Deep learning classification of embryo quality based on Blastocyst

P. Malathi[1], Dr. M. Gomathi[2]

[1]*Research Scholar, Department of Computer Science, Shrimathi Indira Gandhi College, Trichy, & Assistant Professor, Department of Computer Science, Guru Nanak College, Chennai, India*

[2]*Assistant Professor& Research Supervisor, Department of Computer Science, Shrimathi Indira Gandhi College Trichy, India.*

***Abstract*— A vital reproductive technique that has assisted millions in overcoming infertility issues is in vitro fertilization (IVF). The factors like environmental, clinical, and genetic are continue to vary due to complex interaction which affects its success rate. Although significant determinants of IVF success have been found using traditional statistical approaches, these methods frequently fail to capture the complex interactions that exist between these variables. Within this framework, artificial intelligence's deep learning subset provides a potent instrument for IVF decision-making and predictive modelling. This work investigates how deep learning techniques can be applied to enhance IVF outcome prediction by classifying excellent and bad embryos according to the existence of blastocysts. For the purpose of evaluating the viability of embryos, convolutional neural networks (CNNs) are utilized, especially transfer learning models like VGG19, MobileNetV2, InceptionV3 and Xception. Among these models MobileNetV2 performed well with the accuracy of 90% with low loss score of 0.21. Even in confusion metrices, MobileNetV2 achieved low false positive rate. Xception came in second with an accuracy of 87% and a loss score of 0.31. It also had a low false positive rate and a high true negative rate, indicating that it did a good job at classifying embryos of poor quality.**

***Keywords:* In Vitro Fertilization, Embryos, Deep learning, Transfer Learning model.**

I. INTRODUCTION

In-vitro fertilization(IVF) is a reproductive medicine which has undergone a revolution, it offers infertile couples a chance to become pregnant[1]. Even though with advances in technology, the range of clinical, genetic and environmental factors complicate the success rate of IVF[2]. The success rate of IVF has been found to be largely dependent on traditional methods; however, recent developments in artificial intelligence (AI), especially in deep learning, have opened up new avenues for predictive modeling and individualized treatment in the field of reproductive health [3].

Deep learning, a type of machine learning, is especially well-suited for applications in the healthcare industry because it uses neural networks to identify patterns in large and complex datasets [4]. Deep learning algorithms can evaluate a tonne of patient data in the context of IVF, including medical history, test results, lifestyle factors, and even photographs of the embryo, in order to forecast the chance of good outcomes[5]. The capacity to handle multi-dimensional data is essential for identifying the complex interrelationships among the various elements that influence the success or failure of IVF[6].

The ability of deep learning to enhance embryo selection, automate decision-making procedures, and customize treatment regimens presents a promising avenue for IVF research[7]. In order to evaluate the viability of embryos, convolutional neural networks (CNNs) have been used, while recurrent neural networks (RNNs) have been utilized to forecast the success of IVF using patient data collected over time[8]. With these methods, therapy outcomes can be better understood through data-driven insights and evaluation procedures that are less subjective and more predictive[9].

This study investigates how deep learning methods might improve IVF predictive prediction. Our goal is to create deep learning models that can automate the grading of embryos, discover important success factors, and ultimately enhance tailored treatment plans by using extensive datasets that embryo photos. We anticipate that this research will help to optimize assisted reproduction technologies by revealing fresh information about the intricate and non-linear aspects influencing IVF outcomes.

II. LITERATUREREVIEW

Deep learning applications in IVF are a quickly developing discipline that combine cutting-edge computational methods with reproductive medicine to increase therapy personalization and success rates. The majority of early research has been on how conventional variables like patient age, body mass index (BMI), and ovarian reserve affect the results of IVF. However, the limits of conventional statistical approaches, which often fail to capture the complex and non-linear interactions between these factors, have driven the adoption of artificial intelligence (AI) technologies[10].

Regression models and decision trees have been used in the past to forecast the results of IVF using clinical and demographic data. Studies like *Ashish Goyal et all*, emphasized characteristics such as age, past reproductive history, and embryo quality as major determinants. Predictive accuracy is nonetheless limited by these models' dependence on linear connections and inability to deal with the high-dimensionality of IVF data[11].

