finally, the chapter highlights the limitations and challenges associated with current methods, underscoring the need for more objective, accurate, and scalable solutions in embryo evaluation.

2 In Vitro Fertilization and Embryo Selection

2.1 Overview of the clinical IVF process

IVF has changed the world of ART for the rising number of infertile couples seeking to create a family. Globally, it is estimated that 1 in 6 couples experience infertility at some point in their lives [1], which accounts for about 17% of the adult population. This intricate suite of clinical and laboratory methods manipulates human eggs and sperm outside the body in an attempt to facilitate a successful pregnancy. Procedures differ based on each couple's circumstances, but a typical IVF cycle typically follows several major phases, as illustrated in Figure 1[1].

Figure 1: Key Stages in the In Vitro Fertilization Procedure

The steps of IVF are :

[1] https://www.imprimisivfsrinagar.com/what-are-the-five-stages-of-ivf/

- **Ovarian stimulation** : The IVF process begins with controlled ovarian stimulation using injectable gonadotropins to encourage the development of multiple follicles. Unlike a natural cycle that produces one egg, IVF aims to retrieve 10–15 mature oocytes to enhance success rates. The stimulation protocol is individualized based on age, ovarian reserve, and clinical profile to balance efficacy with safety, particularly to reduce the risk of ovarian hyperstimulation syndrome (OHSS).
- **Oocyte and Spemen retrieval** : Once follicles reach optimal maturity, oocyte retrieval is performed under sedation using transvaginal ultrasound guidance. A fine needle is used to aspirate oocytes from the ovarian follicles. On the same day, a semen sample is collected from the partner, preparing them both for fertilization.
- **Fertilization** : In the laboratory, mature oocytes are fertilized with processed sperm via conventional insemination or Intracytoplasmic Sperm Injection (ICSI) a technique often used for male factor infertility. Fertilization is assessed 18 hours later by the presence of two pronuclei, indicating successful formation of a zygote.
- **Embryo culture and transfer** : After fertilization, embryos are cultured in specialized incubators that replicate the uterine environment. Their development is continuously monitored from the cleavage and morula stages through to the blastocyst stage, typically reached by day 5 or 6. This developmental progression enables embryologists to evaluate and select embryos with the highest implantation potential. Selected embryos are then carefully transferred into the uterus using a fine catheter, often guided by ultrasound. A pregnancy test is performed approximately 9–10 days later. High-quality surplus embryos may be cryopreserved for future transfer, thereby increasing the cumulative chances of a successful pregnancy.

2.2 Embryo selection: historical development

Since its introduction over forty years ago, embryo selection has been a major idea in IVF. Two main early events combined to create it: the clinical requirement to lower the number of embryos transferred to prevent multiple pregnancies and the use of gonadotropin stimulation to improve embryo availability. Too many embryos for a safe transfer produced a natural requirement for choosing the most viable embryo.

The early apparent success of simple selection strategies led to a strong belief in the "**selectability**" of embryos [2], therefore motivating the field to seek more sophisticated methods. Support from professional associations and major corporate funding really boosted this perspective and accelerated the research and application of embryo selection technologies in clinical settings.

As time has passed, studies have progressively improved the methods for selecting embryos, evolving from simple visual evaluations to more refined technologies like time-lapse imaging and genetic testing. Even though some advancements have been achieved, current research is focused on enhancing accuracy and customizing selection strategies for each patient.

2.3 Major embryo selection strategies

With the aim of enhancing IVF results, different approaches have been created and put into practice over time to evaluate embryo quality and choose the most suitable candidates for transfer. These can be broadly categorized:

- **Morphological assessment**: This is the most widely adopted grading method for evaluating embryos at the blastocyst stage (typically on Day 5 or 6 of in vitro development). Known as the Gardner grading system [3], it involves the visual assessment of three key morphological features under a microscope: the degree of blastocyst expansion, the quality of the inner cell mass (ICM), and the appearance of the trophectoderm (TE). Each embryo receives a composite score, such as 4AB, where the first digit indicates the expansion stage, the first letter reflects ICM quality, and the second letter indicates TE quality. Table 1 summarizes the Gardner grading scheme used in clinical embryology.

Table 1: Gardner Grading System for Blastocyst-Stage Embryo

Grading Component	Grade	Description
Expansion Stage	1	Early blastocyst: cavity $< 50\%$ of embryo volume
	2	Blastocyst: cavity $> 50\%$ of embryo volume
	3	Full blastocyst: cavity completely fills the embryo
	4	Expanded blastocyst: cavity larger than early embryo, zona pellucida is thinning
	5	Hatching blastocyst: trophectoderm begins to herniate through the zona
	6	Hatched blastocyst: embryo has completely escaped from the zona pellucida
Inner Cell Mass (ICM)	A	Tightly packed, many cells
	B	Loosely grouped, several cells
	C	Very few, poorly organized cells
Trophectoderm (TE)	A	Many cells forming a cohesive epithelium
	B	Few cells forming a loose epithelium
	C	Very few, irregularly shaped cells
Example Grade	4AB	Expanded blastocyst (Grade 4) with Grade A ICM and Grade B TE

- **Preimplantation genetic testing for Aneuploidy (PGT-A)**: This technique involves biopsying a small number of cells from the embryo (usually from the TE at the blastocyst stage) and testing them for chromosomal abnormalities (aneuploidy). The aim is to select chromosomally normal (euploid) embryos for transfer, as aneuploidy is a major cause of implantation failure and miscarriage [4].
- **Morphokinetics using Time-Lapse Imaging (TLI)**: TLI systems incorporate cameras into incubators, allowing for continuous monitoring and recording of embryo development without removing embryos from their stable culture environment. This generates detailed data on the timing of cell divisions and other dynamic developmental events (morphokinetics), which can be used for assessment [5].
- **Omics technologies**: These methods focus on the molecular profile of the embryo or its environment. Examples include metabolomics (analyzing small

molecules consumed or released by the embryo into the culture medium) and proteomics (analyzing proteins secreted by the embryo - the secretome) [6]. These are typically non-invasive approaches seeking molecular biomarkers of viability.

Each of these strategies aims to overcome the limitations of the previous ones, particularly the subjectivity and limited predictive power of traditional morphology. However, as will be discussed in the following section, each method also presents its own unique set of challenges, limitations, and controversies regarding its clinical utility and impact on IVF outcomes.

3 Challenges in Embryo quality assessment

Even though choosing embryos has been a key part of IVF for many years, figuring out how to assess embryo quality and predict how well they will develop is still a big challenge in ART. Even with all the new technology, the assessment methods we have today, from the old-school grading systems to genetic tests, still have major drawbacks that limit how useful and trustworthy they are in clinical settings.

3.1 Limitations of traditional morphological grading

Morphological assessment is about looking at the structure of embryos at certain stages of development, usually on Day 2/3 during the cleavage stage and Day 5/6 when they reach the blastocyst stage. Grades are assigned based on a variety of conditions including number of cells, symmetry, fragmentation and blastocyst expansion, according to grading systems such as those developed by Gardner and Schoolcraft, as described in Section 2.3. However, this technique has some disadvantages:

- **Variability in observations**: Embryologists can produce a expansive variation in results and thus even two observations from the same observer may not be the same. Although international guidelines (e.g., Istanbul Consensus) [7] attempt to standardize criteria, inconsistencies across clinics persist.
- **Limited predictive accuracy**: Morphological grading provides only moderate predictive value for implantation and live birth outcomes, with usual very low Accuracy values.
- **Static nature**: Evaluations rely on observations from a single moment, overlooking significant dynamic developmental occurrences like the timing of

division and synchrony.

Despite these limitations, morphological grading remains widely utilized because it is simple, non-invasive, and cost-effective. Still, its limitations show that we really need better and fairer ways to evaluate things.

3.2 Challenges of emerging technologies

A bunch of ideas have been developed to tackle the issues with outdated methods, but they face some pretty tough challenges too:

- **Time-Lapse Imaging**: allows for ongoing observation of embryo development, recording complex morphokinetic details while ensuring stable culture conditions. Research shows that using TLI for selection doesn't really improve outcomes when compared to the usual morphological assessment [8]. The effort needed for manual analysis and the high cost of equipment make it challenging for many to use it broadly. The primary value of TLI could instead be in the development of AI tools by providing detailed datasets.
- **Omics technologies**: such as Metabolomics and Proteomics have been developed to assess embryo viability non-invasively by evaluating molecular signatures in collected spent culture media. While metabolomic and proteomic profiles can show if an embryo is viable, the current methods are complex, expensive, and not really ready for routine clinical use at this point. Identifying reliable biomarkers that are effective across various patients and circumstances remains a significant challenge.

4 Conclusion

Embryo selection remains one of the most critical and complex challenges in the IVF process. For a long time, scientists would just look at embryos to grade them, but this method isn't always accurate and can be inconsistent. More recent advances, like genetic testing and time-lapse videos offer more information, but they are expensive, complicated, and don't always improve success rates. The key takeaway is that even with new tools, the process of selecting the most viable embryo is still far from being an exact science. Looking ahead, a more effective strategy will likely involve integrating multiple sources of information : morphological, temporal, and genetic, potentially enhanced by artificial intelligence, to support more informed and reliable

clinical decisions.

The next chapter provides a detailed review of recent advances in AI-based embryo selection systems, with particular emphasis on available datasets, model architectures, and the main challenges in developing clinically robust solutions.

Chapter II: Related works of AI-based approaches for Embryo selection in IVF

"If we knew what it was we were doing, it would not be called research, would it?"

– Albert Einstein

1 Introduction

This chapter provides an overview of the current state of research on the use of AI in embryo selection for IVF. While AI holds significant promise for improving decision-making in this context, several critical challenges are holding it back. One major limitation lies in the restricted accessibility of training data. Most datasets used to develop AI models are private, limiting transparency, reproducibility, and the ability to benchmark different approaches. Furthermore, many existing models are trained on data from only a few clinics, so they may not work well for a wider population. Additional obstacles include the reliance of some AI systems on single static images, rather than leveraging the full temporal dynamics available in time-lapse embryo imaging. Moreover, the lack of interpretability—often referred to as the "black box" nature of AI—can lead to reluctance among clinicians to trust and adopt these tools.

In light of these issues, this chapter reviews the most relevant AI-based approaches in the literature, highlights their strengths and limitations, and identifies key methodological gaps. These insights form the basis for the experimental strategies proposed in this thesis, which aim to address current limitations through improved spatial-temporal modeling and more interpretable, clinically relevant solutions.

2 Studies review and gap analysis

The application of artificial intelligence AI in embryo assessment for IVF has seen significant advancements, yet several challenges persist. A primary concern is the limited availability of publicly accessible, annotated datasets. Many studies, such as the early work by Bormann et al.[9] with low-cost imaging systems or the extensive dataset described by Zhylko et al. [10] which, while detailed, remains private, rely on proprietary data. This common practice, as pointed by Kromp et al. [11] (2023) in their appeal for public benchmarks ,makes it difficult to validate externally and limits how well AI models can be applied in different clinical environments. Notably, the dataset provided by Gomez et al. [12] stands out because it includes high-resolution, multi-focal time-lapse videos along with detailed annotations, which makes it a really important resource in a field where data is often limited..

Data diversity remains another critical issue. AI models, like the one for automated blastomere counting developed by Moradi Rad et al. [13], are often trained on data

from a limited number of clinics, raising concerns about their applicability to broader populations. For example, Khosravi et al. [14] developed the STORK model using time-lapse and clinical data, achieving high accuracy but underscoring the necessity of multi-center datasets to ensure robust model generalization when they tested it on external clinic data.

