AI-Based Framework for Automated Cell Cleavage Detection and Timing in Embryo Time-Lapse Videos

Yasmin Alharbi*[1], Sultanah Alshammari[2], Aisha Elaimi[3]
Department of Computer Science-Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia[1,2]
Center of Research Excellence in Artificial Intelligence and Data Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia[2]
Department of Medical Laboratory Technology-Faculty of Applied Medical Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia[3]
Institute of Genomic Medicine Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia[3]
Center of Innovation in Personalized Medicine, King Abdulaziz University, Jeddah 21589, Saudi Arabia[3]

***Abstract*—In vitro fertilization (IVF) has become a primary therapeutic intervention for couples worldwide addressing infertility challenges. IVF success depends critically on embryo quality assessment, where cell cleavage timing serves as a key developmental parameter. Traditional morphological evaluation methods suffer from inter-observer variability and labor-intensive manual analysis. This study presents an automated AI-based framework for cleavage stage detection and cleavage onset timing estimation from Time-Lapse Microscopy (TLM) videos to assist embryologists in embryo selection. The proposed YOLO-based approach addresses significant class imbalance through selective data augmentation and random undersampling strategies. To ensure precise temporal data, an OCR (Optical Character Recognition) library was integrated to automatically read and record the Hours Post-Insemination (HPI) timestamps from the video frames. The proposed framework accurately identifies cell division stages up to the seven-cell stage with 1-2 hours mean timing delay post-insemination. The framework achieves an overall Accuracy of 86.61% , F1-score of 86.24% ,and precision of 86.24% in cleavage stage classification, demonstrating significant improvements over existing methods, particularly in intermediate and later stages (4-cell to 8-cell transitions) where previous research have demonstrated challenges in accurately detecting them. Automated extraction of morphokinetic parameters enables objective embryo assessment, reducing subjectivity in clinical decision-making. The proposed framework demonstrated significant improvements over previous research, which frequently has trouble accurately classifying beyond early cleavage stages. This has implications in improving the selection of good-quality embryos, and thus to help improve the success rate of IVF. This work contributes to advancing assisted reproductive technology by providing reliable, automated embryo quality assessment tools.**

Keywords*—In vitro fertilization; Time-Lapse Microscopy (TLM) videos; AI-based framework; cleavage stage; cleavage onset timing; optical character recognition; Hours Post-Insemination (HPI)

I. Introduction

Infertility affects approximately 17. 5% of adults worldwide each year, and 1 in 6 people are affected by infertility in their lifetime [1]. Serious consequences can negatively affect the lives of infertile couples in terms of physical and psychological well-being, social life, and finances [2]. Among assisted reproductive technology (ART), in vitro fertilization (IVF) is the most popular treatment for infertility since the first successful IVF birth in the UK in 1978. In a typical IVF cycle, the ovaries are stimulated hormonally, mature oocytes are surgically removed, fertilization takes place in a lab, and the embryo is then transferred to the uterus. Currently, there are 35 ART centers in Saudi Arabia that perform more than 20,000 IVF treatment a year [3]. However, the success rates of these treatments are still low. In general, only 20-30% of the transferred embryos result in implant [4] and less than half of the implanted embryos survive to live birth. Thus, selecting the best quality embryo is crucial to ensure the success of the IVF process. Embryologists categorize each embryo based on different morphological criteria and morphokinetic parameters that have been linked to successful embryo development ,including the examination of cell division dynamics [5]. In addition, irregular timings have been associated with genetic abnormalities, implantation failure, and miscarriage that normally manifest during the 8-cell stage [6], [7]. The number of embryonic cells and their subsequent cell division define a cell cleavage stage. It has been demonstrated that the time of cell divisions in human IVF embryos correlates with embryonic viability [6], [8]. Time-lapse studies showed that embryos that cleave in middle time periods have a higher probability of implantation compared to those that develop faster or slower [9], [10]. To assess timing features, embryologists manually annotate the cleavage occurrences of each embryo. Annotating a single embryo usually takes less than two minutes [11], but when examining many embryos (e.g., 5-10), the total time needed can increase significantly - up to 20 minutes [12]. Although manual annotation of embryo morphokinetics is considered efficient, this process is often tiresome and prone to subjectivity or human error. When counting cells in the early stages of embryonic development, cell overlap presents a significant challenge. It becomes harder to identify individual cells as they multiply because they are more densely packed, especially in the stage of 7-8 cells. This overlap makes accurate counting and monitoring more difficult. With rapid technological advancements, artificial intelligence (AI) offers a promising solution to these challenges. AI can help embryologists overcome these obstacles by automating routine tasks such as counting cells, tracking video clips,

and keeping track of division times. The use of artificial intelligence, particularly deep learning models such as convolutional neural networks (CNNs), has expanded significantly in embryo evaluation. These models are capable of automating tasks such as object detection, image segmentation, and optical character recognition (OCR), which contribute to faster and more consistent assessments of embryo development [13]–[16]. In particular, time-lapse imaging combined with AI methods has shown promise in analyzing human embryo development, enabling continuous monitoring of cell division dynamics [17]. Manual annotation of time-lapse microscopy (TLM) videos in embryology is a significant challenge, as it is a time-consuming task and subject to human variation and bias. With the increasing number of cases requiring embryological assessment, the burden on embryologists is increasing. This research develops an AI-based solution to fully automate the annotation of embryo development in TLM videos. The goals include cell detection and counting, cleavage time estimation, and enhancement of precision, productivity, and the reliability of embryo assessment. Achieving these automation objectives will help embryologists in their workflows by facilitating real-time, evidence-based decision-making, thus enhancing the practice of reproductive medicine.

The remainder of this paper is structured as follows. Section II reviews related work. Section III describes the proposed framework in detail. Section IV presents the experimental results, and Section V offers a discussion and analysis of these findings. Finally, Section VI concludes the paper and outlines potential directions for future research.

II. Related Work

Recent decades have witnessed a significant transformation in the medical scene due to technological advancements, with AI emerging as a pivotal factor in transforming healthcare [18]. The incorporation of the two fields has revolutionized various aspects of medical practice, including accurate diagnosis [19], early disease detection [20], [21], and the reading and maintenance of medical records [22]. Medical imaging has benefited greatly from AI, especially deep learning, which can identify intricate patterns and characteristics that the human eye misses [23] By offering fresh viewpoints on pertinent image attributes, these technologies not only improve diagnostic precision but also aid in clinical decision-making. In the context of assisted reproductive technologies (ART), several studies have investigated the potential of employing AI at various stages of the reproductive process. Supervised learning models have been used to evaluate embryo motility [24], assess the cost-effectiveness of egg freezing [25], and classify sperm cells [26]. Artificial neural networks (ANNs), including deep learning models, have also been widely utilized in embryo image segmentation [27], blastocyst assessment [28], and prediction of implantation outcomes, thereby enhancing consistency and scalability in clinical decision-making. Building on these broader applications, recent research has shown increasing interest in exploiting AI, particularly object detection models, to analyze embryo development using TLM. This trend has opened up unprecedented opportunities to automate the detection of critical embryonic developmental stages, such as cell divisions and stage transitions, which are essential criteria for embryo evaluation and selection in the context of assisted reproductive treatments. Studies have varied in their approaches, with an increasing number using AI-based approaches, ranging from probabilistic models to deep learning frameworks.

Among the early efforts that used probabilistic models to automate cleavage detection were [29] proposed a conditional random field (CRF) framework that combines tracking- based and tracking-free elements within a data-driven approach. The model leverages a rich set of discriminative image and geometric features, along with their spatiotemporal context, to enhance the detection of cleavage stages. Using time-lapse embryo microscopy image showed an increased accuracy in identifying mitotic events. Upon assessment of 275 clinical sequences, the suggested model demonstrated an enhanced accuracy of 24.2% compared to traditional tracking-based techniques and 35.7% over tracking-free approaches in measuring division events within the initial 48 hours of development.

