



Student Name: Salma Javid
Enrolment number: 30107961
Module: CS4T702 MSc Project
Branch: Faculty of Computing, Engineering and Science

University of South Wales
Prifysgol De Cymru

MSc. in Artificial Intelligence

Year: 2023-2024

MSc. Project

Applications of AI in IVF — Embryo Classification
AI Architect Design Investigation and Comparative Study

First Supervisor: Dr. Carl Jones

Senior Lecturer - Computer Games Development | Faculty of Computing, Engineering and Science

Second Supervisor: NA

Declaration

STATEMENT OF ORIGINALITY

University of South Wales
Prifysgol De Cymru

Faculty of Computing, Engineering and Science

This is to certify that, except where specific reference is made, the work described in this project is the result of the investigation carried out by the student, and that neither this project nor any part of it has been presented, or is currently being submitted in candidature for any award other than in part for the MSc award, Faculty of Computing, Engineering and Science from the University of South Wales.



Signed: _____
Date: 20th September 2024

Abstract

Embryo classification plays a critical role in the success of in-vitro fertilization (IVF) by determining which embryos have the highest potential for live birth. Traditionally, this assessment is performed manually by embryologists using Gardner's blastocyst grading system which assigns scores based on the blastocyst's degree of expansion (EXP), the quality of the inner cell mass (ICM), and the quality of the trophectoderm epithelium (TE). However, manual grading is subjective, time-consuming, and prone to variability. To overcome these challenges, this project explores the application of artificial intelligence (AI) for automating embryo classification into Gardner's gold and silver standards.

This report presents a comprehensive comparative analysis of various AI architectures for embryo classification using a dataset of 2,344 embryo images. The investigation employs techniques ranging from traditional machine learning models like Support Vector Machines (SVM) and Random Forests to state-of-the-art deep learning techniques such as CNN, ResNet-50, DenseNet-201, VGG16, InceptionV3, and Xception. The study incorporates extensive data preprocessing methods such as feature engineering, feature scaling, normalization, standardization, image resizing, rescaling, data augmentation, SMOTE application, etc. to optimize model performance. Data visualization techniques are used to gain insights into data distribution, model predictions and misclassifications, allowing for a deeper understanding of model behaviour. The models are evaluated using key performance metrics such as accuracy, precision, recall, and F1 score.

The results of this study offer a comparison of traditional machine learning and advanced deep learning models, providing valuable insights into the most effective methods for embryo classification in IVF. Comparative analysis of the models reveals the strengths and weaknesses of each approach, providing insights into their applicability in clinical settings. The findings demonstrate that AI-driven approaches can significantly improve the efficiency, consistency, and scalability of embryo assessment, offering a robust tool for fertility specialists. The significance of this work lies in its potential to contribute to the development of a robust and accurate system for embryo classification.

Our findings indicate that advanced deep learning AI models, particularly CNNs, significantly outperform traditional methods, offering higher accuracy and consistency in embryo classification. This research underscores the potential of AI to revolutionize reproductive medicine, enhancing decision-making processes and improving outcomes in IVF procedures.

Acknowledgements

First of all, I'd like to thank my best friend who happens to be the Lord of all the worlds — the God Almighty, for fulfilling my dream of doing my Masters in the UK, and for helping me sustain through it, Al'humdulillah for everything.

Next, I'd like to express my heartfelt thanks to my course leader Dr. Mabrouka Abuhmida whose support and encouragement have been instrumental throughout the course. She's very inspiring and instils the interest to learn and prove ourselves. Dr Mabrouka is an exceptional educator and mentor. She's very passionate about her subject (Artificial Intelligence) - this has profoundly influenced me.

And finally, I wish to thank Dr. Carl Jones, my supervisor for his invaluable guidance all through the project — his creative insights were constructive and his attention to detail was excellent. I sincerely appreciate his mentorship, encouragement, and directions, which have significantly contributed to the success of this work. His unwavering support has been crucial to the development of this dissertation.

Table of Contents

Declaration	2
Abstract	3
Acknowledgements	4
Chapter 1: Introduction to AI in IVF	8
Understanding IVF [In-Vitro Fertilization]	8
— A History	8
— What is IVF?	8
— Who can undergo IVF?	9
— Why is IVF required?	9
— Current Infertility Rate [Statistics 2024]	9
— Infertility Statistics by Gender	9
— Effects of Infertility	10
— IVF Success and Failure Rate [Statistics]	10
— Effects of IVF Failures	12
— IVF Failure: Reasons	12
— Embryo: An Introduction	13
— Day 5 Embryo: Blastocyst	13
— Benefits of Blastocyst Implantations	14
— Embryo Classification	14
— Gardner Blastocyst Grading System	15
— Challenges in Embryo Grading/Classification	16
Applications of AI in IVF	16
— Image Classification with AI	16
— AI Applications in Embryo Classification	17
— Need for Automated Solutions	17
Chapter 2: Literature Review	17
Chapter 3: Project Details	24
— Scope of the Study	24
— AI Architect Design Investigation	24
— Comparative Study	24
— Methodology	25
— GitHub Link	25
— Tools and Technologies	25
— Programming Language: Python	25
— Environment: Google Colab [Advantages]	26
— Libraries Used	26
Chapter 4: Dataset Description	27

— Database Inspection	27
— Data Visualization	30
— Pie Chart [Data Distribution]	30
— Blastocyst Cavity Expansion (EXP) Distribution	31
— Inner Cell Mass (ICM) Quality Distribution	31
— Trophectoderm (TE) Quality Distribution	32
— Correlation Matrix: EXP, ICM, TE, and Standard	32
— 3D Scatter Plot: EXP vs ICM vs TE coloured by Standard	33
— Data Preprocessing	33
— Checking for Missing Values	33
— K Nearest Neighbors (KNN) Imputation	33
— Database/Feature Engineering	34
— Feature Mapping	34
— Splitting the Dataset	35
— Checking Class Distribution	35
— Feature Scaling	35
— Normalizing, Resizing and Rescaling Images	35
— Image Augmentation	35
— Standardization	36
Chapter 5: Applying AI Algorithms	36
Traditional Machine Learning Methods	36
1 — Random Forest Classification [RFC]	36
2 — Support Vector Machines [SVM]	38
Deep Learning Models	39
3 — Convolutional Neural Networks [CNN]	39
Performance Comparison: RFC, SVM and CNN	41
Transfer Learning	42
4 — ResNet-50	42
5 — DenseNet 201	43
Advanced Algorithms	44
6 — VGG 16, Inception V3, Xception	44
7 — Xception model with SMOTE	45
Chapter 6: Comparative Performance Evaluation — A Study	46
— Random Forest Classification, SVM and CNN	46
— ResNet-50, DenseNet 201, VGG16, InceptionV3, and Xception	48
— Confusion matrix for Xception model with SMOTE	49
Chapter 7: Project Artefact [Prototype]	49
— Features	50
— Scope	51

— Limitations	51
— Future Enhancements	52
Chapter 8: Conclusions, Challenges and Future Work	52
— Conclusion	52
— Challenges	52
— Future Work	53
Chapter 9: Concerns	54
— IVF: Ethical and Moral Issues	54
— Social Stigmas Around IVF	55
References	56
Appendices	60
— Project Plan	60
— Gantt Chart	62
— Ethics Form	62

Chapter 1: Introduction to AI in IVF

Understanding IVF [In-Vitro Fertilization]

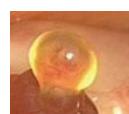
— A History

In the General Hospital of Oldham, UK, at 11:47 pm on July 25, 1978, the world of fertility medicine changed forever with the birth of 5 lb 12 oz Louise Joy Brown, the first baby born after successful in-vitro fertilisation (IVF) — this is why 1978 is known as “The birth year of IVF”. With this IVF was recognized as a viable fertility treatment — all thanks to Sir Robert Edwards. Indeed, for his efforts towards developing IVF and embryo transfer (IVF/ET) to treat infertility he was awarded the 2010 Nobel Prize for Physiology or Medicine (Biggers, 2012).

Since that Tuesday night 46 years ago, the International Committee for Monitoring Assisted Reproductive Technologies, estimates that at least 12 million babies (Adamson et al, 2022) have been born through IVF and other assisted reproductive technologies (ARTs). The history of IVF/ET is extensive and it has been recently documented in part by Johnson (2011) and on the web at www.IVF-Worldwide.com (ref 2).

— What is IVF?

“**In vitro**” is Latin for “**in glass**” (*Definition of IN VITRO*, 2024), and, “**In-Vitro**” also refers to “In the laboratory (outside the body)” which is the opposite of “**In vivo** (in the body)”. In-Vitro Fertilization or IVF refers to fertilization which happens outside the body and in an artificial environment (Merriam Dict.). IVF is one of several Assisted Reproductive Technologies (ARTs) available to help people with fertility problems to have a baby. In this process, eggs taken from a woman are fertilized with a man’s sperm in a laboratory. This fertilized egg cultured for 5-6 days, now called an embryo, is then returned to the woman’s womb to grow and develop into a beautiful baby.



The Egg Excellency

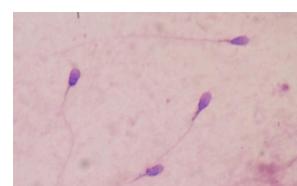
Egg is the largest cell in a woman’s body. It has a central nucleus with genetic material (23 chromosomes) and is protected by a thick, extracellular shell of sugar and protein called the zona pellucida. This zona stops the entry and fusion of more than one sperm.

Image 1: A human egg is caught on camera as it emerges from a woman's ovary. The yellowish egg which is about the size of this dot (-) is seen exiting a fluid-filled follicle on the surface of the 45-year-old woman's ovary. ('Out of the ovary',, 2008)

Note: In IVF, the patient is administered with a higher level of FSH [fertility hormone called follicle stimulating hormone] so that ovaries produce multiple eggs. The extra eggs can be used for fertilisation later, if required, and will provide the clinic with a greater choice of embryos to be used in the treatment.

The Story of Sperm

Image 2: Sperm at 400x Zoom



Sperm is the smallest cell in a man's body. It has 23 chromosomes. The tip of the sperm head called the acrosome enables the sperm to penetrate the egg. The tail moves with whip-like movements back and forth to propel the sperm towards the egg (Sperm: MedlinePlus Medical Encyclopedia, Team B.S., 2015)

Fertilization in IVF

In the laboratory, the eggs are stripped of surrounding cells and prepared for fertilization in a petri dish. Fertilization is done by one of two techniques:

1. First Method: The eggs are incubated with thousands of sperm and fertilization occurs naturally.
2. Second Method: Each egg is individually injected with a single sperm in a process called intra-cytoplasmic sperm injection [ICSI]. This is done to maximize the certainty of fertilization when there are issues with sperm quality.

Embryo Transfer

Upon successful fertilization, the zygotes (fertilized eggs) are allowed to grow in the lab for up to 6 days where they begin developing into embryos. The embryologist selects the best embryo(s) to be transferred to the woman's body. There, it requires about three or so days to implant firmly onto the endometrium (the inner lining of the uterus). Once successfully implanted, the pregnancy is confirmed... ***and that's how life begins!*** (Nassim Assefi et al., 2015)

— Who can undergo IVF?

IVF is recommended to women under the age of 43 who have been unsuccessful in conceiving naturally after trying for about two years (IVF, 2017).

— Why is IVF required?

"1 in 6 people globally are affected by infertility", [WHO, 2023]. Technically, infertility is a disease of the male or female reproductive system. Understandably, this is a major health challenge globally. (National Institute for Health and Care Excellence, *Prevalence / Infertility / CKS / NICE* [2023])

— Current Infertility Rate [Statistics 2024]

- Infertility affects about 1 in 7 couples in the UK (approximately 3.5 million people) [NICE, 2017a].
- An estimated 15% of the global population have trouble conceiving (UCLA Health, 2020).
- Globally, 48.5 million couples experience infertility (Reproductive Biological Endocrinology, 2015) (*Infertility Statistics and Survey 2024 / SingleCare*, [2024])

— Infertility Statistics by Gender

Infertility is nearly as common in men as it is in women. According to reports, 9% of men and 10% of women are affected by infertility (CDC, 2013). In general, about 30% of fertility issues are attributed to the woman, 30% to the man, and 30-40% to unknown causes (Infertility, 2019).

— Effects of Infertility

Infertility and its complications, like miscarriages, can affect a person's overall health and negatively impact their quality of life. They experience psychological and interpersonal distress. Some pointers:

- Infertility is one of the primary reasons for divorce among couples. (International Journal of Reproductive Biomedicine, 2020)
- Up to 60% of infertile individuals report psychiatric symptoms with significantly higher levels of anxiety and depression. (Clinical Therapeutics, 2014)
- Nearly 41 to 87% of infertile women suffer from anxiety and depression. (BMC Women's Health, 2004)

— IVF Success and Failure Rate [Statistics]

Around 55,000 patients had In-Vitro fertilisation (IVF) or donor insemination (DI) treatment at HFEA licensed fertility centres in the UK in 2021 (HFEA, [2021])

Preliminary UK statistics for IVF | Published: June 2023

IVF success rates have tripled over the last 20 years in the UK, with almost a third of all embryo transfers in women under 35 resulting in a baby (Pidd, 2020).

Success and Failure Rate

Despite significant efforts by research groups, the global success rate of IVF remains around 40% in terms of live births. A woman's age is the main factor here. A 2022 study found that the clinical pregnancy rate for women under 30 was 69.4%, whereas for women above 40, it declined to 9.4% (*IVF Success Rates By Age*, 2023). "*Women up to the age of 35 have the highest 47% chance of IVF success.*" — Dr Gorgy, The Fertility and Gynaecology Academy (Sanaz Ghazal, M.D., Forbes). The below table shows the age-wise variation of live births through IVF.

Table 1: IVF — Age and Live Births

Age	Live Birth
< 35	44.5%
35 - 37	32.4%
38 - 40	20.2%
41 - 42	9.6%
> 42	2.9%

Cost of Infertility Treatment

The cost of private treatment for 1 IVF cycle can cost up to £5,000 or more.

— Effects of IVF Failures

IVF failures have a significantly greater impact compared to natural infertility issues, deeply affecting mental and psychosocial well-being. The experience can be painful and emotionally exhausting. Beyond the disappointment of treatment failure, the high costs, along with the physical and psychological toll, can lead to intense stress, stigma, financial strain, and even depression.

— Scope of the problem [The Problem Statement]

With an approximate 35% success rate, there's a huge potential for improvement... ***To achieve success, we first need to fully understand failure.***

— IVF Failure: Reasons

Where there's failure, there's scope for improvement

[In-depth Research and advanced AI Applications required].

IVF failure can result from various factors at different stages of the process. However, our current focus is on failures specifically related to embryos. Let's explore this in more detail.

— Challenges with Embryos

- **Quality of Embryos:** Quality decides the outcome - A good quality embryo can lead to successful pregnancy, while a bad quality may result in IVF failure (15 Reasons For IVF Failure).
- **Egg & Sperm Quality:** Best quality of both are required to have a good quality embryo. As a woman ages, the quality of their eggs decreases (admin, 2020). Poor egg quality can hinder fertilization or embryo development. Sperm quality is equally crucial, as abnormalities affect fertilization success.
- **Chromosomally Abnormal Embryos:** Not all embryos are viable. While an embryo may reach the blastocyst stage on time and have a regular shape on the outside, it may have chromosomal or genetic abnormalities on the inside. A woman's body rejects these kinds of embryos and this results in IVF failure (Shauli, 2021). Older eggs are more prone to contain abnormal chromosomes (15 Common Reasons for IVF Failure).
- **Embryo Implantation Failure:** Embryo implantation can fail if the uterine lining is resistant or incompatible with certain embryos, or if the embryo is unable to successfully embed itself in the uterine wall. Factors contributing to failed implantation include poor embryo quality, a thin endometrium, and conditions like PCOS (kjkadmin, 2022).
- **Unknown Reasons:** Embryos which appear healthy in the lab can fail to implant due to unknown reasons (*scope for research*).