Temporal modeling of embryo development is essential for capturing dynamic biological processes, yet some AI approaches rely on static images or a few discrete time points, potentially missing critical developmental transitions. Borna et al. [15] looked into using three static images taken from various early time points. On the other hand, Liao et al. [16] showed that using longer time-lapse sequences can help capture morphokinetic changes better and enhance predictive performance. the The benefits of including temporal information were also highlighted by Wang et al. [17] who examined different multi-frame architectures using time-lapse data to analyze embryo morphology

The "black box" nature of many deep learning models remains a barrier to clinical adoption, as clinicians are often hesitant to rely on cloudy technology. While explainable AI techniques have been proposed to address this issue, their incorporation into embryo assessment workflows is still limited and and more effort is required to build practitioners' trust. Bormann et al. [18] emphasized the importance of consistency and objectivity that AI can bring, which is crucial for building clinician confidence. The importance of AI in maintaining quality control, As demonstrated by Bormann et al. [19], the significance of AI in upholding quality control is linked to the requirement for trustworthy and transparent models.

Combining imaging data with patient-specific clinical information is a step closer in enhancing embryo assessment. Studies like those by Raef et al. [20] and Bormann et al. [18] have shown that integrating clinical and image-based features can significantly rise predictive accuracy. Patil et al. [21] also showed improved clinical pregnancy prediction by combining static images with clinical records. Other studies have focused purely on clinical or embryologist-derived data, using traditional machine learning techniques as seen in Blank et al. [22] with Random Forests or Uyar et al. [23] with Naive Bayes, further highlighting the diverse data types available for predictive modeling. However, refined multi-modal data fusion, for example combining 3D graph representations from focal stacks as explored by He et al. [24] with clinical data, remains an area with

significant opportunity for future researchs.

To overcome these challenges, there is a pressing need for the development and sharing of diverse, well-annotated public datasets. Initiatives like that of Kromp et al. [11], who utilized transformer-based architectures on publicly available static images, exemplify efforts toward transparency and reproducibility. Continued advancements in temporal modeling and explainable AI, along with strategic multi-modal integration, will be critical to creating reliable, transparent, and clinically applicable embryo assessment tools that can ultimately improve IVF outcomes.

3 Datasets and AI-based Embryo classification studies

Some publicly accessible embryo imaging datasets have enabled recent advances in AI-based embryo viability and morphology prediction. A recent public dataset is from Kromp et al. (2023) [11], which consists of 2,344 static blastocyst-stage images (PNG format) annotated with Gardner grading criteria (expansion, inner cell mass, trophectoderm) These images (available via Figshare [1]) enable training of DL models for blastocyst morphology.another dataset is the one provided by Gomez et al. A Time-Lapse Video Dataset [2] that is among the most prominent public resources. It comprises 704 developing human embryos recorded over 0–6 days post-fertilization with multi-focal imaging [12], providing a rich source of temporal information for morphokinetic analysis. Deep learning models leveraging this dataset have demonstrated strong predictive performance:

- Kalyani et al. (2024) [25] trained a hybrid CNN–RNN model using ResNet50 for feature extraction followed by a GRU, analyzing daily TLM frames from Day 0 to Day 3 to predict blastocyst formation. Their model achieved 93% accuracy on a held-out validation set.
- Barhoun et al. (2025) [26] classified human embryo developmental stages using an enhanced R(2+1)D (3D CNN) model for initial predictions from optimized time-lapse datasets, followed by dynamic programming (Viterbi algorithm) to refine stage sequences across 15 developmental stages. Their model achieved 93.3% overall accuracy on their test set.

However, to the best of our knowledge, the applications of Vision Transformer (ViT)

[1] https://figshare.com/articles/figure/Blastocyst_dataset_zip/20123153/3?file=39348899
[2] https://zenodo.org/records/6390798#:~:text=

and/or other transformer-based architectures to the Gomez dataset remain limited. In addition, several other publicly available and multi-center datasets have also supported AI development for embryo evaluation:

- Tran et al. [27] introduced the IVY model, a CNN trained on 10,638 time-lapse videos from 8 IVF clinics internationally. IVY directly predicts fetal-heart pregnancy from raw TLM videos, achieving a high average AUC of 0.93 across 5-fold cross-validation and comparable performance on external validation sets. This model outperformed traditional morphokinetic scoring and showed robustness across clinics.
- Khosravi et al. [14] developed STORK, an Inception-V1 CNN trained on 1,930 time-lapse images from Weill Cornell. STORK classified good vs. poor quality blastocysts with 96.94% image-level accuracy and AUC 0.987, and 97.5% accuracy at the embryo level using a voting scheme.

Although public datasets are rare, multi-center collaborations and private clinical datasets have been pivotal in advancing AI-based embryo evaluation. Table 2 summarizes the datasets, reported performance metrics, and their availability, highlighting the diversity of AI techniques and clinical outcomes explored—predominantly using private data.

Table 2: Overview of studies employing AI for embryo evaluation using different AI techniques

Author	Data Type	AUC	AI Technique	Output	Available
Embryo assessment at blastocyst stage					
Chavez-Badiola et al. (2020) [28]	Static Images	79	Deep Neural Network	Embryo aneuploidy	No
VerMilyea et al. (2020) [29]	Static Images	64	ResNet-Inception	Embryo morphology	No
Kragh et al. (2019) [30]	Time-lapse	69	CNN-RNN	Embryo morphology	No
Rad et al. (2019) [31]	Static Images + Clinical Record	71	Compact-Contextualize-Calibrate	Clinical pregnancy	No
Bormann et al. (2020a) [18]	Static Images + Clinical Record	90	CNN	Embryo morphology	No
Blank et al. (2019) [22]	Clinical Record	74	Random Forest	Clinical pregnancy	No
Bormann et al. (2020b) [9]	Static Images	52	CNN	Embryo morphology	No
Khosravi et al. (2019) [14]	Time-lapse + Clinical Record	98	STORK (Inception-V1)	Embryo morphology	No
Loewke et al. (2022) [32]	Static Images	75	ResNet ×3	Morphology—Ongoing pregnancy	No
Borna et al. (2024) [15]	Static Images	75	DeepEmbryo	Clinical Pregnancy	No
Embryo assessment at cleavage stage					
Coticchio et al. (2021) [33]	Time-lapse	83	LSTM	Embryo morphology	No
Rad et al. (2018) [13]	Static Images	94	UNet	Embryo morphology	No
Bormann et al. (2021) [19]	Static Images	59	CNN	Embryo morphology	No
Wu et al. (2021) [34]	Static Images	74	CNN	Embryo morphology	No
Liao et al. (2021) [16]	Time-lapse	72	DenseNet201 + Focal Loss + LSTM	Embryo morphology	No
Uyar et al. (2009) [35]	Clinical Record	71	SVM	Clinical pregnancy	No
Uyar et al. (2015) [23]	Clinical Record	80	Multiple Classifiers	Clinical pregnancy	No
Hariton et al. (2021) [36]	Clinical Record	68	LightGBM (Decision Tree)	Clinical pregnancy	No
Raef et al. (2020) [20]	Clinical Record	90	Random Forest	Clinical pregnancy	No
Kromp et al. (2023) [11]	Static Images	74	XCeption + DeiT + Swin Transformer	Embryo morphology	Yes
He et al. (2023) [24]	Static Images	62	GNN	Embryo morphology	No
Wang et al. (2023) [17]	Time-lapse	77	ResNet-R(2+1)D	Embryo morphology	No
Embryo assessment at cleavage and blastocyst stage					
Patil et al. (2019) [21]	Clinical Record + Static Images	86	CNN	Clinical pregnancy	No
Sawada et al. (2021) [37]	Time-lapse	67	ResNet + Attention	Live birth	No
Tran et al. (2021) [27]	Time-lapse	93	CNN	clinical pregnancy	No
Full Developmental Stage Assessment of Embryo					
Gomez et al. (2022) [38]	Time-lapse	73	R2Plus1D + LSTM	Embryo phase	Yes
Zhylko et al. (2024) [10]	Time-lapse + Clinical Record	N/A	N/A	Embryo phase + Clinical Pregnancy + Embryo Morphology	No

4 Gomez et al's Dataset

4.1 Dataset description

In this master thesis, we focus one the dataset introduced by Gomez et al [12], which represents one of the most substantial publicly accessible resources for time-lapse embryo analysis. This dataset consists of 704 time-lapse videos of developing human embryos sourced from ICSI cycles conducted at Nantes University Hospital. Image acquisition was performed using an Embryoscope time-lapse system, which captured images every 10 to 20 minutes with a 635 nm LED light source and Hoffman's contrast modulation optics. For each of the 704 embryos, videos are available from seven distinct focal planes [12], resulting in a total of 2.4 million images. The images are provided as 500×500 pixel, grayscale JPEG files.

A key feature of this dataset is its highly detailed annotations, which were performed by qualified embryologists. These annotations mark the specific timing of 16 distinct cellular events based on the nomenclature proposed by Ciray et al [39]. The events track the entire development from second polar body appearance (tPB2) and pronuclei dynamics (tPNa, tPNf) through all cleavage stages (t2 to t9+), compaction (tM), and finally blastocyst formation, expansion, and hatching (tSB, tB, tEB, tHB) as Illustrated in Figure 2.

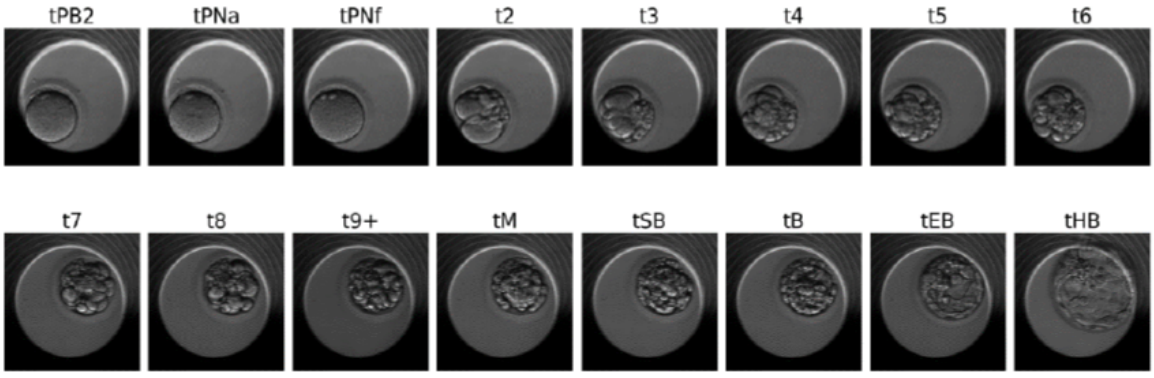

Figure 2: All of the 16 development stages presented in the dataset.

To make the data suitable for deep learning, these event timings were used to assign a classification label to every frame of the videos. A frame showing a specific event is labeled as such, while subsequent frames are assigned the label of the most recent event until a new one occurs. The dataset deliberately includes both 499 embryos considered viable for transfer and other embryos that were discarded due to poor

development, capturing a wide variety of normal and abnormal morphokinetic features. This granular, frame-by-frame labeling is designed to support the development and comparative evaluation of deep learning models for automated embryo assessment.