Building upon this concept, [30]. extended the use of CRF to automate monitoring for human embryonic cells beyond the 4-cell stage using TLM images. Instead of segmenting or tracking individual cells, their method used a conditional random field framework to predict the number of cells in each frame. This gets around the drawbacks of earlier techniques that were limited to 4-cell analysis. The CRF model incorporates temporal dependencies between frames as well as a wide range of discriminative frame-based and cell evolution features. To further enhance predictions, authors also suggested incorporating learned probabilities of cell transitions. Their approach outperformed alternative models, achieving over 92.4% accuracy in counting cells up to 5 or more when tested on a dataset of 33 embryo sequences.

In a subsequent work [31], the same authors introduced a hybrid model that integrates CNN with CRF to automatically count cells in TLM images of human embryos. Cell counting was formulated as a classification problem, and an end-to-end convolutional neural network trained directly on raw microscopy images were utilized, with the only annotation needed being the total cell count. Their deep CNN method significantly outperformed manual feature-based counting techniques by generating hierarchical visual representations from massive amounts of data. It achieved 92.18% accuracy in predicting cell numbers up to the 5-cell stage. Using a conditional random field to model the growth of embryos over time, they introduced temporal constraints and further improved performance. Despite obstacles from cell overlaps, noise, and image variability, evaluation on a dataset of 265 embryo sequences showed that their deep learning framework could count cells in early embryonic development with reliability. The findings indicate the potential of deep learning techniques to automate cell counting in TLM images without performing segmentation steps.

While earlier methods primarily used probabilistic models and hand-crafted features, more recent research utilizes deep learning techniques to automate feature learning straight from raw images. For instance, Malmsten et al. [32] suggested an automated approach employing a CNN to identify and classify cell divisions directly from raw TLM images, with the goal of reducing the subjectivity and manual effort normally associated with embryonic assessment. Two datasets were used to test their approach: a private human embryo dataset (up to the 8-cell stage) and a public mouse embryo dataset (up to the 4-cell

stage). In 93.9% of cases, the model detected human cleavage stage transitions within a five-frame margin, while it attained nearly 100% accuracy for mouse embryos. This work showed how CNN-based systems can improve annotation uniformity and facilitate automated evaluation of embryo viability in IVF situations.

Contributing to this line of research, Sharma et al. [17] used a YOLO-based detection framework instead of traditional CNN classifiers, which enabled the extraction of accurate timing metrics, which can be used to which can be used to make predictions of embryo viability and the real-time detection of cell divisions. Their approach used YOLOv5 to detect embryonic cells and Pytesseract to extract timestamps from TLM videos. Focusing on embryos up to the five-cell stage, the system achieved an average temporal deviation of only 2–3 hours post-insemination. The model effectively detected important features like blastocysts and morulae and separated cells from detritus, despite performance declining in later stages due to increased cell overlap. Their study further emphasizes the expanding significance of automated, data-driven analysis in improving consistency and efficiency in IVF operations by demonstrating the ability to annotate videos at a rate of roughly one minute apiece.

Despite significant advances in automated embryo monitoring using deep learning and computer vision, existing methods continue to have critical limitations in accurately detecting and tracking cleavage events, especially in complex scenarios involving rapid cell divisions, high overlap, and fragmentation. Many previous studies used CRF or CNN for cell counting and stage classification. Still these methods often lack direct cleavage stage detection or rely on suboptimal OCR methods, such as Pytesseract, to extract hpi, leading to reduced accuracy in time annotation. Additionally, the majority of studies have concentrated on human embryos, but our research utilizes mouse embryo datasets, which have distinct morphological traits and developmental dynamics. An AI-based automated framework is proposed to fill these gaps by utilizing YOLOv11 for accurate cell detection and a more sophisticated OCR library tailored for TLM conditions, which guarantees reliable and accurate hpi extraction.

III. Methodology

This research proposes an AI-based automated annotation framework that predicts cell cleavage onset and cell stage from TLM videos. The pipeline is composed of three main components:

1) Object Detection Model: This uses a deep learning approach to localize and count blastomere cells in TLM embryo frames, supporting fast and accurate cleavage detection even in the presence of overlapping cells or fragments.
2) Cleavage Stage Transition Detection: It monitors changes in the number of detected cells and captures the first frame where a new cell count appears, as this corresponds to the earliest moment of morphological transition that indicates cell cleavage. reflecting the onset of a cleavage stage. Biologically, this is consistent with embryologists' approach of identifying the beginning point of cleavage as the key developmental indicator. The primary output of this step is an organized list of key frames that mark the beginning of each cell division stage, with one frame assigned to each stage. This list forms a valuable basis for the subsequent step in analyzing embryonic development.
3) Time Extraction: It uses OCR to read the hpi from each selected frame. When OCR fails due to low contrast or visual artifacts, the timing is estimated using the video's frame rate and fertilization start time.

For object detection, YOLO (You Only Look Once) is used because of its real-time detection capabilities and track record of accuracy in biomedical image analysis [33]. YOLOv11 was selected despite the consideration of competing models because it strikes a compromise between speed, accuracy, and ease of integration with video-based pipelines. For hpi extraction, Three OCR tools—PyTesseract, EasyOCR, and PaddleOCR— are evaluated and compared chosen for their various underlying technology and relevance to TLM video conditions. While PyTesseract is a typical rule-based image processing OCR engine, EasyOCR and PaddleOCR use deep learning-based designs, which provide more robustness when dealing with noisy, low-contrast, or irregular timestamps. Fig. 1 illustrates the workflow of the methodology. The proposed framework produces two primary outputs: 1) annotated frames enriched with relevant metadata, and 2) an Excel file containing cleavage timing data. As shown in the methodology flowchart, the system identifies the cleavage time points t2–t8, where tn represents the hpi time at which the embryo reaches n cells. The reported temporal outputs correspond to sequential inter-cleavage durations between consecutive developmental stages (e.g., t3 - t2, t4 - t3, t5 - t4). This structured output enables embryologists to systematically analyze early embryo developmental dynamics.

A. Dataset Description

Due to ethical considerations and privacy issues, it is challenging to obtain human embryo dataset. Thus, a dataset of mouse embryos was used instead. According to the National Institutes of Health, mice are a model organism with benefits, including genetic homogeneity, affordability, and simplicity of access [34]. Moreover, the limited availability of publicly accessible resources of human embryos emphasizes the importance of using mouse embryos to enhance this field of study. The study utilizes a TLM dataset that has been previously developed and documented in [35]. The dataset consists of video sequences recorded during the development of approximately 100 mouse embryos, taken from the 2-cell stage through to the blastocyst stage. The videos record cell divisions and dynamic changes in embryo morphology throughout important stages of preimplantation development (e.g. 2-8 cell stage, morula, blastocyst). The dataset was captured using EmbryoScope™ (Vitrolife), a specialized culture device designed with 12 individual wells per slide to allow parallel monitoring of multiple embryos under identical conditions. The dataset consisted of 18 videos, each containing 12 embryos, along with an additional video containing 8 embryos, resulting in a total of 224 individual mice embryo development sequences. These videos were recorded under controlled laboratory conditions [35] and include various cleavage events from the 2-cell stage