— Embryo: An Introduction

We already know that a zygote is the single cell formed when an egg and sperm fuse in a process called fertilization. In the first 12 to 24 hours after fertilization, the zygote enters the cleavage stage, during which rapid mitotic cell divisions occur. With each division, the cell count doubles, leading to exponential growth. By the time the zygote reaches the 32-cell stage, it is referred to as a morula. (*Human embryogenesis (article) | Embryology*)

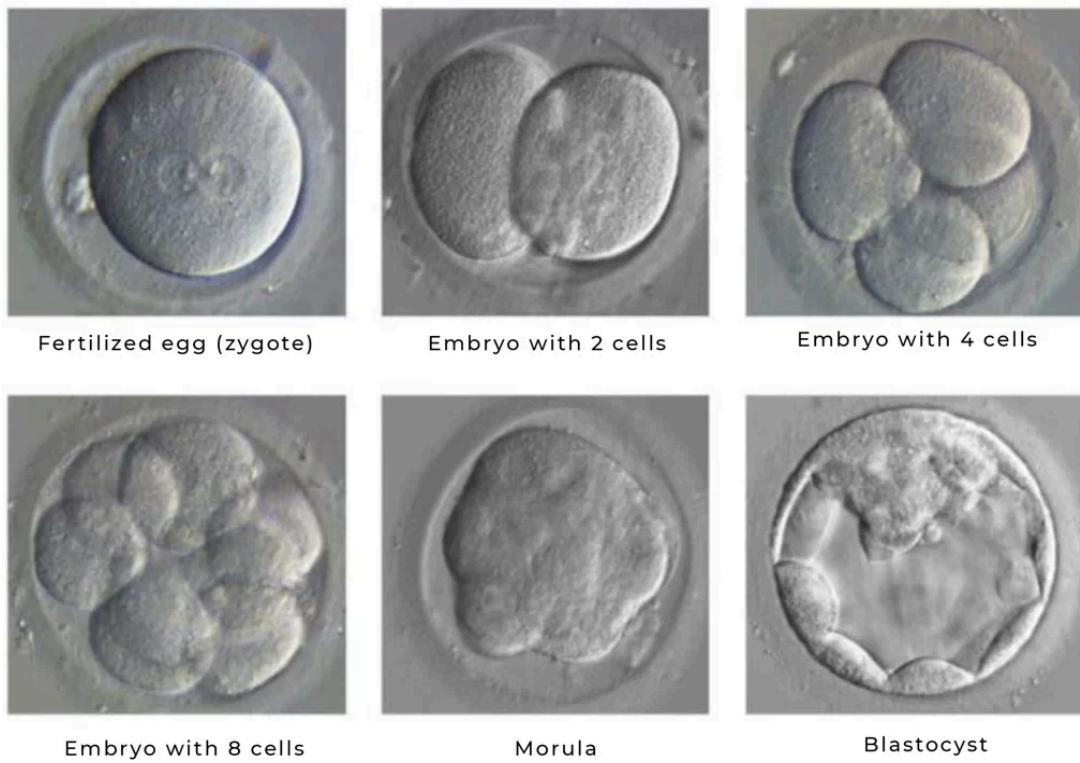


Image 3: Embryo stages and development (*Understanding IVF Embryo Grading Systems | ARC® Fertility*)

— Day 5 Embryo: Blastocyst

BLASTOCYST

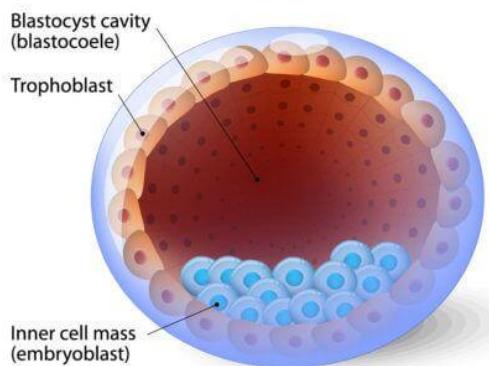


Image 4: Parts of a Blastocyst (Donna, 2024)

Embryos, in labs or naturally developing in a woman's womb, usually reach blastocyst stage by day 5 after fertilization. At this stage, the embryo has developed into a structure consisting of 200-300 cells (MD, 2024). These cells now begin to differentiate and develop more specific forms and functions organized into three distinct parts:

- A central cavity filled with fluid called blastocoel cavity.
- Inner cell mass — Pushed off to one side of the sphere that will develop into the foetus. It eventually turns into the cells of body tissue (muscle, brain, bone, etc).
- Trophoblast — An outer shell layer responsible for initial implantation into the uterine wall of the mother's uterus (MD, 2024). It becomes the placenta and other tissues necessary supporting foetal structures for pregnancy. It also facilitates the exchange of nutrients and waste between the mother and the developing embryo and is essential for the maintenance of the pregnancy.

All of the three are crucial components each serving a different but essential role in development of a baby.

— Benefits of Blastocyst Implantations

In assisted reproduction, predicting a successful outcome relies on selecting embryos with higher implantation potential. Research has shown that transferring embryos at the blastocyst stage can improve implantation, pregnancy, and live birth rates (Gardner & Lane, 1997). Blastocyst-stage transfers are favoured because they offer greater synchrony with the uterine lining (Ahlström et al., 2011), as embryos at this stage tend to be more viable and robust (Sivanantham et al., 2022). In natural pregnancy, the embryo typically implants into the uterus at the blastocyst stage, around day 5, when the uterine lining is most receptive. In IVF, transferring day 5 blastocysts better aligns with this natural timing, potentially increasing the likelihood of a successful pregnancy.

Time Lapse Imaging [TLI]:

In time-lapse imaging (TLI), embryos are cultured in incubators equipped with built-in microscopes that automatically capture high-quality images every 5–20 minutes at a set focus and magnification. This allows embryologists to monitor embryo development dynamically through a screen without needing to remove them from the incubator, maintaining a stable environment. By reviewing the video footage, embryologists can assess the embryos' progress and select the best ones for transfer. TLI enables detailed observation of morphokinetic parameters, which are key indicators of embryo viability (Marcos Meseguer et al., 2011).

TLI technology has greatly enriched embryology by providing valuable insights that improve IVF outcomes. It is now considered one of the most powerful tools for studying embryo development patterns in uninterrupted culture conditions. Due to these benefits, time-lapse systems are increasingly used in IVF labs worldwide.

— Embryo Classification

In IVF, accurate embryo classification is crucial for selecting the most viable candidates for implantation and improving the chances of a successful pregnancy. This process involves assessing embryos using both morphological and morphokinetic parameters, through static images and time-lapse videos.

Embryo Grading (Blastocyst Scoring)

Embryo grading is a key technique used to evaluate embryo quality before transfer, playing a vital role in IVF success rates. For Day 5 embryos, grading focuses on three main factors: blastocyst expansion, inner cell mass quality, and trophectoderm epithelium (IVF Embryo Grading, 2023). Specialists assess and assign grades based on parameters like cell number, uniformity of cell division, fragmentation, and development rate. Accurate grading improves embryo selection, enhancing the likelihood of implantation and pregnancy, reducing transfer cancellations, and alleviating patient stress.

The Most Crucial Step: Embryo Selection

The primary goal of IVF and embryo culture is to produce high-quality embryos with the greatest potential for live births ('Blastocyst Culture, Day 5 IVF Embryo Transfer & In Vitro Fertilization'). Given the rise of single embryo transfer (SET), effective identification methods are more crucial than ever.

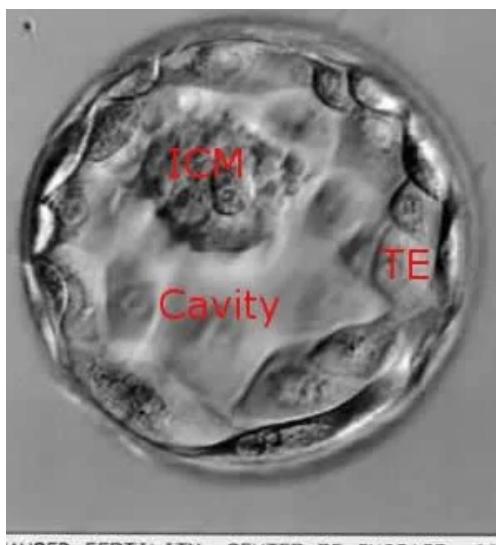
— Gardner Blastocyst Grading System

Gardner's scale is the most common and universally used system for grading blastocysts or embryos.

Prof. David Gardner, Group Director of ART, Scientific Innovation, and Research, is a prominent figure in fertility science. He is a Fellow of the Australian Academy of Science and a Member of the Order of Australia (AM). Prof. Gardner is renowned for his pioneering contributions to embryo culture techniques, particularly for developing methods to grow blastocysts to day five maturity before transfer (*Pioneers of fertility innovation: in conversation with Prof. David Gardner*). He also created the widely used 'Gardner Grade' for assessing blastocyst quality, and his work is considered the gold standard in the field of fertility science.



Image 5: Prof. David Gardner



ANCED FERTILITY CENTER OF CHICAGO 22

Image 6: Blastocyst ('Implantation of Blastocysts in Humans')

In embryo assessment, the blastocyst scoring system based on morphology traditionally uses the standardized Gardner score (Gardner DK, 2000) which rates the three morphological features. This scheme [Gardner et al., 2000] is recommended by an international expert group.

1. **Blastocyst Expansion (EXP):** It grades the degree of expansion of the blastocyst cavity.
Scale: 1 to 6 (6 being the most advanced expansion).

2. Inner Cell Mass (ICM): Graded for size and compactness of cells.
Scale: A, B, or C (A = best).
3. Trophectoderm Epithelium (TE): Graded for cohesiveness and the count of cells
Scale: A, B or C (A being the best).

Embryo grading considers all factors as equally important for a successful pregnancy, recognizing that each plays a crucial role. Morphological scoring of blastocyst characteristics is closely associated with success rates (Sivanantham et al., 2022). This grading system has demonstrated effectiveness in prospective studies and is widely endorsed by international experts to improve IVF outcomes.

The embryo grade is usually expressed as a combination of a number and two letters (e.g., 5AA). A 5AA grade represents the highest quality in both developmental stage and cellular composition. Embryos graded 5AB and 5BA also have a good chance of success in transfer. Generally, a higher quality grade correlates with a greater likelihood of successful pregnancy following embryo transfer.

— Challenges in Embryo Grading/Classification

The traditional method used for embryo classification is manual assessment by embryologists.

Limitations of Manual Methods: Manual assessment is subjective and can vary between practitioners. It is not only time-consuming but also subject to variability and human error. Also, in the images from time-lapse technology, stereoscopic cells of different heights overlap in the images which makes it difficult even for an experienced embryologist to accurately count the number of cells. These limitations highlight the need for automated, more consistent solutions that can assist embryologists in making informed decisions.

Applications of AI in IVF

A crucial step in achieving a successful ART outcome is selecting the highest quality embryos for uterine transfer [Wu et al., 2021]. Consequently, there is a strong demand for highly efficient computer-assisted methods for automated embryo grading. Advances in pattern recognition, machine learning, and time-lapse image analysis have led to the development of several computer-assisted techniques for embryo grading and selection. These methods demonstrate that morphological characteristics can serve as key features for evaluating embryo quality.

Currently, deep neural networks have achieved remarkable results in AI tasks within computer vision. Deep learning algorithms, particularly deep convolutional neural networks (CNNs), have become the preferred approach for medical image analysis. Accurate classification of early embryo development stages using these techniques provides valuable information for embryologists, which is critical for successful IVF outcomes.

— Image Classification with AI

Artificial Intelligence (AI), utilizing machine learning and deep learning technologies, empowers computers to learn from data and make predictions. Image classification, a key application of this technology, involves automatically categorizing images into predefined classes. This advancement has transformed various fields, including medical imaging, where it plays a crucial role in diagnosing and analyzing medical conditions.

— AI Applications in Embryo Classification

AI can be applied to embryo classification, potentially improving accuracy and consistency. In embryo classification, AI analyses images of embryos and classify them with a high degree of precision, accuracy and consistency, thus assisting embryologists in making more informed decisions while also improving the overall reliability of the classification process.

— Need for Automated Solutions

One of the greatest problems in assisted reproduction today is the high multiple pregnancy rate (multifetal pregnancies), which happens when multiple embryos are transferred (Puissant et al., 1987;); which is why single embryo transfer [SET] is considered as an effective method where only a single high-quality embryo is transferred. Multiple embryo transfers increase the likelihood of pregnancy but have reported increased risks of adverse maternal and perinatal consequences. Therefore, single embryo transfer (SET) is advocated, aiming to promote singleton gestation and reduce the number of multiple pregnancies. (Ai et al., 2021)

To enhance embryo evaluation, selection (which are crucial components of the IVF process) and to improve live birth rates from SETs, innovative technologies are needed. Recently, AI-based methods have emerged as promising tools, offering objective, standardized, and efficient evaluations. Imaging is one of the most significant areas of AI application. The current focus of AI applications in embryology is automating embryo classification and selection for implantation.

Chapter 2: Literature Review

Recent advancements in artificial intelligence, particularly deep learning, have shown promise in enhancing the predictive capabilities of time-lapse imaging by automating the analysis of complex morphokinetic data (Tran et al., 2019). Deep learning algorithms, such as convolutional neural networks (CNNs), have been successfully applied to various medical imaging tasks showing great success in computer-assisted clinical applications, including diagnosis, classification and segmentation of medical images [Y. Zhou et al., 2018] demonstrating their ability to learn from large datasets and identify patterns that may be imperceptible to the human eye [Mikkel F Kragh., 2019].

CNNs trained on medical imaging in embryology plays a great role in improving embryo viability prediction and grading. They are increasingly employed to analyze time-lapse images, automating complex morphokinetic data analysis with greater accuracy (Tran et al., 2019) as have shown remarkable success in image recognition, particularly in embryo grading and prediction. Kragh et al. developed a CNN model for real-time embryo detection, while Jorgen et al. (2023) demonstrated the efficacy of 3D CNNs for predicting IVF outcomes, underscoring the importance of diverse and representative data.

The deep neural network (DNN) by Rawat et al., 2017, achieved a 98% accuracy rate in predicting blastocyst quality from raw digital images. Similarly, Tsung Jui Chen et al. used the ResNet-50 model on a large dataset, achieving predictive accuracies of 96.24% for blastocyst development, 91.0% for ICM quality, and 84.42% for TE quality. Pegah K. also applied the Inception-V3 model, reaching 96% accuracy in blastocyst image classification.

In addition to CNNs, other computer-assisted methods like pattern recognition (Manna et al., 2004) and machine learning (Nanni, 2013) have been utilized for embryo grading, demonstrating that morphological characteristics can serve as discriminative features for evaluating embryo quality (Morales et al., 2008).

The paper "*A Deep Learning Framework Design for Automatic Blastocyst Evaluation With Multifocal Images*" focuses on developing a deep learning model for automated assessment of human blastocyst quality. The study designed three novel deep neural network (DNN) models based on the VGG-16 architecture, specifically tailored for a time-lapse imaging (TLI) system. Their purpose was to determine the best performing framework for automatic embryo evaluation and classification based on multi-focal images from two classes classification: good and poor.

The Model I (ensemble model) was motivated by Catalin Buiu's paper [C. Buiu, et al., 2020] in which a MobileNetV2 ensemble was used for classification. A shared VGG-16 model was used to extract features. While Model II (voting machine) adopted a voting mechanism to obtain a classification result. The model used a simple VGG-16 network to obtain the probability. Model III employed a multichannel combination network, selecting the sharpest images from the set and transforming them into grayscale before inputting them into the VGG-16 network. These three models were trained on the same blastocyst image dataset and compared their performances. The study found that the multichannel combination model, which used image pre-selection based on sharpness, achieved the best results.

A Review: Despite its promising outcomes, the study had limitations, notably that all blastocyst images and grading information were sourced from a single reproductive centre, and the results were not correlated with clinical outcomes like implantation or live birth. This highlights the need for multicenter studies to validate the model's performance and ensure its broader applicability.

Limitations: Limited availability of large and diverse datasets, which are crucial for training robust models. Many existing studies in this domain often involve small sample sizes and lack independent validation, underscoring the need for further research and data collection in this field.

Historically, the significance of morphological analysis in embryo selection was first highlighted by Scott et al. in 2000. Subsequently, Paternot et al. (2011) proposed a logistic regression-based embryo classification model, though its complexity limited its clinical application. Conaghan et al. (2013) improved embryo selection by combining time-lapse imaging with morphometric software, and Milewski et al. (2017) advanced this by using an artificial neural network (ANN) for embryo selection.

Further contributions include Jonaitis et al.'s (2016) comparative study of neural networks, support vector machines, and nearest neighbour classifiers for detecting cell division times. Khan et al. (2016) employed a deep CNN to classify cell numbers, while Ng et al. (2018) integrated late fusion networks with dynamic programming (DP) to predict various cell development stages, achieving better results than single-frame models.

Another study, "*Multi-Task Deep Learning With Dynamic Programming for Embryo Early Development Stage Classification From Time-Lapse Videos*" proposes a multi-task deep learning with dynamic programming (MTDL-DP) approach for assessing the embryo quality.

This is one of the first studies that applies the MTDL-DP approach for automatic embryo development stage classification from time-lapse videos. Different CNN models were constructed, e.g., one-to-many, many-to-one, and many-to-many. The one-to-many and many-to-many MTDL frameworks performed the best. Considering the trade-off between training time and classification accuracy, they recommend the one-to-many MTDL framework because it achieves comparable performance with the many-to-many MTDL framework, with much lower computational cost.

Now let's have a look at the research paper titled, "*Deep Learning Classification Integrating Embryo Images with Associated Clinical Information from IVF Treatments*". It is from the 12th congress of the Asia Pacific Initiative on Reproduction at Adelaide Convention Centre.

Three AI models were developed, trained, and tested on a database comprising a total of 1503 international treatment cycles (Thailand, Malaysia, and India): 1) A Clinical Multi-Layer Perceptron (MLP) for patient clinical data. 2) An Image Convolutional Neural Network (CNN) AI model using blastocyst images. 3) A fusion model using a combination of both models. All three models were evaluated against their ability to predict clinical pregnancy and live birth.

Result: The MLP model achieved a strong performance of 81.76% accuracy, 90% average precision and 0.91 AUC, the CNN model achieved a performance of 66.89% accuracy, 74% average precision and 0.73 AUC, the Fusion model achieved 82.42% accuracy, 91% average precision and 0.91 AUC. From the visualization process they found that female age and female BMI to be the most important factors, whilst Trophectoderm to be the most important blastocyst feature... ***this is interesting!***

Review: The fusion AI model integrating clinical features and embryo images made more informed predictions, achieving better performance (than separate models). This study demonstrates that AI for IVF applications can increase prediction performance by integrating blastocyst images with patient clinical information.

Segal et al. (2018) developed a random forest-based tool for predicting blastocyst formation, training the model on 2,744 embryos with a 76.4% accuracy rate. Their study highlighted the impact of different transfer learning strategies on model performance, emphasizing the importance of fine-tuning, feature extraction and fully connected layers.

Morphokinetic algorithms have also been proposed for predicting blastocyst formation and selecting embryos with the highest implantation potential (Wong, 2013). Segal et al. (2018) trained a quality assessment model on 1,012 blastocysts, achieving a 91.74% accuracy rate by preprocessing images to improve dataset quality. Tajbakhsh et al. (2016) further demonstrated how varying transfer learning strategies can influence model performance. Aisha Khan et al. (2016) emphasized the importance of 'big data' in reproductive medicine, particularly for training deep learning models using time-lapse imaging data. Also, Kragh's (2019) innovative work combined temporal information with fixed focal length images for embryo evaluation, offering new insights into the potential of deep learning in embryology.

Now let's discuss another research report which explores the development of a deep learning-based classification system for day 3 human embryos. Using low-resolution microscopic images, the study evaluates the effectiveness of convolutional neural networks (CNNs) and a deep ensemble learning (EL) model in grading embryo quality.

The study utilized 3,601 microscopic images from 1,800 IVF couples. Various CNN architectures, including DenseNet, GoogLeNet (Inception V3), Residual Network (ResNet), and Visual Geometry Group (VGGNet), were applied to these images. All networks were pre-trained with 1.28 million nature images from the ImageNet database. The images were preprocessed, downsized to 512x512 pixels, and subjected to real-time data augmentation to enhance model performance. An ensemble learning (EL) model was then developed by integrating discriminative features from the embryos using logistic regression.