Deep learning has been studied to get around these limitations, with an emphasis on recurrent neural networks (RNNs) and convolutional neural networks (CNNs). CNNs have been shown to be effective in embryo selection [12]. For instance, *Darius Dirvanauskas et al*, improved the accuracy of predictions for embryo viability over standard embryologist assessments by developing a CNN-based algorithm to analyse time-lapse embryo imaging.

Additionally, RNNs have demonstrated promise, particularly in the analysis of longterm data like treatment cycles and hormone levels. RNNs were utilized by *Isaac Glatstein et al* on extensive IVF datasets, yielding superior predicted accuracy in contrast to linear models. One benefit of these models is that they may capture temporal dependencies, which is important when trying to figure out how elements unique to each patient change throughout the course of treatment[13].

Numerous comparative analyses have demonstrated deep learning models' superiority over conventional techniques. In predicting IVF outcomes, *Hakija Bečulić et al,* discovered that machine learning models in particular, deep learning techniques perform better than logistic regression, providing improved sensitivity and specificity. In order to further improve predictions, emerging trends also involve the integration of deep learning models with multi-omics data (genomics, proteomics, etc.)[14]. Through the integration of these many data sources, *Chunyu Huang et al* investigated the combination of genetic and non-genetic data and showed that deep learning models can more accurately predict complicated reproductive outcomes[15].

III. METHODOLOGY

This section delineates a systematic methodology for crafting and implementing deep learning algorithms geared towards classification of embryos into “Good” and “Bad”. Google Colab, a Python Integrated Development Environment (IDE), is used to carry out this entire routine. Figure 1 illustrate the workflow of the model development and implementing classification models for grading embryos.

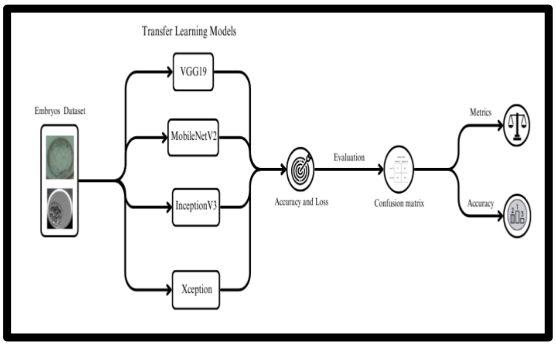


Fig 1: Workflow of proposed method

Data collection and pre-processing

The 1,461 embryo picture samples in the collection are divided into two classes: "Good" (viable embryos) and "Bad" (non-viable embryos). The figure 2 exhibits the distribution between good and bad embryo based on the presence of blastocysts. It was obtained from the Kaggle, an open repository for machine learning[16].

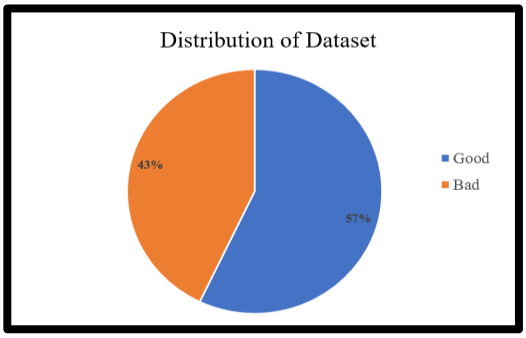


Figure 2: A distribution chart of embryos dataset

Partitioning the data into subsets for testing, validation, and training is one of the preprocessing procedures. Eighty percent of the dataset was used for training, while twenty percent was used for testing. The training set contains 667 images of good embryos and 501 of bad embryos, while test set contains 167 and 126 of good quality and bad quality embryos images, respectively. To achieved the robust model performance assessment, the data was normalized, shrinking the images to a standard dimension, and data augmenting was applied to improve model generalization.

Model development

Convolutional neural networks (CNNs) are the most widely used deep learning technique for image categorization and segmentation. Additionally, CNN is commonly utilized for medical image analysis to provide improved outcomes[17], [18].For image classification and pattern recognition applications, a number of well-known CNN models, including VGG-Net, GoogLeNet, ResNet, Inceptionv3, DenseNet, and AlexNet, have been widely used. The use of these pre-trained models has significantly expanded with the development of Transfer Learning (TL). With a limited amount of training data, higher success models have been created by using the knowledge of the previously trained network [19]. In this study, a classification task was performed using pretrained VGG-19, MobileNetV2, InceptionV3, and Xception transfer learning models. The models' performances were then compared. Ten epochs per fold were used to train the models, and a good degree of classification performance was attained.