4.2 Focal Plane

The dataset provides comprehensive imaging by including seven different focal planes for each of the 704 embryo videos (see Figure 3). An embryo is a three-dimensional object, and the ability to change the focal plane of the microscope allows for a better visualization of its structure. The available focal planes are:

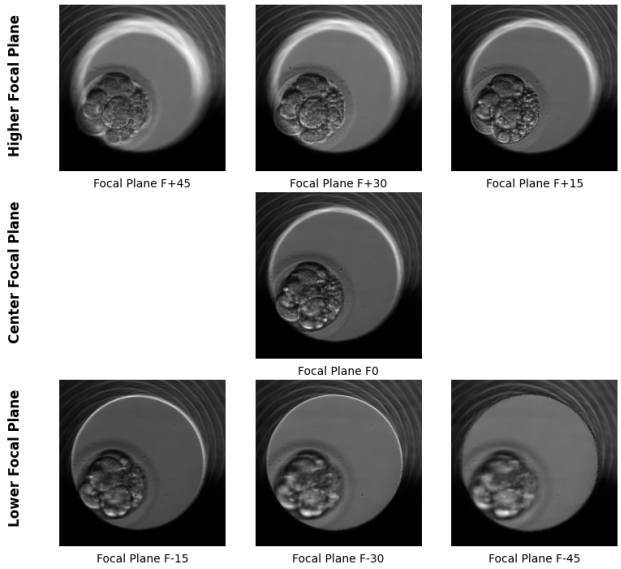

Figure 3: An Illustration Example of the seven focal planes presented in this dataset ranging from F- 45 to F+45.

- F-45, F-30 and F-15: are lower focal planes, capturing deeper layers of the embryo.
- F0: is The central focal plane, providing the most balanced and sharpest image.
- F+15, F+30 and F+45: are higher focal planes, capturing the outer layers of the embryo.

The data is organized into separate compressed folders for each focal plane. The main folder, embryo_dataset.tar.gz, contains the videos from the central focal plane, F0. Six other folders, named embryo_dataset_X.tar.gz (where X is one of the alternative plane labels), contain the videos recorded at those specific focal settings. Each of these folders contains the full set of 704 videos, each captured entirely at that specific focal plane.

4.3 Issues related to Gomez et al. dataset

While the dataset provided by Gomez et al. represents a significant and valuable contribution to the field of embryo analysis, it presents several fundamental challenges that complicate the training of deep learning models—particularly in terms of data imbalance and data quality.

The dataset exhibits two primary forms of imbalance :

1. There is a significant disparity in the number of images associated with each developmental stage (see Figure 4),

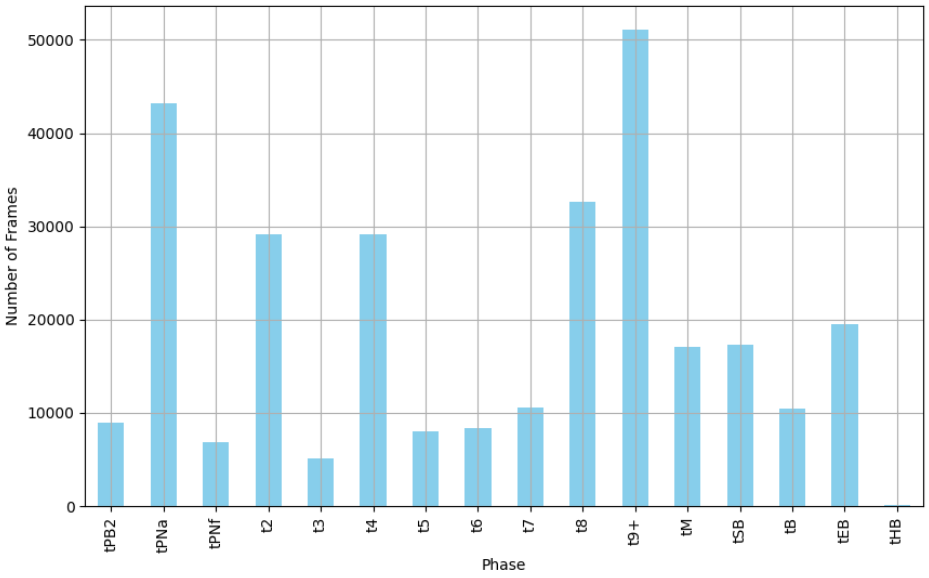

Figure 4: Distribution of Frames Across Phases.

2. The number of annotated stages varies widely between videos, ranging from 6 to 16 (see Figure 5). This can lead to inadequate training for underrepresented stages like t3, t5, and t7. Furthermore, the dataset suffers from data quality problems, including misannotated images and the labeling of blurry or unclear frames. In some videos, all frames were found to be incorrectly annotated as belonging to a single, earlier phase, while in others, frames were simply annotated incorrectly (see Figure 6).

These issues necessitated careful data refinement to ensure the accuracy and reliability of the training process.

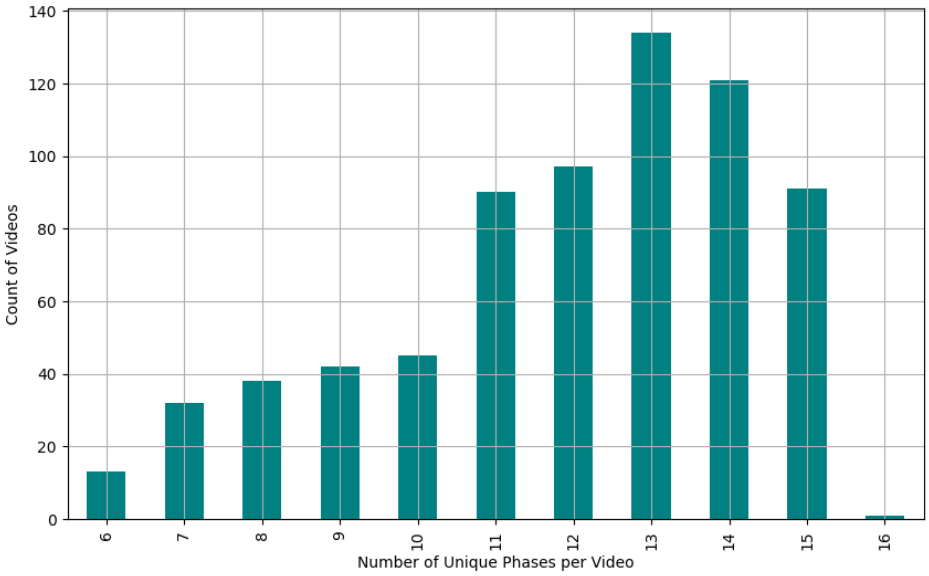

Figure 5: Distribution of Phase Counts Across Videos.

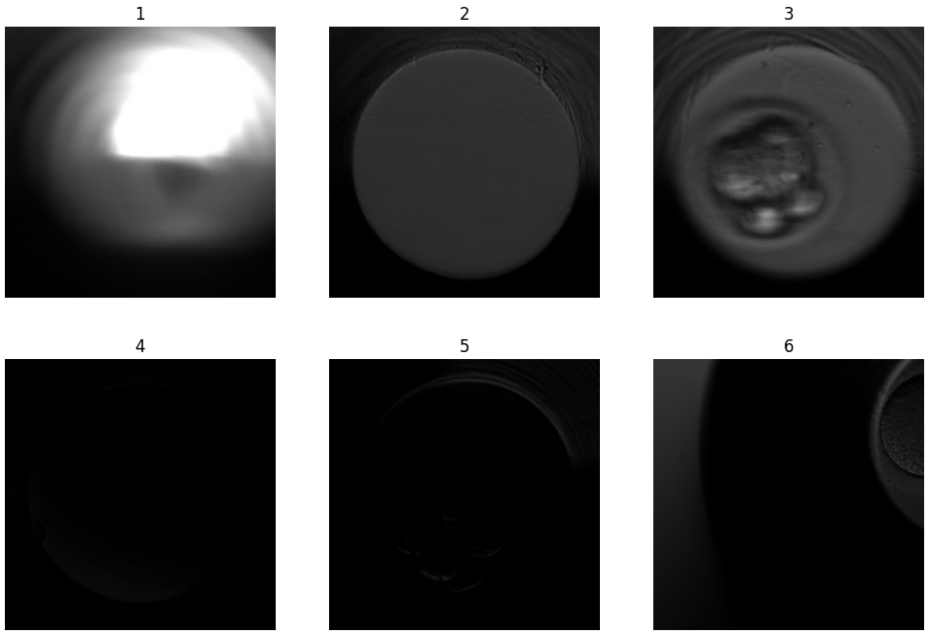

Figure 6: Examples of corrupted images include: (1) overexposed frames, (2) empty images mislabeled as tEB, (3) blurred frames, (4) underexposed (excessively dark) images, (5) unidentifiable images mislabeled as valid stages, and (6) misaligned frames with the camera.

5 Research positioning and proposed approach

In light of the gaps and limitations identified in the current state of the art, particularly regarding data imbalance, static image analysis, lack of temporal modeling, and the opacity of AI systems—this work proposes a multi-faceted experimental framework to advance the field of automated embryo classification in IVF.

Our proposed approach directly addresses the shortcomings of traditional pipelines by introducing more expressive architectures and targeted strategies tailored to the complexity of embryo development. The core methodology is structured around three main objectives:

- **First**, to establish the value of comprehensive spatial information, we will directly compare the performance of a traditional 2D CNN model with a 3D CNN model that can process volumetric data from multiple focal planes.
- **Second**, to address the need for robust temporal modeling, we will evaluate and contrast a baseline temporal model with a state-of-the-art Vision Transformer architecture, the TimeSformer, to better capture the dynamic development of the embryo.
- **Third** and Finally, to tackle the complexity of classifying numerous developmental stages, we will introduce and test a hierarchical approach 'Divide and Conquer' strategy, which aims to improve accuracy by training specialized models on distinct phases of embryonic development.

6 Conclusion

This chapter has outlined the current landscape of AI-based approaches for embryo selection, highlighting both their potential and their limitations. Despite promising advances, challenges related to data availability, generalizability, and model interpretability persist. The dataset proposed by Gomez et al., while valuable, also illustrates the typical issues faced when training AI models on real-world clinical data.

In response to these challenges, we have introduced a structured experimental framework designed to improve embryo classification through enhanced spatial and temporal modeling, as well as stage-specific specialization. The next chapters detail the implementation of these methods, the preprocessing pipeline, and the evaluation strategies used to assess their clinical relevance and robustness.

The next chapter will present a detailed description of the proposed methods, the experimental setup, and the results obtained, along with a critical evaluation of their performance and clinical relevance.

Chapter III: Experiments and results analysis

"If I have seen further, it is by standing on the shoulders of giants."

– Isaac Newton

1 Introduction

This chapter presents the experimental framework and results of our study on automated embryo stages classification using AI-based models. To this end, we begin by introducing the concept of transfer learning and the various deep learning models employed, along with their respective advantages in the context of medical image analysis.

The main goal of this work is to identify the most effective strategies to automatically classify the stages of embryo development from time-lapse images. To address this, we evaluate how different AI methods can capture both spatial characteristics (embryo morphology) and temporal dynamics (development over time). To this end, three main experiments were designed :

- First, we compare 2D and 3D convolutional neural networks (CNNs) to assess their ability to model spatial information from multiple focal planes.
- Second, we evaluate temporal modeling by applying both sequential and transformers-based architectures to video sequences of embryo development.
- Third, we implement a hierarchical "divide and conquer" strategy, that classified general stages and then refined predictions into more detailed sub-stages.

This chapter details the design, implementation and outcomes of each experiment, and highlighting the key insights that pave the way for more accurate and efficient AI-based embryo stage classification in future research.

2 Theoretical background : transfer learning and pretrained models

Transfer learning refers to the process of adapting a pre-trained model, typically trained on a large and general dataset for a new but related task with often limited data. In the context of CNNs, this approach has become essential in domains such as medical imaging, where acquiring large, annotated datasets is challenging due to privacy concerns, expert labeling requirements, and clinical variability [40], [41].