To validate the model, an independent test cohort of 699 images from 350 couples was used. The CNN models were compared with the performance of four experienced embryologists, with 8 to 26 years of experience, who classified the embryos based on established morphological criteria.

The feature maps indicated that the proposed EL model, by integrating key morphological features (in accordance with the clinical criteria), achieved 89.16% significantly outperformed embryologists in classifying. Also, deeper CNN models, such as VGG19 and DenseNet169, demonstrated better classification performance, aligning with the clinical criteria used by embryologists. It was observed that the classification performance of VGG and DenseNet models improved with the increase in layers. The accuracy increased from 0.7099 to 0.7352, when VGG16 was switched to VGG19. Similarly, accuracy improved from 0.8526 to 0.8628, when DenseNet121 was replaced by DenseNet169. Thus, it appears that more layers tend to perform better than fewer layers in embryo analyses.

Discussion: The study highlighted the potential of the EL model as an assistive tool in fertility clinics. The superior performance of the model suggests that deeper CNN architectures are better suited for embryo analysis. However, limitations were noted, including the use of a single hospital's data and a population primarily from Northeast China, which may affect the generalizability of the results.

Conclusion: Such models could enhance embryo selection accuracy and consistency in IVF procedures, although further validation across diverse populations is recommended.

Now let's review a report on the "*Deep Learning based Cleavage-stage Blastocyst Prediction with Time-lapse Images*". This report investigates the application of deep learning techniques for predicting blastocyst development at the cleavage stage using time-lapse imaging. The study explores both single-frame and multi-frame network architectures, aiming to evaluate their effectiveness in predicting embryo outcomes.

Single-frame networks included popular architectures like ResNet, MobileNet, Inception, DenseNet, and Vision Transformers (ViT). And, multi-frame networks incorporated ResNet-50 as the primary image encoder, combined with various temporal models and feature aggregation strategies, such as LSTM, GRU, I3D, ResNet3D, MC, R(2+1)D, and MFNet. These networks were applied to time-lapse images to predict the potential of embryos to develop into blastocysts.

Results: The accuracy of single-frame Networks ranged from 66.42% to 74.72%, with no clear correlation between network complexity and performance. Notably, ViT-large, with 302.78 million parameters, had similar accuracy to ResNet50, which has only 23.56 million parameters. The multi-frame networks consistently outperformed single-frame networks, with ResNet-R(2+1)D achieving the highest accuracy at 77.74%. This suggests that morphokinetic information, which multi-frame networks capture, is more beneficial for predicting blastocyst development than static morphological features.

Discussion: The detailed performance analysis of single-frame and multi-frame networks revealed intriguing insights. The analysis revealed that increasing model complexity in single-frame networks does not necessarily improve accuracy. This raises questions about the need for intricate architectures in scenarios where simpler models suffice. In contrast, multi-frame networks achieved accuracies below 77.74%, with relatively minor variations in performance across different models. Technically multi-frame networks, while achieving higher accuracies, presents a relatively small margin of improvement compared to single-frame counterparts.

Conclusion: The study highlights the importance of balancing model complexity and computational efficiency. The similarities in accuracy between models with vastly different parameter counts also emphasize the need to explore more efficient network architectures.

Now let's have a look at the paper titled "*AI in human in vitro fertilization and embryology*". It discusses that one of the biggest challenges in this kind of work is the quality of the training data and how the machine learns from it. For example, the machine learns from training a set of embryo images that have been evaluated and graded by embryologists. It contemplates that machine learning depends on the quality of the input data and states that it would be ideal if ML could occur without human involvement.

Various attempts have been made to replicate Blastocyst morphological grading by automatic systems. These aimed to predict the ICM and TE grade automatically by using static or TLM images. One approach included training TLM Blastocyst data graded by embryologists on the ICM and TE. The preprocessing of the images was done using cropping and a CNN combined with a recurrent neural network to predict the ICM and TE morphology from the same images within three focal planes. The overall AUC was between 0.63 and 0.65 for the ICM and TE, with a high accuracy of distinguishing A (high) versus C (low) grade ICM and TE (97.8% and 98.1%, respectively).

Another paper has described the ability of a CNN to predict Blastocyst expansion (96% accuracy), ICM (91% accuracy), and TE (84% accuracy) quality grades using single static images taken under an inverted microscope (T.J. Chen et al., 2019). From the report of J. Malmsten, et al 2020: With embryo segmentation performed by the AI algorithm the model achieved an accuracy of 93.9% accuracy.

Now let's review a report on "*Artificial intelligence-enabled system for embryo classification and selection based on image analysis*". In this GoogleNet Inception v3 Convolutional Neural Network (CNN) architecture is employed to classify embryos into either the blastocyst or non-blastocyst stage, automating a critical step in in-vitro fertilization (IVF) practices. The system's performance was benchmarked against classifications made by experienced embryologists.

Results: The AI system achieved a classification accuracy of 95%, closely aligning with the average accuracy of 98.25% achieved by embryologists.

Conclusion: This study represents the first implementation of AI-based image analysis in a fertility practice for automating embryo classification and selection. The findings suggest that CNNs, particularly GoogleNet Inception v3, can reliably automate embryo selection with high accuracy.

Next, let's review the paper titled "*Automated Detection of Human Blastocyst Quality Using Convolutional Neural Network and Edge Detector*," which proposes a CNN model for automating blastocyst quality detection. The study highlights the importance of preprocessing using the Canny edge detector, to enhance model accuracy.

Edge detection is vital in computer vision as it significantly affects subsequent image processing tasks like texture recognition. Canny edge detector is notable for its noise reduction and precise edge detection. It effectively separates background noise from complex images and identifies key edges, making it ideal for applications requiring high accuracy.

Methodology: The CNN model was tested on 249 human blastocyst images, with the Canny edge detector used as a preprocessing step before classification.

Results: Initially, the CNN model achieved a detection accuracy of 64.29%. However, after applying the Canny edge detector, the accuracy improved to 98.32% for training data and 84.62% for testing data.

Conclusion: This research demonstrates that incorporating the Canny edge detector significantly enhances the performance of CNN models in classifying blastocyst quality. This approach is essential for improving the accuracy of image-based deep learning models, making it a valuable tool for automating embryo assessment in fertility clinics.

Another study "*Characterization of an Artificial Intelligence Model for Ranking Static Images of Blastocyst Stage Embryos*" evaluates AI model's effectiveness in ranking embryos to predict clinical pregnancy outcomes. A large and diverse dataset from various U.S. IVF labs were used — The model, an ensemble of ResNet-18 CNNs — achieved an AUC of 0.74, outperforming manual grading. The AI focused on key features like the inner cell mass and trophectoderm but struggled with score granularity, suggesting a hybrid approach combining AI and human expertise via manual grading could enhance decision-making.

This study provides insights into the decision-making process of the AI model, revealing that the features it prioritizes closely align with those used in manual grading systems. The AI models show significant potential for enhancing embryo ranking and improving clinical pregnancy outcomes.

Now let us explore recent advancements in applying Artificial Neural Networks (ANNs) to medical imaging in embryology, particularly for predicting and classifying embryo viability. These technologies have shown significant potential in enhancing the accuracy of embryo selection, particularly through the analysis of time-lapse images and morphological features. ANNs have been widely used in the prediction and classification of medical data, including assessing the quality of embryos.

Rocha et al. (2017) utilized segmentation methods to identify the quality of trophectoderm (TE) and inner cell mass (ICM), achieving accuracies of 86.6% for TE and 91.3% for ICM. Uyar et al. (2015) reported an 80.4% accuracy in predicting embryo implantation using a naïve Bayes model. Khosravi et al. (2019) developed an AI framework that achieved over 98% accuracy in classifying human blastocysts. By generating heat maps and integrating a decision tree with factors like blastocyst quality and maternal age, the model also predicts pregnancy likelihood. Xiaoqin Ye et al. (2020) further demonstrated the potential of CNNs and Xception architecture in assessing time-lapse and static embryo images to predict blastocyst implantation.

Conclusion: Deep learning, particularly CNNs, has revolutionized embryology by improving the accuracy and automation of embryo grading and selection. However, the success of these models relies on high-quality datasets and careful consideration of biases and limitations. Continued research and development in this area promise to further enhance reproductive medicine, making embryo selection more objective, consistent, and effective.

While all the experiments provide valuable insights into the potential of deep learning for embryo assessment, further research is necessary to validate these findings on diverse datasets and clinical scenarios. The translation of these models into real-world clinical practice requires addressing challenges related to interpretability, generalizability, and ethical considerations.

Chapter 3: Project Details

— Scope of the Study

Objectives — Aims to Achieve: To investigate various AI architectures for the task of embryo classification and perform a comparative analysis to identify the most effective model(s).

— AI Architect Design Investigation

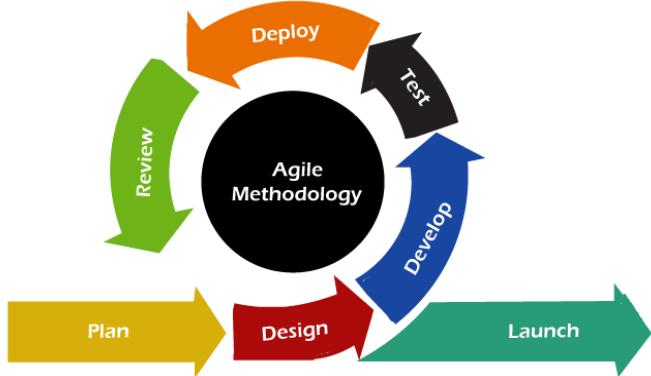
Given the dataset of 2344 images with an imbalance between the classes (300 gold and 2044 silver), choosing the right classification algorithms can significantly impact the model's performance. So this study explores several AI models ranging from traditional machine learning models like Support Vector Machines (SVM) and Random Forests to state-of-the-art deep learning techniques such as CNN, ResNet-50, DenseNet-201, VGG16, InceptionV3, and Xception to determine their suitability for embryo classification. This approach allows us to progressively test models ensuring we find the best-performing algorithm for the classification task.

— Comparative Study

A comparative analysis is crucial as it allows us to evaluate the strengths and weaknesses of different models, ensuring that the best possible method is identified for accurate embryo classification. The comparison is based on four metrics:

- Accuracy: How often the model correctly classifies an embryo.
- Precision: Precision measures the accuracy of the positive predictions
- Recall: Measures the ability to find all relevant instances.
- F1 Score: The harmonic mean of precision and recall.

Conclusions and Recommendations: Identify the best approaches for embryo classification and provide recommendations for future research or clinical application.



— Methodology

For the execution of this MSc project, the Agile methodology was adopted as the guiding framework due to its flexibility and adaptability. Agile is particularly well-suited to the dynamic nature of research and development, where requirements can evolve over time. The project was structured into iterative stages, enabling continuous

progress, regular feedback, and ongoing refinement. This iterative approach allows for effective navigation through uncertainties and complexities, ensuring that adjustments can be made as new insights emerge. By leveraging Agile principles, the aim is to maximize efficiency and deliver valuable, high-quality work.

— GitHub Link

This project (report, code file, prototype and presentation) are posted at this repository:

https://github.com/SalmaJKhan/MSc_Project_30107961

— Tools and Technologies

Listed below are the essential components used for project development:

- Cloud Platforms: Google Cloud
- Development Environment: Google Colab
- Programming Language: Python

— Programming Language: Python

Rationale for Choosing Python:

Python is ideal for this project due to its extensive ecosystem supporting a wide range of machine learning and deep learning models. Python's ease of use and readability simplify handling complex tasks, such as image classification and model comparisons.

Python excels in data processing with libraries like NumPy and Pandas, which handle large embryo datasets, and visualization tools like Matplotlib aid in presenting results. Its versatility allows it to manage everything from data preprocessing to model evaluation.

The GPU acceleration in frameworks like TensorFlow enhances the performance of deep learning models, particularly critical when training resource-intensive networks like ResNet and DenseNet. The availability of pre-trained models simplifies transfer learning, enabling rapid development and fine-tuning of models.

In summary, Python's robust libraries, flexibility, and performance make it the ideal choice for implementing and comparing various AI models in IVF, supporting accurate classification of embryos while facilitating rapid prototyping and experimentation.

— Environment: Google Colab [Advantages]

- Google Colab provides free access to GPUs & TPUs which significantly speed up model training.
- As it is cloud-based, it is accessible everywhere. No worries about hardware limitations or configurations of the local machine.
- Colab comes pre-installed with many popular libraries like TensorFlow, PyTorch, Keras, OpenCV, etc., saving setup time.
- Seamless integration with Google Drive for storing and accessing datasets and other files.
- The interface is simple and user-friendly too.

— Libraries Used

- io: Used for streaming input and output data.
- os: Provides functionality for interacting with the operating system (e.g., listing files).
- numpy (np): Used for numerical operations, handling arrays, and matrices.
- seaborn (sns): Data visualization library for creating attractive statistical graphics.
- pandas (pd): Powerful library for data manipulation and analysis.
- matplotlib.pyplot (plt): Plotting library used to visualize data through plots, charts, etc.
- google.colab.files: Facilitates uploading and downloading files in Google Colab.
- tqdm: Provides progress bars to visually track progress of data processing.
- PIL (Python Imaging Library): Used for working with images.
- sklearn.model_selection.train_test_split: Splits the data into training and testing sets.
- sklearn.preprocessing.LabelEncoder: Converts categorical labels into numerics.
- IPython.display: Used for displaying images.

- Deep Learning Frameworks: TensorFlow/Keras
- Machine Learning Libraries: Scikit-learn
- Data Augmentation and Handling: SMOTE (Synthetic Minority Over-sampling Technique)

Chapter 4: Dataset Description

The dataset utilized in this study originates from the paper, "*An annotated human blastocyst dataset to benchmark deep learning architectures for in vitro fertilization*" (Kromp et al., 2023). It has 2,344 clinically annotated blastocysts images from 837 patients in Portable Network Graphics (PNG) format.

Gardner Score Consortium Annotation: Each image is annotated using the Gardner morphological grading system, which evaluates cell expansion (EXP), the quality of the inner cell mass (ICM), and the trophectoderm Epithelium (TE), by an international consortium of experts, known as Gardner-experts. Each image was reviewed by at least five senior embryologists, and a consensus agreement was calculated to provide the Gardner scores. For validation purposes, a gold-standard test set was created from 300 images, while the remaining images were allocated to the silver-standard training set.

Embryo Development: All oocytes collected during the study were cultured under identical conditions. The embryos were cultured in vitro until day 5, reaching the blastocyst stage.

Blastocyst Imaging: Images were captured using an Olympus IX50 microscope (Vienna, Austria) at 400x magnification. Documentation was facilitated by the Octax EyeWare imaging and archival software (Vitrolife, Sweden).

Data Records: The dataset is available via the Figshare repository (Kromp et al., 2022).

Ethical Compliance: The study adhered to stringent ethical standards and the authors confirmed compliance with all relevant ethical guidelines. Informed consent was obtained from all patients, and the research was approved by the Ethics Committee of the Faculty of Medicine at Johannes Kepler University, Linz, Austria (Nr. 1238/2021). The Local Institutional Review Board (GNEDS) (ethics committee) approved this anonymised database registered under CNIL approval number 1760497.

— Database Inspection

After importing the required libraries, the below steps were taken:

— Listing files and folders from the dataset directory:

```
→ Files and folders in 'Blastocyst_Dataset':  
Blastocyst_Dataset  
Blastocyst_Dataset.csv
```

— Listing the first 7 records from the CSV file using Blastocyst_Dataset.head(7)

```
# Display the first 7 rows of the DataFrame
print("First 7 rows of the dataset:")
Blastocyst_Dataset.head(7)
```

→ First 7 rows of the dataset:

	Image File Name	EXP	ICM	TE	Standard
0	0175_05.png	3.0	1.0	1.0	Silver
1	420_02.png	3.0	0.0	0.0	Silver
2	680_01.png	2.0	0.0	0.0	Silver
3	340_03.png	3.0	0.0	1.0	Silver
4	571_02.png	0.0	3.0	3.0	Silver
5	0064_01.png	3.0	0.0	1.0	Silver
6	478_03.png	3.0	0.0	0.0	Silver

— Listing the last 7 records using Blastocyst_Dataset.tail(7)

```
▶ # Display the last 7 rows of the DataFrame
print("Last 7 rows of the dataset:")
Blastocyst_Dataset.tail(7)
```

→ Last 7 rows of the dataset:

	Image File Name	EXP	ICM	TE	Standard
2337	832_01.png	3.0	1.0	0.0	Gold
2338	832_02.png	2.0	1.0	0.0	Gold
2339	833_02.png	NaN	NaN	1.0	Gold
2340	835_05.png	3.0	0.0	NaN	Gold
2341	836_01.png	3.0	2.0	1.0	Gold
2342	837_01.png	3.0	0.0	1.0	Gold
2343	837_02.png	2.0	1.0	2.0	Gold

— Displaying the first 7 images from the image folder with their corresponding features from the CSV:



File: 0175_05.png
EXP: 3.0, ICM: 1.0, TE: 1.0, Standard: Silver



Investigating the dataframe

- Knowing the Shape:

```
[ ] Blastocyst_Dataset.shape
```

```
→ (2344, 5)
```

Result: The dataset has 2344 rows/records and 5 columns/features.

Displays dimensions (count of rows and columns) of the Blastocyst_Dataset dataframe

- Displaying the Dataset Size:

```
[ ] Blastocyst_Dataset.size
```

```
→ 11720
```

11720 is the overall size of the dataset.

$$2344 \times 5 = 11720$$

- Listing Columns Names

```
[ ] Blastocyst_Dataset.columns
```

→ Index(['Image File Name', 'EXP', 'ICM', 'TE', 'Standard'], dtype='object')

Result: The data frame has the **5** following columns:

1. **Image File Name** has the file name of the image.
2. **EXP** column/feature refers to Blastocyst Cavity Expansion.
3. **ICM** column/feature refers to Inner Cell Mass quality.
4. **TE** column/feature refers to Trophectoderm quality.
5. **Standard** column/feature shows image's Classification: Gold or Silver.

