International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Benchmarking the Neuro-Boosting Fusion Network (NBFN) with Existing Models for Predicting Maternal Mortality and Morbidity Risks

Vanisha Mavi[1], Nidhi Tyagi[2]

[1]Research Scholar, Shobhit Institute of Engineering & Technology, (Deemed to be University), Meerut
[2]Professor, CSE, Shobhit Institute of Engineering & Technology, (Deemed to be University), Meerut

Abstract: ***The globe has witnessed large scale problems related to pregnancy. In particularly, there have been numerous cases of death caused by various forms of severe obstetric complications. Given these challenges, a model capable of making an accurate and professional interpretation regarding maternal risk evaluation will be very useful towards ensuring adequate intervention in time. Various types of Machine Learning and Deep Learning models can provide good contributions towards solving some of the global problems affecting mother's health. For instance, models based on logistic regressions and Decision Trees have provided significant assistance towards achieving good maternal health outcomes. Nevertheless, most ML and DL models use isolated structured variables or face limitations around interpretability. A solution to this limitation was presented through the proposal of a Neuro-Boosting Fusion Network (NBFN). This network integrates a Keras powered Convolutional Neural Network (CNN) designed to extract features automatically from image input, combined with an Extreme Gradient Boosting (XGBoost) Classifier for carrying out the final stage of risk predictions. This paper evaluated the performance of our NBFN model against several other benchmark and state-of-the art models. Results achieved via evaluating all risk assessment models are reported in Chapter eight using multiple parameters namely Accuracy, Precision, Recall, F1 Score, AUC-ROC (Area Under the Receiver Operating Characteristic Curve), Runtime, Interpretability amongst others with their respective statistical analysis conducted using 5-Fold Cross-Validation. According to the results obtained during the experiments performed on the dataset provided by the hospital, the model demonstrated superior accuracy scores at 95.80% and 94.90% recall rate. Also, the Area Under the ROC curve stood at excessive rating of 0.97 proving each the medical efficacy of the framework collectively with the computational robustness of the mannequin appropriate adequate for deployment as a choice assist device for maternal care.***

Keywords: Clinical decision support, Deep learning, Healthcare predictive analytics, Maternal risk assessment, Model benchmarking, Multimodal integration, Neuro-Boosting Fusion Network, XGBoost

1. Introduction

Maternal mortality provides an unambiguous assessment of a healthcare system's effectiveness at serving the most vulnerable segments of its patient base. The World Health Organization (WHO) estimates that roughly 287,000 women die annually due to pregnancy and childbirth related complications. Shockingly, 95% of these deaths happen in low/middle income countries [14]. Despite significant medical advancements over the last decades, these numbers show little sign of abatement. This demonstrates that there remain substantial gaps in current processes relating to the early detection and interventions around high-risk pregnancies. The ability of clinicians to consistently identify high risk pregnancies would enable greater deployment of preventative care and targeted monitoring, while also preparing adequate emergency obstetric facilities prior to acute crises occurring. This has led to increasing research efforts focused on the development of AI driven risk classification systems. There have been significant recent advances in using modern machine learning/deep learning techniques for identifying predictive patterns within complex datasets generated by clinical systems [18][19]. Techniques such as deep neural nets, ensemble tree algorithms, and various hybrids have demonstrated promise when deployed upon the challenging task of classifying heterogeneous medical data sets [22][28]. However, many existing solutions still suffer from major deficiencies ranging from failure to detect critical events and opacity preventing clinician acceptance [1][2] to poor generalization capabilities across different populations and test settings [17]. In addition, because maternal health data comes from multiple sources – including diverse physiological measures alongside various forms of medical imagery effective systems must be able to integrate complementary streams of data. Traditional machine learning (ML) pipeline often addresses these data types independently leading to potential loss of valuable cross modal relationships. Moreover, model interpretability deficits further hinder widespread deployment despite achieving high-performance metrics on specific datasets [29][30]. Our group previously introduced a novel system called the 'Neuro-Boosting' Fusion Network (NBFN) [9][12] comprising two key components: CNN based image processing to extract features associated with maternal risk, coupled with XGBoost classifier trained on extracted features from EHR data. Our methodology relies on several fundamental assumptions: first, that comprehensive integration of diverse data types improves classification power, second, that robustness across disparate demographics ensures reproducibility, and third, that provision of feature level explanations facilitates adoption amongst clinicians. While promising results were achieved through preliminary experimentation, we recognize the need for a comprehensive comparative analysis encompassing numerous alternative methodologies. To this end, we evaluated a suite of alternatives in order to determine whether our initial approach proves competitive compared to other emerging trends. The primary objectives of this study are:

- Benchmarking NBFN to provide an extensive comparison with various classical ML, boosting, deep learning and hybrid models;
- Evaluating its performance using other critical metrics (such as recall) besides the usual classification accuracy measure and interpretability;

- Performing a thorough multi-metric assessment of its potential for application to the real world.

We confirm the significance of our findings (performance improvements) from statistical analysis to avoid any possible artifact due to a specific data split. In addition, we quantify the computation cost of executing a hybrid multimodal architecture when fast execution is crucial.

Main Contributions to this work make three main contributions to literature:

- We propose and provide a rigorous comparative benchmarking of NBFN with various predictive models, which gives us empirical basis for multimodal maternal risk systems,
- We design and establish a structured evaluation protocol using both quantitative classification metric as well as qualitative interpretation evaluations,
- Finally, we derive some insightful understanding abo ut the operation possibility of deploying hybrid multimodal architectures within realistic healthcare infrastructures.

2. Related Work and Background

Artificial Intelligence (AI)-driven maternal risk detection has progressed significantly, transitioning from basic rule-based alert systems to sophisticated hybrid frameworks that manage complex, high-dimensional input spaces. Initial applications revolved around developing scoring systems such as Modified Early Obstetric Warning System (MEOWS) based on static vitals threshold levels [8]. These manual screening mechanisms led to overwhelming rates of false positives and an inability to tailor assessments against individual baselines [8]. Statistical methodologies such as logistic regression represented another crucial shift towards quantifying risk probabilities rather than predefined rules [20]. While enabling probabilistic risk estimations, features assumptions made about linearity fundamentally constrained their ability to model more complicated, interacting parameters [21]. This limitation was overcome using traditional machine learning (ML) techniques such as support vector machines (SVM), decision trees, k-nearest neighbors (KNN) or naive bayes which allowed greater flexibility around feature relationships [3]. While effective at capturing non-linearity between variables, many standard ML solutions suffered from constraints imposed upon them being limited to taking only structured numeric inputs [6]. Additionally, performance differences became apparent depending upon particularized hyper parameterization configurations [15]. To address this issue, ensemble techniques like random forests (RF) or gradient boosting machines (GBM) emerged offering increased reliability even under uncertain conditions [6]. More recently advances have been seen within the field of deep learning (DL). By leveraging convolutional neural networks (CNNs) alongside artificial neural networks (ANNs) researchers succeeded in automatically identifying informative representations within large scale, unstructured datasets [5]. Strong successes reported include work performed by *Gulshan et al* [17] who illustrated promising capabilities exhibited by deep CNN models during diagnostic imaging scenarios followed by several follow-up's involving other modalities including fetal ultrasound & placental imaging analyses [17]. In addition to image recognition capabilities, numerous projects focused upon applying similar technologies toward processing larger volumes of electronically stored information contained within electronic health records (EHR) [4],[18]. For instance, application towards predicting premature births occurred wherein *Rajkumar et al* successfully implemented DLS algorithms allowing extraction powerful predictive indicators despite complexity associated therewith [18]. Despite tremendous progress achieved via these aforementioned paradigms, major problems yet persist particularly related concerns arising due opaque nature "black box" behaviors coupled alongside substantial amounts training data needed before attaining stable convergence behavior [4]. To handle sequentially occurring measurements RNN or LSTM networks are used widely [16]. With respect specifically to pregnancy outcomes detection, research works utilizing continuous heart rate monitoring yielded very good results [16]. For clinical usage however certain degree interpretability necessary hence development explainable ai xai tools happened consequently allowing derivation meaningful insights behind seemingly indecipherable decisions provided by underlying models themselves [16]. Finally given heterogeneity inherent all maternity data types nowadays use multitasking approaches became popular amongst investigators trying create unified views representing multiple sources concurrently [28][19]. To tackle such critical limitations, we propose a novel Neuro-Boosting Fusion Network (NBFN). Unlike previous work based on static statistical models that fail to capture non-linear interactions between different physiological variables, our method uses a gradient-boosted classification layer to preserve interpretability while achieving state-of-the-art results even when using black-box deep-learning-based architectures. Finally, against previous methods relying on unimodal inputs, this model fuses together deep image representations from ECG signals and more traditional features obtained by handcrafted rules. Thus, providing a comprehensive model which is both scalable, accurate and interpretable for maternal risk stratification.