At its core, transfer learning assumes that knowledge acquired from one domain (the source domain) can be reused to improve learning in another domain (the target domain). Typically, a CNN pretrained on ImageNet (which contains over 1.2 million

images across 1,000 categories) serves as the starting point (see Figure 7). The early convolutional layers in such models learn low-level generic features such as edges, corners, and textures—features that are widely applicable across visual tasks. As one moves deeper into the network, the learned features become increasingly abstract and task-specific, but they can still provide useful representations for related domains, including medical imaging [40].

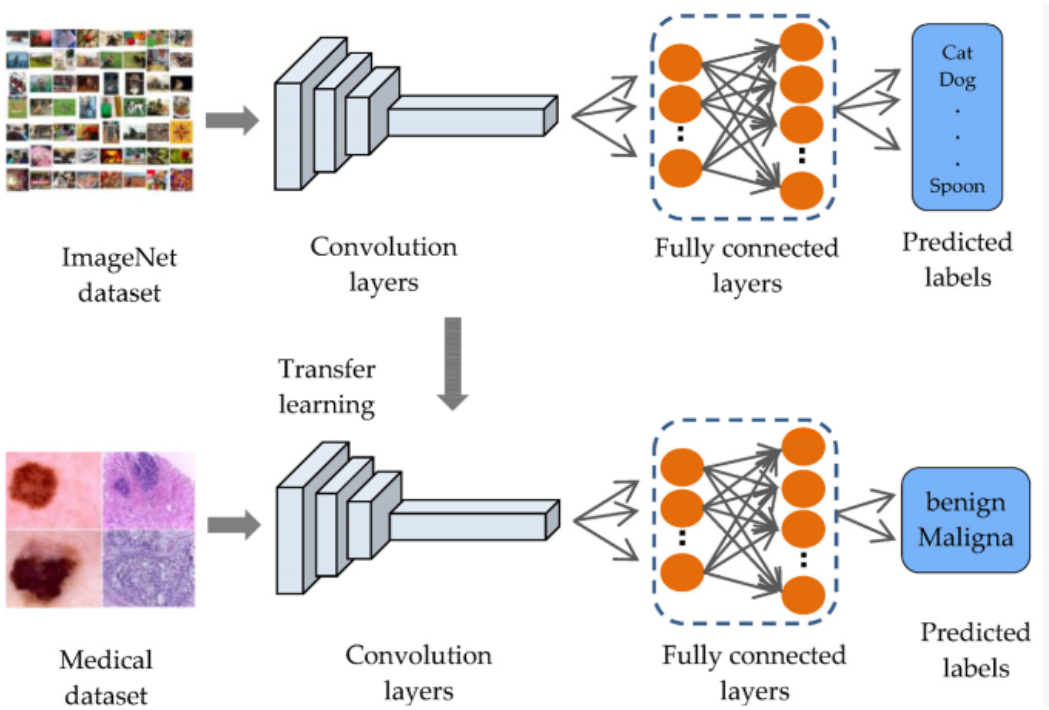

Figure 7: transfer learning from ImageNet [42]

2.1 Classical transfer learning with CNNs

Transfer learning with CNNs remains a widely adopted strategy in medical image analysis. Given that medical datasets are typically small, imbalanced, or difficult to annotate, leveraging pretrained CNNs such as VGGNet [43], ResNet [44], DenseNet [45], and Inception [46] enables the reuse of low-level visual features learned from large-scale datasets like ImageNet [47]. These features often generalize well to grayscale modalities like X-ray, CT, or MRI, despite domain discrepancies.

Two principal strategies dominate classical transfer learning:

- **Feature extraction**: Only the pretrained convolutional base is used to encode the input image into a feature vector, which is then fed into a new classifier (e.g., a fully connected layer).

- **Fine-tuning**: The entire network or a subset of layers is retrained on the medical dataset to adapt the learned features to domain-specific patterns [41].

Studies show that fine-tuning deeper layers often yields superior performance, especially when enough domain-specific labeled data is available [48]. Nonetheless, CNNs inherently focus on local spatial features due to the limited receptive fields of convolutional kernels.

2.2 Vision transformers (ViTs) and hybrid architectures

Vision Transformers (ViTs) have revolutionized the field by modeling global image dependencies using self-attention instead of convolution. ViTs segment an image into fixed-size patches (e.g., 16×16 pixels), linearly embed them, and treat the sequence of patches like words in natural language processing [49].

In medical imaging, ViTs have demonstrated promising performance in domains where long-range contextual relationships are essential, such as in histopathological image analysis, tumor localization in high-resolution 3D scans, or whole-slide image classification [50].

However, pure ViT models require large amounts of labeled data and significant computational resources. As a compromise, hybrid models like TransUNet [51], UNETR [52] use CNNs to extract low-level features and Transformers to model high-level semantic interactions, achieving a balance between performance and data efficiency [52]. For tasks involving temporal image sequences, such as cardiac cine MRI, endoscopic videos, or fetal ultrasound, temporal context is vital. The TimeSformer architecture [53] extends Vision Transformers to video data by decomposing spatiotemporal self-attention into spatial and temporal components. In clinical applications, TimeSformer has proven effective in detecting motion anomalies, classifying cardiac function, and interpreting ultrasound videos frame-by-frame [54].

2.3 Advantages and challenges of transfer learning

Transfer learning offers multiple benefits in the context of deep CNNs and medical image analysis:

- Reduced Data Requirements: Since the model starts with pretrained weights, high performance can be achieved with smaller labeled datasets [40].
- Improved Generalization: The pretrained features often improve classification

accuracy, especially when data is limited or noisy [41].

- Lower Computational Cost: Requires significantly less training time and hardware resources compared to training from scratch [40].

but despite its benefits, transfer learning is not universally beneficial. Key challenges include:

- Domain Shift: Significant differences in data distributions (e.g., natural images vs. grayscale medical scans) can hinder performance, especially if not addressed via domain adaptation techniques [41].
- Architectural Constraints: Pretrained models may not be well-suited to the specific needs of the target task (e.g., input size, number of output classes, or modality differences) [40].
- Fine-Tuning Complexity: Deciding which layers to freeze and which to retrain, as well as setting appropriate learning rates, requires domain expertise and empirical tuning [55].

3 Data preparation

As mentioned earlier in Chapter 2, Section 4.3, the dataset suffers from several issues, including significant data imbalance, critical data quality concerns, and misannotated images. To solve these problems, a comprehensive cleaning and balancing approach was implemented.

First, each video was explored frame-by-frame. If a video contained blurry, underexposed, or overexposed frames, all frames from the beginning of the problematic segment to the end of the video were systematically removed. For example, if an issue was identified starting at frame 239 in a 400-frame video, frames 239 through 400 were discarded. Due to its extreme scarcity, the hatched blastocyst (tHB) phase was also excluded from the analysis, reducing the total number of classes to 15. However, videos with misannotated images were kept, as we did not have the embryological expertise to re-annotate them accurately. The final number of frames remaining after this cleaning process is presented in Table 3.

Subsequently, the dataset was prepared for experiments using two distinct methods, each utilizing a 70% training, 15% validation, and 15% testing split. The detailed distribution of frames for each phase under both approaches is shown in Table 4.

- **Phase-Level split**: In this approach, a random sampling strategy was used to construct balanced training, validation, and testing sets. This method ensures that each of the 15 classes is equally represented, which mitigates model bias caused by the original dataset's imbalance.
- **Embryo-Level split**: In this second approach, the dataset was divided by video. The 704 embryos were split into distinct training, validation, and testing sets, ensuring that all frames from a single embryo belong exclusively to one set, which prevents data leakage between sets.

Table 3: Summary of frame counts Before and After data cleaning

Phase	**Raw Data**	**Deleted Frames**	**New Data**
tPB_2	8895	6	8889
tPNa	43244	1	43243
tPNf	6793	10	6783
t2	29189	136	29041
t3	5127	232	4895
t4	29177	641	28536
t5	8045	51	7994
t7	10531	118	10413
t8	32602	1387	31215
$t9^+$	51112	1138	49974
tM	17084	716	16368
tSB	17244	843	16401
tB	10387	1666	8721
tEB	19535	3161	16374
tHB	97	51	46
Total	**297428**	**10281**	**287147**

Table 4: Distribution of Frames per Phase for Phase-Level and Embryo-Level Splits

Phase	**Phase-Level Split**			**Embryo-Level Split**		
	Train	**Val**	**Test**	**Train**	**Val**	**Test**
tPB_2	3427	734	734	6110	1459	1320
tPNa	3426	735	734	30257	6197	6789
tPNf	3427	734	734	4774	944	1065
t2	3427	734	734	20252	4379	4422
t3	3426	734	735	3357	510	1028
t4	3427	734	734	20243	4179	4114
t5	3426	734	735	5581	1070	1343
t6	3426	735	734	5927	1021	1294
t7	3427	734	734	6922	1176	2315
t8	3426	735	734	21994	5253	3968
$t9^+$	3426	735	734	35582	7641	6751
tM	3426	734	735	11773	2539	2056
tSB	3426	734	735	11178	2611	2612
tB	3427	734	734	6038	1309	1374
tEB	3427	734	734	11612	2190	2572
Total	**51398**	**11013**	**11014**	**201600**	**42478**	**43023**

4 Experimental approaches and models design

This section outlines the experimental strategies and model architectures developed to address the challenges of automated embryo stage classification. Each approach was designed to explore specific aspects of spatial, temporal, and hierarchical modeling within the context of time-lapse embryo imaging. Figure 8 summarizes our workflow.

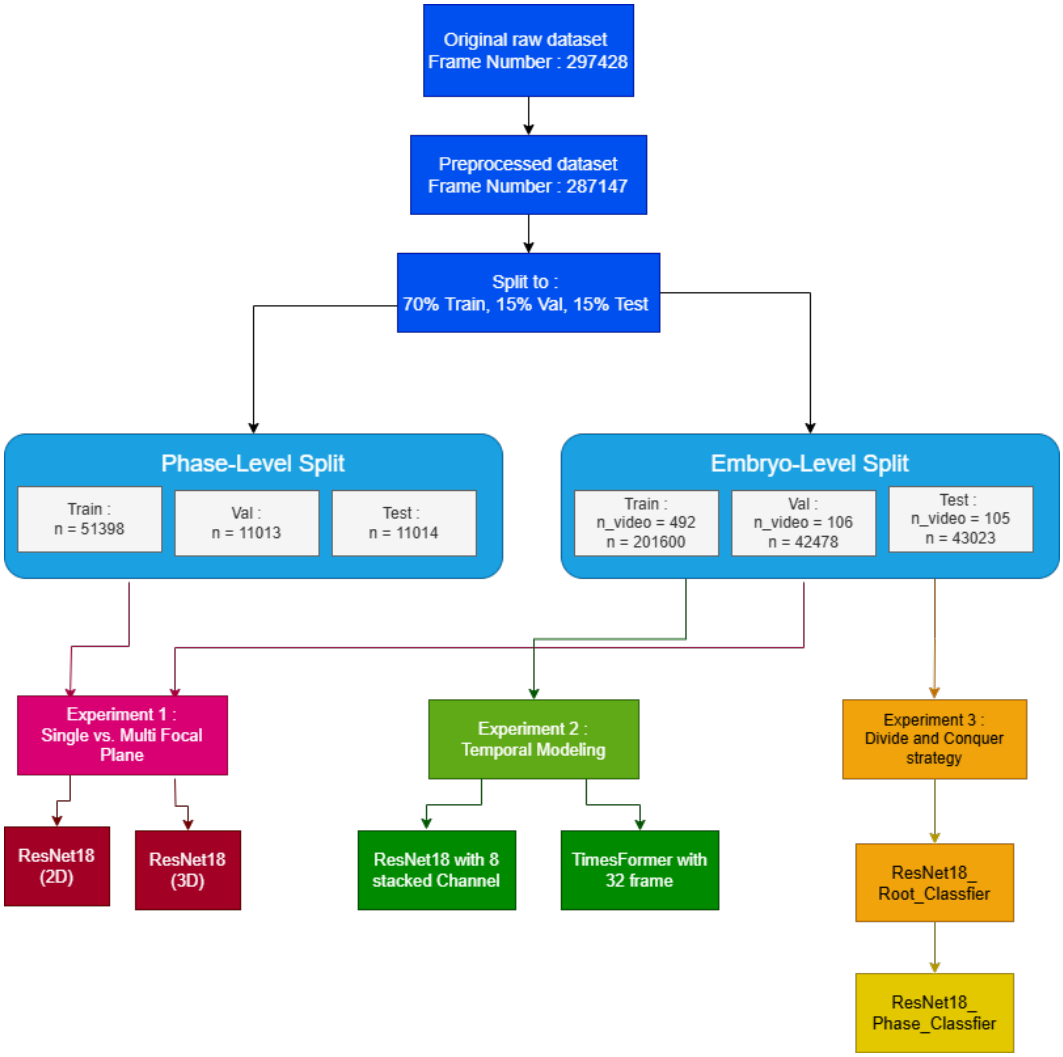

Figure 8: Global work-flow from raw dataset to our experimental approach.