— Counting the Columns

```
[ ] len(Blastocyst_Dataset.columns)
```

→ 5

— Listing the Length of the Dataset

```
▶ len(Blastocyst_Dataset)
```

→ 2344

Result: The dataframe is 2344 rows/records.

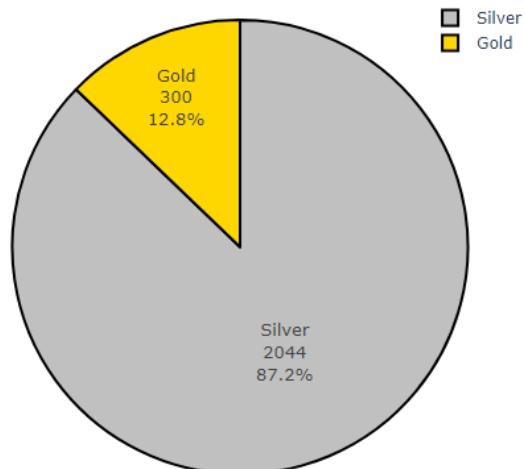
— Listing Unique Values in Target Column

```
[ ] Standards = Blastocyst_Dataset['Standard'].unique()  
print("Standard:", Standards)
```

→ Standard: ['Silver' 'Gold']

Result: The dataset has two unique values in the Standard column, which are Silver and Gold.

— Data Visualization

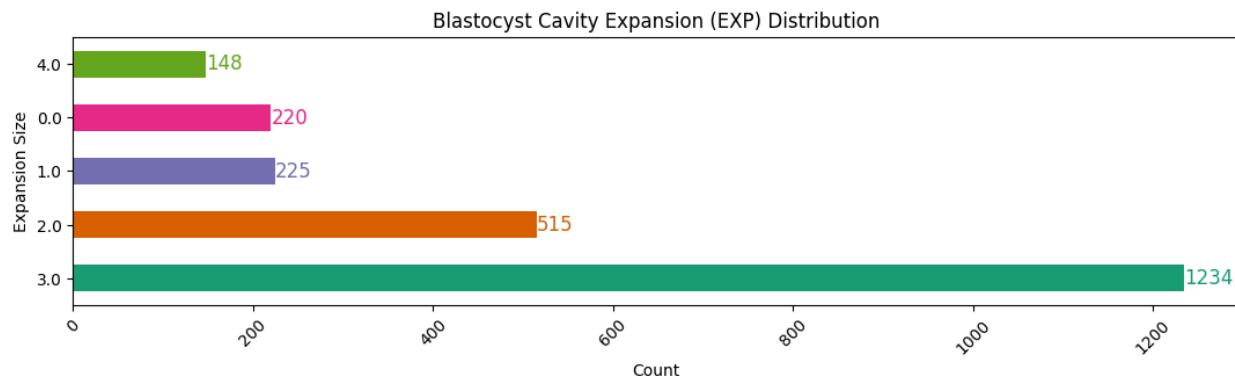


Let's create graphical representations of data to identify patterns, trends, and insights effectively. It simplifies complex data, making it easier to interpret and communicate findings.

— Pie Chart [Data Distribution]

As stated in the dataset description, the dataframe has 2044 silver standard embryo images which account for 87.2% and 300 gold standard images which account for 12.8% of the total images.

— Blastocyst Cavity Expansion (EXP) Distribution



Result:

148 embryos have 4.0 expansion size

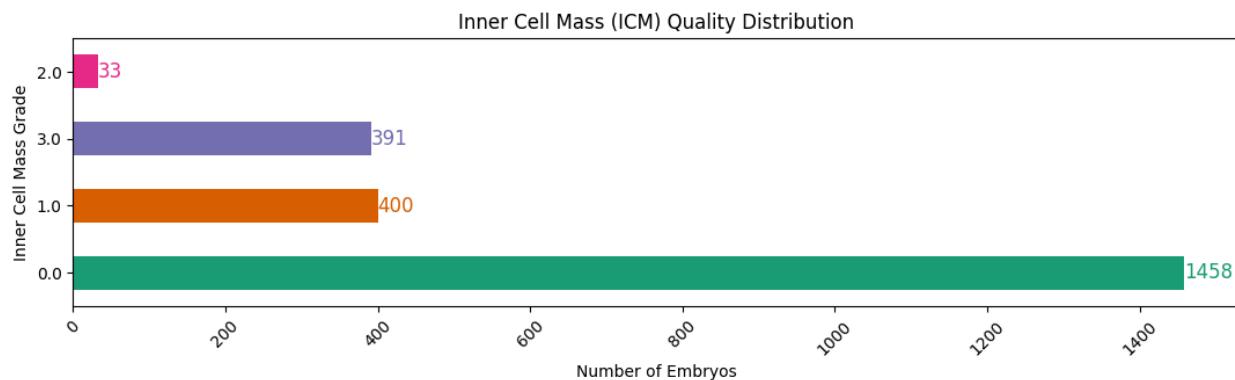
220 have 0.0 expansion size

225 have 1.0 size

515 have 2.0 size, and

1234 have 3.0 expansion size

— Inner Cell Mass (ICM) Quality Distribution



Result:

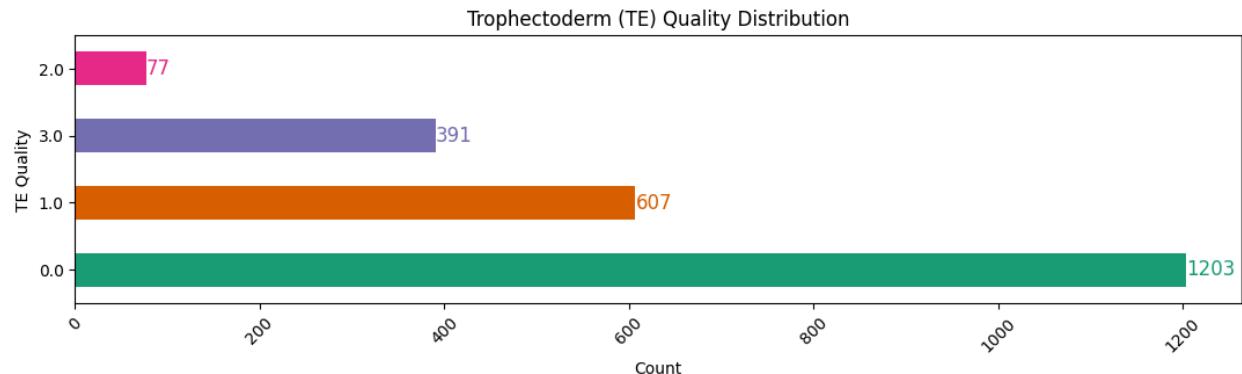
1458 embryos have 0.0 Inner Cell Mass Grade

400 embryos with 1.0 ICM Grade

391 with 3.0 ICM Grade and

just 33 with 2.0 Grade

— Trophectoderm (TE) Quality Distribution



Result:

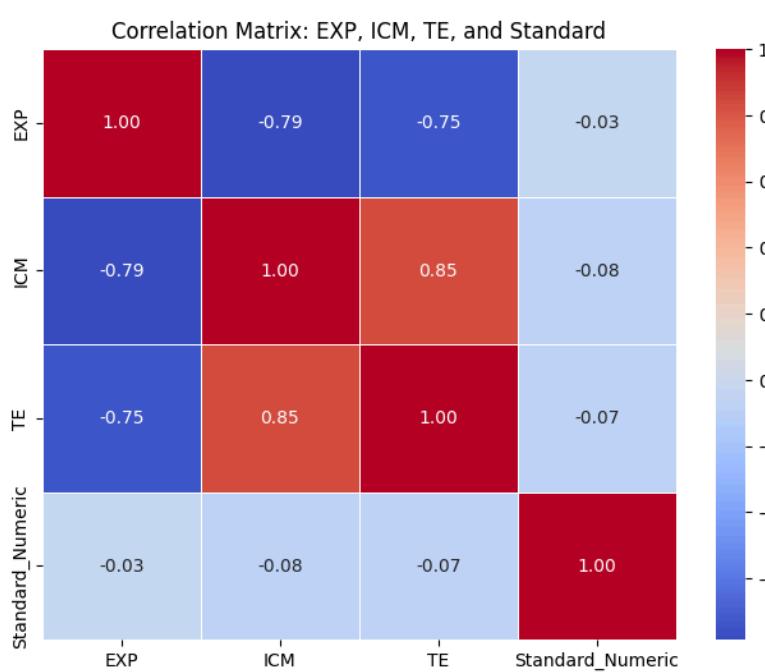
1203 embryos with 0.0 TE Quality Grade

607 with 1.0 TE Grade

391 with 3.0 Grade, and

77 with 2.0 grade

— Correlation Matrix: EXP, ICM, TE, and Standard

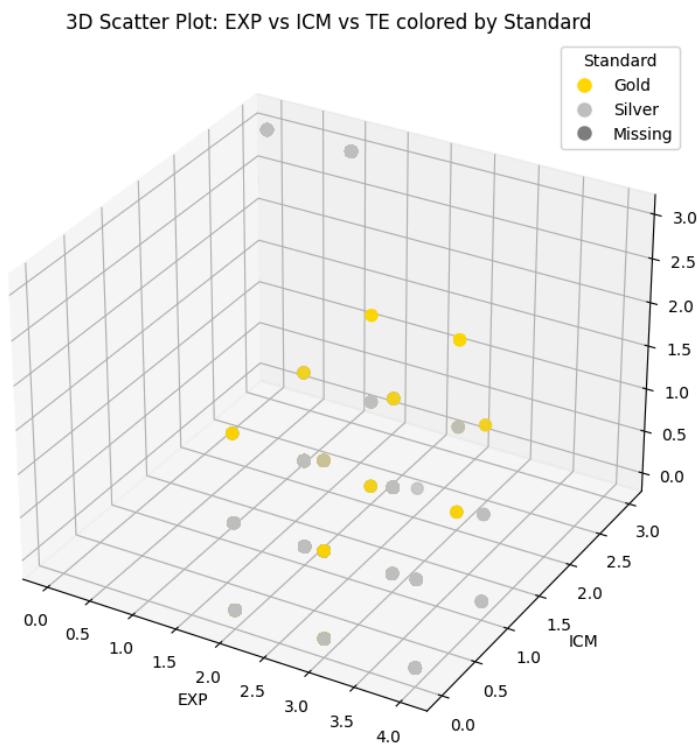


Heatmap Interpretation:

- Diagonal Values (1.00):** These are always 1 because a variable is perfectly correlated with itself.
- EXP vs. ICM (-0.82):** There is a strong negative correlation between EXP and ICM, meaning that as EXP increases, ICM decreases.
- EXP vs. TE (-0.77):** There is a strong negative correlation between EXP and TE, meaning that as EXP increases, TE decreases.
- ICM vs. TE (0.85):** A strong positive correlation is seen between ICM & TE, meaning they increase or decrease together.

- Standard_Numeric vs. Other Variables (All values close to 0):** Indicating that the "Standard_Numeric" variable has little to no linear correlation with EXP, ICM, and TE.

— 3D Scatter Plot: EXP vs ICM vs TE coloured by Standard



This **3D scatter plot** shows distribution of **EXP**, **ICM**, and **TE**, with points coloured according to the classification of "Standard" (Gold, Silver, Missing).

X-axis shows "EXP" denoting embryo expansion grade.

Y-axis shows "ICM" representing the measure of inner cell mass.

Z-axis labelled "TE" refers to the variable trophectoderm Epithelium.

Colours represent different classes:

Yellow: Gold standard embryos.

Light Grey: Silver standard embryos.

Dark Grey: Missing classification.

Purpose of the Plot: This scatter plot is used to visualize embryos

classified as "Gold" and "Silver" and missing (classification) distribution across a three-dimensional space defined by **EXP**, **ICM**, and **TE**. This visualization provides insight into the data that is crucial for the analysis to observe any **patterns** or **clusters** between the different classifications.

— Data Preprocessing

— Checking for Missing Values

```
[ ] Blastocyst_Dataset.isnull().sum()
```

	0
Image File Name	0
EXP	2
ICM	62
TE	66
Standard	0

dtype: int64

Missing Values Result Summary:

- Image File Name: 0 — No missing values.
- EXP: 2 — 2 missing values.
- ICM: 62 — 62 missing values.
- TE: 66 — 66 missing values.
- Standard: 0 — No missing values.

— K Nearest Neighbors (KNN) Imputation

- For ICM and TE we have 62 and 66 missing values respectively. As the dataset in the CSV file shows continuous silver or gold values so using KNN ‘nearest neighbour’ imputation is best suited here – it takes into account the similarity of 5 near instances.
- As EXP has only 2 missing values, simple mean or median imputation is used.

Next, verification was done to check if any missing values remain:

	Blastocyst_Dataset.isnull().sum()
	0
Image File Name	0
EXP	0
ICM	0
TE	0
Standard	0

Result: Success! No missing values.

— Database/Feature Engineering

Feature engineering transforms raw data into a format suitable for machine learning algorithms. Effective feature engineering can significantly impact the success of a model, often leading to improved accuracy, robustness, and interpretability.

— Feature Mapping

Converting the 'Standard' column to numerical values: Gold = 1, Silver = 0

Let's check the head and tail of the dataframe to verify if feature mapping is successful. Result:

	Blastocyst_Dataset.head()																																										
	<table border="1"> <thead> <tr> <th></th> <th>Image File Name</th> <th>EXP</th> <th>ICM</th> <th>TE</th> <th>Standard</th> <th>Standard_Numeric</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0175_05.png</td> <td>3.0</td> <td>1.0</td> <td>1.0</td> <td>Silver</td> <td>0</td> </tr> <tr> <td>1</td> <td>420_02.png</td> <td>3.0</td> <td>0.0</td> <td>0.0</td> <td>Silver</td> <td>0</td> </tr> <tr> <td>2</td> <td>680_01.png</td> <td>2.0</td> <td>0.0</td> <td>0.0</td> <td>Silver</td> <td>0</td> </tr> <tr> <td>3</td> <td>340_03.png</td> <td>3.0</td> <td>0.0</td> <td>1.0</td> <td>Silver</td> <td>0</td> </tr> <tr> <td>4</td> <td>571_02.png</td> <td>0.0</td> <td>3.0</td> <td>3.0</td> <td>Silver</td> <td>0</td> </tr> </tbody> </table>		Image File Name	EXP	ICM	TE	Standard	Standard_Numeric	0	0175_05.png	3.0	1.0	1.0	Silver	0	1	420_02.png	3.0	0.0	0.0	Silver	0	2	680_01.png	2.0	0.0	0.0	Silver	0	3	340_03.png	3.0	0.0	1.0	Silver	0	4	571_02.png	0.0	3.0	3.0	Silver	0
	Image File Name	EXP	ICM	TE	Standard	Standard_Numeric																																					
0	0175_05.png	3.0	1.0	1.0	Silver	0																																					
1	420_02.png	3.0	0.0	0.0	Silver	0																																					
2	680_01.png	2.0	0.0	0.0	Silver	0																																					
3	340_03.png	3.0	0.0	1.0	Silver	0																																					
4	571_02.png	0.0	3.0	3.0	Silver	0																																					
	Blastocyst_Dataset.tail()																																										
	<table border="1"> <thead> <tr> <th></th> <th>Image File Name</th> <th>EXP</th> <th>ICM</th> <th>TE</th> <th>Standard</th> <th>Standard_Numeric</th> </tr> </thead> <tbody> <tr> <td>2339</td> <td>833_02.png</td> <td>2.369342</td> <td>0.0</td> <td>1.0</td> <td>Gold</td> <td>1</td> </tr> <tr> <td>2340</td> <td>835_05.png</td> <td>3.000000</td> <td>0.0</td> <td>0.2</td> <td>Gold</td> <td>1</td> </tr> <tr> <td>2341</td> <td>836_01.png</td> <td>3.000000</td> <td>2.0</td> <td>1.0</td> <td>Gold</td> <td>1</td> </tr> <tr> <td>2342</td> <td>837_01.png</td> <td>3.000000</td> <td>0.0</td> <td>1.0</td> <td>Gold</td> <td>1</td> </tr> <tr> <td>2343</td> <td>837_02.png</td> <td>2.000000</td> <td>1.0</td> <td>2.0</td> <td>Gold</td> <td>1</td> </tr> </tbody> </table>		Image File Name	EXP	ICM	TE	Standard	Standard_Numeric	2339	833_02.png	2.369342	0.0	1.0	Gold	1	2340	835_05.png	3.000000	0.0	0.2	Gold	1	2341	836_01.png	3.000000	2.0	1.0	Gold	1	2342	837_01.png	3.000000	0.0	1.0	Gold	1	2343	837_02.png	2.000000	1.0	2.0	Gold	1
	Image File Name	EXP	ICM	TE	Standard	Standard_Numeric																																					
2339	833_02.png	2.369342	0.0	1.0	Gold	1																																					
2340	835_05.png	3.000000	0.0	0.2	Gold	1																																					
2341	836_01.png	3.000000	2.0	1.0	Gold	1																																					
2342	837_01.png	3.000000	0.0	1.0	Gold	1																																					
2343	837_02.png	2.000000	1.0	2.0	Gold	1																																					

— Splitting the Dataset



Number of training images: 2109

Number of testing images: 235

The dataset was split into training and testing sets to evaluate the model's performance in the ratio of 90:10. 90% training data and 10% to testing. 2109 images are in the training set and the testing set has 235 images.

— Checking Class Distribution



Training Set Class Distribution:

Gold (1): 270

Silver (0): 1839

Testing Set Class Distribution:

Gold (1): 30

Silver (0): 205

— Feature Scaling

Real-world data often has features with varying scales. For machine learning models to interpret these features accurately, they must be brought to the same scale or range. Feature scaling is a preprocessing technique which standardizes or normalizes dataset features — it adjusts the range of independent variables (features) on a similar scale to ensure that no feature dominates others, allowing machine learning models to interpret them accurately.

— Normalizing, Resizing and Rescaling Images

All images are resized to `128x128`. All `image_data` and `labels` are converted to a NumPy array, which is essential for working with machine learning models as they require input data in array form. Image pixel values are normalized to the range `[0, 1]`.

