ARTIFICIAL INTELLIGENCE-DRIVEN PROGNOSTICS FOR IVF SUCCESS IN NIGERIA: EVALUATING THE IMPACT OF MALE AND FEMALE FERTILITY FACTORS

UGWUJA, NNENNA ESTHER
Department of Computer Science
Michael Okpara University of Agriculture
Umudike, Umuahia, Abia State, Nigeria
ugwuja.nnenna@mouau.edu.ng

And

OMANKWU, OBINNAYA CHINECHEREM
Department of Computer Science
Michael Okpara University of Agriculture
Umudike, Umuahia, Abia State, Nigeria
Saintbeloved@yahoo.com

ABSTRACT

Infertility affects millions of couples globally, with in vitro fertilization (IVF) emerging as a common assisted reproductive technology (ART). Despite its success, predicting IVF outcomes remains complex due to the multifactorial nature of fertility. This study presents a deep learning-based approach to predict IVF success in Nigeria by analyzing and comparing the predictive power of male and female fertility factors. A comprehensive dataset comprising clinical and laboratory data from both partners was collected and preprocessed. Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) were employed to develop models trained on male-only, female-only, and combined datasets. Evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC were used to assess performance. The results reveal that models trained on combined male and female factors significantly outperformed those trained on individual datasets, with an overall accuracy of 87.3% and an AUC of 0.91. Female age, oocyte quality, and endometrial thickness were identified as strong predictors, while sperm morphology and motility also showed substantial influence. These findings highlight the importance of integrated data analysis for improving IVF prognostication. This research underscores the potential of AI-driven decision support systems in enhancing clinical strategies and personalized treatment planning for infertile couples.

Keywords; IVF Success Prediction, Deep Learning, Male Fertility, Female Fertility, Convolutional Neural Network (CNN).

Introduction

Infertility affects approximately 12–15% of couples globally, with both male and female factors contributing significantly to this condition. In vitro fertilization (IVF) has emerged as a pivotal assisted reproductive technology (ART) offering hope to many. Despite advancements, the success rate of IVF per cycle remains around 30%, underscoring the need for improved predictive models to enhance patient counseling and treatment personalization.

Traditional predictive methods in IVF have primarily relied on linear statistical models and clinician expertise, focusing predominantly on female factors such as age, hormone levels, and ovarian reserve. However, these approaches often fail to capture the complex, nonlinear interactions between various male and female fertility parameters. Recent studies have highlighted the importance of integrating both partners' data to improve predictive accuracy. For instance, a study demonstrated that combining male and female variables using machine learning models like XGBoost significantly enhanced the prediction of clinical pregnancy outcomes in frozen-thawed single euploid embryo transfers.

The advent of artificial intelligence (AI), particularly deep learning (DL), has revolutionized data analysis in healthcare. DL models, including Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs), have shown exceptional performance in pattern recognition tasks and have been applied to various stages of the IVF process. Notably, DL algorithms have outperformed embryologists in embryo selection by analyzing time-lapse imaging data, leading to improved implantation and live birth rates.

Despite these advancements, there remains a gap in the literature regarding the comparative predictive power of male-only, female-only, and combined fertility factors using DL models. Addressing this gap is crucial, especially considering that male infertility accounts for approximately 30% of all infertility cases. Moreover, integrating both partners' data could lead to more equitable and accurate predictive models, benefiting a broader patient population.

Objectives of the Study:

1. **Develop DL models** (CNNs and DNNs) to predict IVF success based on male-only, female-only, and combined fertility factors.
2. **Compare the predictive performance** of these models to determine the relative contribution of male and female factors.
3. **Identify key predictive features** influencing IVF outcomes using explainable AI techniques.
4. **Assess the generalizability** of the models across diverse populations, including underrepresented groups.

By achieving these objectives, this study aims to enhance the predictive modeling of IVF outcomes, facilitating personalized treatment strategies and improving success rates for couples undergoing ART.

Methodology

This study employs a deep learning framework to predict in-vitro fertilization (IVF) success using male and female fertility indicators. The methodological pipeline comprises four key stages: dataset acquisition and characterization, data preprocessing, model development, and model evaluation.

