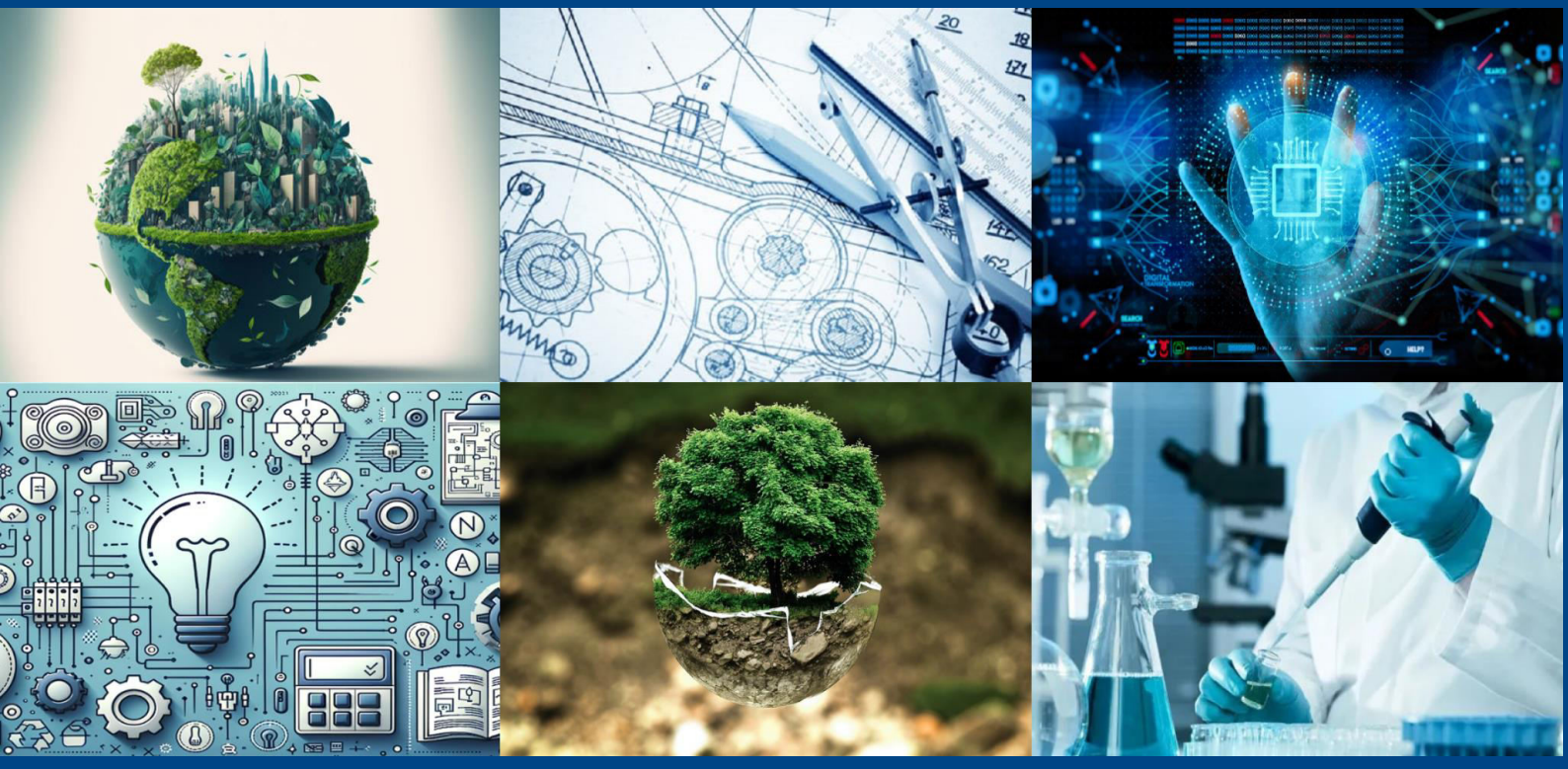
ISSN: 2582-7219

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 5, May 2025**

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805043

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)

Image Processing Approach for Grading IVF Blastocysts using Machine Learning

Rashmi N

Assistant Professor, Dept. of CSE, Shridevi Institute of Engineering and Technology, Tumkur, Karnataka, India

Manu S, Nagapooja M, Bhavana M R, Bindhushree R

Dept. of CSE, Shridevi Institute of Engineering and Technology, Tumkur, Karnataka, India

ABSTRACT: In vitro fertilization (IVF) is an advanced treatment for infertility problems. As the number of couples seeking in vitro fertilization (IVF) grows, infertility has become a global health issue. For the purpose of evaluating patient records, our model provides an economic approach for selecting minimal number of features for the IVF dataset, The IVF dataset comprise of 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 worked with some of the methods of feature selection which includes feature importance and a correlation matrix with 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 of 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 proper decision-making model that support 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