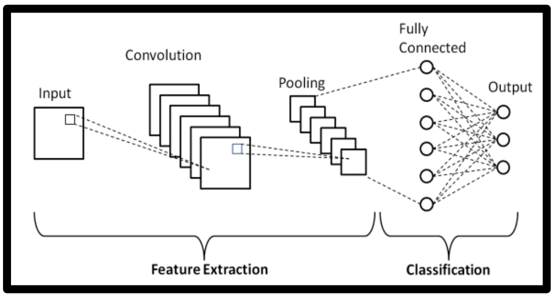


Figure 3: Basic workflow of CNN architecture[20]

VGG19

CNN architecture, VGG19 is renowned for its efficiency and simplicity in image classification applications. It was developed by the Visual Geometry Group at Oxford University and consists of 19 layers, including 3 fully connected layers and 16 convolutional layers. It has been widely used in many computer vision applications and uses small receptive fields (3x3) to capture high-level properties [21].

MobileNetV2

An effective deep learning architecture for mobile and edge devices is called MobileNetV2. It keeps high accuracy while lowering the number of parameters and computing cost by the use of depth wise separable convolutions. It is the ideal option for model deployment in environments with limited processing power because it is particularly well-suited for real-time applications [22].

InceptionV3

A multi-branch structure is used by the cutting-edge convolutional neural network architecture InceptionV3 to capture different feature sizes. Through the simultaneous combination of several kernel sizes, the model is able to learn rich feature representations. Compared to conventional CNNs, InceptionV3 is renowned for its exceptional accuracy and high performance in image classification tasks. It can accomplish these goals with less parameters[23].

Xception

The most sophisticated transfer learning model is Xception. It also expands on the idea of depthwise separable convolutions like MobileNet.It is made up of 36 convolutional layers and is intended to maximize model efficiency while achieving cutting-edge results in image classification applications. Because of its ability to extract characteristics from complex images, Xception is used for a variety of computer vision [24].

Performance and Evaluation

The effectiveness of a machine learning model is evaluated using performance metrics, which are numerical measurements. In this study, we use metrics like accuracy, precision, recall, and F1-score to assess how well the model performs in correctly classifying embryos into "Good" and "Bad" categories. Accuracy evaluates the model's overall correctness, whereas precision and recall provide details about the model's performance on each class. A single measure that fairly incorporates both recall and precision is provided by the F1-score [25].

The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) that the model generates on the test data is displayed in a confusion matrix: FP stands for wrongly predicted positive class, FN for incorrectly predicted negative class, TP for correctly predicted positive class, and TN for correctly predicted negative class. By calculating the metrics accuracy, recall, precision, and F1-score from the values in the table, a confusion matrix aids in comparing and assessing the performance of machine learning models. These metrics were then used to rank the learner models [25], [26].

IV. RESULT

The model performance was evaluated using various metrics, including confusion matrix, loss function, and scoring metrics. With a training accuracy of 90% and a low loss score of 0.21, the MobileNetV2 model outperformed the other models. The accuracy scores for VGG19, InceptionV3, and MobileNetV2 were 83%, 87%, and 88%, respectively. The accuracy and loss scores for the training sets are shown graphically in Figure 3.

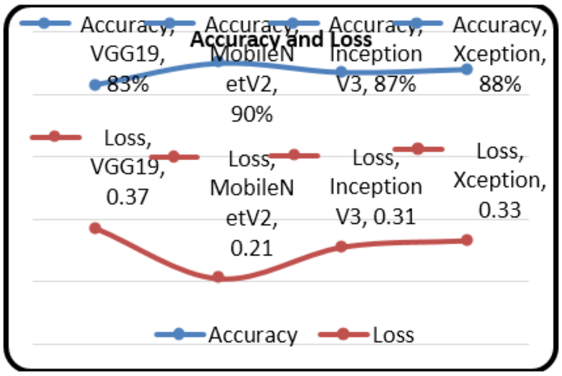
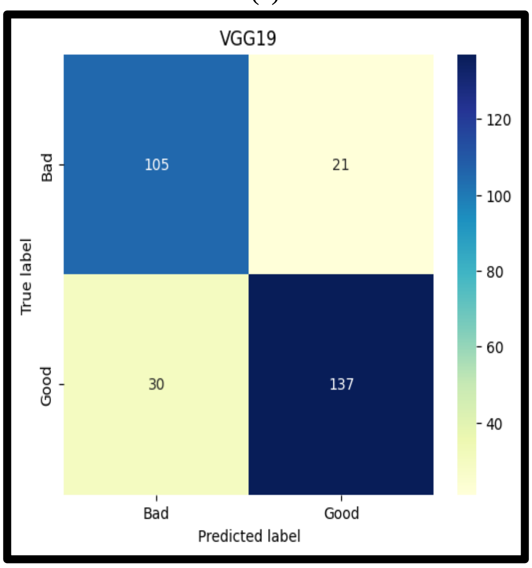