— Image Augmentation

ImageDataGenerator from TensorFlow/Keras is used for this. It generates variations of existing images by applying random transformations (like rotation, width/height shift, shearing, zoom, flip, fill, etc.). It:

- Improves model performance (preventing overfitting)
- Addresses class imbalance
- Increases Dataset Size
- Improving its generalization to new, unseen data
- Enhances Model Robustness

```
# Initialize the ImageDataGenerator with augmentation options
datagen = ImageDataGenerator(
    rotation_range=30,          # Randomly rotating images 30 degrees
    width_shift_range=0.3,       # Randomly shift the image horizontally to 30% of the width
    height_shift_range=0.3,      # Randomly shift the image vertically to 30% of the height
    shear_range=0.3,            # Shear intensity (angle in counter-clockwise direction in degrees)
    zoom_range=0.3,             # Randomly zoom 30% into images
    horizontal_flip=True,       # Randomly flip images horizontally
    fill_mode='nearest'         # Fill in newly created pixels after a transformation
)
```

Display the first 7 augmented images



— Standardization

Standardization scales features so that they have a mean of 0 and a standard deviation of 1 ensuring that each feature contributes equally to the model's learning process.

Chapter 5: Applying AI Algorithms

Traditional Machine Learning Methods

1 — Random Forest Classification [RFC]

A **Random Forest Algorithm** is a traditional supervised machine learning algorithm used for classification tasks. It is an ensemble learning method that builds a forest of decision trees by selecting random subsets of data and features during training and combining their predictions for results. This approach helps reduce overfitting and increases the model's robustness.

In this project, Random Forest serves as the first model architecture employed to classify embryos into gold and silver categories. It is well-suited for this task due to its ability to provide feature importance insights, and deliver strong baseline performance.

Architecture details:

First, the image preprocessing is done. Images are loaded using PIL and converted into arrays. Each image is flattened into a one-dimensional array of pixel values, transforming them into features that can be used by the Random Forest model. The model uses the RandomForestClassifier from the sklearn library. The flattened image arrays (features) and their corresponding labels are fed into the classifier for training, enabling the model to learn the patterns distinguishing "Silver" from "Gold" embryos. Once trained, the model makes predictions on a separate test set of 10% embryo images. Key performance metrics such as accuracy, precision, recall, and F1 score are calculated to assess the model's performance.

- Accuracy: 0.85
- Precision: 0.00
- Recall: 0.00
- F1 Score: 0.00

Classification Report:

	precision	recall	f1-score	support
Silver (0)	0.87	0.98	0.92	205
Gold (1)	0.00	0.00	0.00	30
accuracy			0.85	235
macro avg	0.43	0.49	0.46	235
weighted avg	0.76	0.85	0.80	235

Results

- Accuracy: 0.85 — The model correctly classified 85% of the images overall.

- Precision: 0.00 — Precision is the ratio of true positives to the sum of true positives & false positives. 0.00 score means that the model did not correctly identify any instances of Gold class.

- Recall (or sensitivity): 0.00: Recall is the ratio of true

positives to the sum of true positives and false negatives. The model did not identify any Gold class instances at all.

- F1 Score: 0.00: It is the harmonic mean of precision and recall. 0.00 score means the model's F1 score for the Gold class is zero, which is consistent with zero precision and recall.

Classification Report for Silver (0):

Precision: 0.87: Means 87% of the predicted Silver instances being true positives.

Recall: 0.98: The model identifies 98% of the actual Silver class instances correctly.

F1 Score: 0.92: The harmonic mean of precision and recall for the Silver class.

Classification Report for Gold (1):

Precision: 0.00: The model failed to correctly identify any Gold class instances.

Recall: 0.00: The model missed all actual Gold class instances.

F1 Score: 0.00: This score reflects the lack of precision and recall for the Gold class.

Interpretation

- **Class Imbalance:** The results suggest there is significant imbalance in the dataset, with only 300 Gold class samples compared to Silver 2044. The model is heavily biased toward the majority class (Silver) and fails to learn to identify the minority class (Gold).

- **Model Performance:** While the overall accuracy seems high, it is misleading because the model performs poorly on the minority class. The lack of precision, recall, and F1 score for the Gold class indicates that the model is not effectively distinguishing between the two classes.

2 — Support Vector Machines [SVM]

A support vector machine (SVM) is a machine learning algorithm that uses supervised learning models to solve complex classification (particularly good at binary) problems. Its strength lies in finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space which effectively separates classes, making it well-suited for tasks like embryo classification where clear decision boundaries are crucial. In this project, SVM is used as the second model architecture to classify embryos into gold and silver categories.

Architecture Explained:

- All images are resized to 128x128 pixels, converted to grayscale, and normalized by scaling the pixel values to a range between 0 and 1. The images are then flattened into 1D arrays.
- The dataset is split into training and testing sets using an 90/10 split, ensuring a balanced representation of both "Gold" and "Silver" categories through stratified sampling.
- Model Training: A Support Vector Classifier (SVC) with a linear kernel is used as the classifier.
- Prediction and Evaluation: The trained SVM model makes predictions on the test set.

→ SVM Accuracy: 0.81
 SVM Precision: 0.25
 SVM Recall: 0.23
 SVM F1 Score: 0.24

SVM Classification Report:

	precision	recall	f1-score	support
Silver (0)	0.89	0.90	0.89	205
Gold (1)	0.25	0.23	0.24	30
accuracy			0.81	235
macro avg	0.57	0.57	0.57	235
weighted avg	0.81	0.81	0.81	235

Results:

SVM Accuracy: 0.81 — The model correctly classified 81% of the images overall.

Precision: 0.25 — Low precision indicates many false positives. Among the instances that the model predicted as Gold, only 25% were correctly identified.

Recall: 0.23: The model correctly identified only 23% of the actual Gold instances. Recall is low, showing that the model missed many Gold instances.

F1 Score: 0.24: Reflects poor performance in both precision and recall for the Gold class.

Classification Report for Silver (0):

Precision: 0.89: Meaning 89% of the predicted Silver instances are true positives.

Recall: 0.90: The model identifies 90% of the actual Silver instances correctly.

F1 Score: 0.89: The harmonic mean of precision and recall for the Silver class.

Classification Report for Gold (1):

Precision: 0.25: This reflects that only 25% of the instances predicted as Gold were actually Gold.

Recall: 0.23: Indicates that only 23% of the actual Gold instances were correctly identified.

F1 Score: 0.24: Low score for the Gold class indicates poor performance.

Interpretation

Class Imbalance: The SVM model performs well on the Silver class, with high precision, recall, and F1 score but performs poorly on the Gold class - this indicates a strong bias toward the majority class.

Deep Learning Models

Deep learning is a subset of machine learning that focuses on using artificial neural networks with multiple layers to model complex patterns and representations in data. Inspired by the structure and function of the human brain, deep learning networks — often called deep neural networks — are designed to automatically learn hierarchical features from raw data, making them particularly powerful for tasks involving large and complex datasets.

Deep learning has revolutionized the field of artificial intelligence, enabling significant advances in areas such as computer vision, natural language processing, and reinforcement learning. Its ability to model complex patterns and handle large data makes it a foundational technology for modern AI applications.

3 — Convolutional Neural Networks [CNN]

A Convolutional Neural Network (CNN) is a deep learning model designed for processing structured grid data, particularly images. It employs convolutional layers to automatically and adaptively learn spatial hierarchies of features, such as edges, textures, and complex shapes to capture patterns. Typically used for image classification, object detection, and segmentation tasks, CNNs have become fundamental in computer vision due to their ability to learn and recognize intricate visual patterns.

Model Architecture:

- Input layer: Images are resized to 128x128 pixels, with a single grayscale channel. The pixel values are normalized to fall within the range [0, 1]. The model begins with two convolutional layers: The first uses 32 filters of size 3x3, followed by a ReLU activation function which captures low-level features such as edges and textures. The second has 64 filters of size 3x3, further extracting more complex features from the images. After each convolutional layer, a max-pooling operation with a 2x2 window is applied, reducing spatial dimensions and focusing on important features. Lastly the output is flattened into a 1D vector, preparing it for fully connected layers.

- Fully Connected Layers: Dense layer contains 128 units, activated using ReLU. A dropout layer with 50% rate is added to prevent overfitting by randomly disabling half of the neurons during training. The output layer has 2 units (for two classes: Gold-Silver) activated by softmax function.
- Compilation: Adam optimizer is used for efficient gradient descent, and the sparse categorical cross-entropy loss function is suitable for multi-class classification when labels are integer-encoded. The model also tracks accuracy as a performance metric.
- Training: The model is trained on the dataset for 7 epochs.
- Evaluation: Metrics such as accuracy, precision, recall, and F1 score are calculated to assess the quality of the predictions.

The CNN architecture is well-suited for image classification tasks, allowing the model to automatically learn important features of embryo images and provide highly accurate predictions. The dropout regularization ensures the model generalizes well without overfitting the training data.

Model Performance:

Accuracy: 87.23%

Precision: 0.00%

Recall: 0.00%

F1 Score: 0.00%

Classification Report:

	precision	recall	f1-score	support
Silver (0)	0.87	1.00	0.93	205
Gold (1)	0.00	0.00	0.00	30
accuracy			0.87	235
macro avg	0.44	0.50	0.47	235
weighted avg	0.76	0.87	0.81	235

Summary of Results

1. **Accuracy: 87.23%:** The model correctly classified 87.23% of the images — high accuracy.
2. **Precision 0.00%:** None of the instances predicted as Gold were actually Gold. This indicates a severe issue with false positives for the Gold class.
3. **Recall 0.00%:** The model did not correctly identify any of the actual Gold instances. This suggests the model is failing to detect Gold instances altogether.
4. **F1 Score 0.00%:** Zero because both precision and recall are zero for the Gold class.

Classification Report

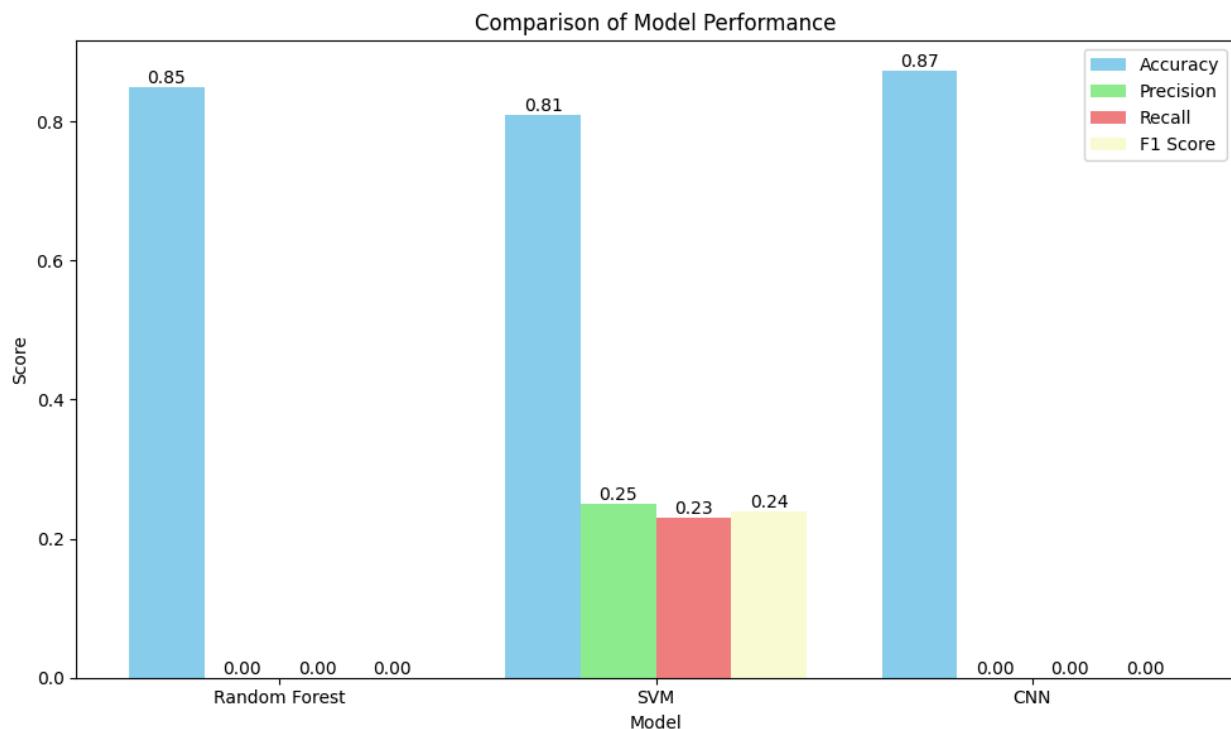
- **Silver (0):**
 - **0.87 Precision:** 87% of the predicted Silver instances are true positives.
 - **1.00 Recall:** The model identifies all actual Silver instances correctly.
 - **0.93 F1 Score:** Indicates strong performance in precision and recall.

- **Gold (1):**
 - **0.00 Precision:** None of the instances predicted as Gold were correct.
 - **0.00 Recall:** None of the actual Gold instances were correctly identified.
 - **0.00 F1 Score:** Reflects poor performance in both precision and recall.

Interpretation

- **Class Imbalance:** The CNN model struggles with the minority class (Gold). Although the overall accuracy is high, the performance on the Gold class is extremely poor.
- **Performance on Silver Class:** The CNN model performs well on the Silver class, with high precision, recall, and F1 score.

Performance Comparison: RFC, SVM and CNN



The bar chart compares the performance of Random Forest, SVM (Support Vector Machine), and CNN (Convolutional Neural Network) models using accuracy, precision, recall, and F1 score as metrics.

Observations:

Random Forest:

- Accuracy: 0.85 (85%)
- Precision, recall, and F1 score are all 0.00.

This suggests the model is not effectively classifying both classes or only predicts one class.

SVM:

- Accuracy is 0.81 (81%).
- Precision is 0.25, recall is 0.23, and F1 score is 0.24. These scores are low but present, indicating some prediction of both classes, though the performance is poor overall.

CNN:

- Accuracy is 0.87 (87%).
- Precision, recall, and F1 score are all 0.00, indicating the same issue as the Random Forest model.

Conclusions:

- This is an issue of class imbalance. The CNN and Random Forest models might be biased toward predicting only one class, leading to a recall and precision of 0.
- The SVM model, while showing lower accuracy, has non-zero precision, recall, and F1 scores, indicating it does predict both classes but does so with relatively poor performance.

Transfer Learning

4 — ResNet-50

ResNet, a groundbreaking CNN architecture, won the 2015 ImageNet competition, demonstrating superior performance in processing individual video frames (He et al., 2016). The model's pre-trained weights on ImageNet (Deng et al., 2009) are particularly beneficial for reducing overfitting in small datasets. Technically it is deep CNN with 50 layers. It addresses the degradation problem by introducing "skip connections" or shortcuts that bypass one or more layers, allowing the model to learn residual functions. This enables effective training of much deeper networks. It balances depth and complexity well and can handle imbalanced datasets by fine-tuning or using techniques like class weighting. ResNet-50 has become a popular architecture for image classification and other computer vision tasks due to its high accuracy and ease of training.

Model Architecture:

The model uses ResNet-50, a powerful deep learning architecture pre-trained on the ImageNet dataset, for classifying embryo images into two categories: Gold (high quality) and Silver (lower quality). The ResNet-50 model is well-suited for this task due to its ability to effectively handle complex image recognition through residual learning, which allows for the training of very deep networks without vanishing gradient issues.

1. The ResNet-50 model is used without the top (classification) layers, retaining the convolutional base for feature extraction. The weights of the pre-trained ResNet-50 layers are frozen, ensuring that the model's learned representations from ImageNet are not modified during training.
2. Custom Layers: After the convolutional base, the architecture includes:
 - o A GlobalAveragePooling2D layer to reduce the spatial dimensions.
 - o A Dense layer with 256 units and a ReLU activation function to introduce non-linearity and help in learning complex patterns from the extracted features.

- A 50% Dropout layer to prevent overfitting by randomly setting half of the nodes to zero.
 - A final Dense layer with a single output unit and a sigmoid activation function for binary classification (either Gold or Silver).
3. Compilation: Adam optimizer with a low learning rate of 0.0001 is used to ensure stable updates to weights. The loss function used is binary crossentropy (for binary classification) and accuracy is tracked as the performance metric.
 4. Data Augmentation: ImageDataGenerator is used to improve the model's robustness and prevent overfitting. This technique increases the diversity of the training data without collecting new samples.
 5. Training: The model is trained for 7 epochs, with 10% of the data reserved for testing.
 6. Evaluation: Metrics such as accuracy, precision, recall, and F1 score are calculated to assess the performance of the ResNet-50 model in distinguishing between Gold and Silver blastocysts.

Results: Same as CNN

ResNet-50 Accuracy: 87.23%
ResNet-50 Precision: 0.00
ResNet-50 Recall: 0.00
ResNet-50 F1 Score: 0.00

5 — DenseNet 201

DenseNet-201 is a very deep convolutional neural network with 201 layers. Its dense connectivity pattern is advantageous for learning complex patterns in the data. In this architecture, each layer is directly connected to every other layer, ensuring maximum information flow and efficient feature reuse. This design helps mitigate the vanishing gradient problem. Despite its depth, DenseNet-201 is parameter-efficient, as it reduces redundancy by reusing features. This architecture excels in image classification tasks, offering high performance while maintaining computational efficiency.

Model Architecture:

1. The DenseNet201 model (from the Keras library) is loaded with pre-trained weights from the ImageNet dataset, which allow the model to utilize knowledge from a large and diverse image dataset. The DenseNet201 model is used as a feature extractor without the final classification layers (include_top=False). The base layers are frozen, meaning they are not updated during training, which preserves the learned features.
2. **Custom Layers:**
 - Input_shape defines the input dimensions (128x128 pixels, 3 colour channels).
 - **GlobalAveragePooling2D** is applied to reduce the spatial dimensions, summarizing the learned information from the DenseNet layers.
 - A fully connected **Dense layer** with 256 units and **ReLU activation** is added, introducing non-linearity and helping the model learn complex patterns from the images.
 - A **Dropout layer** (with 0.5 dropout rate) is employed to mitigate overfitting by randomly setting half of the neurons to zero during each training iteration.