4.1 Experiment 1: Single focal plane vs. Multiple focal plane

The experiment was designed to compare two types of deep learning models 2D and 3D for classifying embryo development stages. The goal was to see how adding depth (3D information) affects model performance. We used the popular ResNet-18 architecture for both approaches (see Figure 9). ResNet was chosen because it solves the vanishing gradient problem, which used to make training deep networks difficult. It does this using "skip connections" that help the model train more effectively by allowing gradients to flow through the network more easily [44]. This makes ResNet a strong choice for challenging image classification tasks.

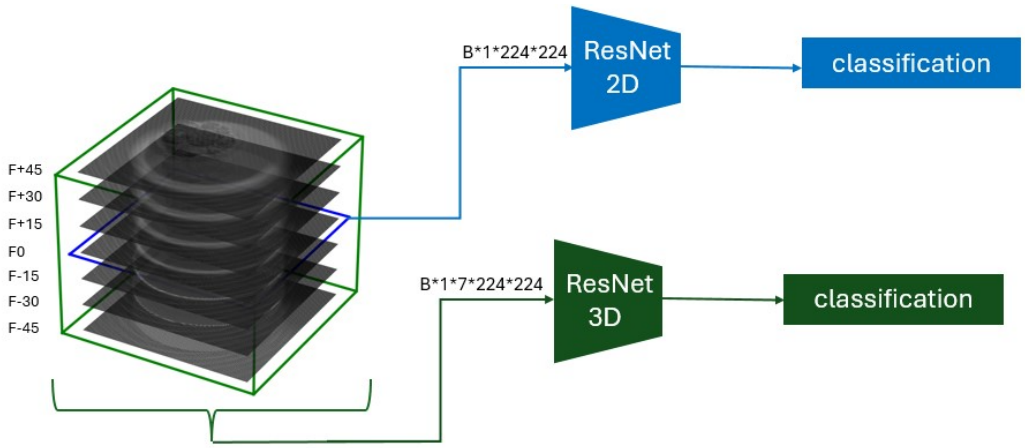

Figure 9: First experiment design.

4.1.1 The Single-Focal Plane (2D) configuration

This approach served as the baseline, mirroring traditional 2D image analysis techniques in embryology.Where we employed a ResNet-18 a well-know architecture that was pre-trained on the ImageNet dataset [56]. Using pre-trained models leverages features learned from a massive dataset, which can improve performance and training efficiency. These model were configured to process images exclusively from the central focal plane (F0). The F0 plane was specifically selected as it consistently provides the most balanced and sharpest image of the embryo's morphology.

4.1.2 The Multi-Focal Plane (3D) configuration

This configuration was designed as the advanced approach, engineered to process volumetric data and fully exploit the embryo's spatial structure.ResNet-18 were transformed into 3D CNNs by replacing their 2D convolutional kernels with 3D equivalents to integrate data from the lower, central, and higher focal planes simultaneously (see Figure 10). Critically, unlike their 2D counterparts, these 3D models were initialized with random weights, meaning they were trained from scratch on the embryo dataset without leveraging pre-training.The input for the 3D model was a $1\times7\times224\times224$ tensor, created by stacking the images from all seven focal planes (from F-45 to F+45). The first dimension represents the channel of the image, and the second dimension represents the focal stack's depth.

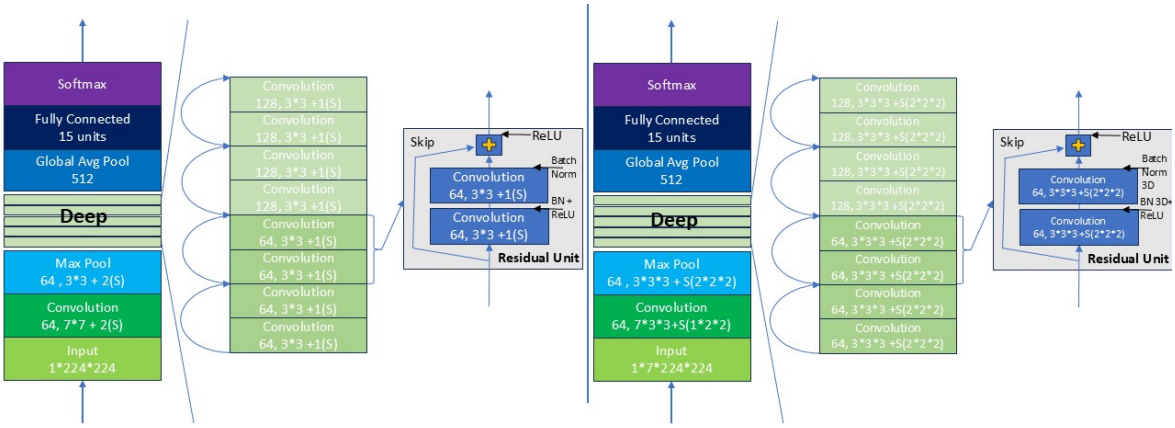

Figure 10: 2D and 3D ResNet architectures where : **Left** is the 2D architecture, **right** is the 3D architecture.

4.2 Experiment 2: Evaluating temporal modeling architectures

This second experiment is designed to conduct a comprehensive evaluation of deep learning architectures engineered to model temporal dynamics. The core objective is to move beyond static image analysis and compare how different models interpret information embedded in sequences of frames over time and detect if there is a changing en embryo morphology As illustrated in Figure 11. The experiment contrasts two distinct approaches, each representing a different strategy for learning from spatio-temporal data: a 2D and ResNet architecture, and a state-of-the-art TimeSformer model [53].

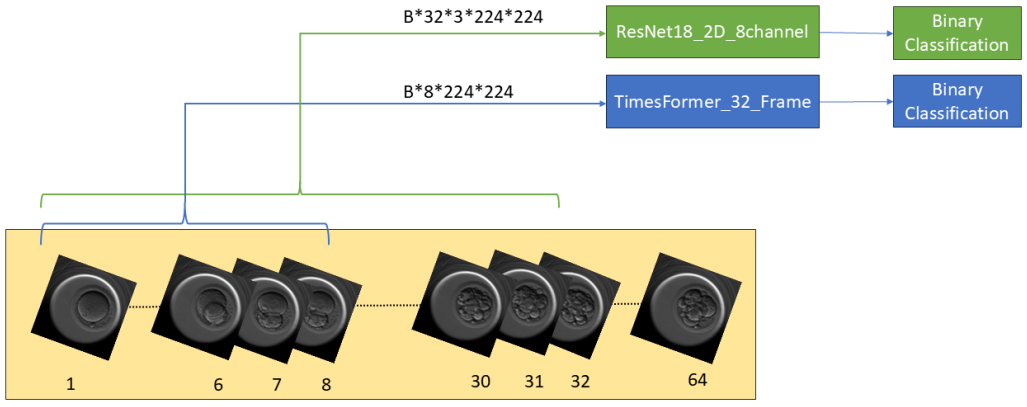

Figure 11: Second experiment design.

4.2.1 2D CNN with temporal frame stacking

This method serves as a baseline for incorporating temporal information into a traditional 2D architecture. here we used a standard, pre-trained 2D CNN ResNet-18 without altering its fundamental architecture. It's a common technique to provide a 2D model with a sense of time by manipulating the input data format rather than the model itself. A sequence of 8 consecutive grayscale frames is extracted from the video. These frames are then "stacked" along the channel dimension to create a single multi-channel tensor. This stacked input (e.g., an 8-channel tensor) replaces the standard 3-channel (RGB) input that the 2D CNN expects. so The input Would be $8\times224\times224$.

4.2.2 TimeSformer model

TimeSformer [53] is a state-of-the-art video classification model based on the transformer architecture, which replaces convolution with self-attention mechanisms. Unlike traditional CNNs that extract local features, TimeSformer is designed to capture long-range dependencies across both spatial and temporal dimensions, making it particularly effective for video data. In this study, it was selected to move beyond static image analysis and better capture the dynamic information inherent in embryo development sequences.

Architecturally, TimeSformer is a video-adapted variant of the Vision Transformer (ViT) [50]. Each video frame is divided into a grid of patches, which are flattened into a sequence of tokens—similar to words in a sentence. The model then applies a spatio-temporal self-attention mechanism: spatial attention captures relationships within each frame, while temporal attention models dependencies across frames. This decoupling allows for efficient and scalable video processing without the high computational cost typical of full 3D attention.

In our implementation, we used the pretrained TimeSformer model available on Hugging Face [1], provided by Facebook and originally trained on the Kinetics-400 dataset [57]. The input to the model is a tensor of shape $32\times3\times224\times224$, where 32 represents consecutive, non-overlapping frames extracted from embryo development videos. This longer temporal window is crucial for capturing morphokinetic patterns subtle timing and structural changes that characterize different developmental phases. By analyzing an extended sequence, the model can learn to distinguish between

[1] https://huggingface.co/facebook/timesformer-base-finetuned-k400

phases that may appear visually similar in individual frames. This ability to model temporal dynamics is hypothesized to enhance classification accuracy in tracking embryo development progression.

4.3 Experiment 3: the "Divide and Conquer" strategy for Hierarchical Classification

This third experiment investigates a "Divide and Conquer" strategy, a hierarchical approach to embryo classification (see Figure 12). Instead of compelling a single model to classify all 15 granular phases simultaneously, this methodology breaks the complex problem down into smaller, more manageable sub-problems. This experiment utilizes the same foundational models as Experiment 4.1.

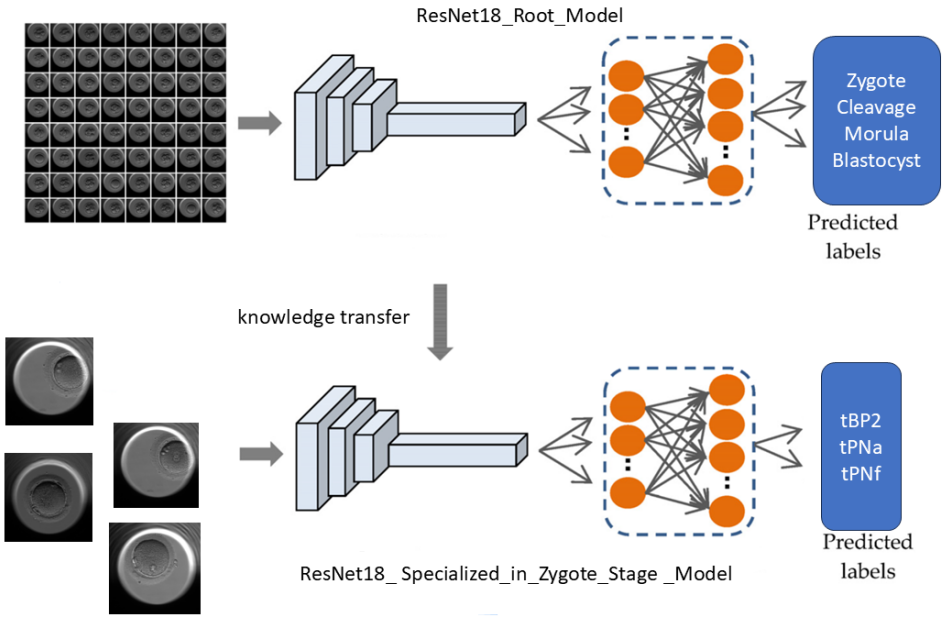

Figure 12: Third experiment design.