A form of fertility treatment is in vitro fertilization. The World Health Organization defines infertility as a year of regular, unprotected intercourse that does not result in pregnancy [1]. IVF is among assisted reproductive methods that can assistcouples in having a child. Damaged or obstructed Fallopian tubes, endometriosis, and hormonal imbalance can all lead to infertility in women. A low sperm count or sperm content can cause infertility in men. When it comes to embryo selection (in terms of good and bad blastocyst) and embryo transfer, the human blastocyst plays an important part in IVF treatment [2]. The main goal of the research is to create a model that selects the minimal number of features from the IVF patient record dataset by first collecting the patient records, pre-processing 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. For image generation and classification of blastocyst photos, the research uses the deep learning techniques GAN and CNN. Data cleansing is essential before machine learning algorithms can deliver high-quality results. When data is incomplete, it means that some values are

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)

missing or erroneous, and the database contains faulty data. 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 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.

A. Data Collection

The Human Fertilization and Embryology Authority (HFEA) provided information for the IVF dataset in order to identify and collect theinformation 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 dataset for blastocyst images were collected from the named authors published paper: Pegah Khosravi, Ehsan Kazemi, Qian sheng Zhan, et al., "Deep learning offers rigorous assessment and selection of human blastocysts following in vitro fertilization" (2019) [6].

B. Data pre-processing and Feature Selection

To deal with some of the missing instances in the dataset, data pre-processing with mean median computations is performed. The Lasso regression model was employed in the feature selection procedure. To reduce redundant features from the dataset, the Lasso regression model (Logistic Regression with L1-regularization) can be employed. L1- regularization introduces sparsity to the dataset and reduces the values of redundant feature coefficients to 0. It is a highly beneficial strategy for reducing the dataset's dimensionality by deleting extraneous features. It's also good at picking the most important columns (features) in the training dataset for predicting the target variable. The project even works with the different methods of feature selection and feature extraction for the IVF records. When working with feature importance the implementation was done by applying Random Forest, Decision tree, Extra tree classifier and XG boost classifier. The accuracy and the number of features selected were considerably good.

C. Feature Extraction

Feature extraction is a technique for transforming raw data into numerical features that may be processed while retaining the original data set's content. When applied directly to raw data, 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 into a 2D PCA and when applying LDA the dimensionality of the dataset was reduced into 1feature out of the original number of 27 features.

D. Image Generation and Classification

The study focuses on image generation and classification for the blastocyst photographs. GANs, a powerful class of unsupervised learning networks, are used to create images. GANs are two neural network models that can evaluate, capture, and repeat changes in a dataset. In data science, CNN are widely used in applications like computer vision and picture categorization [4]. This deep learning concept was implemented for the blastocyst images, where the GAN method was used to generate images for good and poor blastocyst and worked on 500 epochs For a CNN model a dataset was created where it comprised of the actaul 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 worked from 50 to 500 number of epochs and tabulated the results accordingly.

III. PROPOSED SYSTEM

The project's proposed method begins with data collection for the IVF dataset and the blastocyst pictures. The dataset contains several missing instances and categorical values that must be converted to numerical values. 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 the image generation as the collected blastocyst images are very few in number, by creating our own dataset using GAN the accuracy of the model can be considerable good. By keeping this approach image classification for the blastocyst images are being done and considerably good accuracies are being

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)

generated when applying different algorithms for the model.

Further, the number of features obtained and the image generated can be verified with an expertized gynecologist and with 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 embryologist as selection of embryo plays a vital role in IVF treatment.