- The final **Dense layer** uses a **sigmoid activation** function to output a probability score for binary classification (Gold or Silver).
3. The model is compiled with the **Adam optimizer** with a learning rate of 0.0001 to control the step size for weight (stable) updates. The **binary crossentropy** loss function is used and the model's performance is evaluated using **accuracy**.
 4. **Data Augmentation** is applied using ImageDataGenerator.
 5. **Data Split:** 90% training 10% testing
 6. **Training:** Model is trained for 7 epochs using a batch size of 33.
 7. **Evaluation:** After training, predictions are made on the test set.

```
DenseNet201 Accuracy: 87.23%
DenseNet201 Precision: 0.00
DenseNet201 Recall: 0.00
DenseNet201 F1 Score: 0.00
```

Results: Same as CNN

Advanced Algorithms

6 — VGG 16, Inception V3, Xception

VGG16: Introduced by the Visual Geometry Group, VGG16 is known for its simplicity and uniform architecture, consisting of 16 layers with small 3x3 convolutional filters and 2x2 max pooling layers - effective for learning initial patterns.

InceptionV3: Developed by Google, InceptionV3 combines multiple convolutional filters of different sizes to capture various feature scales within a single layer. It handles varying feature sizes well and can be a good choice for balancing performance and efficiency. This architecture enhances efficiency and reduces computational cost compared to VGG16, while maintaining high performance.

Xception: An extension of InceptionV3, Xception replaces the standard convolutional layers with depthwise separable convolutions, further optimizing efficiency, making it well-suited for complex tasks with limited computational resources.

Each architecture balances depth, complexity, and computational efficiency, offering trade-offs suited to different applications in computer vision.

Model Architecture:

The data is loaded and the 'Standard' column values ('Gold' or 'Silver') are converted into numeric form where Gold is mapped to 1 and Silver to 0. The images are resized to 128x128, and the pixel values are normalized to a range of [0, 1]. All images and their corresponding labels are converted into NumPy arrays.

The dataset is split into training [90%] and testing [10%] subsets. Augmentation is applied to the training images. The transformations include random rotations, shifts, shearing, zooming, and horizontal flipping, which helps prevent overfitting and makes the model more robust. Now pre-trained models are created using VGG16, InceptionV3, or Xception. All layers of the base model are frozen, meaning their weights are not updated during training (transfer learning).

Custom Layers like `GlobalAveragePooling2D()` are added which reduces the spatial dimensions of the output from the base model, averaging the feature maps for each filter. The 'Dense layer with 256 units with Relu activation and a fully connected layer with 256 neurons is added for non-linear transformation. A `Dropout (0.5)` is added to prevent overfitting by randomly dropping out neurons during training and lastly another single Dense layer with 'sigmoid' activation` for binary classification (Gold or Silver) is added.

The model is compiled with Adam optimizer and a learning rate of 0.0001 is used to update model weights. Binary cross-entropy is used as the loss function for binary classification. The model is trained on augmented data (`datagen.flow()`) and evaluates its performance on the test set. Next predictions are made on the test set using `model.predict()`.

The performance evaluation metrics are accuracy (percentage of correct predictions), precision (proportion of true positive predictions out of all positive predictions), recall (proportion of true positive predictions out of all actual positive samples) and F1 Score (harmonic mean of precision and recall, balancing these two metrics).

Conclusion: Three models (VGG16, InceptionV3 and Xception) are built and trained for 7 epochs and then evaluated on the test set. Their performance metrics are displayed below:

Xception Results:	InceptionV3 Results:	VGG16 Results:
Accuracy: 87.23%	Accuracy: 87.23%	Accuracy: 87.23%
Precision: 0.00	Precision: 0.00	Precision: 0.00
Recall: 0.00	Recall: 0.00	Recall: 0.00
F1 Score: 0.00	F1 Score: 0.00	F1 Score: 0.00

Result: Same as CNN

7 — Xception model with SMOTE

Addressing Class Imbalance: To improve detection of the underrepresented "Gold" class, let's use SMOTE (Synthetic Minority Oversampling Technique). SMOTE generates synthetic samples thus balancing the dataset and prevents bias towards the majority class. SMOTE is applied to the training set for oversampling the "Gold" class. Since SMOTE requires 2D data, we first flatten the image data (X_train), and after oversampling, reshape it back to its original dimensions for model compatibility.

The `train_and_evaluate` function trains the Xception model on the balanced dataset, helping it classify both "Gold" and "Silver" effectively. Finally, we evaluate the model on the original test set to see how well it generalizes to unseen data, especially for the "Gold" class.

Xception with SMOTE Results:

```
Accuracy: 76.17%
Precision: 0.04
Recall: 0.03
F1 Score: 0.03
(0.7617021276595745,
 0.03571428571428571,
 0.03333333333333333,
 0.03448275862068965)
```

Result:

Accuracy: 76.17% – the model correctly predicted 76.17% of the total cases.

Precision: 0.09 — quite low, meaning that only 9% of the samples the model predicted as "Gold" were actually gold. This indicates many false positives, the model predicted gold when the sample was silver.

Recall: 0.10 — quite low again, meaning it only correctly identified 10% of the actual "Gold" embryos.

This suggests a high number of false negatives: The model failed to detect many actual "Gold" embryos.

F1 Score: 0.10 — Low again; the model is not performing well in balancing both precision and recall.

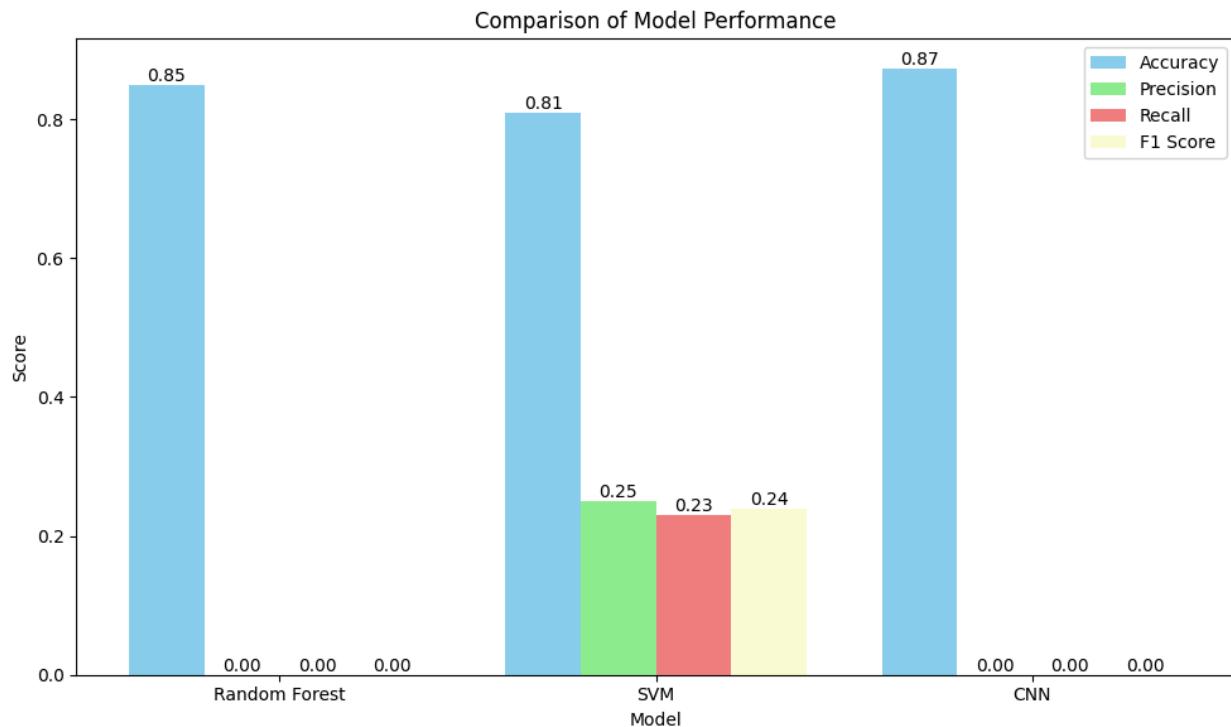
This indicates class imbalance: The model appears biased toward predicting "Silver," leading to higher accuracy but poor performance in detecting "Gold."

Misclassification: The low precision and recall show that the model is misclassifying a significant number of samples, both by incorrectly labelling non-Gold embryos as Gold (false positives) and by failing to identify actual Gold embryos (false negatives).

Chapter 6: Comparative Performance Evaluation — A Study

Performance Comparison:

— Random Forest Classification, SVM and CNN



This **bar chart** compares the performance of three different models: **Random Forest**, **SVM (Support Vector Machine)**, and **CNN (Convolutional Neural Network)** based on four metrics: **Accuracy**, **Precision**, **Recall**, and **F1 Score**.

Results:

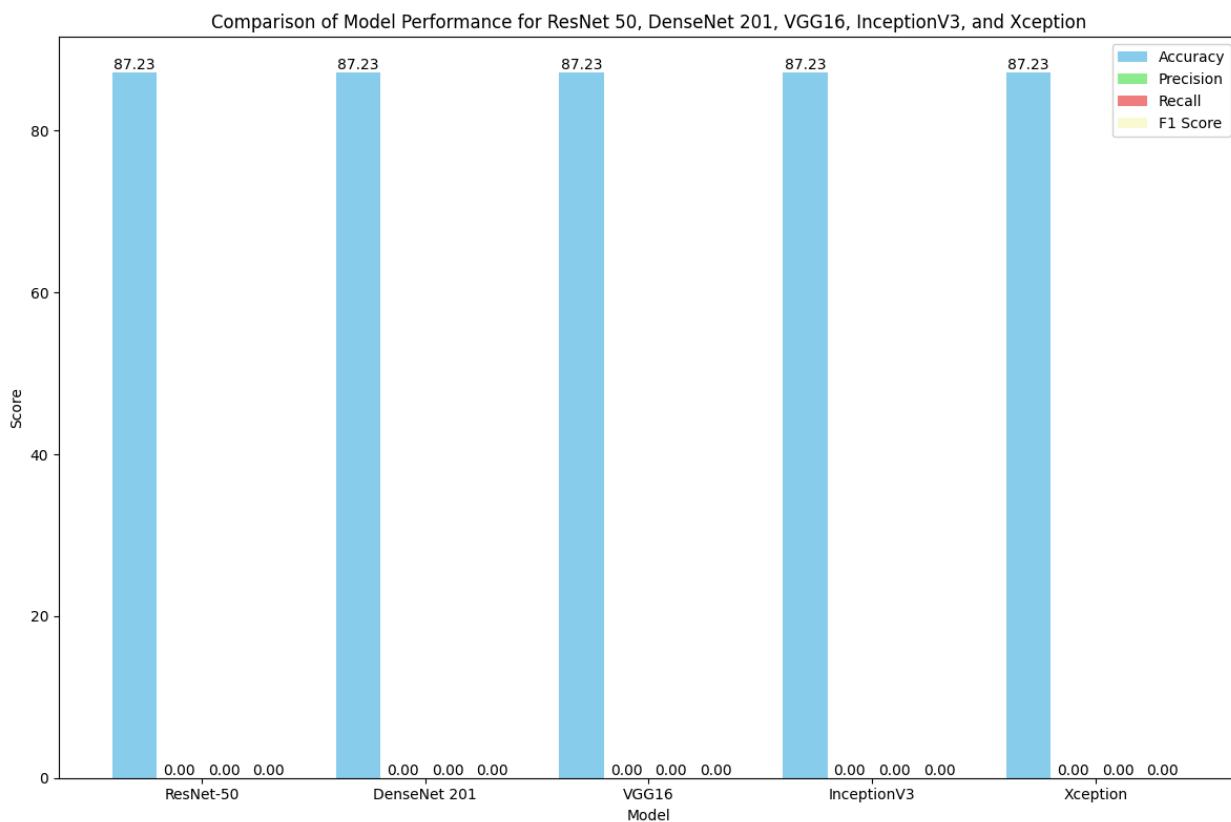
1. **Random Forest:**
 - **Accuracy:** 0.85.
 - **Precision, Recall, and F1 Score** are showing **0.00**, indicating potential issues with performance, such as possible imbalanced classes or misclassification.
2. **SVM:**
 - **Accuracy:** 0.81.
 - **Precision, Recall, and F1 Score** are low:
 - Precision: 0.25
 - Recall: 0.23
 - F1 Score: 0.24
 - Model has reasonable accuracy but struggles with false positives and false negatives.
3. **CNN:**
 - **Accuracy:** 0.87 (highest among the three).
 - Precision, Recall, and F1 Score are showing 0.00 (similar to Random Forest) indicating potential issues with performance, such as possible imbalanced classes or misclassification.

Observations:

- **Accuracy** seems to be the strongest metric for Random Forest and CNN, but their performance on **Precision, Recall, and F1 Score** is poor.
- The **SVM** model has a balanced, but generally low performance across all metrics.
- There is an issue of **class imbalance** in the dataset, where one class (silver) is dominant, leading to the high accuracy but low performance in the other (gold) metric.

2nd Performance Comparison:

— ResNet-50, DenseNet 201, VGG16, InceptionV3, and Xception



This bar chart compares the performance of five different models—ResNet-50, DenseNet 201, VGG16, InceptionV3, and Xception—using metrics such as accuracy, precision, recall, and F1 score.

In the chart:

- Accuracy is represented by light blue bars.
- Precision, Recall, and F1 Score all show values of 0.00 for each model.

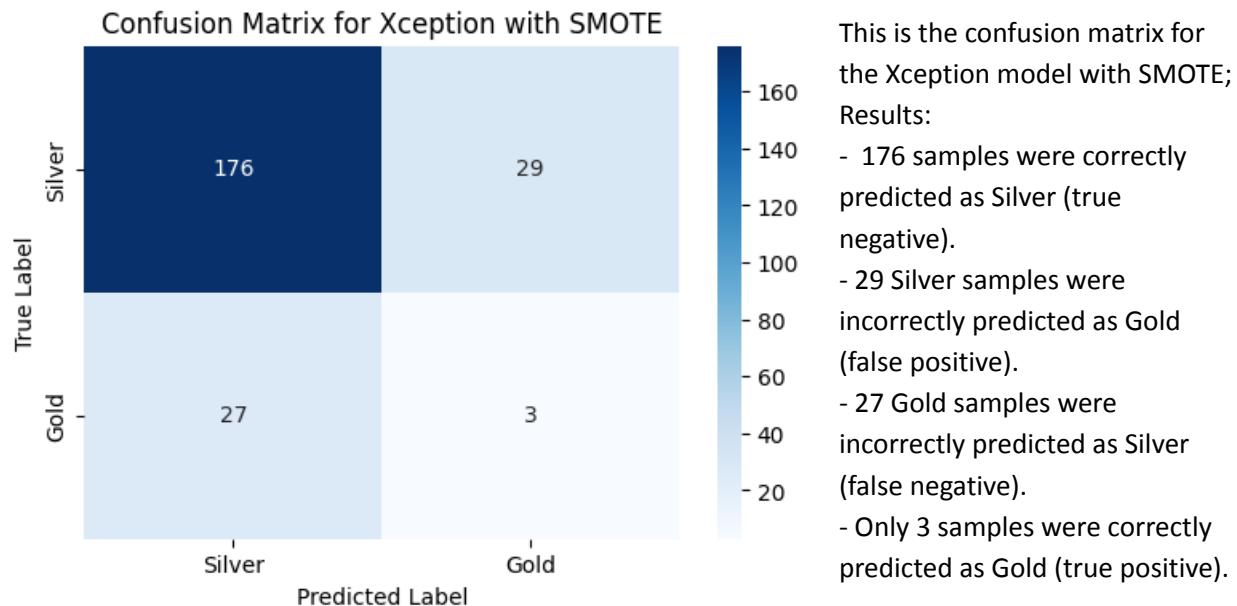
Observations:

- Accuracy for each model (ResNet-50, DenseNet 201, VGG16, InceptionV3, Xception) is 87.23%.

- Metrics for precision, recall, and F1 score are all 0.00 indicate a possible issue with class imbalance. This is happening as the models are predicting only one class for the entire test set (either all Gold or all Silver), leading to poor scores for precision, recall, and F1 score.

Note: Improving the detection of the "Gold" class requires addressing the class imbalance issues that may be causing the model to miss these instances.

— Confusion matrix for Xception model with SMOTE



Key Observations:

The model performs well in identifying Silver embryos, with 176 correctly classified out of a total of 205 Silver samples (176 + 29), leading to high precision and recall for the Silver class. However, the model struggles significantly in classifying Gold embryos, correctly predicting only 3 out of 30 (3 + 27) Gold embryos. This is reflected in the low precision and recall for the Gold class.

Possible Implications:

- **Class Imbalance:** Despite using SMOTE to balance the classes, the model still struggles to correctly identify the minority class (Gold), which indicates that the model is still biased toward the majority class (Silver).
- **Low Precision and Recall for Gold:** As seen in the confusion matrix, the number of false negatives (27) is high for the Gold class, which leads to low recall. Similarly, the number of false positives (29) impacts the precision for the Gold class.

