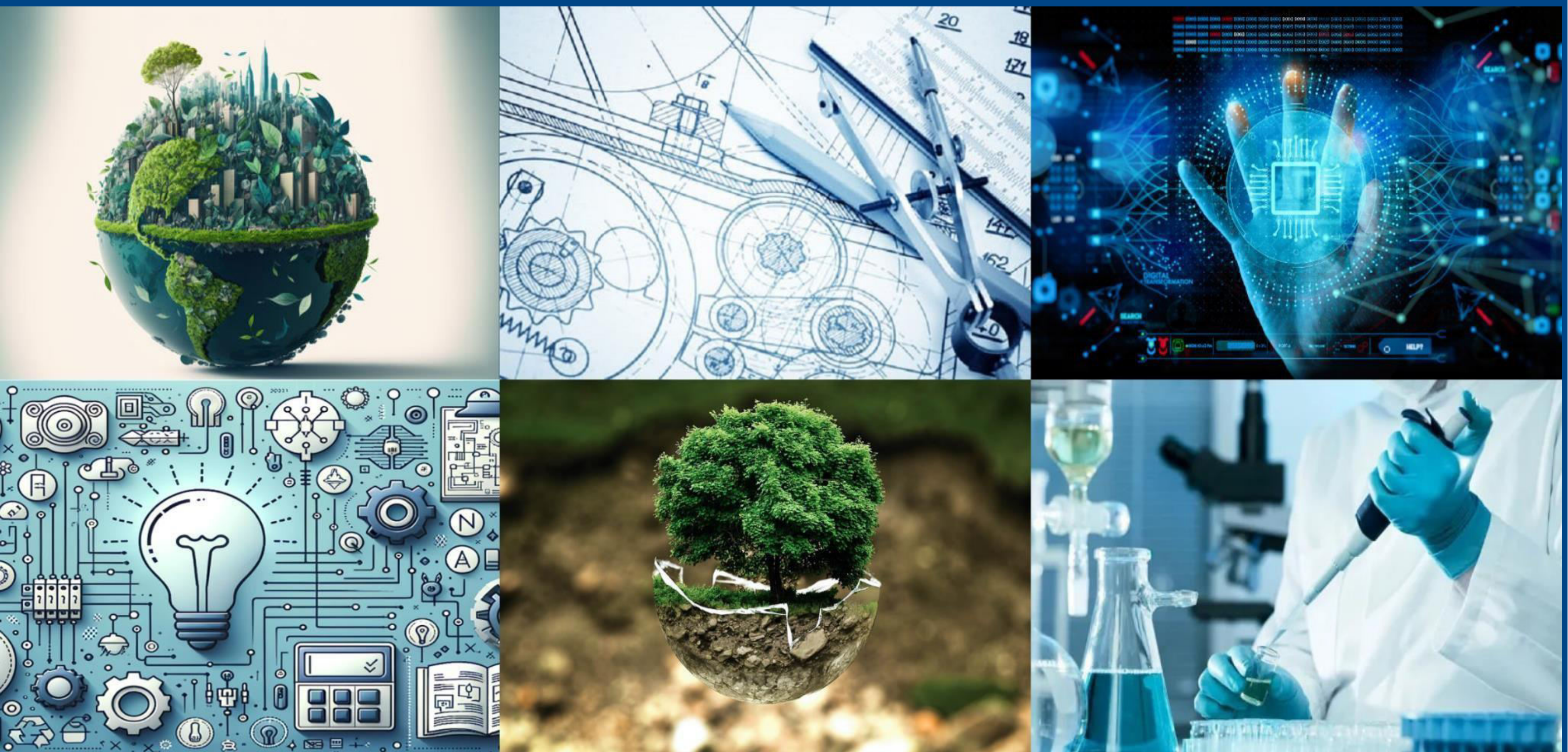
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 7, July 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