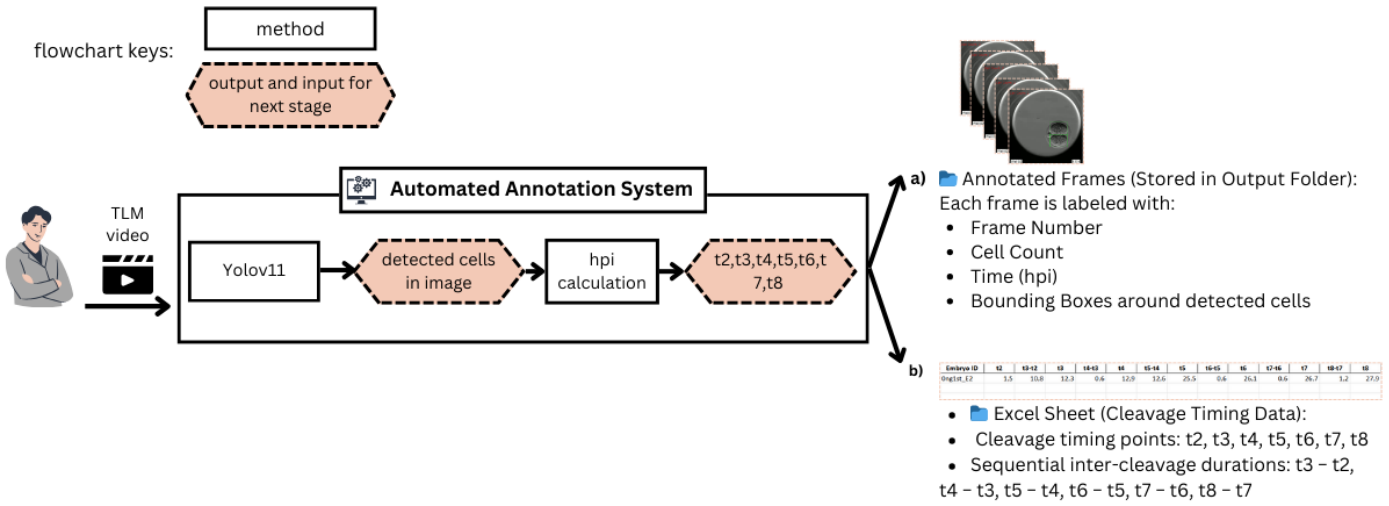

Fig. 1. AI-based automated framework for predicting cell cleavage stages and timing from Time-Lapse Microscopy videos.

to later stages, such as 8-cell or blastocyst formation. All of the videos were saved in AVI format and were captured at a constant 5 frames per second (FPS). The videos also contained time information that was embedded and crucial to our study: the elapsed time since fertilization in hpi is indicated by a visible timestamp in the bottom-right corner of each frame, which is shown as white digits on a black rectangular background.

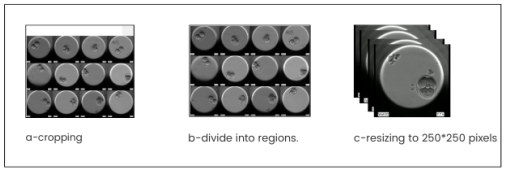

Fig. 2. Preprocessing steps for embryo isolation: (a) cropping of the original TLM field of view, (b) division into fixed regions corresponding to individual embryos, and (c) resizing of each isolated embryo sequence to 250 × 250 pixels.

B. Preprocessing Pipeline

As a first step, each TLM video undergoes basic preprocessing to ensure consistency in frame dimensions, format, and temporal alignment. Using Python and the OpenCV (cv2) package, the videos are cropped to remove unnecessary white space at the top, ensuring that the focus remains on the embryos. For every video, this results in a consistent resolution of 1000x752 pixels and contains 12 embryo wells arranged in a fixed 3-row by 4-column grid. To isolate individual embryos for independent analysis, the video frame is uniformly divided along both axes—width divided by 4 and height divided by 3 resulting in a consistent crop size of 250×250 pixels per well. This standardized resolution ensures that each embryo is spatially separated, uniformly scaled, and appropriately formatted for subsequent deep learning-based object detection and OCR-based timing analysis. As a result of this preprocessing stage, the framework extracts 224 individual embryo videos from the original dataset, each corresponding to a single embryo well and standardized at a resolution of 250×250 pixels. For each video, video frames are extracted at 5 FPS, matching the dataset's native recording rate. This uniform sampling ensured consistent data representation across recordings, supporting reliable analysis and classification of embryos at varying developmental stages. Fig. 2 illustrates the preprocessing steps used for spatial isolation of individual embryos.

C. Data Annotation

Annotation was conducted using 100 TLM videos to provide spatial and temporal oversight of the model development. An experienced embryologist from King Abdulaziz University Hospital was involved to ensure the accuracy of the labeled data and provide the biological "ground truth" needed to effectively train the AI model. First, a frame-level annotation file was created for each embryo video, recording three attributes for every frame: the frame number, the hours post-insemination (hpi), and the cell stage. The dataset included frames from the following cell stages: two-, three-, four -, five -, six -, seven-, and eight-cells. Bounding boxes were manually drawn around each individual blastomere(cell) using the Roboflow platform. Roboflow's annotation tool offers an easy-to-use interface for effective annotation, to ensure consistency and high-quality labeling throughout the dataset [36], [37]. This discrete object detection strategy was adopted over global embryo localization to provide the model with the high-resolution morphological data required to differentiate between consecutive cleavage states (e.g., 2-cell vs. 3-cell). By identifying each cell as an independent instance, the proposed framework calculates the developmental stage through the aggregate count of detected bounding boxes per frame, ensuring that subtle internal boundaries are recognized as primary features for classification .

D. Data Augmentation and Class Distribution Balancing

The dataset comprises frames capturing various stage of cell development, such as the 2-, 3-, 4-, 5-, 6-, 7-, and 8-cell stages. With earlier stages (like 2-cell and 4-cell) being over- represented and later stages (like 6-cell, 7-cell, and 8-cell) being relatively underrepresented, in Fig. 3 a thorough distribution analysis shows a notable imbalance across the stages. The observed imbalance in the number of frames across cleavage stages—characterized by a higher proportion of frames in earlier stages and fewer in later stages—can be attributed to two main factors. The first is **short Later Stage Duration**: Compared to earlier stages, later stages—like 6-cell and 7-cell—occur over significantly shorter time intervals, resulting in fewer frames being recorded. The second factor is **cell Overlap in Advanced Stages**: Individual cells spatially overlap as cell division advances, making precise cell differentiation more difficult. Both data collection and annotation are made more difficult by this phenomenon. All images were scaled to a consistent resolution of 416 x 416 pixels to ensure consistency throughout the dataset. To guarantee sufficient representation of every cell stage for model development and assessment, the dataset is divided into 70% for training, 20% for validation, and 10% for testing. Some of the dataset's original images showed blurring and focus issues, which are common problems with (TLM) generated images. These flaws can occur as a result of lens focus errors or image noise, which, if not addressed, can have a negative impact on the model's performance. The dataset was artificially expanded while maintaining biological relevance through the use of images augmentation. The changes consist of:

1) Flipping both horizontally and vertically: To take into consideration variations in imaging angles.
2) Brightness Adjustment (±8%): To improve resilience to changes in lighting, exposure changes are simulated.
3) Random Gaussian Blur (sigma = 1 px): To improve generalization and simulate noise, making the model more tolerant to minor image flaws and obstacles experienced in TLM imaging.

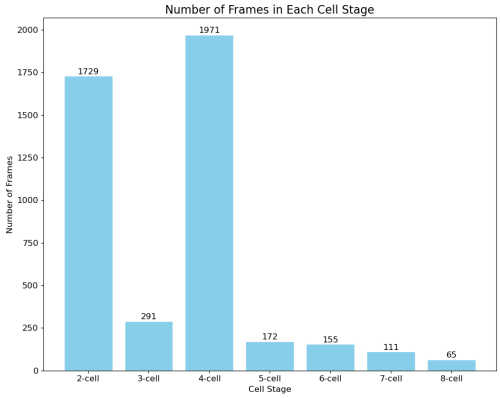

Fig. 3. Dataset distribution.

