Ensemble Learning Technique for Artificial Intelligence Assisted IVF Applications

George Vergos
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
gvergos@physics.auth.gr

Lazaros Alexios Iliadis
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
liliadis@physics.auth.gr

Paraskevi Kritopoulou
Dept. of Applied Informatics
University of Macedonia
Thessaloniki, Greece
pkritopoulou@uom.edu.gr

Achilleas Papatheodorou
Embryolab Fertility Clinic
Thessaloniki, Greece
a.papatheodorou@embryolab.eu

Achilles D. Boursianis
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
bachi@physics.auth.gr

Konstantinos-Iraklis D. Kokkinidis
Dept. of Applied Informatics
University of Macedonia
Thessaloniki, Greece
kostas.kokkinidis@uom.edu.gr

Maria S. Papadopoulou
Dept. of Information and Electronic Eng.
International Hellenic University
Sindos, Greece
mspapa@ihu.gr

Sotirios K. Goudos
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
sgoudo@physics.auth.gr

2023 12th International Conference on Modern Circuits and Systems Technologies (MOCAST) | 979-8-3503-2107-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/MOCAST57943.2023.10176690

***Abstract*—In-vitro fertilization (IVF) is considered one of the most effective assisted reproductive approaches. However, it involves a series of complex and expensive procedures, that result in approximately 30% success rate. New technological concepts, like deep learning (DL), are required to increase the success rate while simplifying the procedure. DL techniques can automatically extract valuable features from the available data, can be flexible, and can work efficiently on multiple problems. Motivated by these characteristics, in this work, a DL ensemble model is trained on a private dataset to classify blastocysts' images. Further analysis is conducted to provide insights for future research.**

***Index Terms*—Deep learning, IVF, convolutional neural network, ensemble learning**

I. Introduction

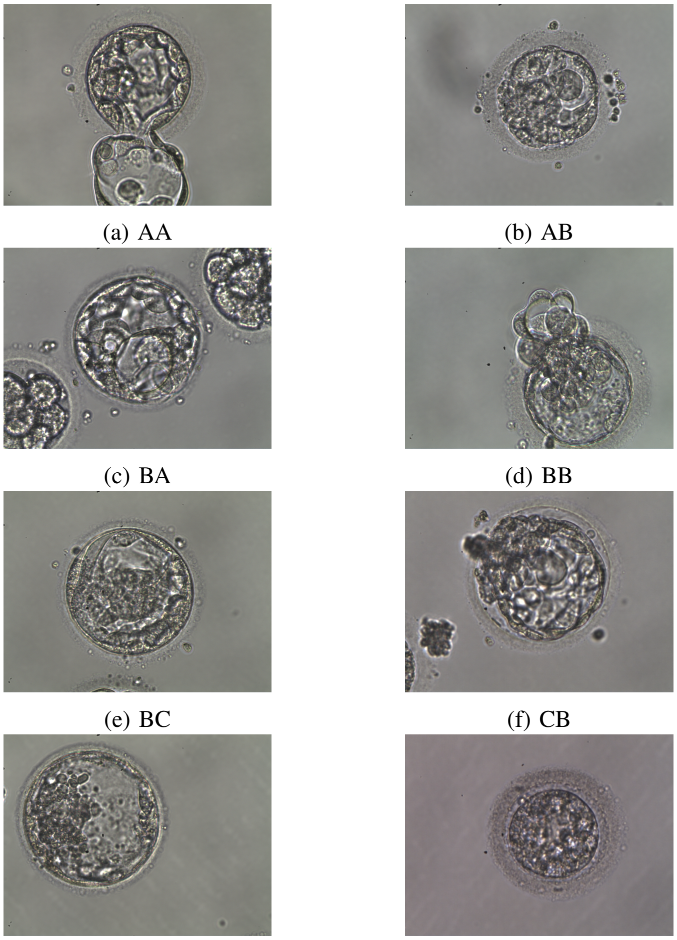
Thanks to in-vitro fertilization (IVF) technology, millions of kids have been born. However, only one-third of the couples, who resort to IVF, manage to have a child, despite the long and expensive procedures. To overcome several challenges, such as the age, the quality of the embryo, and the limitations of the required technology [1], embryologists and researchers search for new tools and methods that will provide better results. Time-lapse imaging incubators (TLI) were introduced as an IVF method more than ten years ago. Setting regular time intervals, TLI takes photographs, which form a video of the embryo's development. TLI enables controlled culture conditions, while the developmental events are annotated in a dynamic procedure. Such events, which are called morphokinetic parameters, include cell divisions, blastocyst formation, and expansion [2].