An Image Processing Method for Machine Learning-Based IVF Blastocysts Grading

Harshitha T A, Mr. Suthan R

M. Tech Student, Dept. of CSE, SIET, Tumkur, India

Assistant Professor, Dept. of CSE, SIET, Tumkur, India

ABSTRACT: In vitro fertilization (IVF) is a treatment for infertility. As more couples choose IVF, infertility has become a global health issue. Our model helps choose the fewest important features from the IVF dataset, which has 28 features. The project works on two machine learning techniques, feature selection where the number of features selected out of 27 is 6 using the lasso regression model, when it comes to feature extraction, the model used Principal component analysis (PCA), which had a 74% accuracy, and Linear discriminant analysis (LDA), which had a 76% accuracy which allows for in-depth analysis of the important features required for IVF treatment and dimensionality reduction for the IVF dataset. The project focused on working with some of the methods of feature selection, which include feature importance and a correlation matrix with a heatmap. When working with feature importance using four various algorithms, such as Random Forest, Decision Tree, Extra Tree classifier, and XG boost classifier, Random Forest obtained the best result of about 97.47% accuracy with 9 features selected, and Extra Tree classifier gave an accuracy of 97.45% with 9 features selected. The project even works on the collection of blastocyst images, which is achieved by using deep learning models. The blastocyst dataset is divided into good and poor images, which comprise 42 images each, and for these images, applying image generation using Generative Adversarial Networks (GAN) and image classification using CNN, which gave an accuracy of about 85% for 500 epochs. More research is required on a proper decision-making model that supports patient wellness.

KEYWORDS: In vitro fertilization (IVF), Data collection, Data pre-processing, Machine learning, Deep learning, Feature selection, Feature extraction, Generative adversarial networks (GAN), and Convolutional neural networks (CNN).

I. INTRODUCTION

In vitro fertilization is a fertility treatment. The World Health Organization defines infertility as a year of regular, unprotected intercourse that does not result in pregnancy [1]. IVF is among the assisted reproductive methods that can assist couples in having a child. Damaged or obstructed Fallopian tubes, endometriosis, and hormonal imbalance can all lead to infertility in women. Men may face infertility due to low sperm count or quality. When it comes to embryo selection (in terms of good and bad blastocysts) and embryo transfer, the human blastocyst plays an An important part of IVF treatment [2]. The main goal of this research is to create a model that selects the fewest important features from IVF patient record dataset by first collecting the patient records, pre-processing them, and then using machine learning techniques for feature selection and extraction. The initiative tries to avoid the "curse of dimensionality," which involves working with redundant features, which can be time-consuming and undermine the model's accuracy [3]. Using deep learning models such as GAN and CNN, the research also works on image generation and image classification for the blastocyst dataset. GAN is used to assess, capture, and copy the dataset's variations. CNN will recognize the relevant features without the need for human intervention [4].

II. PROPOSED METHODOLOGY

The purpose is to determine the most fundamental aspects of IVF treatment. The project uses GAN and CNN for generating and classifying blastocyst images. Data cleansing is essential before machine learning algorithms can deliver high-quality results. Incomplete data may have missing or wrong values, which can affect research. This could cause issues with research. As a result, the data that was initially obtained must go through data pre-processing and handling the missing instances with the help of mean and median computation. Additional tasks such as feature selection and extraction using PCA and LDA, as well as other machine learning algorithms, are subsequently performed on the dataset.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A. Data Collection

The Human Fertilization and Embryology Authority (HFEA) provided information for the IVF dataset to identify and collect the information needed to diagnose infertility in women. This set has 28 features, namely: Patient age at treatment, Total number of prior IVF cycles, Type of infertility, Cause of infertility, Fresh eggs collected, Embryos transferred, and so on. The blastocyst image dataset was collected from a published paper by Pegah Khosravi, Ehsan Kazemi, Qian Sheng Zhan, et al. 4., titled "Deep learning offers rigorous assessment and selection of human blastocysts following in vitro fertilization" (2019) [6].