Undersampling was used to lessen the dominance of over-represented stages following augmentation. The goal was to maintain adequate representation for efficient training while achieving a more consistent distribution of 700 frames across all cell stages. Fig. 4 shows the distribution of the dataset for all cell stages after undersampling and augmentation.

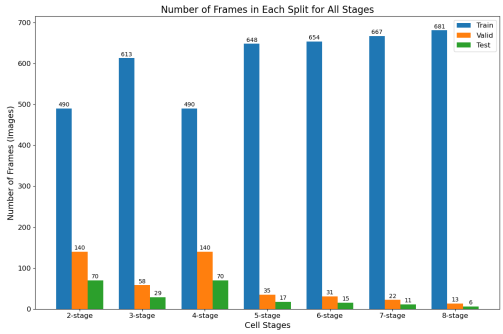

Fig. 4. Distribution of dataset frames for each cell stage: 70% for training, 20% for validation, and 10% for testing after handling class imbalance.

E. Cell Detection

Object Detection is one of the principal application areas in computer vision. The aim of object detection is to detect and localize objects in images through class labels and bounding boxes [38], [39]. There are two main types of detection methods: two-stage detectors like Faster R-CNN and Mask R-CNN that are guaranteed to produce an accurate answer, and one-stage detectors like YOLO and SSD that produce fast answers. There is a growing cadre of methods, including DETR and EfficientDet, that are focused on efficiency and scalability. Among these detection methods, YOLO is a prominent method for real-time detection. YOLO divides images into grids, predicting the bounding boxes and confidence scores for the object(s) contained in each of the grid cells. Using layers that are pre-trained among other bounding box layers, YOLO is able to produce good performance across a range of objects.

YOLOV11 is the most recent model in Ultralytics' YOLO series of real-time object detectors. it comes in a variety of model sizes, from nano to extra-large, making it suitable for a wide range of applications, from resource-limited edge devices to powerful computing systems. It adds new architectural features such as the C3k2 block, SPPF, and C2PSA, which improve its ability to extract and process features more efficiently. These enhancements enable YOLOv11 to better analyze and interpret complex visual data, potentially improving detection accuracy in a variety of scenarios. The integration of advanced spatial attention mechanisms, such as the C2PSA, is a significant improvement because it allows the model to focus on important regions of an image. YOLOV11 was selected for cell detection due to its ability to efficiently process large-scale time-lapse microscopy (TLM) datasets while maintaining reliable localization performance. Unlike two-stage object detectors, YOLO's single-stage architecture enables rapid frame-wise inference, which is essential for analyzing long TLM video sequences at high temporal resolution. Furthermore, the proposed framework uses YOLO exclusively for detecting individual cells rather than directly classifying cleavage stages. Developmental stages are subsequently inferred by counting detected cells across consecutive frames, thereby avoiding reliance on subtle morphological distinctions between stages and improving robustness under variable illumination and cell

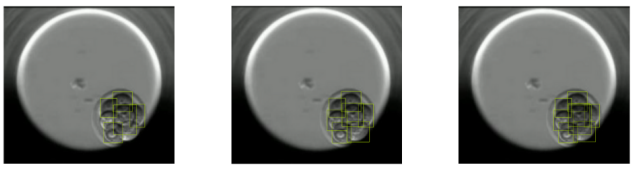

Fig. 5. Impact of IoU threshold on detection accuracy.

overlap conditions. Our proposed model operated as follows: Detection, Identifying and locating every instance of the "cell" class in the frame. Stage Assignment: The overall count of detected cells within a frame determines the cleavage stage label for that frame. For example, the existence of five unique cells signifies the 5-cell stage. The YOLOv11 model was set up on Ultralytics Hub [40]. To improve accuracy and shorten training time, a pre-trained model served as a foundation for the training process. The training procedure was set up with 100 epochs using 416×416 resized input images, a 100-epoch patience threshold to avoid overfitting. The batch size of 22 is determined by the AutoBatch mechanism, which automatically determines the maximum batch size that can be accommodated in the GPU memory during training to adjust to the available computational resources. The optimizer was configured with a momentum of 0.9 and a learning rate of 0.002. To train the YOLOv11 model efficiently, two NVIDIA T4 GPUs were used in a Kaggle environment.

YOLOv11 efficiently locates and categorizes overlapping objects, such as cells in our case, by employing anchor boxes to predict bounding boxes and their associated confidence scores in a single forward pass. During detection, many bounding boxes may be proposed for the same object. In order to overcome this, YOLO employs Non-Maximum Suppression (NMS), which selects the box with the highest confidence score after suppressing overlapping boxes based on their Intersection over Union (IoU) values. This ensures precise localization of individual cells, even when they significantly overlap, and sustains a high detection performance as indicated by measures like mean Average Precision (mAP). accuracy. Fig. 5 illustrates the impact of Intersection over Union (IoU) thresholds on detection accuracy. A higher IoU value (e.g., 0.70) facilitates more precise and reliable cell localization, whereas lower IoU values (e.g., 0.20 or 0.40) are associated with increased misdetections or overlapping prediction errors.

F. Cleavage Stage Identification and HPI Determination

Using a fixed confidence threshold of 0.55 across all cleavage stages, a YOLOv11 model is employed to detect individual cells within each frame. Additionally, an IoU threshold of 0.70 is applied by the model to provide precise bounding box predictions. The cell count is shown on the frame, and bounding boxes are drawn around any cells that are found. Only cleavage stages containing between 2 and 8 cells are considered in this analysis. A frame is saved to the output folder only when it corresponds to the initiation of a new cleavage stage, defined by the following criteria:

1) Uniqueness: if the number of detected cells is unique, meaning it hasn't been recorded before.
2) Progression: The current cell count is higher than the previous count that was noted.

Because of the inherent heterogeneity of embryo growth, certain division phases might not be captured in the video, necessitating this second requirement. For instance, the three-cell stage might not be apparent in the video, or some embryos might split straight from two to four cells without going through it. This criterion reduces the possibility of errors resulting from overlapping cells that could otherwise be misinterpreted for example, falsely recognizing three cells when only two are present by guaranteeing that only major stage transitions are recorded. Determining the corresponding hpi of the saved frames is the next step. Accurate hpi calculation is essential to monitor embryo development over time. OCR is a technology that transforms handwritten or printed text in images into digital formats that can be read by machines [41]. OCR is essential to biomedical research because it allows the extraction of quantitative data from medical images, including Medical Records [42], Ocular Biometry reports [43], text from figures in full-text biomedical articles [44], and experimental annotations. Numerous efficient Python libraries have been developed for OCR, including Tesseract, a popular open-source engine; Pytesseract [45], a Python wrapper for Tesseract; OpenCV, a flexible computer vision library; and EasyOCR [46], an open-source OCR tool based on deep learning. EasyOCR uses long short-term memory (LSTM) architectures and CNNs to detect and recognize text. PaddleOCR, deep learning-based OCR system created by the PaddlePaddle group. For high text detection and recognition accuracy, PaddleOCR utilizes CNNs, attention mechanisms, and sequence models like recurrent neural networks (RNNs) [47], [48]. Deep learning-based OCR tools were chosen for their robustness and ability to handle the complex imaging conditions common in biomedical TLM data, thus making sure that the annotations of hpi are accurately extracted.