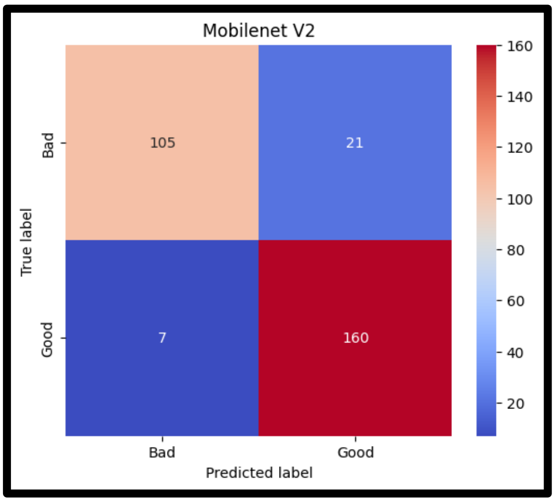
Figure 3: Training: accuracy and loss graph

Confusion matrix

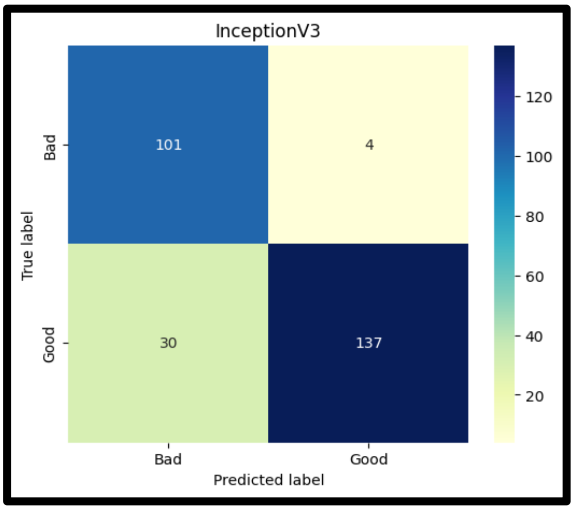
(a)



(b)



(c)



(d)

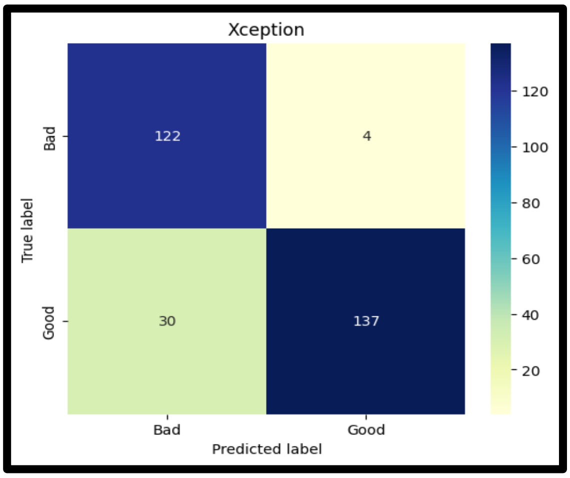


Figure 4: Confusion matrix of transfer learning models: VGG19, MobileNetV2, InceptionV3 and Xception

The confusion metrices(Figure 4) illustrates the classification performance of four transfer learning models. Among these models, MobileNetV2 performed well, with a notable low false negative rate(7), which denotes it effectively identified most of the "Good" embryos accurately. Followed by Xception model also performed well, with a low number of false positive(4) and a high true negative count(132). Indicating its ability to accurately classify the "Bad" embryos. In contrast, VGG19 and InceptionV3 exhibited high false negative rates(30) suggests a tendency to misclassification of "Good" embryos as "Bad". Overall MobileNetv2 and Xception showed superior accuracy and reliability in classification, making them strong models for further optimization and deployment in real-world application.