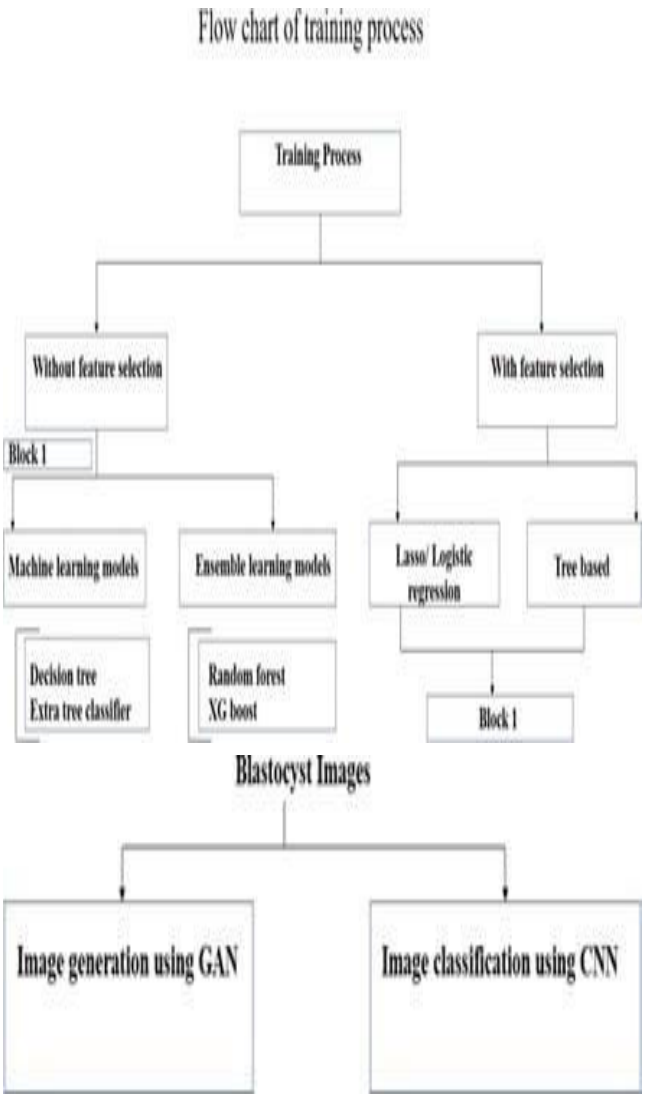


Fig 1. Represents the flow chart of training process.

IV. RESULTS AND DISCUSSIONS

In this section discuss about 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 Lasso regression model: Out of 27 features 6 features have been selected and the features selected are total number of previous IVF cycles, cause of infertility ovulatory disorder, cause of infertility male factor, stimulation used, fresh egg collected and embryos transferred.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805043

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)

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 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

TABLE II. FEATURES EXTRACTION FOR PCA AND LDA

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

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805043

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)

Fig 3. Accuracy of CNN model.

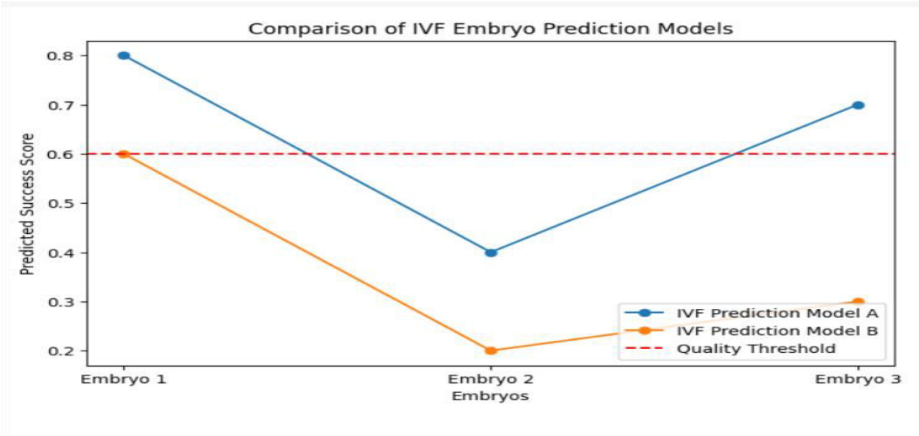


Fig 4. Loss and validation loss of CNN model.

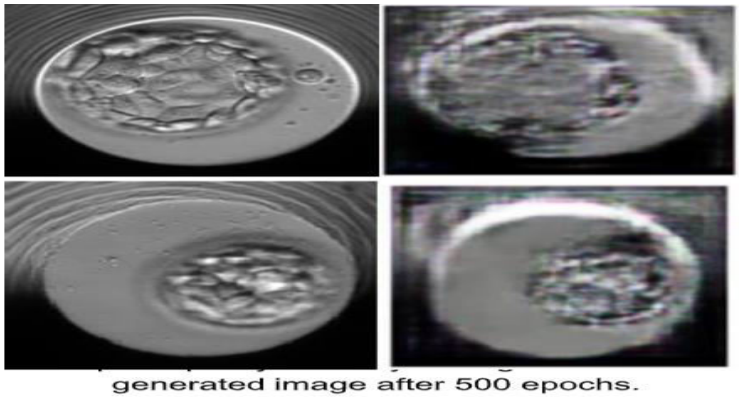


Fig 5. Comparing original good quality and poor-quality blastocyst image and the generated image after 500 epochs.

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 gynaecologists and a survey can be conducted 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 that can further create the dataset when the dataset is very minimal to calculate the accuracy. There can be use of any other algorithms apart from the algorithms that are being used in the project.

VI. 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 is some time consuming and reduces the accuracy of the model. The feature extraction for the IVF dataset is used to avoid the "curse of dimensionality". The project even works on the image generation using GAN and image classification using CNN model for the blastocyst images. The results obtained when applying machine learning and deep learning models are considerable good. When working with feature importance Random Forest has given a good accuracy of about 97.47%, and feature extraction LDA gave an accuracy of about 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 the 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.
5. Wiweko, "Human Blastocyst Classification after In Vitro Fertilization Using Deep Learning," 2020 7th International Conference on Advance 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 STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

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

www.ijmrset.com