The timestamp was extracted by cropping a fixed region of interest corresponding to the white rectangular overlay containing the hpi value, which appears at a consistent location across all time-lapse images. The cropping coordinates (x = 300, y = 350, width = 200, height = 200) were predefined and automatically applied to all frames, without manual intervention or sequence-specific adjustment, ensuring standardized OCR input across the dataset. To ensure the reliability of the extracted hpi values, a simple numerical validation step was applied to the OCR output. Since all embryos in the dataset fall within a biologically plausible range below 73 hours post-insemination, any extracted value exceeding 73 hpi was treated as an OCR artifact caused by an erroneous leading digit and was automatically corrected by removing the leading digit, restoring the expected temporal range. This lightweight validation step effectively filtered OCR misreads while preserving correct timestamps for downstream cleavage timing analysis. Due to variability in image brightness or quality, OCR is not always flawless and sometimes misreads numbers. Three OCR libraries—PyTesseract, EasyOCR, and PaddleOCR—were tested for performance. The hpi value is allocated to the frame if OCR is successful in extracting it. When OCR is unable to retrieve the timestamp. Several reasons can lead to OCR failures:

- creation of out-of-bound values, where valid times-

tamps (e.g., 69.9 h) are misinterpreted as wrong values like '4569,' which exceeds the system's maximum allowed hpi range (mice embryo typically reaches the 8-cell stage around 48 to 72 hpi).

- misrecognition of digits, where the algorithm finds or interprets numbers improperly.

An alternate approach has to be used. In these situations, our dataset of TLM videos' empirical observations of cleavage timing serve as the basis for this alternate approach. After examining several embryos, it was observed that the cleavage stages occurred at a comparatively constant interval of about 0.6 hpi each frame. Previous studies have employed a similar approach, in which, in the absence of precise hpi values, the time annotations for cleavage stages were obtained using frame rate and time-lapse system setups [17].

$$t_n = t_1 + 0.6 \times (F_n - 1)$$

where:

- t_1 – The hpi of the first frame.
- F_n – The frame number for which the time is being calculate.
- 0.6 – The frame interval.

IV. Results

This section presents the results of the proposed methodology on two major tasks: cell-cleavage stage detection and OCR evaluation that used for cleavage onset time prediction. For clarity, the findings are divided into two corresponding subsections.

A. Cell Stage Detection

This section presents a detailed evaluation of the YOLOv11 model in detecting embryo cell stages (two to eight cells), using a test set of 218 images sized (416x416 pixels). The YOLOv11 model achieved strong performance, with a precision of 0.9793, a recall of 0.9565, and an F1-score of 0.9678. The mAP@0.5 (mean Average Precision at an IoU threshold of 0.5) was 0.9767. This shows that our model successfully identified and localized most cells, achieving a strong alignment between the predicted and ground-truth bounding boxes. while mAP@0.5:0.95, which evaluates precision across various IoU thresholds from 0.5 to 0.95 in increments of 0.05, was measured at 0.74. was 0.7407. Fig. 6 shows the overall precision and recall, as well as the mean Average Precision (mAP) at thresholds 50 (mAP50) and 50-95 (mAP50-95), for the validation set throughout training epochs. Fig. 7 shows the confusion matrix for the model in detecting and classifying cell stages generally. For the majority of cell stages, the matrix shows high classification accuracy; this is especially true for the 2-cell and 4-cell stages, when most predictions match the actual labels. Compared to the earlier cleavage stages in the matrix, there were more incorrect classifications between the seven-cell stage and the nearby phases of five, six, and eight cells.

A set of key metrics was used to evaluate the model's performance: accuracy, precision, recall, F1-score, and specificity.

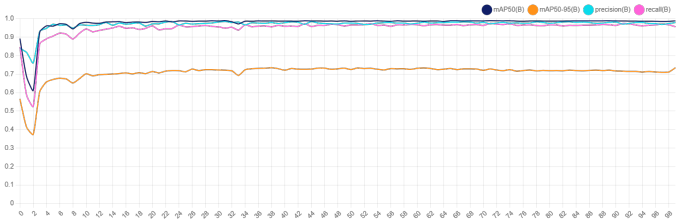

Fig. 6. Validation metrics over epochs.

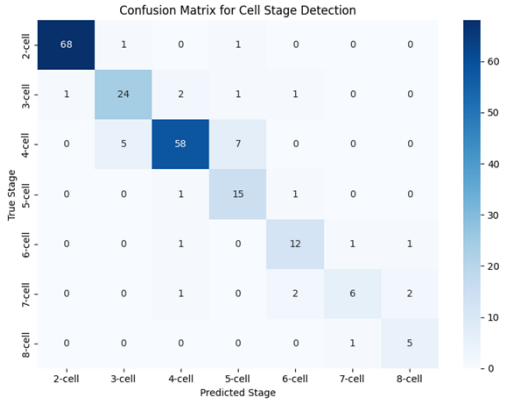

Fig. 7. Confusion matrix for automated cell stage detection. Rows: true stages (embryologist annotations); Columns: predicted stages (model output).

TABLE I. Evaluation Metrics for Every Cell Stage

Category	Precision	Recall	F1-Score	Specificity	Accuracy
2-cell	98.55%	97.14%	97.84%	99.32%	98.53%
3-cell	80%	82.76%	81.35%	96.83%	94.86%
4-cell	92.06%	82.86%	87.25%	96.62%	92.02%
5-cell	62.50%	88.24%	73.17%	95.52%	94.86%
6-cell	75%	80%	77.42%	98.03%	96.97%
7-cell	75%	54.55%	63.27%	99.04%	96.88%
8-cell	62.50%	83.33%	71.43%	98.60%	98.26%
Overall	**86.24%**	**86.24%**	**86.24%**	**97.71%**	**86.61%**

These metrics offer a thorough grasp of the model's capacity to accurately identify and categorize every cell stage based on the number of detected cells in the frame. Precision measures the percentage of all predicted cell stages correctly identified as instances of the target cell stage. Recall (also known as sensitivity) measures the ratio of true positives (correctly detected instances of the target stage) to the sum of true positives and false negatives (i.e., all frames that truly belong to that stage). The precision and recall weighted average is known as the F1-score. Specificity is calculated as the ratio of true negatives (i.e., frames that are correctly identified as not belonging to the target stage) to the total number of frames that truly do not represent that stage. It demonstrated the ability of the model to prevent misclassifications for non-target stages. The accuracy reflects the overall effectiveness of the model by measuring the proportion of correctly classified instances (both positives and negatives) out of all predictions made. Table I summarizes these metrics for each cell stage using input images of size (416x416 pixels).

Although the table presents the raw quantitative results,

the subsequent list presents a brief overview of the model's performance at each stage, emphasizing both its advantages and limitations:

- Two-cell stage: The model demonstrated strong efficacy in this category by classifying 2-cell stages with high precision 98.55%, recall 97.14% , and F1-score 97.84%. These results indicate robust and consistent detection of two-cell instances with a low misclassification rate.
- Three-cell and four-cell stages: With F1-scores of 81.35% and 87.25% , respectively, both stages demonstrated high performance, although with somewhat reduced precision and recall when compared to the 2-cell stage.
- Five-cell and Six-cell stages: The model's performance showed moderate accuracy. For the 5-cell stage, the F1-score of 73.17% reflects low precision 62.50% and high recall 88.24%, indicating frequent false positives despite successfully identifying most true cases. In contrast, the 6-cell stage achieved a more balanced outcome (F1-score 77.42%; precision 75%; recall 80%), suggesting that most instances were correctly classified with a moderate level of misclassification.
- seven-cell and eight-cell stages: The model struggled most with the 7-cell stage, hitting an F1-score of just 63% and recall at 55%. This happened mostly because there weren't many examples in the dataset, and the cells tended to overlap a lot, which led to confusion with the 6- or 8-cell stages. On the other hand, the 8-cell stage did better, landing an F1-score of 71% The recall was strong 83% so the model caught most of the real 8-cell stages. But the precision dropped to 63%, which means it also flagged quite a few that weren't actually 8-cell, so false positives were still a problem.
- Overall Performance: With an overall F1-score of 86.24% and specificity of 0.9771, demonstrating a strong ability to distinguish between target and non-target classes and consistent classification performance across all cell stages.