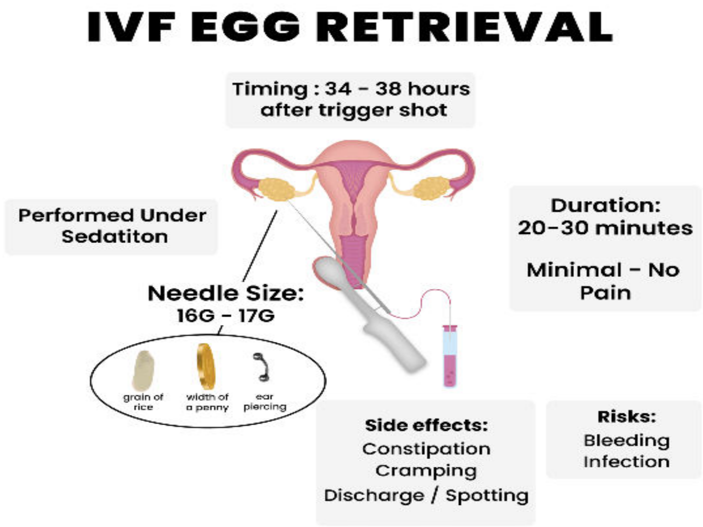


Fig. Represent the process of egg retrieval.

B. Data Pre-Processing and Feature Selection

Missing data in the dataset was handled using mean and median calculations. The Lasso regression model was used for feature selection. To reduce redundant features from the dataset, the Lasso regression model (Logistic Regression with L1-regularization) can be employed. This model, which uses L1-regularization, introduces sparsity by setting redundant feature coefficients to zero, reducing dataset complexity. It is a highly beneficial strategy for reducing the dataset's dimensionality by deleting extraneous features. The project also used various methods for feature selection and extraction of IVF records. When working with feature importance, the implementation was done by applying Random Forest, Decision Tree, Extra Tree classifier, and XGBoost classifier. The accuracy and the number of features selected were considerably good.

C. Feature Extraction

Feature extraction is a method used to turn raw data into numbers that can be used in calculations while keeping the original content of the data When applied directly to raw data, of AI and the Machine learning produces better results. Feature extraction is a technique for reducing a dataset's dimensionality by repurposing earlier data. To speed up training, lower the danger of overfitting, and improve data presentation, this work employs principal component analysis (PCA) and linear discriminant analysis (LDA). When working with PCA, the dimensionality of the dataset was reduced to a 2D PCA, and when applying LDA, the dimensionality of the dataset was reduced to 1feature out of the original number of 27 features

D. Image Generation and Classification

This study is focused on generating and classifying images of blastocysts. GANs, a powerful class of unsupervised learning networks, are used to create images. The model was trained for 500 epochs For a CNN model a dataset was created where it comprised of the actual dataset (42) and generated images (20) a total of 62 images were created for good and poor images each and outputted an accuracy of 85% for 500 epochs. The project tested the model from 50 to 500 epochs and recorded the results.

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807070

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. PROPOSED SYSTEM

The project starts with collecting data from the IVF dataset and blastocyst images. The dataset has some missing data and categorical values that need to be converted into numbers. Data pre-processing, which converts/transforms raw data into an understandable and efficient format, can help with this. In the training process, feature selection and its methods, such as feature importance and correlation matrix with heatmap, are now being used. These methods are implemented using supervised machine learning algorithms. Feature extraction uses both supervised and unsupervised machine learning algorithms.

The project even focuses on working with image generation, as the collected blastocyst images are very few in number. By creating our dataset using a GAN, the accuracy of the model can be considerably good. Image classification for blastocyst images is then performed, and good accuracy is achieved with various algorithms. Further, the number of features obtained and the image generated can be verified with an expert gynecologist, and these features can further predict whether there can be a successful IVF treatment or not. The images generated for blastocyst can be further consulted by an embryologist, as the selection of the embryo plays a vital role in IVF treatment.