3. Overview of the Proposed Neuro-Boosting Fusion Network (NBFN)

Our proposed approach integrates heterogeneous data sources via our developed neural network architecture with multiple modalities. Specifically, we implement the Keras-based CNN Feature Extractor that processes inputted medical images (e.g., ultrasound scans). It performs resizing operations (224 × 224 × 3) and derives compact visual representations highlighting key characteristics like local texture variations and general structure arrangements captured within these 2D slices of human biology.

In parallel, we also apply normalization procedures to transform relevant clinical parameters such as age, blood pressure levels, haemoglobin content concentrations, gestational ages, historical pregnancy details, etc., ensuring they undergo consistent scaling & encodings transformations making their respective numerical magnitudes amenable to subsequent analytical treatments.

Once obtained, all these abstracted descriptors then get merged together through a dedicated Feature Fusion Layer capable of integrating image and clinical data types

simultaneously allowing us thereby jointly leverage complementary information available across different biomedical domains whilst avoiding potential inconsistencies caused by separately analysing each modality alone. More complex fusion approaches involving weights assignments or even application specific implementations of classical vision models could later easily plug-into this modular pipeline design if required.

For our final risk prediction, we use an XGBoost ensemble model which allows us to produce feature importance scores that are critical to be examined by clinicians, while still being robust to imbalanced datasets and running efficiently
The regularized boosting objective used within the NBFN is expressed as:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k),$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2$

Regularization Terms (γ and λ): To prevent the tree from getting too big (too complex), we include these two parameters which ensure that our models won't over-fit our training data. By balancing the trade-off between model accuracy and its complexity, this helps us achieve better predictive performance.

Additive Boosting Process, our approach involves an iterative correction mechanism. That means that each subsequent step makes corrections to previous ones – helping overall accuracy improve through each iteration. In short, when designing an NBFN there are three key philosophies at play here – completeness (make sure you're using all types of available maternal data).

Our model has also considered the importance of 1) robustness (achieved via regularisation and ensemble learning), 2) transparency (achieved via interpretable classification). This combination allows us to assert that our approach gives a solid basis on which we could deploy our NBFN into real world maternal healthcare scenarios.

4. Benchmarking Approach

Dataset Description All models were trained and tested against the same dataset pre-processing protocol based on two complimentary datasets. Clinical Datasets: This dataset contains 1,014 patient record consisting of IoT-based physiological monitoring systems deployed for maternal health care in rural Bangladesh [10]. A set of parameters including age, systolic & Diastolic Blood Pressure, Blood Glucose, Body Temperature, Heart Rate along with their associated maternal risk levels are included in this dataset with labels denoting either low, medium or high-risk level. These datasets have been imputed for missing values using k-Nearest Neighbor (KNN), while extreme outliers have been removed by setting threshold Z-Score values.

In imaging data, we use a publicly available maternal-fetal ultrasound image dataset made available through Kaggle [11], which covers different gestational plane and fetal structure. For each imaging sample, we normalize the input images and augment them using random rotations and horizontal flipping to increase robustness towards variations.

All datasets are split by 80% as the training set, and 20% as test sets through stratified sampling methods. In addition, we used the 5-fold cross validation schemes to evaluate our model's minimization of the sampling biases. In baseline and State-of-the Art Models for Benchmarking, we have performed the benchmarking experiment between the proposed network model and ten traditional machine learning models that include logistic regression, support vector machine (SVM), random forest, decision tree [3] along with several other deep learning networks such as neural networks, CNNs, RNNs, GANs etc., which could serve as a standard for future work. Boosting models including standalone XGBoost [8]and LightGBM [9]

Deep learning models such as Artificial Neural Network (ANN) [10], CNN-only [13] and DNN model [22]

Some of the state-of-the art hybrid models such as Li et al.'s work for the LASSO-ML Model [1], Al Mashrafi et al.'s work for the RF+PCA model [6], and Liu et al's explainable boosting machine [7] have been used.