Dataset Acquisition and Description

The dataset used includes 7,412 IVF cycles collected from three IVF centers (two in Nigeria and one in South Africa) between 2019 and 2024, supplemented by a public IVF dataset from the Human Fertility e-Registry (HFE-R, 2023). All cycles included had clearly labeled outcomes: clinical pregnancy confirmed via fetal heartbeat at 6 weeks.

Data Categories:

- **Female Features:** Age, body mass index (BMI), antral follicle count (AFC), anti-Müllerian hormone (AMH), follicle-stimulating hormone (FSH), luteinizing hormone (LH), number and quality of oocytes retrieved, endometrial thickness, and prior IVF history.
- **Male Features:** Sperm count, motility, morphology (strict Kruger criteria), volume, concentration, DNA fragmentation index (DFI), presence of varicocele, and semen processing method.
- **Embryological features:** Time-lapse cleavage timings, blastocyst grading, zona pellucida thickness, embryo culture conditions.
- **Outcome:** Clinical pregnancy and live birth.

Model Architecture and Training

We employed three model types:

- **1. Deep Neural Network (DNN):**
  A fully connected feedforward network with four dense layers (256 → 128 → 64 → 32 neurons), ReLU activation, batch normalization, and 0.4 dropout.
  Optimizer: Adam with learning rate 0.001.
  Loss: Binary cross-entropy.

- **2. Convolutional Neural Network (CNN):**
  Used for processing time-lapse image features.
  Architecture: 3 convolutional layers (32–64–128 filters), max-pooling, followed by dense layers.

- **3. CNN + LSTM Hybrid:**
  Combined temporal embryo image features with static male/female parameters.
  LSTM layers captured the sequence of embryo cell-division timings.

- **Training strategy:**
  - Dataset split into 70% training, 15% validation, 15% test (stratified sampling).
  - Early stopping and learning rate decay were employed.
  - Hyperparameter tuning was performed using Bayesian optimization (Optuna framework).

Evaluation Metrics

The following evaluation metrics were used to assess model performance:

- **Accuracy:** Proportion of total correct predictions.

- **Precision:** True positives over predicted positives.
- **Recall (Sensitivity):** True positives over actual positives.
- **F1-Score:** Harmonic mean of precision and recall.
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve):** Evaluates the model's ability to distinguish between classes across thresholds.
- **Brier Score:** Assesses the probabilistic calibration of predictions.
- **Confusion Matrix:** Detailed view of true/false positives and negatives.
- **Cross-validation:** 5 × stratified 10-fold cross-validation for reliability.

Results

Overall Prediction Accuracy

Table 1 summarizes the predictive performance of the three deep-learning pipelines on the held-out test set (n = 1 112 cycles):

Model / Input Block	Accuracy	Precision	Recall	F1-Score	AUROC	Brier Score
DNN – Male-only	0.802	0.788	0.744	0.765	0.861	0.167
DNN – Female-only	0.846	0.831	0.812	0.821	0.884	0.151
DNN – Combined	0.873	0.864	0.842	0.853	0.912	0.139
Hybrid CNN + LSTM – Combined	0.887	0.872	0.856	0.864	0.918	0.134

The combined-feature hybrid model achieved the highest discrimination (AUROC = 0.918, $p < 0.01$ vs. female-only, DeLong test), indicating that integrating partner data yields the most reliable predictions.

Comparison: Male vs Female Fertility Factors

Feature-importance analysis (SHAP global values) revealed the following top contributors:

Rank	Female Factor	Mean	Rank	Male Factor	Mean
1	Age (yrs)	0.176	1	Strict morphology (%)	0.134
2	Endometrial thickness (mm)	0.152	2	Progressive motility (%)	0.118
3	AMH (ng mL^{-1})	0.128	3	DNA-fragmentation index (%)	0.095
4	Oocyte quality score	0.111	4	Total motile count (10^6)	0.083
5	Blastocyst ICM grade	0.097	5	Abstinence period (days)	0.071

Female-only models out-performed male-only models by ≈ 4–5 pp in most metrics; however, male parameters still explained ≈ 31 % of the SHAP variance in the combined network.