Chapter 7: Project Artefact [Prototype]

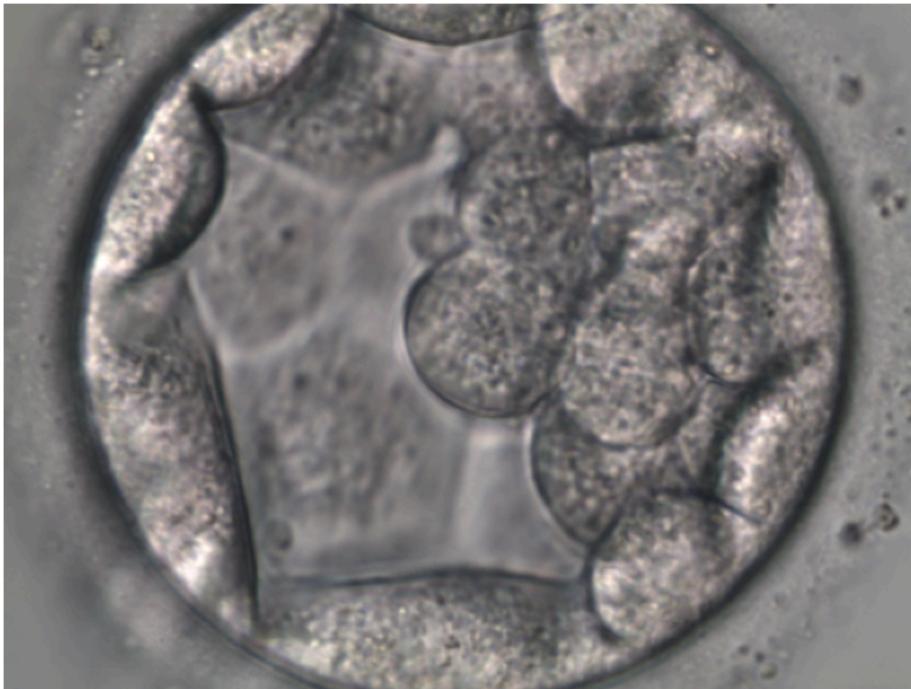
— Features

This prototype demonstrates an AI-based embryo classification tool, which employs the Xception model to classify embryo images into Gardner's Gold or Silver standard. The primary features of the prototype include:

1. **User-Friendly:** Users can upload an embryo image directly through the interface.
2. **Image Preprocessing:** Uploaded images are automatically resized, normalized, and converted into the appropriate format for input into the neural network.
3. **Classification:** The prototype utilizes the pre-trained Xception model, known for its high accuracy in image classification tasks. Custom layers have been added to tailor the model for binary classification specific to embryo grading.
4. **Prediction Output:** Gardner's Gold or Silver standard. This immediate feedback is crucial for clinicians or researchers seeking to quickly evaluate embryo quality.
5. **Visual Feedback:** The uploaded image and its preprocessed version are displayed, offering a visual confirmation to the user.

The embryo image uploaded is: 341_05.png

Uploaded Image: 341_05.png



Processed Image: 341_05.png



```
WARNING:tensorflow:5 out of the last 17 calls to <function TensorArrayWrite</>
1/1 ━━━━━━━━ 2s 2s/step
Predicted Standard: Gold
```

— Scope

This prototype serves as a proof-of-concept for using AI in embryo classification within IVF practices. Its primary application is to assist embryologists and clinicians in:

- **Rapid Assessment:** Providing a quick and objective evaluation of embryo quality.
- **Efficient Tool:** For medical professionals to understand AI applications in embryo classification.
- **Clinical Decision Support:** Helping clinicians make informed decisions about which embryos to select for transfer, thereby potentially improving IVF success rates.

— Limitations

Despite its promising functionality, the prototype has several limitations; few of them are listed below:

1. **Limited Dataset and Generalization:** The prototype may not generalize well to embryos from different patient populations or clinical settings due to variability in imaging conditions, patient demographics, and laboratory protocols.
2. **Need for Further Validation:** This prototype requires extensive testing and validation on diverse and clinically annotated datasets before it can be integrated into routine clinical workflows.

3. **Absence of Clinical Correlation:** The prototype currently classifies embryos based solely on image data without correlating with clinical outcomes like implantation or live birth rates. Thus, its practical utility in improving IVF success rates is yet to be established.
4. **Ethical and Regulatory Considerations:** Implementation in a real-world clinical setting would require careful consideration of ethical issues and regulatory approval.

— Future Enhancements

This prototype represents a step forward in leveraging AI for embryo classification in IVF, offering a glimpse into the potential for automated, objective embryo evaluation. However, further refinement and validation are essential for its transition from a prototype to a reliable clinical tool. To enhance the prototype's clinical applicability and reliability, future work could focus on:

- **Integration with Clinical Data:** Incorporating patient-specific information and clinical outcomes to improve the model's predictive power and provide more comprehensive decision support.
- **Training on Larger, Diverse Datasets:** Utilizing larger datasets from multiple clinical centres to improve model generalization and robustness across different populations.
- **Automated Segmentation:** Enhancing the preprocessing pipeline to include automated segmentation of key embryo features, improving classification accuracy.
- **Real-time Feedback and User Interface:** Developing a more user-friendly interface with real-time feedback capabilities to facilitate ease of use in a clinical environment.
- **Validation Studies:** Conducting multicenter validation studies to ensure the model's applicability and reliability across different laboratory settings and patient demographics.

Chapter 8: Conclusions, Challenges and Future Work

— Conclusion

While Convolutional Neural Networks (CNNs) have shown significant promise in embryo classification, several challenges remain. A primary concern is the need for high-quality image datasets and the risk of overfitting when working with small datasets (Shen et al., 2017). Enhancing AI models with traditional methods, such as integrating CNNs with manual grading by embryologists, could potentially improve embryo selection processes. However, current studies often rely on data sourced from single reproductive centres, limiting the generalizability of findings. The lack of correlation with clinical outcomes like implantation or live birth further underscores the need for multicenter studies to validate AI models and ensure their broader applicability. Future research should focus on integrating these models into clinical workflows and further refining their accuracy through larger, more diverse datasets.

— Challenges

The characteristics needed for human blastocyst assessment are still not fully discovered yet. To date, assessment has mainly been based on morphological characteristics. Limited by experience and subjective factors, the evaluation results of the same embryo may be very different among multiple embryologists, or multiple observations by the same embryologist. An automatic and objective evaluation system will greatly help embryologists perform their daily work. Although video analysis has been a highly active topic in AI, the study of embryo videos still faces significant difficulties due to the appearance variation and occlusion in cell division, which changes continuously and is difficult to track.

Various attempts are required to efficiently incorporate deep learning algorithms and TLM videos to provide reliable methods for embryologists to select good-quality embryos, and thus to help improve the success rate of IVF Cultivation, selection and transplantation of embryos are the key steps in determining a successful implantation during IVF. During the development of embryos, the morphological characteristics and kinetic characteristics are highly correlated with the outcome of transplantation.

Automation in time-lapse imaging using artificial intelligence (AI), either by computer vision technology or by deep learning frameworks serves as an embryo ranking tool. Though it helps considerably in laboratory decisions of choice of embryos, there is a need for prospective trials and long-term follow-up of babies born (Kragh & Karstoft, 2021; Mihdi Afnan et al., 2021) and their health concerns.

— Future Work

To develop a “universal AI embryo evaluation and selection tool” which will be useful across multiple laboratories to efficiently classify embryos and help the clinicians choose the best one for a successful single embryo transfer resulting in a live birth. All local laboratory conditions worldwide, their culture media, type of incubators used, and methods of embryo assessment (which have a significant impact on data interpretation) should be considered. The universal AI model should accommodate all differences of ethnicities, region, religion, colour, heights, weights, ages, medical conditions, IVF cycles, embryo stages, etc. The model should also consider the patient's personal profile, medical history, and lifestyle data. Further studies are required that incorporate the myriad factors necessary for IVF success.

It is important that the tool should be trained on many clinical parameters including the above, along with age of patient, ovarian reserve, stimulation protocols used, and ovarian response, among others, in an effort to incorporate as many factors as are associated with successful IVF treatment.

To validate these methodologies, it will be important to use vastly diverse data sets from multiple laboratories for AI training. It should be a system that uses advanced algorithms incorporating multiple factors to predict IVF outcome. The diversity in geography, patient demographics, and differences in laboratory practices, such as the use of various culture media, incubators, microscopes, and embryo assessment methods, significantly impact data interpretation. To achieve this, vastly diverse datasets from multiple laboratories must be used for training. Therefore, creating universal AI embryo evaluation and selection paradigms that accommodate these differences is essential.

An AI algorithm developed in one clinic may not perform equally well in another, highlighting the necessity for models trained on a variety of clinical parameters, including patient age, ovarian reserve, stimulation protocols, and ovarian response.

Moreover, the field must address the need for robust datasets representative of the various ethnicities, collected from reproductive centres from multiple varied populations, including various developmental stages and outcomes. The challenges associated with annotating datasets should be addressed. Future research should focus on testing on larger datasets, integrating automated segmentation procedures, and combining data from oocytes, pronuclei, and embryos to enhance the system's reliability and applicability. Additionally, incorporating clinically relevant patient and cycle variables is crucial for enhancing AI systems' accuracy and applicability in assisted reproductive technology (ART) practices.

There's a need for multicenter studies to validate the model's performance and ensure its broader applicability. Additionally, advanced AI techniques should be explored for identifying and segmenting inner cell mass (ICM) from the trophectoderm (TE) on blastocyst images, and evaluating the use of AI methods in embryo biopsy and genetic screening.

Efforts should be made to segment, evaluate, and predict individual expansion [EXP], Inner Cell Mass [ICM] and Trophectoderm [TE] quality in order to accurately and automatically predict blastocyst quality. It demands the universal tool should be tested on the most advanced state-of-the-art techniques, and the most suitable should be employed for development.

A universal AI system should also accommodate a range of clinical factors associated with successful IVF treatments. Developing models that can analyze personal profiles, patient medical history, and lifestyle data could predict and improve IVF outcomes. As automation in time-lapse imaging using AI, whether through computer vision technology or deep learning frameworks, advances, there remains a critical need for prospective trials and long-term follow-up to assess the health outcomes of babies born through AI-assisted embryo selection. Addressing these challenges will be pivotal in refining AI's role in assisted reproductive technology.

Chapter 9: Concerns

— IVF: Ethical and Moral Issues

Ethical and Moral issues with embryo classification for IVF include:

Bias and Fairness: AI models can inadvertently introduce biases, potentially favouring certain genetic traits over others. This raises concerns about "designer babies" and the ethical implications of selecting embryos based on perceived genetic advantages.

Data Privacy and Consent: Embryo classification involves sensitive personal and genetic information. Ensuring data privacy and obtaining informed consent for using such data in AI models is critical, especially when the data may be used to make decisions about future lives.

Equity of Access: Advanced AI technologies in IVF may only be accessible to those who can afford them, potentially widening the gap in healthcare equity. This raises questions about fairness and the right to access such reproductive technologies.

Interference in the Reproduction Process: In vitro fertilization (IVF) involves a significant level of medical intervention in the natural reproduction process. This raises moral questions about the extent to which it is acceptable to interfere with natural processes, including concerns about potential long-term effects on children born through these technologies.

Moral and Religious Considerations: Some cultures and religions have specific beliefs about the beginning of life. Using AI to create life may conflict with these beliefs, raising religious moral questions.

Moral Status of the Embryo: One of the most contentious issues is the moral and legal status of the embryo. Since IVF often involves the creation of multiple embryos, with some being selected for implantation and others being discarded or used for research, this raises complex ethical questions about when life begins and the rights of the embryo.

- These dilemmas require careful consideration of both individual rights and broader societal values, balancing the potential benefits of assisted reproduction with respect for moral and ethical principles.
- Addressing these ethical concerns requires careful consideration, transparent communication with patients, and the establishment of clear guidelines and regulations to ensure that AI in embryo classification is used responsibly and ethically.

— Social Stigmas Around IVF

Designer Babies: Using AI for embryo selection could lead to the creation of "designer babies," where parents might choose donor embryos based on desired traits like appearance, intelligence, or athletic ability. This evokes fears of eugenics and societal pressure to select specific characteristics.

Playing God: Some people believe we're interfering with God's plan and view AI in IVF as an unnatural interference with the reproductive process, believing it crosses moral or religious boundaries by "playing God." This stigma is tied to the ethical debate about whether it is appropriate to select/discard embryos.

Value of Human Life: The use of AI in embryo selection can lead to the perception that some lives are valued over others, especially when it comes to discarding embryos that do not meet certain criteria. This can create discomfort and debate about the ethical implications of valuing potential life based on genetic or developmental qualities.

Inequality and Access: The use of advanced AI in IVF might be perceived as a privilege accessible only to those who can afford it, potentially exacerbating social inequalities. This stigma revolves around the idea that AI in reproductive technologies could widen the gap between different socioeconomic groups.

Distrust in Technology: Some individuals harbour a distrust of AI and advanced technologies, especially when applied to something as personal and sensitive as reproduction. Concerns about the accuracy, safety, and long-term effects of using AI in IVF can fuel scepticism and stigma around its use.

— *These social stigmas highlight the need for public education, ethical guidelines, and transparent communication about the benefits and limitations of AI in IVF to address concerns and foster informed discussions.*

References

- Siobhan “At least 12 Million Babies Since the First IVF Birth”. Available at: <https://planer.com/blog/331-2023/1480-at-least-12-million-babies-since-the-first-ivf-birth-in-1978.html> (Accessed: 11 September 2024).
- IVF-Worldwide. Available at: <https://ivf-worldwide.com/> (Accessed: 11 September 2024).
- Biggers, J.D. (2012) ‘IVF and embryo transfer: historical origin and development’. *Reproductive BioMedicine Online* 25(2), pp. 118–127. Available at: <https://www.sciencedirect.com/science/article/pii/S1472648312002921> (Accessed: 11 September 2024).
- *Definition of IN VITRO* (2024). Available at: <https://www.merriam-webster.com/dictionary/in+vitro> (Accessed: 11 September 2024).
- ‘Out of the ovary’. (2008) *Nature* 453(7198), pp. 965–965. Available at: <https://www.nature.com/articles/453965a> (Accessed: 11 September 2024).
- Sperm: MedlinePlus Medical Encyclopedia Image ([no date]). Available at: <https://medlineplus.gov/ency/imagepages/19471.htm> (Accessed: 19 September 2024).
- Team, B.S. (2015) ‘Sperm Fascination’. *Bimek SLV* 20 October. Available at: <https://bimek.com/2015/10/20/sperm-fascination/> (Accessed: 11 September 2024).
- *How in vitro fertilization (IVF) works - Nassim Assefi and Brian A. Levine - YouTube* ([no date]). Available at: <https://www.youtube.com/watch?v=P27waC05Hdk> (Accessed: 11 September 2024).
- *In vitro fertilisation (IVF) | HFEA* ([no date]). Available at: <https://www.hfea.gov.uk/treatments/explore-all-treatments/in-vitro-fertilisation-ivf/> (Accessed: 11 September 2024).
- *Prevalence | Background information | Infertility | CKS | NICE* ([no date]). Available at: <https://cks.nice.org.uk/topics/infertility/background-information/prevalence/> (Accessed: 11 September 2024).
- *Infertility Statistics and Survey 2024 | SingleCare* ([no date]). Available at: <https://www.singlecare.com/blog/news/infertility-statistics/> (Accessed: 11 September 2024).

- *Infertility* ([no date]). Available at: <https://www.who.int/news-room/fact-sheets/detail/infertility> (Accessed: 11 September 2024).
- *IVF* (2017). Available at: <https://www.nhs.uk/conditions/ivf/availability/> (Accessed: 11 September 2024).
- Team, S. (2020) *See the most recent fertility stats*. Available at: <https://www.singlecare.com/blog/news/infertility-statistics/> (Accessed: 11 September 2024).
- *Fertility treatment 2021: preliminary trends and figures* / HFEA ([no date]). Available at: <https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2021-preliminary-trends-and-figures/#main-points> (Accessed: 12 September 2024).
- *IVF Success Rates By Age In 2024* (2023). Available at: <https://www.forbes.com/health/family/ivf-success-rates-by-age/> (Accessed: 12 September 2024).
- Wu, C. et al. (2021) 'A classification system of day 3 human embryos using deep learning'. *Biomedical Signal Processing and Control* 70, p. 102943. Available at: <https://www.sciencedirect.com/science/article/pii/S1746809421005401> (Accessed: 18 September 2024).
- *IVF* (2017). Available at: <https://www.nhs.uk/conditions/ivf/> (Accessed: 12 September 2024).
- Pidd, H. (2020) 'UK's IVF success rate has tripled in the last 20 years'. *The Guardian* 30 June. Available at: <https://www.theguardian.com/society/2020/jun/30/uk-s-ivf-success-rate-has-tripled-in-last-20-years> (Accessed: 12 September 2024).
- *1 in 6 people globally affected by infertility*: WHO ([no date]). Available at: <https://www.who.int/news/item/04-04-2023-1-in-6-people-globally-affected-by-infertility> (Accessed: 12 September 2024).
- *Failed IVF - 15 Reasons For IVF Failure & What To Do next* ([no date]). Available at: <https://fertilityfoundation.org/ivf-failure-why-did-my-ivf-cycle-fail/> (Accessed: 12 September 2024).
- kjkadmin (2022) 'Top 8 reasons for the failure of IVF'. *KJK Hospital* 29 July. Available at: <https://www.kjkhospital.com/reasons-for-the-failure-of-ivf/> (Accessed: 12 September 2024).
- *Understanding the 15 Common Reasons for IVF Failure* ([no date]). Available at: <https://www.cloudninefertility.com/blog/understanding-the-15-common-reasons-for-ivf-failure> (Accessed: 12 September 2024).
- Shauli, N. (2021) *Why IVF Sometimes Fails*. Available at: <https://www.elite-ivf.com/why-ivf-sometimes-fails/> (Accessed: 12 September 2024).
- admin (2020) 'Why does IVF fail?'. *The Fertility & Gynaecology Academy* 3 September. Available at: <https://www.fertility-academy.co.uk/blog/why-does-ivf-fail/> (Accessed: 12 September 2024).
- *Human embryogenesis (article) / Embryology* ([no date]). Available at: <https://www.khanacademy.org/test-prep/mcat/cells/embryology/a/human-embryogenesis> (Accessed: 12 September 2024).
- *IVF Embryo Grading* (2023). Available at: <http://www.medparkhospital.com/en-US/disease-and-treatment/ivf-embryo-grading> (Accessed: 12 September 2024).