B. OCR Evaluation

Three OCR libraries were assessed to determine their performance in extracting HPI timestamps: PyTesseract, EasyOCR, and PaddleOCR. The objective text was a time stamp in the format xx.x h, placed in a small white rectangle with black numbers in the bottom right corner of each frame. Due to the large variation in contrast between the images in the dataset, no pre-processing was applied, rendering the generic enhancement techniques unreliable. Instead, each frame was cropped to the region of interest to isolate the timestamp and minimize noise. The performance of PyTesseract, EasyOCR, and PaddleOCR was evaluated using Manual Recognition Accuracy. To establish a ground truth, To perform this, a dataset of frames was randomly selected, and the timestamps visible in the images were manually annotated as ground truth. The output of each OCR engine was manually compared against the visual timestamp visible in the Time-Lapse Monitoring (TLM) images. An "Exact Match" criterion was utilized, where a detection was marked as correct only if it perfectly mirrored the visual data. The following list summarizes the OCR libraries' performance, highlighting the progression from less to more effective approaches depending on digit recognition under varying image conditions in terms of accuracy, dependability, and robustness.

- PyTesseract: Initially, PyTesseract was tested with various configurations but showed unsatisfactory results. The library had difficulties in accurately reading digits, either misreading them or failing to extract them at all. For example, "13.7" was incorrectly read as "3.7".
- EasyOCR: Next, EasyOCR was subsequently evaluated and demonstrated improved performance compared to PyTesseract.However, the library continued to have challenges in accurately recognizing some numbers. In particular, the library failed to read the tens place in some cases, such as reading "31.5" as "1.5", or completely ignoring the tens place and reading "1.5" as "5".
- PaddleOCR: PaddleOCR was evaluated last and demonstrated the highest accuracy. All values extracted by Paddle were accurately matched to the annotations provided by embryologists. Due to this consistent and reliable performance, Paddle was adopted as the preferred tool in our methodology for automating the hpi annotation process.

Table II summarizes the performance of the three OCR libraries in extracting hpi values, compared to the expert-annotated ground truth.

TABLE II. COMPARISON OF EXPERT-ANNOTATED HPI VALUES WITH PREDICTIONS FROM THREE OCR TOOLS

Expert (hpi)	PaddleOCR	EasyOCR	PyTesseract	Error Description
1.5	1.5	5	–	Misread digit (EasyOCR) ,Missing value (PyTesseract)
35.1	35.1	35.4	–	Misread digit (EasyOCR) ,Missing value (PyTesseract)
46.5	46.5	646.5	–	Extra digits (EasyOCR) ,Missing value (PyTesseract)

V. DISCUSSION

This study aimed to automate the process of tracking embryonic development by identifying cleavage stages and accurately determining the onset times of cell division. Deep learning and image processing techniques were used to analyze TLM videos. This was achieved by constructing an efficient methodology that annotates hpi of each cleavage stage while accurately locating cells within the corresponding stage images. This study showcases the power of deep learning and computer vision in automating the identification of crucial cleavage stages in embryo development. By providing precise, frame-by-frame predictions and organized data on cleavage timing, the proposed approach gives embryologists valuable developmental insights that exceed those gained from traditional manual observation. This level of accuracy in stage detection allows for more objective and consistent evaluations of embryos. Our methodology demonstrated the ability to accurately identify the onset of cell division stages up to the seven-cell stage, with an average time delay of 1–2 hpi. However, at the eight-cell stage, a significant time delay was observed compared to the embryologists' annotations. This disparity is explained by the more significant overlap between cell borders, which intensifies with increasing cell count. This

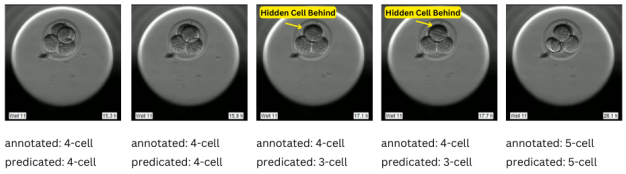

Fig. 8. Time-series of frames illustrating cell occlusion: A 4-cell stage can be seen in the first frame, but as the series goes on, one cell trails behind another, giving the impression that the stage contains only three cells. This illustrates how difficult it is to find overlapping cells.

overlap between cells was also observed in [14,17], and the inferiority of their methodology caused a delay of 2-3 hours in detecting the starting time of cell cleavage. The occurrence of overlapping cells and visual obstructions in certain frames, particularly during the later stages (after the six-cell stage), led to a decrease in detection accuracy. This was especially evident in the counting and differentiation of closely packed cells. This illustrates a challenge with the dataset, which contains overlaps that occur naturally in the later stages of cleavage. Adding a wider variety of overlapping cell configurations to the dataset could help overcome this restriction. Future enhancements, such as the integration of temporal context or refined segmentation techniques, could significantly bolster performance in these challenging scenarios. Our proposed methodology also faces challenges in detecting hidden cells: cells with dark shadows or those that are partially hidden behind other cells are not recognized. This problem is particularly important in the transition state between one stage and another, as shown in Fig. 8. More occluded cell samples should be included in the dataset in future studies so that the model can better learn and generalize these patterns.

The proposed model operated directly on raw TLM frames to ensure that the training data closely matched real-world scenarios. This design choice aims to enhance the robustness and generalizability of the method by avoiding transformations that may alter biologically relevant visual cues. Given the heterogeneous illumination conditions present in the TLM dataset, applying global image normalization or contrast enhancement could introduce inconsistencies across samples and potentially distort meaningful morphological information. As part of our research, the time differences between subsequent cleavage stages were calculated. This feature helps to improve the assessment of embryo quality by giving embryologists comprehensive data on developmental dynamics. Two complementary techniques were combined to address the dataset's class imbalance issue. Selective data augmentation was applied to increase the number of samples in minority classes. Then, random undersampling is used to decrease the number of samples in the majority classes. This combined strategy yielded satisfactory evaluation metrics, demonstrating its effectiveness in addressing class disparity. In order to put the performance of the proposed YOLOv1-based cell stage detection model into perspective, the obtained results are compared with those of earlier studies that used YOLOv5 for the detection of human embryo cell stage. Applying data augmentation techniques [17], the model performed well on early cleavage stages (e.g., F1-score of 95% for 1-cell and 88% for 2-cell), but performance significantly declined on later stages, with F1-scores falling to 0.32 for 3-cell, 0.36 for 5-cell, and 17% for 6-cell. Our method, on the other hand, produced consistently high overall F1-scores of 0.8624 and accuracy of 86.61% when applied to mouse embryo time-lapse videos. The F1-scores were relatively stable across stages, ranging from a minimum of 63.27% (7-cell) to a maximum of 97.84% (2-cell). Higher specificity (97.71% overall) and a better balance between sensitivity and precision at every stage further supported this performance. The superiority of the proposed framework is quantitatively validated through a comparative analysis with the recent study by Sharma et al. [17], as detailed in Table III. Furthermore, our approach corrects for class imbalance by undersampling and targeted data augmentation, which probably helped to improve minority-stage class detection. The framework demonstrates high adaptability to diverse clinical environments through a calibratable OCR pipeline. By utilizing coordinate-based ROI localization. The system can be seamlessly adjusted to different incubator display formats. Direct quantitative comparison with prior time-lapse embryo analysis studies is constrained by the absence of publicly available, standardized datasets and by differences in embryo species, imaging systems, and annotation protocols. This limitation is explicitly acknowledged in the present study. Accordingly, performance claims are restricted to improvements observed within the same experimental setting and under a consistent evaluation protocol. Where comparisons with related work are discussed, they are intended to illustrate relative gains in robustness and detection capability—particularly at later cleavage stages—rather than to assert absolute superiority across heterogeneous datasets. The comparative analysis highlights the generalizability and robustness of our deep learning pipeline, especially in managing class imbalance and handling transitional stages with greater stability, despite the differences in the dataset.