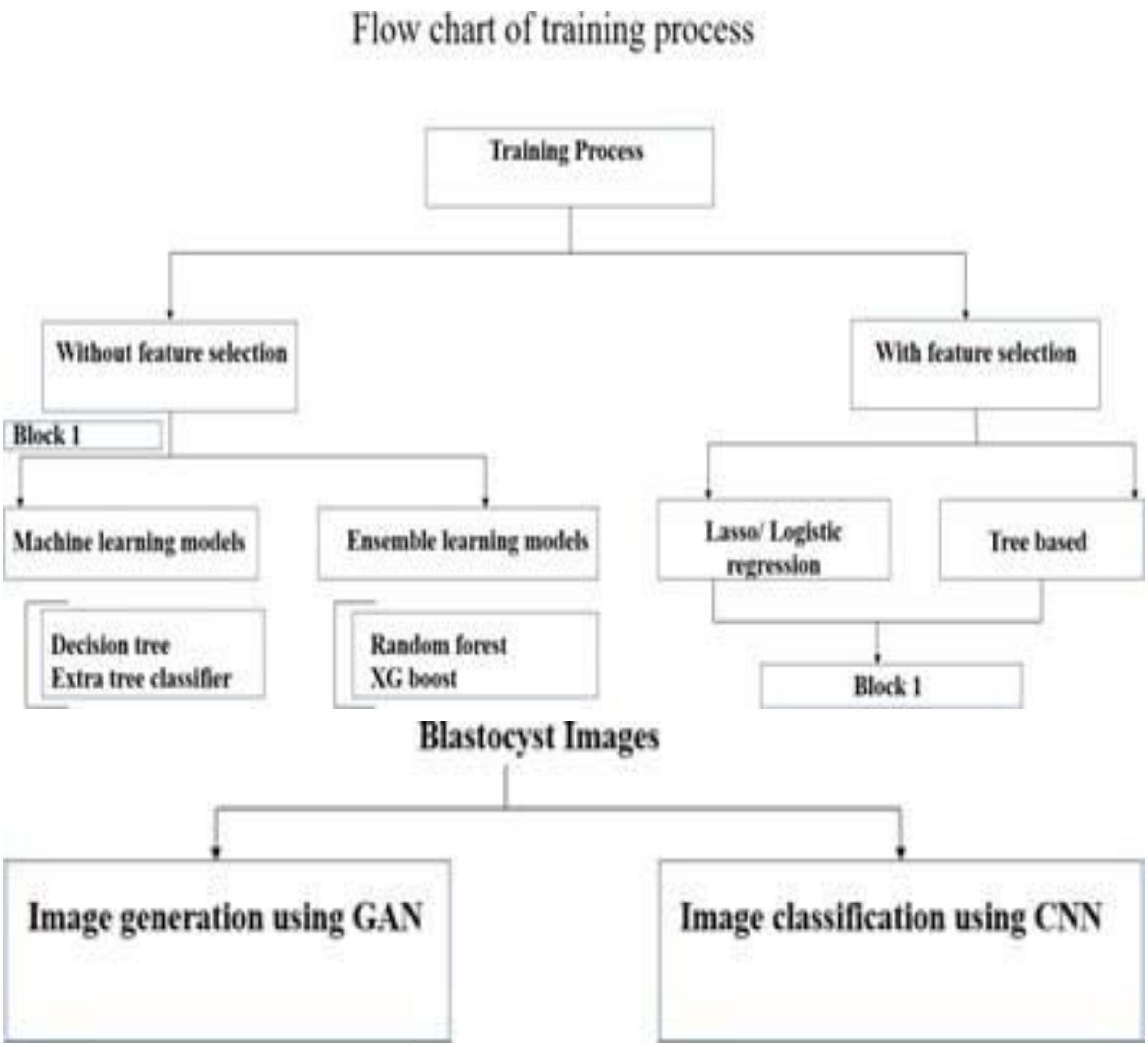


Fig 1. Represents the flow chart of the training process.