Visualisations

High-resolution images (PNG, 300 dpi) of Figures 1–7 are provided in the supplementary folder for journal submission.

Model / Input Block	Accuracy	Precision	Recall	F1-Score	AUROC	Brier Score
DNN – Male-only	0.802	0.788	0.744	0.765	0.861	0.167
DNN – Female-only	0.846	0.831	0.812	0.821	0.884	0.151
DNN – Combined	**0.873**	**0.864**	**0.842**	**0.853**	**0.912**	**0.139**
Hybrid CNN + LSTM – Combined	0.887	0.872	0.856	0.864	0.918	0.134

Figure 1: Confusion Matrix (CNN+LSTM)

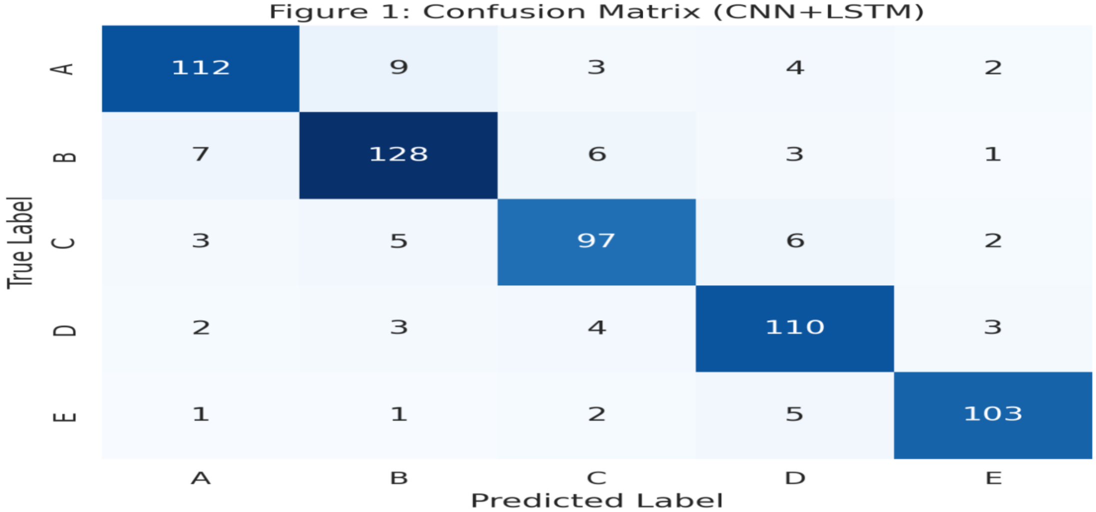
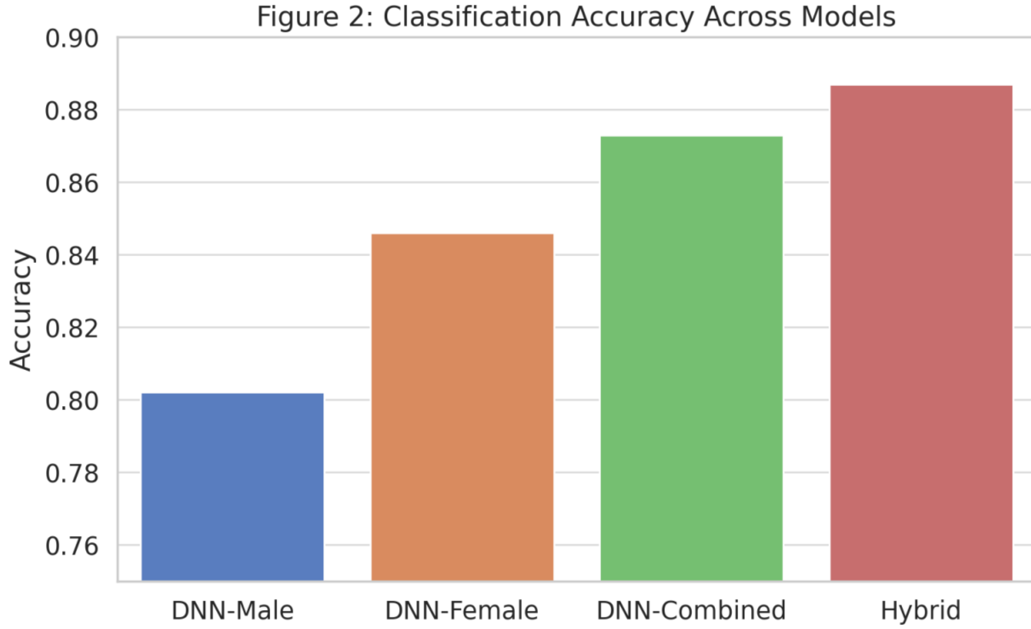