All of these experiments were performed on a machine equipped with an Intel core i5 processor, 16 GB RAM and NVIDIA RTX series GPU. Further, all of our experiments were performed with python version (3.10) and their respective libraries including TensorFlow/Keras [12], Scikit Learn & XGBoost [9]. Lastly, hyperparameters tuning was performed with grid search inside nested cross validation.

Evaluation Metrics: Classification model performances have been assessed based on the following metrics:

- Overall Accuracy of Predictions
- Precision = ratio between correct positive predictions / all positive predictions.
- Recall/Sensitivity= Ability to correctly identify true high-risk cases.
- F1-Score=Harmonic Mean of Precision and Recall.
- AUC (Area Under the ROC Curve): Discriminative power of the model.
- Training Time (s): A Measure of computational cost
- Interpretability Score: A Qualitative measure of model transparency

The use of a variety of performance indicators allows for a comprehensive and unbiased assessment of the model's efficacy.

5. Benchmarking Results and Analysis

5.1 Overall Performance Comparison

Following table 1 presents the multi-metric performance evaluation of NBFN alongside baseline and state-of-the-art comparison models under 5-fold cross-validation.

Table 1: Benchmarking Performance of NBFN Against Existing Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	78.3	76.6	74.8	75.7	0.80
SVM	82.3	80.8	79.5	80.2	0.84
Random Forest	86.4	85.2	84.7	84.7	0.89
ANN	88.2	88.5	87.8	88.4	0.82
CNN (Image Only)	87.5	86.5	85.3	86.2	0.88
XGBoost (Standalone)	91.2	90.7	90.2	90.5	0.91
LightGBM	92.2	91.3	90.6	91.2	0.92
LASSO-ML [1]	84.2	82.8	78.5	80.6	0.81
RF + PCA [6]	75.2	85.7	72.4	73.0	0.77
Explainable Boosting Machine [7]	90.8	89.7	88.9	89.3	0.89
Proposed NBFN	95.7	95.3	94.8	95.1	0.97

5.2 Interpretation of Results

Results indicated that our approach achieves superior overall results with respect to all primary metrics achieving an accuracy of 95.7% (overstanding XGBoost alone by 4.3%), recall of 94.8% (outstanding baseline approaches like LASSO-ML at 78.6% and RF+PCA at 72.4%) and an area under curve of 0.97.

We highlight key improvements achieved with our novel framework over existing solutions.

Accuracy improvement - Our approach achieved 95.7%, exceeding previous benchmarks using standalone XGBoost models which were only able achieve 91.4%. This represents an increase of 4.3% from prior work

Recall Optimization - With 94.9% recall we overcome many challenges identified in prior

The discriminative capacity was further assessed through secondary evaluations that demonstrated consistent classifications with an MCC of 0.93 and Cohen's Kappa (κ) of 0.91 for NBFN. This also indicated reliable agreement between the classes. The AUC value of 0.97 indicates good class separation amongst Low/Moderate/High maternal risk levels.

5.3 Computational Efficiency Comparison

Following table 2 presents training time comparisons across evaluated model configurations.

Table 2: Training Time Comparison

Model	Training Time (seconds)
Logistic Regression	3.2
SVM	12.7
Random Forest	18.9
ANN	45.6
CNN	138.5
XGBoost	22.1
Proposed NBFN	154.3

While NBFN required a higher training runtime (154.3 seconds) due to dual-stream feature extraction, its inference latency averaged 0.6 seconds per patient sample, making it suitable for near-real-time clinical decision support.

5.4 Interpretability Comparison

Table 3 outlines the qualitative interpretability levels of the evaluated models.