Deep learning (DL) methods are part of machine learning methodologies and have been established as the most successful set of approaches in the field of computer vision (CV) [3]. Convolutional neural networks (CNNs), after their success in various image classification tasks [4], constitute the dominant approach in CV and have been widely utilized in several CV tasks, including medical imaging [5]. In the last five years, a growing research trend for DL-empowered IVF approaches can be identified [6]. Motivated by the above, in this paper, an ensemble DL model is trained on a private dataset to classify blastocysts' images.

This work is part of the "Smart Embryo" project which aims to develop artificial intelligence (AI) based software to assist embryologists to evaluate the blastocysts' quality before transfer. This research project aims to develop novel AI methods for DL-powered IVF.

This work is structured in the following way: Section II describes the problem, whereas Section III provides the details of our DL model. The experiments and the results of our approach are discussed in Section IV. The conclusions are presented in Section V.

This research was carried out as part of the project "Classification and characterization of fetal images for assisted reproduction using artificial intelligence and computer vision" (Project code: KP6-0079459) under the framework of the Action "Investment Plans of Innovation" of the Operational Program "Central Macedonia 2014 2020", that is co-funded by the European Regional Development Fund and Greece.



(a) AA (b) AB (c) BA (d) BB (e) BC (f) CB (g) CC (h) DD

Fig. 1: Blastocyst images obtained from our private dataset.

II. Problem description

The different aspects of the blastocysts' grading system are presented in this Section. This morphological evaluation system has been developed to evaluate the quality of a blastocyst before the transfer and can be exploited by DL models. The number of cells inside the embryo begins to outgrow the space inside the zona pellucida. Once this zone breaks, the blastocyst can hatch. Each blastocyst contains two types of cells; those that form the placental tissue, and those that form the fetus. Two key parameters that help embryologists decide which blastocyst has more viability chances [7] are:

- Inner cell mass (ICM) score, (A) through (D)
- Trophectoderm (TE) score, (A) through (D)

In Fig. 1, samples of blastocyst images are depicted. The quality scores are presented in Tables I and II. Since the blastocysts' grading system is based only on morphological characteristics, their classification may be formulated into a CV problem.

In this work, the dataset is constructed in the following way. First, images, accompanied by their respective metadata, are collected from a TLI. Then, the images are categorized into two classes; class 0 and class 1. Class 0 represents the images that have one of the following ICM and TE scores combined: AA, AB, BA. Class 1 represents all the other images.

TABLE I: ICM score

ICM grade	*ICM quality*
A	Large number of cells that are tightly packed
B	Many cells that are loosely grouped
C	Small number of cells

TABLE II: TE score

TE grade	*TE quality*
A	Large number of cells, which form a compact layer
B	Several cells, which form a loose epithelium
C	Small number of large cells

III. DL methodologies

To address the problem of blastocysts' quality classification, three different DL architectures are utilized to construct an ensemble learner. All three models are based on the convolution filters (kernels) that slide along the images to provide the respective feature maps.

A. DL models

AlexNet [8] is the first CNN that outperformed the classical CV algorithms in the Imagenet large-scale visual recognition challenge (ILSVRC) [9]. AlexNet utilizes five convolutional layers, while down-sampled feature maps are created by the three max-pooling layers. The flattening of the data and the outcomes are the results of the three fully-connected layers. The activation function in the convolutional layers is the rectified linear unit (ReLU). ReLU is simply defined as

$$g(z) = \max(0, z) \tag{1}$$

Another popular CNN model is the VGG11 [10]. VGG11 has eight convolution layers, five max-pooling layers, and three fully-connected layers. As in the AlexNet architecture, the activation function is taken to be the ReLU function. Both AlexNet and VGG11 use dropout layers, which randomly set input units to 0. In this way, DL models avoid over-fitting.