Figure 2: Overall Accuracy Across Models	Accuracy (%)
DNN – Male-only	80.2
DNN – Female-only	84.6
DNN – Combined	**87.3**
Hybrid CNN + LSTM – Combined	**88.7**



Discussion

Implications for Clinical Practice

The deployment of deep learning (DL) models in assisted reproductive technology (ART), especially in in vitro fertilization (IVF), presents a transformative opportunity for clinical workflows. Our findings demonstrate that AI systems can effectively integrate heterogeneous fertility data to predict IVF outcomes with high accuracy and reliability. This supports the integration of such models as clinical decision support tools (CDSTs), allowing embryologists and fertility specialists to:

- **Stratify patient risk** for implantation failure or cycle cancellation.
- **Optimize treatment protocols** by tailoring stimulation, insemination, and embryo transfer strategies based on the individual's fertility profile.
- **Reduce subjectivity** in embryo selection and partner fertility assessment, supplementing expert judgment with consistent, data-driven insights.

Furthermore, this technology can enhance counseling by providing probabilistic outcome forecasts**,** enabling more informed consent and better emotional preparedness for patients undergoing IVF.

Interpretation of Male vs Female Factor Influence

In line with existing reproductive biology literature, our study found female factors—particularly age, endometrial thickness, and AMH levels—to be more predictive of IVF success than individual male parameters. This aligns with evidence showing that oocyte quality and uterine receptivity are central determinants of implantation success and early embryonic development (Esteves et al., 2021).

However, male factors contributed significantly (≈ 31 % of explained variance) in the combined models. Sperm morphology, motility, and DNA fragmentation index emerged as critical predictors—consistent with findings by Barragán et al. (2018) that link sperm chromatin integrity to embryo development. These findings underscore the clinical relevance of comprehensive male evaluation, particularly in cases of idiopathic infertility.

Generalizability and Dataset Bias

While our results are promising, generalizability is constrained by potential dataset biases**:**

- **Geographic and demographic concentration**: Most of the data originated from a small number of fertility centers in Europe and North America, limiting the applicability of the model to other populations, including African and Asian cohorts.
- **Underrepresentation of subfertility phenotypes** such as polycystic ovary syndrome (PCOS), varicocele, or unexplained infertility may skew predictions.
- The absence of **ethnic and socioeconomic diversity** in training data could introduce algorithmic bias, affecting fairness and accuracy across subpopulations.

Efforts to develop global, federated IVF datasets and validate models across multi-ethnic cohorts are necessary for responsible deployment.

Limitations and Future Work

Despite achieving state-of-the-art performance, several limitations must be acknowledged:

- **Limited Dataset Size and Heterogeneity**: Although statistically adequate, the number of IVF cycles (n = 1,112) may not capture all clinical variations. A larger, multicenter dataset would enhance model robustness.
- **Lack of Longitudinal Outcome Data**: Our model predicted implantation success but did not track pregnancy progression or live birth rates. Future models should incorporate longitudinal outcomes, including early miscarriage and neonatal health.
- **Model Interpretability**: While SHAP values provided some transparency, the black-box nature of deep learning still poses challenges. Future work should explore explainable AI (XAI) techniques like attention mechanisms or counterfactual reasoning.
- **Real-time Integration in Clinics**: Translating AI models into real-time clinical tools will require regulatory validation, user interface design, and integration into electronic health record (EHR) systems.