4.3.1 Step 1: Foundational "Root" stage classification

The initial step moves away from classifying all granular phases directly. Instead, the models are first trained on a higher-level, more general task: classifying the main developmental "root" stages of the embryo. This involves grouping the granular phases into broader categories, such as the Pronuclei stage, Cleavage stage, Morula, and Blastocyst.

The objective of this step is to teach the models to first recognize the defining features of these major developmental milestones. The best-performing model from

this foundational training is then saved and serves as a new, custom "pre-trained" base model for the next step of the experiment.

4.3.2 Step 2: Fine-Tuning for specific phase classification

This step utilizes the principle of transfer learning. The best model from Step 1, which is now pre-trained on the general root stages, is used as a starting point for a more detailed classification task.

The model is then fine-tuned to classify the specific, individual phases within each of the broader stages. For example, after the model has learned to identify the general "Cleavage Stage," it would be subsequently fine-tuned using only the data for phases t2, t3, t4, t5, t6, t7,t8 and t9+ to learn to distinguish between these closely related sub-stages. This process is repeated for each root stage, creating a set of specialized classifiers that are expert at differentiating the phases within their specific developmental category.

5 Hyperparameters settings

All experiments were conducted using the PyTorch framework (version 2.8) and Python 3.11. The computational workload was managed by an NVIDIA RTX 6000 Ada Generation GPU, a powerful graphics processing unit well-suited for demanding deep learning tasks. In contrast to some related works, no data augmentation techniques were applied in this study.

To ensure a fair and direct comparison between models, a consistent set of hyperparameters was used across all experiments. This standardized approach is crucial for maintaining the reliability and reproducibility of the results, and the general hyperparameter configurations are summarized in Table 5.

However, a notable exception was made for the TimeSformer model in Experiment 2 due to its significant computational cost. With each training epoch requiring approximately 14 hours to complete, this specific model was trained for a reduced **10 epochs** with a **batch size of 1** and a **learning rate of** 1×10^{-4}.

6 Evaluation metrics

To quantitatively evaluate and compare the performance of Our experiments, four standard classification metrics were utilized: accuracy, precision, recall, and

Hyperparameter	Value/Description
Image Size	224×224 pixels
Normalization	Mean and Standard Deviation
Optimizer	Adam
Loss Function	Cross-Entropy Loss
Learning Rate	1×10^{-3}
Batch Size	16
Number of Epochs	100

Table 5: Hyperparameter settings for models training

F1-score. These metrics assess the effectiveness of the models in correctly classifying the developmental phases of the embryos or transition detection. In the following equations, TP, TN, FP, and FN represent the counts of true positives, true negatives, false positives, and false negatives, respectively.

Accuracy Measures the overall proportion of correct predictions among all instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Precision Measures the proportion of positive predictions that were actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall Measures the proportion of actual positives that were correctly identified by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1-Score The harmonic mean of precision and recall, providing a balanced measure between the two.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

7 Results and discussion

The following subsections present the outcomes of each experimental setup, followed by a discussion of the key findings.

7.1 Results experiment 1

To evaluate the role of spatial representation in embryo phase classification, we compared the performance of 2D and 3D CNN models using both phase-level and

embryo-level data splits. Table 6 summarizes the classification report obtained by each model.

Table 6: Performance metrics of 2D and 3D CNNs on Phase-Level vs. Embryo-Level Splits

Model	**Phase-Level Metrics (%)**				**Embryo-Level Metrics (%)**			
	Acc	**Precision**	**Recall**	**F1-score**	**Acc**	**Precision**	**Recall**	**F1-score**
2D ResNet18	88.85	88.93	88.86	88.87	60.30	58.16	60.30	59.21
3D ResNet18	88.90	88.93	88.91	88.91	59.78	57.24	59.78	58.48

As shown in Table 6, both the 2D and 3D models achieved similar accuracy (∼89%) when evaluated using a phase-level split. However, this high performance is largely attributed to **data leakage**, as frames from the same embryo may appear in both training and testing sets. This allows the models to memorize specific visual characteristics unique to individual embryos, which does not reflect **real-world generalizability**. Consequently, we have reservations regarding the validity of the 93% accuracy reported by Barhoun et al [26], as it may similarly be affected by this issue.

In contrast, when the evaluation was performed using the **embryo-level split**, where no embryo appears in more than one subset, the accuracy dropped significantly to ∼60% for both models. This substantial decrease indicates the **true generalization ability** of the models and highlights the challenge of inter-embryo variability.

To further analyze the predictions, confusion matrices were generated for both models under the embryo-level setting. As illustrated in Figure 13, the majority of misclassifications occur between **adjacent developmental phases**, such as confusing `t8` with `t9+` or `t7`. This pattern indicates that the models have partially captured the **temporal continuity of embryo development**, although they struggle with fine-grained distinctions. Additionally, these errors could be partly due to **labeling inaccuracies** in the dataset, as some frames may have been misannotated by embryologists or belong to ambiguous transitional states.

Furthermore, despite its ability to process volumetric data, the 3D ResNet18 did not outperform the 2D model. This may be attributed to several factors as the 3D model was trained from scratch without pretrained weights, focal planes beyond F0 may introduce redundant or noisy information.

Overall, this experiment demonstrates that **embryo-level evaluation is essential**

to avoid misleading results due to data leakage and that **2D CNNs with strong pretrained features can be competitive**, even against more complex volumetric models.

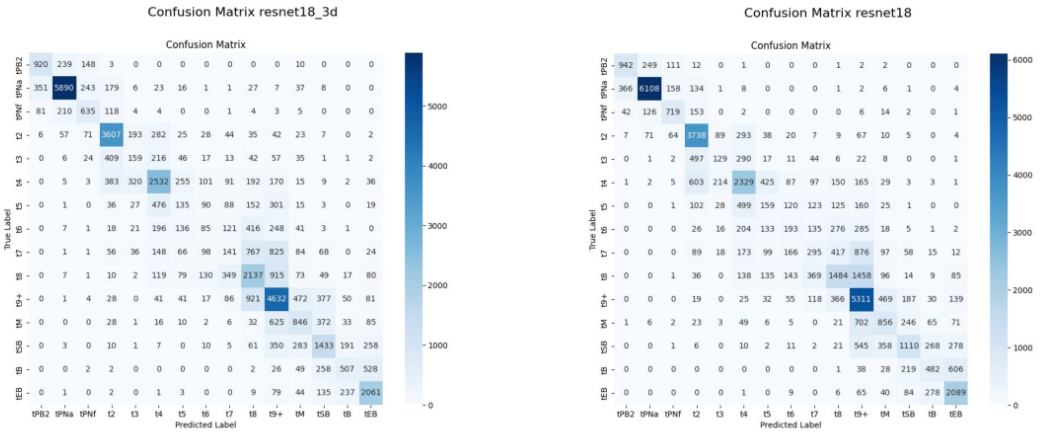

Figure 13: Confusion matrices for both ResNet18 and ResNet18_D models trained on the embryo-level settings

7.2 Results experiment 2

This experiment was designed to assess the importance of temporal information in classifying embryo development phase transition by comparing a baseline 2D convolutional model with a transformer-based spatiotemporal model. Specifically, we evaluated two architectures:

- **ResNet18 (8-channel):** a standard 2D CNN pretrained on ImageNet, modified to accept an 8-channel input created by stacking 8 consecutive grayscale frames along the channel dimension.
- **TimeSformer:** a Vision Transformer-based model specifically adapted for video data, capable of modeling both spatial and temporal relationships using a dual attention mechanism.

Both models were trained on non-overlapping sequences of 8 frames for ResNet18 and 32 frames for TimesFormer. The classification performance of both models is summarized in Table 7.

Table 7: Performance metrics of ResNet18 and TimeSFormer

Model	**ResNet18 Metrics (%)**				**TimeSFormer Metrics (%)**			
	Acc	**Precision**	**Recall**	**F1-score**	**Acc**	**Precision**	**Recall**	**F1-score**
Results	71.04	72.60	71.05	71.82	83.36	84.67	97.06	90.44

The TimeSformer, as shown in Table 7, performed significantly better than the baseline ResNet18 model, realizing an F1-Score of 90% to 71% for the stacked-frame ResNet18. This result emphasizes the prospect of transformer-based architectures for modeling temporal dynamics in time-lapse embryo videos. Where the ResNet18 baseline implicitly captures some temporal features via stacked frames, the TimeSformer is designed explicitly to retain covers spatiotemporal dependencies with self-attention mechanisms of the kind necessary to remember progression patterns. (Differentiating between stages, which might look visually identical, requires working with the kind of resources the TimeSformer has.) The next factor to consider is that the TimeSformer processes 32-frame sequences, comparatively, against 8 frames handled by ResNet18 in the temporal domain. So, the TimeSformer has, quite literally, a wider window into not just identifying phase transitions but remembering and scoring them for what they are.

the TimeSformer's computational cost is a significant limitation. Because of its high memory consumption, it was only possible to train the model for 10 epochs with a batch size of 1. In fact, training the TimeSformer on GPUs with less than 24 GB of VRAM is practically impossible, and this makes the model a poor candidate for many researchers. In contrast, the ResNet18 model underwent training for 100 epochs with a batch size of 16. This highlights a significant trade-off: whereas transformer-based algorithms may extract highly discriminative temporal characteristics from embryo sequences, their substantial resource demands provide practical obstacles to wider deployment.

7.3 Results experiment 3

In this third experiment, we implemented a hierarchical classification strategy based on the principle of *Divide and Conquer*, aiming to reduce the complexity of full 15-class embryo phase classification. The classification task was split into two stages:

1. **Step 1 (Root Stage Classification):** Identify the broader developmental stage the embryo belongs to Pronuclei (Z), Cleavage (C), Morula (M), or Blastocyst (B).
2. **Step 2 (Sub-Phase Classification):** Use fine-tuned models to distinguish between the granular sub-phases within each root stage (e.g., t2–t9+ for Cleavage).

For this experiment, we used the 2D ResNet18 architecture across all classifiers.

Unlike Experiment 4.1, where 3D models were explored, we opted here for ResNet18 due to its strong performance, lower computational demand, and the observation that 3D ResNet18 provided no significant benefit over 2D in earlier tests. Additionally, because this hierarchical setup involves multiple smaller classification tasks, using simpler, pretrained models helps avoid overfitting and maintains faster training and inference times.

The classification performance for the first step — root-level stage classification — is summarized in Table 8.