TABLE III. PERFORMANCE BENCHMARKING AND COMPARATIVE ANALYSIS: PROPOSED YOLOV11 FRAMEWORK VS. STATE-OF-THE-ART BASELINE OF SHARMA ET AL.[17] (YOLOV5)

Cell Stage	Sharma et al.[17] - yolov5 Accuracy (%)	Proposed Model Accuracy (%)
2-cell	85%	98%
3-cell	57%	94%
4-cell	55%	92%
5-cell	51%	94%
6-cell	39%	96%
7-cell	31%	96%
8-cell	47%	98%

VI. CONCLUSION

Given that the timing of cell cleavage is a critical parameter of embryo viability and developmental potential, it certainly plays a major role in embryo assessment. Detecting and extracting the onset of these cleavage events automatically can help embryologists Assess the quality of embryos and select candidates with better chances of implantation. In this work, YOLOv11 and OCR-based timing extraction are used to automatically annotate mouse embryos' cleavage stage and onset time. The suggested approach, despite the challenging class imbalance present in the dataset, achieves high performance in cleavage stage detection and cleavage onset timing extraction. Through the combination of selective data augmentation and random undersampling. The problem of class imbalance was

successfully addressed, which enhanced the model's robustness and evaluation metrics. Our methodology demonstrated the ability to accurately identify the onset of cell division stages up to the seven-cell stage, with an average time delay of 1–2 hpi, and showed excellent performance in differentiating between various developmental phases in mouse embryo TLM videos, achieving Accuracy of 86.61% , F1-score of 86.24% ,and precision of 86.24% in identifying and classifying cleavage stages. One of the key difficulties encountered in this work is the significant cell overlap, which affected the model's performance, particularly at the later stages of cleavage. As it influenced the accuracy of detection and caused minimal delays in determining the onset of division. In addition, the approach was having issues identifying cells that were occluded partially or overshadowed, particularly with fast changes in developmental phases. There are several promising avenues for future research in this field to overcome the present limitations. Accumulating larger and well-annotated TLM datasets is a crucial step to increase model robustness and promote more balanced learning across all cleavage stages. Future work could utilize 3D imaging techniques or incorporate temporal context using video-based models that analyze frame sequences rather than individual frames to address issues such as partially hidden and overlapping cells.

Data Availability

While the dataset is not publicly available due to clinical privacy and institutional ethical constraints, it can be shared upon reasonable request to support reproducibility, subject to applicable data governance protocols.

References

[1] World Health Organization. *Infertility Prevalence Estimates, 1990–2021*. Geneva: World Health Organization, 2023. Licence: CC BY-NC-SA 3.0 IGO

[2] Iqbal, I., Mustafa, G., and Ma, J. 2020. "Deep Learning-Based Morphological Classification of Human Sperm Heads." *Diagnostics* 10 (5): 325–332.

[3] Almaslami, F., Aljunid, S. M., and Ghailan, K. "Demographic Determinants and Outcome of In Vitro Fertilization (IVF) Services in Saudi Arabia." *Journal of International Medical Research* 46, no. 4 (April 2018): 1537-1544. https://doi.org/10.1177/0300060517749329.

[4] Lockhart, L. 2021. "Automating Assessment of Human Embryo Images and Time-Lapse Sequences for IVF Treatment." Master's thesis, Simon Fraser University.

[5] Meseguer, Maria, et al. "The Use of Morphokinetics as a Predictor of Embryo Implantation." *Human Reproduction* 26, no. 10 (2011): 2658–2671.

[6] Wong, Connie C., Kevin E. Loewke, Nancy L. Bossert, Barry Behr, Christopher J. De Jonge, Thomas M. Baer, and Renee A. Reijo Pera. 2010. "Non-invasive Imaging of Human Embryos Before Embryonic Genome Activation Predicts Development to the Blastocyst Stage." *Nature Biotechnology* 28 (10): 1115–1121. Nature Publishing Group US New York.

[7] Rai, R., and L. Regan. "Recurrent Miscarriage." *The Lancet* 368, no. 9535 (2006): 601–611.

[8] Wong, C., AA Chen, B. Behr, and S. Shen. 2013. "Time-lapse Microscopy and Image Analysis in Basic and Clinical Embryo Development Research." *Reproductive BioMedicine Online* 26 (2): 120–129. Elsevier.

[9] Hlinka, D., B. Kal'atová, I. Uhrinová, S. Dolinská, J. Rutarová, J. Řezáčová, S. Lazarovská, and M. Dudáš. 2012. "Time-Lapse Cleavage Rating Predicts Human Embryo Viability." *Physiological Research* 61 (5): 581–589.

[10] Cetinkaya, Murat, Caroline Pirkevi, Hakan Yelke, Yesim Kumtepe Colakoglu, Zafer Atayurt, and Semra Kahraman. 2015. "Relative Kinetic Expressions Defining Cleavage Synchronicity Are Better Predictors of Blastocyst Formation and Quality Than Absolute Time Points." *Journal of Assisted Reproduction and Genetics* 32: 27–35. Springer.

[11] Senechkin, I. V., L. V. Hilkevich, and M. A. Kurcer, "Cleavage of human embryos: Options and diversity," Acta Naturae, vol. 8, no. 3, pp. 88–96, 2016.

[12] Raudonis, Vidas, Agne Paulauskaite-Taraseviciene, Kristina Sutiene, and Domas Jonaitis. 2019. "Towards the Automation of Early-Stage Human Embryo Development Detection." *Biomedical Engineering Online* 18: 1–20. Springer.

[13] Nakaguchi, Victor Massaki, and Tofael Ahamed. 2022. "Development of an Early Embryo Detection Methodology for Quail Eggs Using a Thermal Micro Camera and the YOLO Deep Learning Algorithm." *Sensors* 22 (15): 5820. MDPI.

[14] Sharma, Akriti, Mette H. Stensen, Erwan Delbarre, Momin Siddiqui, Trine B. Haugen, Michael A. Riegler, and Hugo L. Hammer. 2022. "Detecting Human Embryo Cleavage Stages Using YOLO V5 Object Detection Algorithm." In *Symposium of the Norwegian AI Society*, 81–93. Springer International Publishing Cham.

[15] Jia, Nan, Bin Li, Yuliang Zhao, Shijie Fan, Jun Zhu, Haifeng Wang, and Wenwen Zhao. 2023. "Exploratory Study of Sex Identification for Chicken Embryos Based on Blood Vessel Images and Deep Learning." *Agriculture* 13 (8): 1480. MDPI.

[16] Yang, Ruixin, and Yingyan Yu. 2021. "Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis." *Frontiers in Oncology* 11: 638182. Frontiers Media SA.

[17] Sharma, Akriti, Ayaz Z. Ansari, Radhika Kakulavarapu, Mette H. Stensen, Michael A. Riegler, and Hugo L. Hammer. 2023. "Predicting Cell Cleavage Timings from Time-Lapse Videos of Human Embryos." *Big Data and Cognitive Computing* 7 (2): 91. https://doi.org/10.3390/bdcc7020091.

[18] Pinto-Coelho, L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering 2023, 10, 1435. https://doi.org/10.3390/bioengineering10121435

[19] Ghaffar Nia, N.; Kaplanoglu, E.; Nasab, A. Evaluation of Artificial Intelligence Techniques in Disease Diagnosis and Prediction. Discov. Artif. Intell. 2023, 3, 5.