Performance Metrices

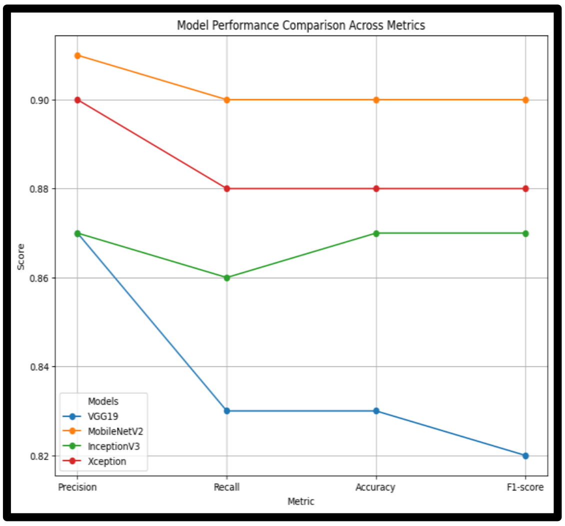


Figure 5: Performance metrices score of the transfer learning models

Figure 5 displays the performance rankings of the four models, with MobileNetV2 as the top performing model with precision of 0.91, recall of 0.90, F1-score of 0.90 and accuracy of 0.90. Followed by Xception with precision of 0.90, recall of 0.88, F1-Score of 0.88, and accuracy of 0.88. Whereas VGG19 and InceptionV3 demonstrates moderate performance with precision of 0.82 and 0.87, recall of 0.83 and 0.86, F1-Score of 0.82 and 0.87 and accuracy of 0.83 and 0.87 respectively.

V. CONCLUSION

This research utilizes deep learning models to identify and classify good and bad embryos based on the presence of blastocysts. The classification process was evaluated using four alternative models with the measures of accuracy, recall, precision, and F1-score. The MobileNetV2 model excelled with an accuracy of 90%. The Xception and InceptionV3 model showed slight changes in results, with accuracy of 88% and 87% respectively. Consistent with related studies, the VGG19 model had the lowest success rate in this study. Recent similar research that have been published in the literature were compared to the results. This study showed that even with little data and few training epochs, the transfer learning method yields good results.

Future scope of this research could concentrate on overcoming the constraints and delving into the utilization of deep learning models in various factors of IVF i.e., genetic, lifestyle, food, clinical factors, and time-lapse photos of the embryo .Moreover, creating more understandable deep learning models could greatly improve our understanding of IVF by revealing their biological mechanisms.

VI. REFERENCE

[1] V. A. Kushnir, G. D. Smith, and E. Y. Adashi, "INFERTILITY: PERSPECTIVE, OPINIONS AND COMMENTARIES The Future of IVF: The New Normal in Human Reproduction," vol. 1, p. 3, doi: 10.1007/s43032-021-00829-3.

[2] J. Wang and M. V. Sauer, "In vitro fertilization (IVF): a review of 3 decades of clinical innovation and technological advancement," *Ther Clin Risk Manag*, vol. 2, no. 4, p. 355, 2006, doi: 10.2147/TCRM.2006.2.4.355.

[3] D. J. X. Chow, P. Wijesinghe, K. Dholakia, and K. R. Dunning, "Does artificial intelligence have a role in the IVF clinic?," *Reproduction & Fertility*, vol. 2, no. 3, p. C29, Jul. 2021, doi: 10.1530/RAF-21-0043.

[4] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 6, pp. 1–20, Nov. 2021, doi: 10.1007/S42979-021-00815-1/FIGURES/11.

[5] M. R. Borna, M. M. Sepehri, and B. Maleki, "An artificial intelligence algorithm to select most viable embryos considering current process in IVF labs," *Front Artif Intell*, vol. 7, p. 1375474, 2024, doi: 10.3389/FRAI.2024.1375474.

[6] L. Shingshetty, N. J. Cameron, D. J. Mclernon, and S. Bhattacharya, "Predictors of success after in vitro fertilization," *Fertil Steril*, vol. 121, no. 5, pp. 742–751, May 2024, doi: 10.1016/J.FERTNSTERT.2024.03.003.