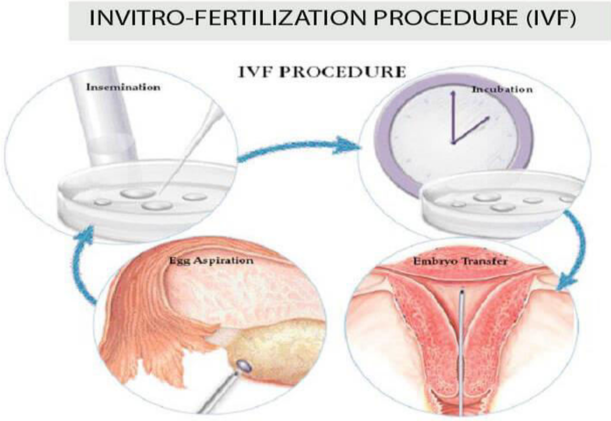


Fig. Represents the procedural model.

IV. RESULTS AND DISCUSSIONS

In this section, we discuss the results being obtained when working with different machine learning and deep learning models.

A. Results obtained for Feature selection
The results obtained when applying the Lasso regression model: Out of 27 features, 6 features have been selected, and the features selected are the total number of previous IVF cycles, cause of infertility ovulatory disorder, cause of infertility male factor, stimulation used, fresh egg collected, and embryos transferred.

B. Figures and Tables
When working with one of the methods of feature selection i.e., feature importance the best accuracy obtained when applying Random Forest was 97.48% which selected 9 features namely: Patient age at treatment, Total no. of previous IVF cycles, Type of infertility female primary and secondary, Type of infertility male primary and secondary, Fresh eggs collected, Embryos collected and Stage Day of embryo transfer. When applying the XG boost classifier, the accuracy obtained was 97.44% and the number of features selected was just 3, namely: Type of infertility, female primary, secondary, and male primary.

The accuracy of principal component analysis and linear discriminant analysis while employing feature extraction to avoid dimensionality reduction was 74% and 76%, respectively.

TABLE I. **FEATURES SELECTED AFTER APPLYING FEATURE IMPORTANCE**

SI no.	Feature importance		
	Algorithms used	Accuracy	No. of features selected
1.	Decision Tree	97.29%	7
2.	Random forest	97.47%	9
3.	XG boost Classifier	97.44%	3
4.	Extra tree classifier	97.45%	9

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807070

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

TABLE II. **FEATURES EXTRACTION FOR PCA AND LDA**

SI no.	Algorithms used	Accuracy
1.	PCA	74%
2.	LDA	76%

Fig 3. Accuracy of the CNN model.

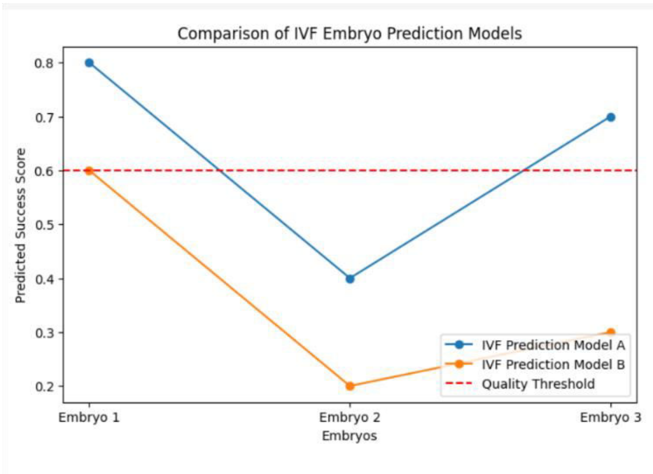


Fig 4. Loss and validation loss of the CNN model.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

With further research towards IVF, the selected features when applying feature selection can be verified with the expertise of gynecologists, and a survey can be conducted to determine whether these features can be considered important when undergoing IVF treatment, and proper predictions can be made for the patient. The generated images when applying GAN can be verified by an embryologist whether the images can be classified into a good or a poor image, which can further create This helps create a dataset when the original data is limited, which makes it possible to calculate accuracy. There can be the use of any other algorithms apart from the algorithms that are being used in the project.