[20] Parmar, U.P.S.; Surico, P.L.; Singh, R.B.; Romano, F.; Salati, C.; Spadea, L.; Musa, M.; Gagliano, C.; Mori, T.; Zeppieri, M. Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases. Medicina 2024, 60, 527. https://doi.org/10.3390/medicina60040527

[21] Tanveer, Hassan, Muhammad Faheem, Arbaz Haider Khan, and Muhammad Ali Adam. "AI-Powered Diagnosis: A Machine Learning Approach to Early Detection of Breast Cancer." INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH 13, no. 2 (2025): 153-166.

[22] Bongurala, Archana Reddy, Dhaval Save, Ankit Virmani, and Rahul Kashyap. "Transforming health care with artificial intelligence: redefining medical documentation." Mayo Clinic Proceedings: Digital Health 2, no. 3 (2024): 342-347.

[23] Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F. Artificial Intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework and Future Research Agenda. J. Ambient. Intell. Humaniz. Comput. 2023, 14, 8459–8486.

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

[25] Kragh, M.F.; Karstoft, H. Embryo selection with artificial intelligence: How to evaluate and compare methods? J. Assist. Reprod. Genet. 2021, 38, 1675–1689.

[26] Chen, F.; Chen, Y.; Mai, Q. Multi-Omics Analysis and Machine Learning Prediction Model for Pregnancy Outcomes After Intracytoplasmic Sperm Injection–in vitro Fertilization. Front. Public Health 2022, 10, 924539.

[27] Zhao, M., Xu, M., Li, H., Alqawasmeh, O., Chung, J. P. W., Li, T. C., Lee, T. L., Tang, P. M., & Chan, D. Y. L. (2021). Application of convolutional neural network on early human embryo segmentation during in vitro fertilization. Journal of cellular and molecular medicine, 25(5), 2633–2644. https://doi.org/10.1111/jcmm.16288

[28] Rajendran, Suraj, Matthew Brendel, Josue Barnes, Qiansheng Zhan, Jonas E. Malmsten, Pantelis Zisimopoulos, Alexandros Sigaras et al. "Automatic ploidy prediction and quality assessment of human blastocysts using time-lapse imaging." Nature Communications 15, no. 1 (2024): 7756.

[29] Moussavi, Farshid, Yu Wang, Peter Lorenzen, Jonathan Oakley, D. Russakoff, and Stephen Gould. "A Unified Graphical Models Framework for Automated Human Embryo Tracking in Time-Lapse Microscopy." In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 314–320. IEEE, 2014.

[30] Khan, Aisha, Stephen Gould, and Mathieu Salzmann. "Automated Monitoring of Human Embryonic Cells up to the 5-Cell Stage in Time-Lapse Microscopy Images." In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 389–393. IEEE, 2015.

[31] Khan, Aisha, Stephen Gould, and Mathieu Salzmann. "Deep Convolutional Neural Networks for Human Embryonic Cell Counting." In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part I 14*, 339–348. Springer, 2016.

[32] Malmsten, Jonas, Nikica Zaninovic, Qiansheng Zhan, Zev Rosenwaks, and Juan Shan. "Automated Cell Division Classification in Early Mouse and Human Embryos Using Convolutional Neural Networks." *Neural Computing and Applications* 33 (2021): 2217–2228. Springer.

[33] Ragab, Mohammed Gamal, Said Jadid Abdulkadir, Amgad Muneer, Alawi Alqushaibi, Ebrahim Hamid Sumiea, Rizwan Qureshi, Safwan Mahmood Al-Selwi, and Hitham Alhussian. "A comprehensive systematic review of YOLO for medical object detection (2018 to 2023)." IEEE Access 12 (2024): 57815-57836.

[34] Quinn, Patrick, and Frederick C. Horstman. "Is the Mouse a Good Model for the Human with Respect to the Development of the Preimplantation Embryo In Vitro?" *Human Reproduction* 13, no. suppl_4 (December 1998): 173–183. https://doi.org/10.1093/humrep/13.suppl_4.173.

[35] Sabhnani, Tanya V., Aisha Elaimi, Hanan Sultan, Adel Alduraihem, Paul Serhal, and Joyce C. Harper. "Increased Incidence of Mosaicism Detected by FISH in Murine Blastocyst Cultured In Vitro." *Reproductive BioMedicine Online* 22, no. 6 (2011): 621–631.

[36] Lebumfacil, Aaron Joaquin, and Patricia Angela Abu. "Traffic Sign Detection and Recognition Using YOLOv5 and Its Versions." In *2022 IEEE 1st International Conference on Cognitive Mobility (CogMob)*, 11–18. IEEE, 2022.

[37] Aydin, Burchan, and Subroto Singha. "Drone Detection Using YOLOv5." *Eng* 4, no. 1 (2023): 416–433. MDPI.

[38] Brownlee, Jason. 2019. "A Gentle Introduction to Object Recognition with Deep Learning." *Machine Learning Mastery*. May 2019. https://machinelearningmastery.com/object-recognition-with-deep-learning/.

[39] Soviany, Petru, and Radu Tudor Ionescu. 2018. "Optimizing the Trade-Off Between Single-Stage and Two-Stage Deep Object Detectors Using Image Difficulty Prediction." In *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 209–214. https://doi.org/10.1109/SYNASC.2018.00041.

[40] Ultralytics. *Ultralytics HUB*. Accessed April 8, 2025. https://hub.ultralytics.com/.

[41] Islam, Noman, Zeeshan Islam, and Nazia Noor. 2017. "A Survey on Optical Character Recognition System." *arXiv preprint* arXiv:1710.05703.

[42] Zaryab, M. A., and C. R. Ng. "Optical Character Recognition for Medical Records Digitization with Deep Learning." In *2023 IEEE International Conference on Image Processing (ICIP)*, 3260–3263. IEEE, October 2023.

[43] Salvi, A., L. Arnal, K. Ly, G. Ferreira, S. Y. Wang, C. Langlotz, V. Mahajan, and C. A. Ludwig. "Ocular Biometry OCR: A Machine Learning Algorithm Leveraging Optical Character Recognition to Extract Intraocular Lens Biometry Measurements." *Frontiers in Artificial Intelligence* 7 (2025): 1428716. https://doi.org/10.3389/frai.2024.1428716.

[44] Kim, D., and H. Yu. "Figure Text Extraction in Biomedical Literature." *PLOS ONE* 6, no. 1 (2011): e15338.

[45] madmaze. "GitHub - madmaze/pytesseract: A Python Wrapper for Google Tesseract." *GitHub*. Accessed March 2025. https://github.com/madmaze/pytesseract.

[46] JaidedAI. "GitHub - JaidedAI/EasyOCR: Ready-to-Use OCR with 80+ Supported Languages and All Popular Writing Scripts Including Latin, Chinese, Arabic, Devanagari, Cyrillic, and More." *GitHub*. Accessed March 2025. https://github.com/JaidedAI/EasyOCR.

[47] Qin, Hao, Jiajie Jin, Wei Sun, Yucheng Qian, Haigang Wang, and Zhi Wang. "An Analysis Method for Recognizing and Analyzing Multi-Type Documents in Electric Power Operation Site Based on PaddleOCR." In *2024 8th International Conference on Smart Grid and Smart Cities (ICSGSC)*, 363-367. 2024. https://doi.org/10.1109/ICSGSC62639.2024.10813769.

[48] Feng, Lei, Zongwu Ke, and Na Wu. "ModelsKG: A Design and Research on Knowledge Graph of Multimodal Curriculum Based on PaddleOCR and DeepKE." In *2022 14th International Conference on Advanced Computational Intelligence (ICACI)*, 186-192. 2022. https://doi.org/10.1109/ICACI55529.2022.9837529.