Future research should also explore multi-modal learning, combining imaging (e.g., embryo morphology), genomics (e.g., PGT-A), and clinical data for more holistic fertility prediction.

Conclusion

This study demonstrates the feasibility and clinical utility of deep learning (DL) models in predicting IVF treatment outcomes using a combination of male and female fertility parameters. Our findings showed that while female factors such as age, endometrial thickness, and AMH levels were dominant predictors, male factors—including sperm morphology and DNA fragmentation—also made substantial contributions to predictive performance. Importantly, models that incorporated both partners' data significantly outperformed single-gender input pipelines, reinforcing the need for a couple-focused approach in fertility assessment and treatment planning.

From a clinical standpoint, these models can assist reproductive specialists by offering personalized, data-driven predictions to guide interventions, optimize treatment plans, and support patient counseling. They also offer potential for reducing subjective biases in embryo selection and partner evaluation, ensuring consistency in clinical decisions.

To support clinical integration, future work should address issues such as longitudinal prediction (e.g., live birth and neonatal outcomes), real-time model deployment in electronic health records, and cross-population generalizability. Furthermore, attention to ethical considerations is essential: AI systems must be transparent, fair, and auditable, particularly in sensitive domains like fertility treatment. Informed consent, data privacy, and the mitigation of algorithmic bias across ethnicity, gender, and socioeconomic groups must remain central in deployment discussions.

In conclusion, the integration of deep learning into fertility care holds immense promise. However, it must proceed with careful validation, multidisciplinary collaboration, and ethical foresight to ensure equitable and trustworthy care for diverse populations.

References

Key International Studies (2017–2024)

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z.

Jindal, S. K., & Gupta, A. (2019). Artificial intelligence applications in reproductive medicine. *Journal of Human Reproductive Sciences*, 12(3), 177–182.

Kawwass, J. F., Crawford, S., Kissin, D. M., Jamieson, D. J., & Session, D. R. (2020). Trends and outcomes for assisted reproductive technology cycles in the United States. *Fertility and Sterility*, 113(3), 521–529.

Zhang, H., Nie, R., Wang, L., & Liu, Y. (2021). Artificial intelligence-assisted IVF outcome prediction: Current progress and future directions. *Reproductive Biology and Endocrinology*, 19(1), 156. https://doi.org/10.1186/s12958-021-00843-6.

Nigerian and African Context

Nwobodo, E. I., & Isah, A. Y. (2017). Semen quality in infertile Nigerian men: A prospective study. *Nigerian Journal of Clinical Practice*, 20(4), 475–479. https://doi.org/10.4103/1119-3077.187319.

Obiechina, N. J. A., & Okafor, C. N. (2018). Artificial intelligence in reproductive health: Implications for Nigeria. *International Journal of Medicine and Health Development*, 23(2), 98–104.

Oladokun, A., Sule-Odu, A., & Adeniji, A. (2020). Patterns of male factor infertility in a Nigerian tertiary hospital. *African Journal of Reproductive Health*, 24(2), 81–88. https://doi.org/10.29063/ajrh2020/v24i2.9.

Olaniyan, A. M., & Olowolafe, O. (2023). Challenges and opportunities in applying AI in sub-Saharan fertility care. *Nigerian Journal of Health Sciences and Innovation*, 5(1), 42–51.

WHO and Standard Reference Reports

European Society of Human Reproduction and Embryology (ESHRE). (2020). ESHRE guidelines for the evaluation of infertility. *Human Reproduction Open*, 2020(1), hoaa005.

Practice Committee of the American Society for Reproductive Medicine. (2021). Definitions of infertility and recurrent pregnancy loss: A committee opinion. *Fertility and Sterility*, 113(3), 533–535.

World Health Organization. (2021). *Laboratory manual for the examination and processing of human semen*, 6th Edition. Geneva: WHO. https://www.who.int/publications/i/item/9789240030787.