- Sivanantham, S., Saravanan, M., Sharma, N., Shrinivasan, J. and Raja, R. (2022) ‘Morphology of inner cell mass: a better predictive biomarker of blastocyst viability’. *PeerJ* 10, p. e13935. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9422976/> (Accessed: 12 September 2024).
- Ai, J., Jin, L., Zheng, Y., Yang, P., Huang, B. and Dong, X. (2021) ‘The Morphology of Inner Cell Mass Is the Strongest Predictor of Live Birth After a Frozen-Thawed Single Embryo Transfer’. *Frontiers in Endocrinology* 12, p. 621221. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7943864/> (Accessed: 12 September 2024).
- *Pioneers of fertility innovation: in conversation with Prof David Gardner* ([no date]). Available at: <https://www.ivf.com.au/blog/pioneers-of-fertility-innovation-in-conversation-with-prof-david-gardner> (Accessed: 12 September 2024).
- *Prof David Gardner*. Available at: <https://www.ivf.com.au/specialists/scientists/prof-david-gardner> (Accessed: 12 September 2024).
- Donna (2024) *Understanding Embryo Grading - Genesis Fertility* New York. Available at: <https://www.genesisfertility.com/blog/understanding-embryo-grading/> (Accessed: 12 September 2024).
- *Understanding IVF Embryo Grading Systems | ARC® Fertility* ([no date]). Available at: <https://www.arcfertility.com/understanding-embryo-grading/> (Accessed: 12 September 2024).
- *A. Degree of expansion*. Available at: <http://atlas.eshre.eu/es/1455006798884190> (Accessed: 12 September 2024).
- MD, A.P.D.A.O. (2024) ‘Embryo Grading Systems During IVF Cycles’. *North Cyprus IVF* 19 April. Available at: <https://www.lowcostivf.net/latest-news-from-north-cyprus-ivf-center/embryo-grading-systems-during-ivf-cycles/> (Accessed: 12 September 2024).
- ‘Implantation of Blastocysts & IVF Embryos in Humans’. ([no date]). Available at: <https://advancedfertility.com/ivf-in-detail/embryo-implantation/> (Accessed: 12 September 2024).
- MD, A.P.D.A.O. (2024) ‘Embryo Grading Systems During IVF Cycles’. *North Cyprus IVF* 19 April. Available at: <https://www.lowcostivf.net/latest-news-from-north-cyprus-ivf-center/embryo-grading-systems-during-ivf-cycles/> (Accessed: 12 September 2024).
- ‘Blastocyst Culture, Day 5 IVF Embryo Transfer & In Vitro Fertilization’. ([no date]). Available at: <https://advancedfertility.com/ivf-in-detail/blastocyst-culture/> (Accessed: 12 September 2024).
- Wang, S., Zhou, C., Zhang, D., Chen, L. and Sun, H. (2021) ‘A Deep Learning Framework Design for Automatic Blastocyst Evaluation With Multifocal Images’. *IEEE Access* PP, pp. 1–1. doi: <https://ieeexplore.ieee.org/document/9328791>
- Liu, Z. et al. (2019) ‘Multi-Task Deep Learning With Dynamic Programming for Embryo Early Development Stage Classification From Time-Lapse Videos’. *IEEE Access* PP, pp. 1–1. doi: <https://doi.org/10.1109/ACCESS.2019.2937765>

- Salih, M. (2023) '#233 : Deep Learning Classification Integrating Embryo Images with Associated Clinical Information from IVF Treatments'. *Fertility & Reproduction* 05(04), pp. 322–322. Available at: <https://www.worldscientific.com/doi/abs/10.1142/S2661318223741334> (Accessed: 28 August 2024).
- Wu, C. et al. (2021) 'A classification system of day 3 human embryos using deep learning'. *Biomedical Signal Processing and Control* 70, p. 102943. Available at: <https://www.sciencedirect.com/science/article/pii/S1746809421005401> (Accessed: 28 August 2024).
- Loewke, K. et al. (2022) 'Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos'. *Fertility and Sterility* 117(3), pp. 528–535. Available at: [https://www.fertstert.org/article/S0015-0282\(21\)02223-8/fulltext](https://www.fertstert.org/article/S0015-0282(21)02223-8/fulltext) (Accessed: 28 August 2024).
- Irmawati, Basari and Gunawan, D. (2019) 'Automated Detection of Human Blastocyst Quality Using Convolutional Neural Network and Edge Detector'. in *2019 1st International Conference on Cybernetics and Intelligent System (ICARUS)*, pp. 181–184. Available at: <https://ieeexplore.ieee.org/document/8874925> (Accessed: 28 August 2024).
- Wang, S., Fan, J., Li, H., Zhao, M., Li, X. and Chan, D.Y.L. (2023) 'A Dataset for Deep Learning based Cleavage-stage Blastocyst Prediction with Time-lapse Images', p. 2023.12.26.573382. Available at: <https://www.biorxiv.org/content/10.1101/2023.12.26.573382v1> (Accessed: 28 August 2024).
- Nasiri, N. and Eftekhari-Yazdi, P. (2015) 'An Overview of The Available Methods for Morphological Scoring of Pre-Implantation Embryos in In Vitro Fertilization'. *Cell journal* 16, pp. 392–405. doi: <https://doi.org/10.22074/cellj.2015.486>
- Zaninovic, N. and Rosenwaks, Z. (2020) 'Artificial intelligence in human in vitro fertilization and embryology'. *Fertility and Sterility* 114(5), pp. 914–920. Available at: <https://www.sciencedirect.com/science/article/pii/S0015028220323992> (Accessed: 28 August 2024).
- Manna, C., Nanni, L., Lumini, A. and Pappalardo, S. (2013) 'Artificial intelligence techniques for embryo and oocyte classification'. *Reproductive BioMedicine Online* 26(1), pp. 42–49. Available at: <https://www.sciencedirect.com/science/article/pii/S147264831200586X> (Accessed: 28 August 2024).
- Dimitriadis, I. et al. (2019) 'Artificial intelligence-enabled system for embryo classification and selection based on image analysis'. *Fertility and Sterility* 111(4), p. e21. Available at: [https://www.fertstert.org/article/S0015-0282\(19\)30179-7/fulltext](https://www.fertstert.org/article/S0015-0282(19)30179-7/fulltext) (Accessed: 28 August 2024).
- Gomez, T. et al. (2022) 'A time-lapse embryo dataset for morphokinetic parameter prediction'. *Data in Brief* 42, p. 108258. Available at: <https://www.sciencedirect.com/science/article/pii/S2352340922004607> (Accessed: 13 May 2024).
- IVF (2017). Available at: <https://www.nhs.uk/conditions/ivf/> (Accessed: 13 May 2024).
- Kesari, G. Here's How AI Is Helping Make Babies By Revolutionizing IVF. Available at: <https://www.forbes.com/sites/ganeskesari/2022/06/30/heres-how-ai-is-helping-make-babies-by-revolutionizing-ivf/> (Accessed: 13 May 2024).

- Goyal, A., Kuchana, M. and Ayyagari, K.P.R. (2020) ‘Machine learning predicts live-birth occurrence before in-vitro fertilization treatment’. Scientific Reports 10(1), p. 20925. Available at: <https://www.nature.com/articles/s41598-020-76928-z> (Accessed: 13 May 2024).

Appendices

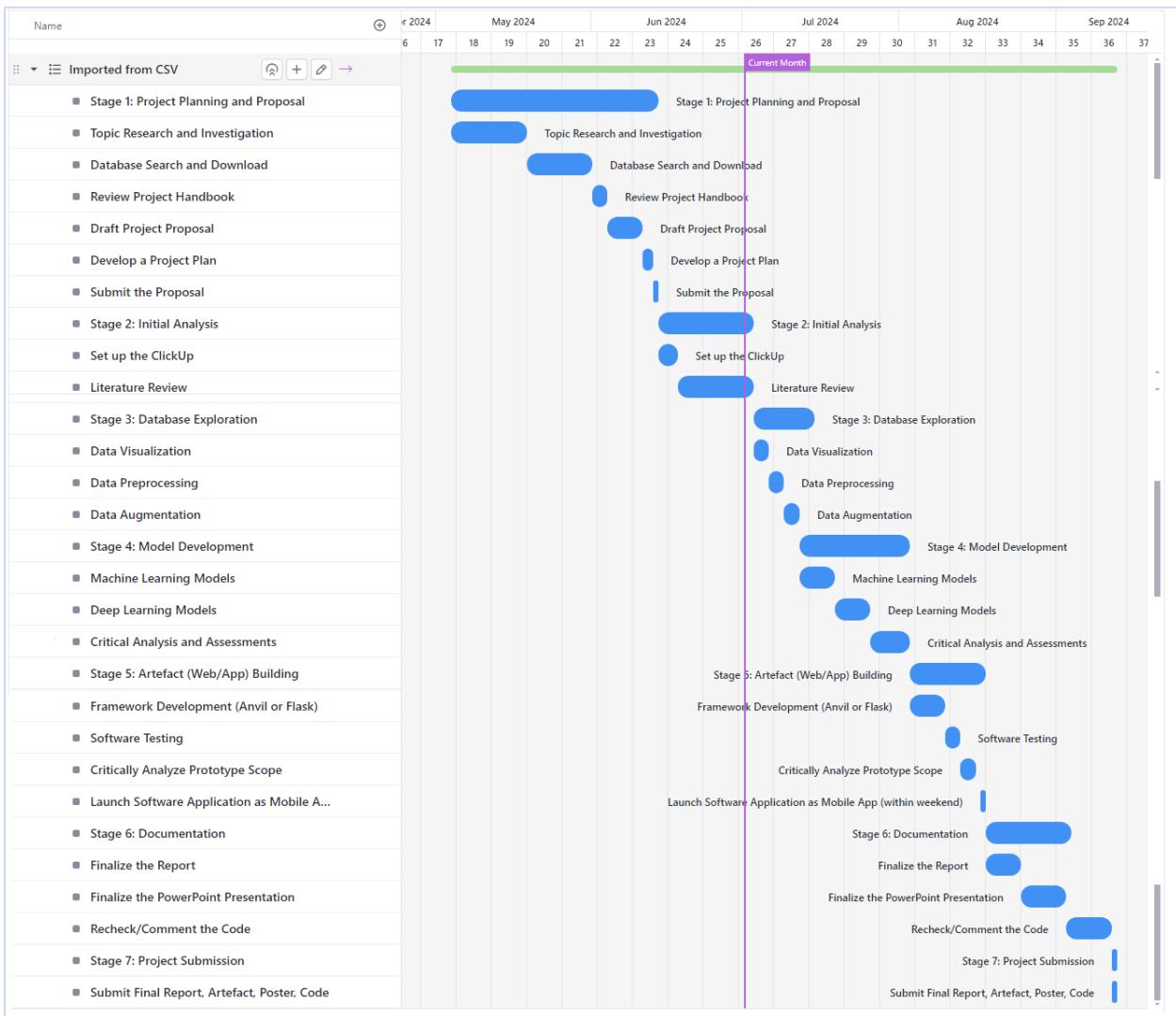
— Project Plan

Tasks with Dates and Deadlines [Time-Table]

Tasks / Activity	Start Date	End Date	Duration (days)
Stage 1: Project Planning and Proposal	4/5/2024	13/6/2024	41
Topic Research and Investigation	4/5/2024	18/5/2024	15
Database Search and Download	19/5/2024	31/5/2024	13
Review Project Handbook	1/6/2024	3/6/2024	3
Draft Project Proposal	4/6/2024	10/6/2024	7
Develop a Project Plan	11/6/2024	12/6/2024	2
Submit the Proposal	13/6/2024	13/6/2024	1
Stage 2: Initial Analysis	14/06/2024	2/7/2024	18
Set up the ClickUp	14/06/2024	17/06/2024	3
Literature Review	18/06/2024	2/7/2024	15
Stage 3: Database Exploration	3/7/2024	14/7/2024	9
Data Visualization	3/7/2024	5/7/2024	3
Data Preprocessing	6/7/2024	8/7/2024	3
Data Augmentation	9/7/2024	11/7/2024	3
Stage 4: Model Development	12/7/2024	2/8/2024	22
Machine Learning Models	12/7/2024	18/7/2024	7
Deep Learning Models	19/7/2024	25/7/2024	7
Critical Analysis and Assessments	26/7/2024	2/8/2024	8
Stage 5: Artefact (Web/App) Building	3/8/2024	17/8/2024	15
Framework Development	3/8/2024	9/8/2024	7
Software Testing	10/8/2024	12/8/2024	3

Critically Analyze Prototype Scope	13/8/2024	15/8/2024	3
Launch Software Application as Mobile App	16/9/2024	17/8/2024	2
Stage 6: Documentation	18/8/2024	19/9/2024	34
Finalize the Report	18/8/2024	01/9/2024	15
Finalize the PowerPoint Presentation	02/9/2024	08/9/2024	7
Recheck/Comment the Code	09/9/2024	18/9/2024	11
Stage 7: Project Submission	19/9/2024	20/9/2024	2
Submit Report, Code and PPT	19/9/2024	20/9/2024	2

— Gantt Chart



— Ethics Form

SECTION A: Project Definition

FOR UNDERGRADUATE & TAUGHT POSTGRADUATE ONLY

Complete the following table with full and relevant information relating to your research.

Student Name	Salma Javid
Student Number	30107961
Student Email Address (University email please)	30107961@students.southwales.ac.uk
Name of Principal Project Supervisor	Dr Carl Jones
Project Title	Embryo Classification: AI Architect Design Investigation and Comparative Study of Multi-Variant Image Classification
Briefly describe the project, being sure to identify any aspects that are relevant to the Ethical Evaluation in Section B. NOTE: A project determined to be High Risk will need to include additional information in Section B to fully-specify the risks and mitigations.	<p>Introduction: This project aims to develop a tool using artificial intelligent algorithms that can classify embryo images for In-Vitro-Fertilization (IVF) treatments.</p> <p>Brief Description: IVF is an assisted reproductive technology (ART) that helps in achieving pregnancy when natural conception is not happening. During IVF, mature eggs are taken from the woman's ovaries and fertilized with a man's sperm in a laboratory. This fertilized egg, now called an embryo, is returned to the woman's womb to grow and develop into a baby.</p> <p>The embryos undergo continuous morphological changes as they develop. So, embryo classification in IVF plays a pivotal role in determining the success or failure of fertility treatments. Traditional methods of embryo assessment rely heavily on manual inspection by embryologists, which can be subjective and labour-intensive. To address these challenges, this project aims to develop an AI based model for automated embryo classification.</p>
Please add an explanation of your study in plain English, with particular focus on any parts of your study which involve human participants. No more than 100 words. This is to help the Faculty Research Ethics Committee (FREC) to understand the project.	<p>The project plan is to build an AI tool to classify embryo images for IVF treatments. IVF is a medical process which helps couples conceive by fertilizing eggs with sperm in a lab and then placing the best embryo in the woman's womb. Embryo classification is crucial for the success of IVF, but traditional methods are manual, time-consuming and prone to errors. This project seeks to automate the process with AI, making it faster, more accurate and reliable. <u>Technically the project only classifies jpg/png images.</u> <u>No human participation is required and I will be using a publicly available dataset.</u></p>

SECTION B: Ethical Evaluation FOR UNDERGRADUATE & TAUGHT POSTGRADUATE ONLY

Consider the following points to determine the level of ethical risk your research presents:

1. Involves those who are considered vulnerable such as:
 - Children under 16.
 - Adults with learning difficulties.Unless in an accredited setting, accompanied by a career or professional with a duty of care.
2. Involves those who are considered highly vulnerable such as:
 - Adults or children with diagnosed mental illness/terminal illness/dementia/in a residential care home.
 - Adults or children in emergency situations.
 - Adults or children with limited capacity to consent
3. Involves those who are “dependent” on others (such as teacher or lecturer to student).
Unless in an accredited setting associated with normal working conditions or routines and within normal operating hours, such as a cultural institution, pre-school, school, or youth club where the research is carried out as part of professional practice such as curriculum development.
4. Requires full NHS ethical approval via the Integrated Research Application System.
5. Requires a Human Tissue Act license.
6. Involves “covert” procedures as in covert observation studies.
7. Involves anything considered “sensitive”. For example, does not carry a risk of those involved disclosing information which compromises the research (e.g., illegal activities; activities where moral opinion may differ, potential professional misconduct – work errors).
8. Induces significant psychological stress or anxiety, or produce humiliation or cause more than fleeting harm / negative consequences beyond the risks encountered in the normal life of the participants (and where the potential for fleeting “harm” is clearly detailed in the participant information sheet). If in doubt regarding definition of the above terminology please contact the research governance office.
9. Involves administration of drugs, placebos or other substances (such as food substances or vitamins) as part of this study.
10. Involves invasive procedures (not limited to blood sampling, collection of biological samples, or passing current through a participant’s body, etc.).
11. Offers any financial inducements to participate in the study.
12. Intends to recruit serving prisoners or serving young offenders via Her Majesty’s Prison & Probation Service.

For your course, there may be specific requirements in **addition** to these, depending on the nature of the subject and how your project is assessed. You must also complete those requirements.

If **none** of the 12 points above apply, then the research can be considered **Low Risk**, unless your course identifies additional criteria relevant to your subject that would render it **High Risk**. This Section is then signed off by yourself and your supervisor, and held on file for review by FREC.

If **any** of the 12 points applies, then the research is considered **High Risk** and students must bring the matter to the attention of their research supervisor immediately. **Research cannot then**

commence until mitigations for the risk are agreed by FREC. Seek advice from your Supervisor, who can help you identify mitigations of the risk or redesign as a Low Risk project.

All students must complete the section below, in collaboration with their supervisor.

Please strike through the statement that **does not apply**.

1. An ethics review has been completed, and the project has been identified as Low Risk.
- ~~2. An ethics review has been completed, and a High Risk was identified. I agree to explain how they may be mitigated below, and agree to abide by any conditions identified at this stage, by my Project Supervisor, the School or the Faculty. I understand that High Risk projects can only proceed with approval from the Faculty Research Ethics Committee.~~

Issues: (Include as much information as possible to help FREC members to understand the issues. Extend onto additional pages as necessary.)

Proposed mitigations: (Include as much information as possible to help FREC members to understand the mitigations. Extend onto additional pages as necessary.)

Student's Signature:



Date: 11th June 2024

Supervisor's statement: I have ensured due diligence and accountable decision making by the student. I have sought appropriate advice where required to support my judgment in this.

Supervisor's Signature:



Date:

11/06/2024

Any false or mis-represented information contributing to this Ethical Evaluation, including attempting to pass off a High Risk project as a Low Risk project, is subject to the Student Misconduct Regulations and may also have legal repercussions.

Both signatures are **required** for all projects, both Low Risk and High Risk.

Thank you!