A variant of VGG is constructed for this particular problem. This model has seven convolutional layers, four max-pooling layers, and three fully-connected layers. In contrast to AlexNet and VGG11, no dropout layer is utilized.

B. Ensemble learning

Ensemble learning is frequently utilized to achieve better results. In this work, VGG11, AlexNet, and the variant of VGG constructed an ensemble classifier. By combining the probability outputs of each base model into two new probabilistic outcomes, the ensemble can predict with greater confidence the blastocyst's class. The fact that two of the DL models contain dropout layers results in a large variance. To stabilize and further improve the DL models' performance, five instances of this learner formed a voting classifier that takes the average over the predictions of all estimators.

IV. Experiments and Results

In this work, an ensemble learning approach is taken for the task of blastocysts image classification. The private dataset consists of 2269 images categorized into two classes; class 0 (suitable for transfer) and class 1 (unsuitable for transfer). The dataset is split into a training set (80% of the initial set) and a test set (20% of the initial set).

To update the models' weights, Adam (adaptive moment estimation) algorithm [11] is utilized. The training is performed on an NVIDIA RTX 3080 Ti GPU. The GPU processor is a chip with a die area of 628 mm^2 and 28,300 million transistors, making it suitable for DL applications. Each base model is trained for 40 epochs with a mini-batch size of 8. The initial learning rate's value is set to 10^{-4} and it is divided by 10 when the error reaches a plateau. VGG11 performs the best with an accuracy of 76% while AlexNet and the VGG Variant have an accuracy score of 73% and 72% respectively. By combining the outputs of the base models into an ensemble the accuracy is further increased to 78%. Finally, to further improve the model's predictive capability, a voting classifier is implemented by taking five instances of the ensemble model and it is trained for 5 epochs. The whole algorithmic procedure is depicted in Fig. 2. The final accuracy score is improved to 81%. The loss and accuracy graphs are shown in Fig. 3 and Fig. 4 respectively.

The accuracy score itself cannot provide enough information about the classifier's learning capabilities. Considering our problem (and binary classification problems in general), a data sample (instance) is classified as either belonging to class 0 or class 1. The following four outcomes are possible:

1) True Positive (T_p): The data sample belongs to class 0 and it is classified as class 0
2) False Negative (F_n): The data sample is in class 0 and it is classified as class 1
3) True Negative (T_n): The data sample belongs to class 1 and it is classified in class 1
4) False Positive (F_p): The data sample is in class 1 and it is classified as part of class 0

The above outcomes are better formatted in the confusion matrix configuration. In our case, the voting classifier's performance is illustrated in Fig. 5. Using two other scores, namely *Precision* and *Recall*, one can define the f1-score which is a measure of the proposed algorithm's accuracy. Recall shows how many of the 0 class instances were predicted correctly