Table 8: F1-Scores for Root Stage and Fine-Grained Sub-Phase Classification

Level 1: Root Stage Classification	
Root Stage	**F1-Score (%)**
Pronuclei (Z)	95.34
Cleavage (C)	93.63
Morula (M)	40.00
Blastocyst (B)	85.65
Level 2: Fine-Grained Sub-Phase Classification	
Sub-Phase	**F1-Score (%)**
tPB2 (Z)	72.73
tPNa (Z)	92.22
tPNf (Z)	73.76
t2 (C)	82.73
t3 (C)	10.36
t4 (C)	61.25
t5 (C)	11.28
t6 (C)	11.87
t7 (C)	11.80
t8 (C)	47.24
t9+ (C)	74.55
tM (M)	40.00
tSB (B)	80.94
tB (B)	48.60
tEB (B)	59.90

The model achieved high performance in classifying the Pronuclei (Z) and Cleavage (C) stages, with F1-scores above 90%, indicating robust discrimination of early developmental periods. The Blastocyst (B) stage was also well-identified, achieving 85.65%, though still more challenging than the earlier stages. Contrary to earlier expectations, the Morula (M) stage had the weakest performance at 40.00%, due to being represented by a single sub-phase (tM), which may limit the model's ability to generalize.

Sub-phase classification results were mixed. The Pronuclei sub-phases performed well, particularly tPNa (92.22%), benefiting from their distinct visual features.

In the Cleavage stage, performance varied widely. While t2 (82.73%) and t9+ (74.55%) which represent the beginning and ending of the cleavage stage, showed promising results, mid-cleavage stages (t3–t7) had extremely low F1-scores (10–12%), suggesting difficulty distinguishing visually similar transitions that occur rapidly.

In the Blastocyst phase, performance was stronger overall, especially for tSB (80.94%), while tB (48.60%) and tEB (59.90%) show that late development stages remain moderately challenging, due inter-observer annotation variability, subtle morphological shifts and from inconsistent annotation as discussed in Section 4.3.

Despite these limitations, this experiment demonstrates the potential of hierarchical classification for embryo staging: by first separating broadly distinct developmental periods, the model avoids learning all 15 granular phases at once, reducing classification complexity. Future enhancements could involve using temporal models (like TimeSformer) at the second step of fine classification, or training stage-specific CNNs with augmented data to better model challenging transitions between phases.

8 Synthesis of experimental findings

Our experiments provided clear insights into how deep learning can be used for embryo classification. We learned that simply using a 3D model isn't necessarily better than a well-trained 2D model, especially when considering the practical challenges of training. Our most significant finding was that models designed to understand time and sequence, like the TimeSformer, are far more effective at correctly identifying embryo phase transition from video. This shows that the dynamic development of the embryo is a crucial piece of information. Finally, our 'divide and conquer' approach proved to be a promising strategy for making the complex task of classifying all 15 stages more manageable. Overall, the results show that embryo-level evaluation is essential to get a true sense of a model's performance and that advanced temporal models, despite their high computational cost, hold the key to building more accurate classification systems. Table 9 summarizes our experimentals approaches.

Table 9: Summary and Comparison of Experimental Approaches

Experiment	Objective	Models	Results
1: Spatial Analysis (2D vs. 3D CNN)	Evaluate the contribution of 3D information (multi-focal planes) compared to a classical 2D approach; Single vs. Multi-Focal Imaging.	2D ResNet-18 (F0 plane) vs. 3D ResNet-18 (7 planes). Training on "Phase-Level" and "Embryo-Level" splits. The 3D CNN showed no significant improvement over the 2D CNN. Both achieved around 60% accuracy with embryo-level splitting. Phase-level split reached 89%, likely due to data leakage.	Phase-Level: ∼89% accuracy for both models (biased by data leakage). Embryo-Level: ∼59% (2D) and ∼58% (3D). A well-pretrained 2D CNN provides a strong baseline. Using multi-focal volumetric data did not yield gains, possibly due to lack of pretraining and redundancy. Embryo-level splitting is essential for robust evaluation.
2: Temporal Modeling (CNN vs. TimeSformer)	Compare architectures to capture temporal dynamics of embryonic development (Temporal Feature Extraction).	ResNet-18 (stack of 8 images) vs. TimeSformer (sequence of 32 images). TimeSformer achieved a significantly higher F1-score (90.44%) compared to the temporal 2D CNN (71.82%), benefiting from full spatio-temporal attention.	ResNet-18: 71.82%. TimeSformer: 90.44%. Models explicitly modeling time (TimeSformer) perform much better. Temporal dynamics provide richer information than static morphology. Limitation: very high computational cost of TimeSformer.
3: Hierarchical Classification ("Divide & Conquer")	Simplify classification of 15 phases by decomposing into simpler steps (Multistage Strategy).	1. "Root" classification (Z, C, M, B) with ResNet-18. 2. Fine-tuning specialized ResNet-18 models for sub-phases. Hierarchical approach achieved high F1-scores (>93%) for macro-stage detection (e.g., pronuclei and cleavage), but poor performance (∼11%) on subtle, rapid sub-stages (t3–t7).	Root: High scores for Pronuclei (95%) and Cleavage (93%). Sub-phases: Highly variable results, low scores for intermediate cleavage stages (t3–t7, ∼11%). Hierarchical modeling reduces complexity and improves generalization, but sub-phase transitions require temporal or specialized models.

9 Conclusion

This chapter has outlined the experimental framework designed to assess various AI models for embryo stage classification and has reported the outcomes of each approach. Despite promising results from advanced architectures, the experiments revealed persistent challenges, namely the significant computational cost of the most effective temporal models and the inherent difficulty in distinguishing between visually similar and rapid developmental sub-phases. The results also illustrated the critical importance of a rigorous evaluation methodology, as only a strict embryo-level data split, in contrast to a phase-level split, can provide a true measure of a model's generalization performance. In response to the complexity of classifying all 15 stages at once, this work has demonstrated that a hierarchical "Divide and Conquer" strategy, combined with the power of temporal models, provides a promising path forward, even if further refinement is needed.

The final section of our work will now provide a comprehensive synthesis of the key findings presented throughout the research, thoroughly discuss the specific objectives that were successfully achieved, critically address the various limitations encountered during the study, and finally propose promising directions for future research to build upon the foundation established here.

Conclusion and future research directions

This master thesis makes several key contributions to the field of automated embryo assessment for IVF. Firstly, we developed and rigorously evaluated an intelligent system, establishing that temporal models like TimeSformer, which analyze the entire developmental sequence of an embryo, are significantly more accurate than static models that assess only single images or spatial data. This underscores the critical importance of capturing morphokinetic dynamics. Secondly, we highlighted a critical methodological issue in the field, demonstrating that improper data splitting (phase-level vs. embryo-level) can lead to data leakage and overly optimistic performance metrics, thereby establishing a more robust evaluation protocol. Finally, we introduced a 'Divide and Conquer' hierarchical classification strategy that successfully simplifies the complex task of identifying 15 distinct developmental stages, making the problem more tractable and improving performance on broader category recognition.

Despite its contributions, this research faced several limitations. The primary constraint was the quality of the public dataset used, which suffered from significant class imbalance, misannotated frames, and poor image quality that required extensive data cleaning and may have impacted the models' ultimate performance. Another major limitation was the high computational demand of the superior TimeSformer model; its heavy memory and processing requirements restricted our ability to perform extensive hyperparameter tuning

and poses a practical barrier for its adoption in clinical settings with limited resources. Lastly, while the hierarchical approach was successful at a high level, the performance in classifying certain fine-grained sub-phases, particularly in the mid-cleavage stages, remained low, indicating that the models still struggle to differentiate between subtle and rapid morphological transitions.

The findings and limitations of this study open up several promising avenues for future research. A clear next step is to combine our most successful approaches by integrating powerful temporal models like the TimeSformer into the second, fine-grained stage of the hierarchical classification framework. This could significantly improve the accuracy of distinguishing between the more challenging sub-phases. Further work should also focus on developing stage-specific models, potentially using data augmentation techniques to create more robust classifiers for the difficult transitional periods. Looking ahead, the integration of multi-modal data, combining the time-lapse imaging with non-invasive data sources such as patient clinical records or metabolomic profiles of the culture medium, could lead to a holistic assessment model with even greater predictive power. Ultimately, the goal is to refine these AI-driven systems into reliable, transparent, and clinically applicable tools that can assist embryologists and improve IVF success rates for patients.

References

[1] World Health Organization. Infertility prevalence estimates, 1990–2021, 2023.

[2] Rocío Nuñez-Calonge, Nuria Santamaria, Teresa Rubio, and Juan Manuel Moreno. Making and selecting the best embryo in in vitro fertilization. Archives of Medical Research, 55(8):103068, 2024.

[3] David K. Gardner and William B. Schoolcraft. In vitro culture of human blastocysts. pages 378–388, 1999.

[4] A Pantou, K Sfakianoudis, E Maziotis, P Tsioulou, S Grigoriadis, A Rapani, P Giannelou, M Asimakopoulou, K Nikolettos, T Kalampokas, et al. Pgt-a: Who and when? a systematic review and network meta-analysis of rcts. In HUMAN REPRODUCTION, volume 35, pages I374–I375. OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND, 2020.

[5] Yasmin Magdi, Ahmed Samy, Ahmed M Abbas, Mohamed Ahmed Ibrahim, Yehia Edris, Ayman El-Gohary, Ahmed M Fathi, and Mohamed Fawzy. Effect of embryo selection based morphokinetics on ivf/icsi outcomes: evidence from a systematic review and meta-analysis of randomized controlled trials. Archives of gynecology and obstetrics, 300:1479–1490, 2019.

[6] Soraia Pinto, Bárbara Guerra-Carvalho, Luís Crisóstomo, António Rocha, Alberto Barros, Marco G Alves, and Pedro F Oliveira. Metabolomics integration in assisted reproductive technologies for enhanced embryo selection beyond morphokinetic analysis. International Journal of Molecular Sciences, 25(1):491, 2023.

[7] The istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. Human reproduction, 26(6):1270–1283, 2011.

[8] Qingxia Meng, Yunyu Xu, Aiyan Zheng, Hong Li, Jie Ding, Yongle Xu, Yan Pu, Wei Wang, and Huihua Wu. Noninvasive embryo evaluation and selection by time-lapse monitoring vs. conventional morphologic assessment in women undergoing in vitro fertilization/intracytoplasmic sperm injection: a single-center randomized controlled study. Fertility and Sterility, 117(6):1203–1212, 2022.

[9] C. L. Bormann, P. Thirumalaraju, M. K. Kanakasabapathy, H. Kandula, I. Souter, I. Dimitriadis, R. Gupta, R. Pooniwala, and H. Shafiee. Consistency and objectivity of

automated embryo assessments using deep neural networks. Fertility and Sterility, 113: 781–787.e1, 2020b.

[10] Dmytro Zhylko, Raquel Del Gallego, Sarah Pardo, Rameen Mahmood, Ya Tung Hsieh, Salma Selim, Daniela Nogueira, Ibrahim El-Khatib, Barbara Lawrenz, Human M Fatemi, and Farah E Shamout. Assisted reproductive technology dataset of embryo time-lapse images and clinical data. medRxiv, 2024. doi: 10.1101/2024.11.01.24316563. Preprint.

[11] Florian Kromp, Raphael Wagner, Basak Balaban, Véronique Cottin, Irene Cuevas-Saiz, Clara Schachner, Peter Fancsovits, Mohamed Fawzy, Lukas Fischer, Necati Findikli, et al. An annotated human blastocyst dataset to benchmark deep learning architectures for in vitro fertilization. Scientific data, 10(1):271, 2023.

[12] Tristan Gomez, Magalie Feyeux, Justine Boulant, Nicolas Normand, Laurent David, Perrine Paul-Gilloteaux, Thomas Fréour, and Harold Mouchère. A time-lapse embryo dataset for morphokinetic parameter prediction. Data in Brief, 42:108258, 2022.

[13] R. Rad, P. Saeedi, J. Au, and J. Havelock. Blastomere cell counting and centroid localization in microscopic images of human embryo. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 1–6, Vancouver, BC, Canada, 2018. IEEE.