[7] M. Salih *et al.*, "Embryo selection through artificial intelligence versus embryologists: a

systematic review," *Hum Reprod Open*, vol. 2023, no. 3, 2023, doi: 10.1093/HROPEN/HOAD031.

[8] C. M. Louis, A. Erwin, N. Handayani, A. A. Polim, A. Boediono, and I. Sini, "Review of computer vision application in in vitro fertilization: the application of deep learning-based computer vision technology in the world of IVF," *J Assist Reprod Genet*, vol. 38, no. 7, p. 1627, Jul. 2021, doi: 10.1007/S10815-021-02123-2.

[9] F. Cascini, F. Santaroni, R. Lanzetti, G. Failla, A. Gentili, and W. Ricciardi, "Developing a Data-Driven Approach in Order to Improve the Safety and Quality of Patient Care," *Front Public Health*, vol. 9, p. 667819, May 2021, doi: 10.3389/FPUBH.2021.667819.

[10] R. Wang *et al.*, "Artificial intelligence in reproductive medicine," *Reproduction*, vol. 158, no. 4, p. R139, 2019, doi: 10.1530/REP-18-0523.

[11] A. Goyal, M. Kuchana, and K. P. R. Ayyagari, "Machine learning predicts live-birth occurrence before in-vitro fertilization treatment," *Sci Rep*, vol. 10, no. 1, p. 20925, Dec. 2020, doi: 10.1038/S41598-020-76928-Z.

[12] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data 2021 8:1*, vol. 8, no. 1, pp. 1–74, Mar. 2021, doi: 10.1186/S40537-021-00444-8.

[13] I. Glatstein, A. Chavez-Badiola, and C. L. Curchoe, "New frontiers in embryo selection," *J Assist Reprod Genet*, vol. 40, no. 2, p. 223, Feb. 2023, doi: 10.1007/S10815-022-02708-5.

[14] H. Bečulić *et al.*, "Sensitivity and specificity of machine learning and deep learning algorithms in the diagnosis of thoracolumbar injuries resulting in vertebral fractures: A systematic review and meta-analysis," *Brain and Spine*, vol. 4, p. 102809, Jan. 2024, doi: 10.1016/J.BAS.2024.102809.

[15] C. Huang *et al.*, "Using Deep Learning in a Monocentric Study to Characterize Maternal Immune Environment for Predicting Pregnancy Outcomes in the Recurrent Reproductive Failure Patients," *Front Immunol*, vol. 12, p. 642167, Apr. 2021, doi: 10.3389/FIMMU.2021.642167.

[16] "Embryo Dataset." Accessed: Oct. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/alaasotohy/embryo-dataset

[17] O. N. Belaid and M. Loudini, "Classification of Brain Tumor by Combination of Pre-Trained VGG16 CNN," *Journal of Information Technology Management*, vol. 12, no. 2, pp. 13–25, Jun. 2020, doi: 10.22059/JITM.2020.75788.

[18] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif Intell Rev*, vol. 57, no. 4, pp. 1–43, Apr. 2024, doi: 10.1007/S10462-024-10721-6/FIGURES/33.

[19] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[20] "Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network | upGrad blog." Accessed: Oct. 04, 2024. [Online]. Available: https://www.upgrad.com/blog/basic-cnn-architecture/

[21] K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," 2015, Accessed: Oct. 04, 2024. [Online]. Available: http://www.robots.ox.ac.uk/

[22] L. Zhao, L. Wang, Y. Jia, and Y. Cui, "A lightweight deep neural network with higher accuracy," *PLoS One*, vol. 17, no. 8, p. e0271225, Aug. 2022, doi: 10.1371/JOURNAL.PONE.0271225.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 2818–2826, Dec. 2015, doi: 10.1109/CVPR.2016.308.

[24] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, Oct. 2016, doi: 10.1109/CVPR.2017.195.

[25] D. Abadjieva *et al.*, "Machine Learning Approach for Muscovy Duck (Cairina moschata) Semen Quality Assessment," *Animals*, vol. 13, no. 10, May 2023, doi: 10.3390/ANI13101596/S1.

[26] "(PDF) Confusion Matrix." Accessed: Oct. 04, 2024. [Online]. Available: https://www.researchgate.net/publication/355096788_Confusion_Matrix