$$Rec = \frac{T_p}{T_p + F_n} \tag{2}$$

and Precision is defined as

$$Prec = \frac{T_p}{T_p + F_p} \tag{3}$$

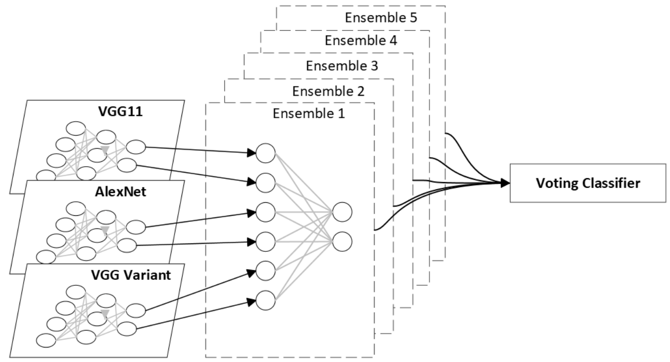


Fig. 2: Model Architecture

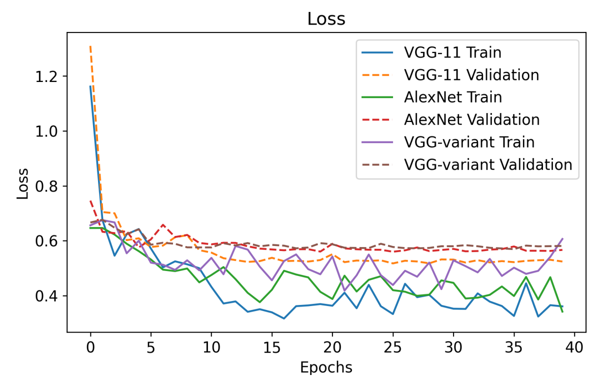


Fig. 3: Loss

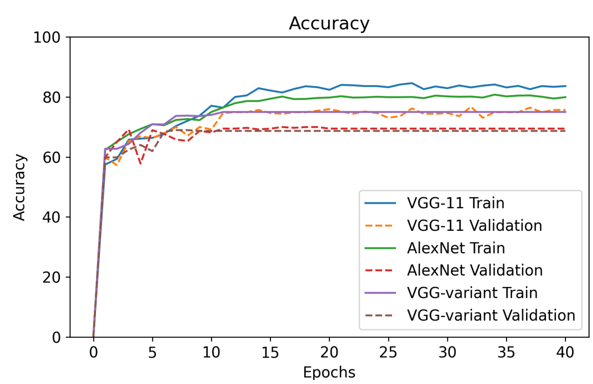


Fig. 4: Accuracy

The f1-score is the harmonic mean of the above measures, and it is defined as

$$\text{f1-score} = 2 \cdot \frac{Rec \cdot Prec}{Rec + Prec} \tag{4}$$

In our case, the final f1-score = 85.5%, which validates our model's good performance.

V. Conclusions

In this work, an ensemble learner is trained on a private dataset to classify blastocysts' images. Three different DL models were trained from scratch, and then they constructed

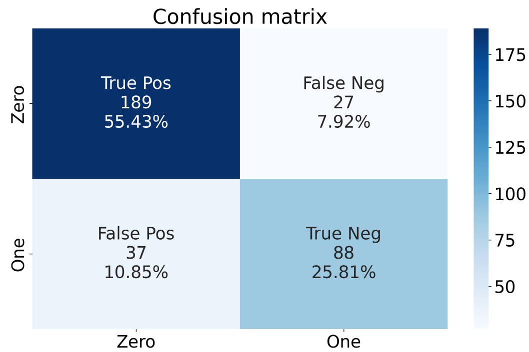


Fig. 5: Confusion matrix of the final ensemble learner

an ensemble learner. Five different instances of this learner formed the final classifier, in a voting framework. This DL method achieved over 81% prediction accuracy, thus it is considered quite satisfactory. Future research includes the extension to a multi-class classification formulation and the utilization of other DL architectures.

References

[1] S. Dyer, G. Chambers, J. de Mouzon, K. Nygren, F. Zegers-Hochschild, R. Mansour, O. Ishihara, M. Banker, and G. Adamson, "International Committee for Monitoring Assisted Reproductive Technologies world report: Assisted Reproductive Technology 2008, 2009 and 2010," *Human Reproduction*, vol. 31, no. 7, pp. 1588–1609, Apr. 2016, doi:10.1093/humrep/dew082.

[2] C. Pribenszky, A.-M. Nilselid, and M. H. M. Montag, "Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis." *Reproductive biomedicine online*, vol. 35 5, pp. 511–520, Nov. 2017.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, ser. Adaptive computation and machine learning. MIT Press, Nov. 2016.

[4] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, pp. 1–98, 06 2017, doi:https://doi.org/10.1162/NECO_a_00990.

[5] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. D. F. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 31–40, 01 2017, doi:https://doi.org/10.1109/JBHI.2016.2635663.

[6] K. Sfakianoudis, E. Maziotis, S. Grigoriadis, A. Pantou, G. Kokkini, A. Trypidi, P. Giannelou, A. Zikopoulos, I. Angeli, T. Vaxevanoglou, K. Pantos, and M. Simopoulou, "Reporting on the value of artificial intelligence in predicting the optimal embryo for transfer: A systematic review including data synthesis," *Biomedicines*, vol. 10, p. 697, 03 2022, doi:https://doi.org/10.3390/biomedicines10030697.

[7] T. Baczkowski, R. Kurzawa, and W. Głabowski, "Methods of scoring in in vitro fertilization," *Reproductive biology*, vol. 4, pp. 5–22, 04 2004.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015, doi:https://doi.org/10.1007/s11263-015-0816-y.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.