[14] Pegah Khosravi, Ehsan Kazemi, Qiansheng Zhan, Jonas E Malmsten, Marco Toschi, Pantelis Zisimopoulos, Alexandros Sigaras, Stuart Lavery, Lee AD Cooper, Cristina Hickman, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. NPJ digital medicine, 2(1):21, 2019.

[15] Mahdi-Reza Borna, Mohammad Mehdi Sepehri, and Behnam Maleki. An artificial intelligence algorithm to select most viable embryos considering current process in ivf labs. Frontiers in Artificial Intelligence, 7:1375474, 2024.

[16] Q. Liao, Q. Zhang, X. Feng, H. Huang, H. Xu, B. Tian, J. Liu, Q. Yu, N. Guo, Q. Liu, and et al. Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring. Communications Biology, 4:415, 2021.

[17] Sijia Wang, Jing Fan, Hanhui Li, Mingpeng Zhao, Xuemei Li, and David Yiu Leung Chan. A dataset for deep learning based cleavage-stage blastocyst prediction with time-lapse images. bioRxiv, pages 2023–12, 2023.

[18] C. L. Bormann, M. K. Kanakasabapathy, P. Thirumalaraju, R. Gupta, R. Pooniwala, H. Kandula, E. Hariton, I. Souter, I. Dimitriadis, L. B. Ramirez, and et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. eLife, 9:e55301, 2020a.

[19] C. L. Bormann, C. L. Curchoe, P. Thirumalaraju, M. K. Kanakasabapathy, R. Gupta, R. Pooniwala, H. Kandula, I. Souter, I. Dimitriadis, and H. Shafiee. Deep learning early

warning system for embryo culture conditions and embryologist performance in the art laboratory. Journal of Assisted Reproduction and Genetics, 38:1641–1646, 2021.

[20] B. Raef, M. Maleki, and R. Ferdousi. Computational prediction of implantation outcome after embryo transfer. Health Informatics Journal, 26:1810–1826, 2020.

[21] S. N. Patil, U. V. Wali, and M. K. Swamy. Selection of single potential embryo to improve the success rate of implantation in ivf procedure using machine learning techniques. In 2019 International Conference on Communication and Signal Processing (ICCSP), pages 0881–0886, Chennai, India, 2019. IEEE.

[22] C. Blank, R. R. Wildeboer, I. DeCroo, K. Tilleman, B. Weyers, P. de Sutter, M. Mischi, and B. C. Schoot. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. Fertility and Sterility, 111:318–326, 2019.

[23] A. Uyar, A. Bener, and H. N. Ciray. Predictive modeling of implantation outcome in an in vitro fertilization setting: an application of machine learning methods. Medical Decision Making, 35: 714–725, 2015.

[24] Chloe He, Neringa Karpavičiūtė, Rishabh Hariharan, Céline Jacques, Jérôme Chambost, Jonas Malmsten, Nikica Zaninovic, Koen Wouters, Thomas Fréour, Cristina Hickman, and Francisco Vasconcelos. Embryo graphs: Predicting human embryo viability from 3d morphology. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Springer Nature Switzerland, 2023. Open Access version provided by the MICCAI Society.

[25] Kanak Kalyani and Parag S Deshpande. A deep learning model for predicting blastocyst formation from cleavage-stage human embryos using time-lapse images. Scientific Reports, 14 (1):28019, 2024.

[26] Abbas Barhoun, Mohammad Ali Balafar, Amin Golzari Oskouei, and Leila Sadeghi. Human embryo stage classification using an enhanced r (2+ 1) d model and dynamic programming with optimized datasets. Biomedical Signal Processing and Control, 107:107841, 2025.

[27] Dimitry Tran, Simon Cooke, Peter J Illingworth, and David K Gardner. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. Human reproduction, 34(6):1011–1018, 2019.

[28] A. Chavez-Badiola, A. Flores-Saiffe-Farias, G. Mendizabal-Ruiz, A. J. Drakeley, and J. Cohen. Embryo ranking intelligent classification algorithm (erica): artificial intelligence clinical assistant predicting embryo ploidy and implantation. Reproductive BioMedicine Online, 41: 585–593, 2020.

[29] M. VerMilyea, J. M. M. Hall, S. M. Diakiw, A. Johnston, T. Nguyen, D. Perugini, A. Miller, A. Picou, A. P. Murphy, and M. Perugini. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during ivf. Human Reproduction, 35:770–784, 2020.

[30] M. F. Kragh, J. Rimestad, J. Berntsen, and H. Karstoft. Automatic grading of human blastocysts from time-lapse imaging. Computers in Biology and Medicine, 115:103494, 2019.

[31] R. M. Rad, P. Saeedi, J. Au, and J. Havelock. Predicting human embryos' implantation outcome from a single blastocyst image. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 920–924, 2019.

[32] K. Loewke, J. H. Cho, C. D. Brumar, P. Maeder-York, O. Barash, J. E. Malmsten, N. Zaninovic, D. Sakkas, K. A. Miller, M. Levy, and et al. Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. Fertility and Sterility, 117:528–535, 2022.

[33] G. Coticchio, G. Fiorentino, G. Nicora, R. Sciajno, F. Cavalera, R. Bellazzi, S. Garagna, A. Borini, and M. Zuccotti. Cytoplasmic movements of the early human embryo: imaging and artificial intelligence to predict blastocyst development. Reproductive BioMedicine Online, 42: 521–528, 2021.

[34] C. Wu, W. Yan, H. Li, J. Li, H. Wang, S. Chang, T. Yu, Y. Jin, C. Ma, Y. Luo, and et al. A classification system of day 3 human embryos using deep learning. Biomedical Signal Processing and Control, 70:102943, 2021.

[35] A. Uyar, A. Bener, H. Ciray, and M. Bahceci. A frequency based encoding technique for transformation of categorical variables in mixed ivf dataset. In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 6214–6217, 2009.

[36] E. Hariton, E. A. Chi, G. Chi, J. R. Morris, J. Braatz, P. Rajpurkar, and M. Rosen. A machine learning algorithm can optimize the day of trigger to improve in vitro fertilization outcomes. Fertility and Sterility, 116:1227–1235, 2021.

[37] Y. Sawada, T. Sato, M. Nagaya, C. Saito, H. Yoshihara, C. Banno, Y. Matsumoto, Y. Matsuda, K. Yoshikai, T. Sawada, et al. Evaluation of artificial intelligence using time-lapse images of ivf embryos to predict live birth. Reproductive BioMedicine Online, 43:843–852, 2021.

[38] Tristan Gomez, Magalie Feyeux, Nicolas Normand, Laurent David, Perrine Paul-Gilloteaux, Thomas Fréour, and Harold Mouchère. Towards deep learning-powered ivf: A large public benchmark for morphokinetic parameter prediction. arXiv preprint arXiv:2203.00531, 2022.

[39] H Nadir Ciray, Alison Campbell, Inge Errebo Agerholm, Jesus Aguilar, Sandrine Chamayou, Marga Esbert, and Shabana Sayed. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. Human reproduction, 29(12): 2650–2660, 2014.

[40] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5):1285–1298, 2016.

[41] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding

transfer learning for medical imaging. Advances in neural information processing systems, 32, 2019.

[42] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience, 19(3):356–365, 2016.

[43] Karen Simonyan. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[45] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4700–4708, 2017.

[46] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.

[47] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging, 35(5):1299–1312, 2016.

[48] Veronika Cheplygina, Marleen De Bruijne, and Josien PW Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical image analysis, 54:280–296, 2019.

[49] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[50] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

[51] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis, 97: 103280, 2024.

[52] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image

segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 574–584, 2022.

[53] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In ICML, volume 2, page 4, 2021.

[54] Md Mostafa Kamal Sarker, Vivek Kumar Singh, Mohammad Alsharid, Netzahualcoyotl Hernandez-Cruz, Aris T Papageorghiou, and J Alison Noble. Comformer: Classification of maternal–fetal and brain anatomy using a residual cross-covariance attention guided transformer in ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 70(11):1417–1427, 2023.

[55] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27, pages 270–279. Springer, 2018.

[56] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.

[57] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. The kinetics human action video dataset. arXiv preprint arXiv:1705.06950, 2017.

Abstract

This master thesis addresses the challenge of embryo selection in IVF, aiming to improve pregnancy outcomes through more objective methods. It proposes an intelligent system for automatically classifying human embryo developmental stages using time-lapse imaging. The study compares 2D and 3D CNNs to assess spatial information and explores temporal models like TimeSformer to capture embryo dynamics. A hierarchical classification strategy is also introduced to handle 15 developmental phases. Results show that while 2D CNNs provide a solid baseline, temporal models significantly outperform them by leveraging morphokinetic features, highlighting the potential of AI to enhance accuracy and consistency in embryo selection.
keywords: In Vitro Fertilization (IVF), Embryo Classification, Deep Learning, Time-Lapse Imaging, Convolutional Neural Networks (CNN), Vision Transformer, TimeSformer, Temporal Modeling, Hierarchical Classification

Résumé

Ce mémoire de Master aborde le défi de la sélection d'embryons en FIV, visant à améliorer les résultats de grossesse grâce à des méthodes plus objectives. Elle propose un système intelligent de classification automatique des stades de développement de l'embryon humain par imagerie accélérée. L'étude compare les réseaux neuronaux convolutif 2D et 3D pour évaluer les informations spatiales et explore des modèles temporels comme TimeSformer pour capturer la dynamique embryonnaire. Une stratégie de classification hiérarchique est également introduite pour gérer 15 phases de développement. Les résultats montrent que si les réseaux neuronaux conjoncturels 2D constituent une base solide, les modèles temporels les surpassent nettement en exploitant les caractéristiques morphocinétiques, soulignant ainsi le potentiel de l'IA pour améliorer la précision et la cohérence de la sélection d'embryons.
mots-clés: Fécondation in vitro (FIV), classification des embryons, apprentissage profond, imagerie accélérée, réseaux de neurones convolutifs (CNN), transformateur de vision, TimeSformer, modélisation temporelle, classification hiérarchique.

مُلخّص

تتناول مذكرة الماستر هذه تحدي اختيار الأجنة في التلقيح الاصطناعي (IVF) ، بهدف تحسين نتائج الحمل من خلال أساليب أكثر موضوعية ودقة. تقترح المذكرة نظامًا ذكيًا للتصنيف الآلي لمراحل تطور الجنين البشري باستخدام التصوير بالفيديو بالفاصل الزمني (time-lapse) . تقارن الدراسة بين الشبكات العصبية الالتفافية ثنائية الأبعاد (2D CNN) وثلاثية الأبعاد (3D

CNN) لتقييم أهمية المعلومات المكانية في تحليل مراحل التطور، كما تستكشف النماذج الزمنية مثل نموذج (TimeSformer) لالتقاط الديناميكيات التطورية للجنين عبر الزمن. بالإضافة إلى ذلك، تم تقديم استراتيجية تصنيف هرمية لمعالجة ١٥ مرحلة تطورية مختلفة للجنين. تُظهر النتائج أنه بينما توفر شبكات (2D CNN) خط أساس قويًا، تتفوق النماذج الزمنية بشكل ملحوظ من خلال استغلال الخصائص المورفوكينية والزمنية، مما يؤكد إمكانات الذكاء الاصطناعي في تحسين دقة واتساق عملية اختيار الأجنة.

الكلمات المفتاحية: الإخصاب في المختبر (IVF) ، تصنيف الأجنة، التعلم العميق، التصوير بالفاصل الزمني، الشبكات العصبية الالتفافية (CNN) ، المحول البصري (Vision Transformer) ، (TimeSformer) ، النمذجة الزمنية، التصنيف الهرمي.