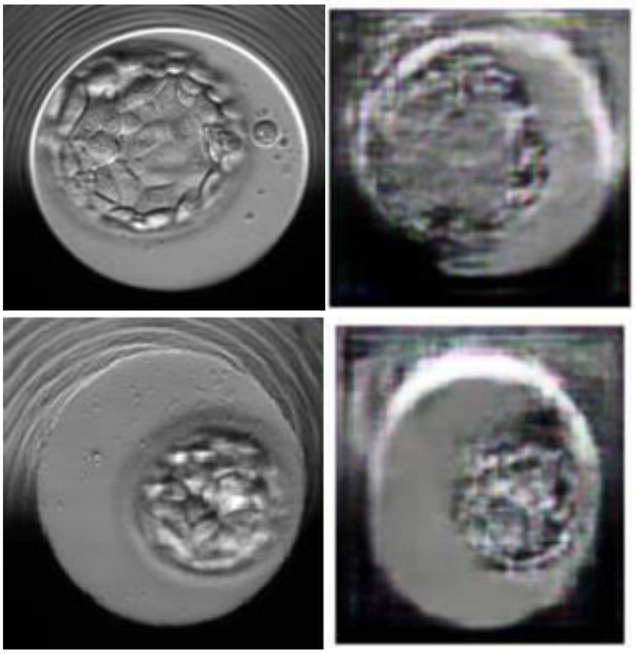


Fig. 5. Comparing original good-quality and poor-quality blastocyst images and the generated image after 500 epochs.

V. CONCLUSIONS AND FUTURE WORK

The project works on feature selection and some of its methods, such as feature importance and correlation matrix with heatmap for the IVF dataset, in order to avoid working with irrelevant features, which can sometimes be consuming and reduce the accuracy of the model. The feature extraction for the IVF dataset is used to avoid the "curse of dimensionality". The project also utilizes image generation via a GAN and image classification with a CNN model for the blastocyst images. The results obtained when applying machine learning and deep learning models are considerably good. When working with feature importance, Random Forest achieved a good accuracy of approximately 97.47%, while feature extraction using LDA yielded an accuracy of roughly 76%. When working with CNN, the accuracy obtained is 85% for 500 epochs.

REFERENCES

[1]. Bo Zhang, Yuqi Cui, Meng Wang, Jingjing Li, Lei Jin, and Dongrui Wu, Senior Member, “In Vitro Fertilization (IVF) Cumulative Pregnancy Rate Prediction from Basic Patient Characteristics”, IEEE 2019.

[2]. Tadepalli, S., Lakshmi, P.V. (2021). Deep Learning in IVF to Predict Embryo Infertility from Blastocyst Images. In: Kumar, A., Mozar, S. (eds) ICCCE 2020.

[3]. S. N. Patil, U. V. Wali, and M. K. Swamy, "Selection of Single Potential Embryo to Improve the Success Rate of Implantation in IVF Procedure using Machine Learning Techniques," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019.

[4]. M. Abbasi, P. Saeedi, J. Au, and J. Havelock, "A Deep Learning Approach for Prediction of IVF Implantation Outcome from Day 3 and Day 5 Time-Lapse Human Embryo Image Sequences," 2021 IEEE International Conference on Image Processing (ICIP), 2021.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[5]. Wiweko, "Human Blastocyst Classification after In Vitro Fertilization Using Deep Learning," 2020 7th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2020.
[6]. Khosravi, P., Kazemi, E., Zhan, Q. et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. npj Digit. Med. 2, 21 (2019).
[7]. Irmawati, Basari, and D. Gunawan, "Automated Detection of Human Blastocyst Quality Using Convolutional Neural Network and Edge Detector," 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS), 2019.
[8]. DeCherney AH, Barnett RL. In Vitro Fertilization Research is Translational Research. Reprod Sci. 2016.
[9]. Tadepalli, Satya Kiranmai, and P.V. Lakshmi. "Application of Machine Learning and Artificial Intelligence Techniques for IVF Analysis and Prediction." IJBDAH vol.4, no.2 2019.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com