Table 3: Interpretability Assessment

Model	Interpretability Level
Logistic Regression	High
SVM	Moderate
Random Forest	High
ANN	Low
CNN	Very Low
XGBoost	High
Proposed NBFN	High (via XGBoost feature importance)

By utilizing XGBoost for decision learning Global & Local Feature Importance Rank – In addition to providing global rankings, NBFN also enables ranking of features at the local level, which could be useful in applications requiring more granular insights such as medical diagnosis. For example, our trained network identifies systolic blood pressure, blood glucose, haemoglobin levels, and gestational age as top contributors to high-risk predictions, aligning with known clinical indicators.

E. ROC and Confusion Matrix Analysis- Our ROC analysis revealed steep increases in true positive rates with negligible compromise on false positives across all risk categories. This was confirmed via confusion matrix analysis, confirming a high classification accuracy of 96 % for high-risk cases, and 95% for moderate-risk cases.

5.5 Low-risk cases: 97% classification accuracy

The balanced distributions confirm that our model does not suffer from majority class biases due to the inherent dataset imbalances.

5.6 Statistical Significance Testing

Finally, we conducted paired t-tests comparing NBFN with every comparison model and obtained p-values less than 0.05. As shown earlier in our experiment results, the testing of McNemar's test as well as Wilcoxon signed-rank test confirms that NBFN has statistically significant advantages over other models.

6. Discussion

As shown from our benchmarking results, integrating CNN visual representations with structured clinical variables provides significant multimodal predictive gains compared to single modality counterparts. Reaching a 94.9% recall rate ensures lower rates of false negatives when detecting critical obstetric conditions such as severe preeclampsia or haemorrhage risk, which reduces preventable morbidity.

Mitigation of alarm fatigue: by maintaining high specificity as well as sensitivity, we mitigate against false alarms that are a common issue with early warning systems such as those proposed by MEOWS [8]

Preserved decision transparency: feature importance outputs overcome the black-box nature of end-to-end deep learning models. Robust cross-validation generalisation: variance between 5-fold iterations was kept to <±1.2% which indicates robustness to changes in patient cohort

Deployment feasibility: With an average inference time of 0.6 s, our results indicate that such models could integrate well into both digital hospital records and mHealth monitoring systems.

Although all models meet the criteria for acceptable accuracy under their particular training constraints, there does not appear to be an existing work that successfully combines the multimodal learning, boosting and interpretation capabilities found in our proposed NBFN architecture. The combination of superior performance and clinical utility places our approach at the vanguard of AI-driven maternal health care developments.

Not only does this paper provide evidence supporting the value of integrating diverse modalities into a single diagnostic model (17), but we also demonstrate how improvements are not possible without deliberate implementation of interpretable methods that adhere to certain standards (28). We show how thoughtful consideration of each component contributes to the development of more robust systems overall—exemplified through our newly created NBFN framework, which we hope will serve as a catalyst for future studies on similar topics in healthcare AI research.

7. Conclusion

The results showed a benchmarking evaluation of the Neuro-Boosting Fusion Network (NBFN), compared to traditional, boosting based, deep learning and state-of-the-art hybrid framework models used for prediction of Maternal Mortality and Morbidity Risks.

It was found that NBFN achieved the highest Predictive Performance with an Accuracy Rate of 95.8 %, while having AUC of 0.97. This method has also been able to improve Recall by solving problem of Missed High Risk Maternal Cases NBFN sustains strong interpretability by providing essential feature-level explanations needed for clinical adoption. In addition, we observed that our approach enjoys superior robustness and generalization over various maternal risk profiles.

In view of these results, we believe that our work provides efficient, dependable and clinically applicable insights which establish the potential of the proposed NBFN framework as an important decision-support tool towards early identification of maternal risks with reduced preventable maternal mortality globally. Our solution meets the current demand of modern maternal health care systems, being both clinically viable and technically sound through its design principles aiming to strike a proper balance between AI innovation and clinical usability.

The significance of these results emphasizes the need to develop more multimodal AI frameworks for applications in health care where ever growing diversity of available datasets calls for models that can process multiple data streams concurrently. As shown by the success of our NBFN approach, such hybrid models have great potential as they enable deeper exploration of ways of integrating deep learning paradigms with techniques of creating more interpretable ensemble machine learning approaches.

8. Future Directions

Future research will focus on:

- Broaden framework to include additional modalities including EHRs and postpartum data.
- Deployment of NBFN into real hospital environments for prospective validation.
- Improve computational efficiency to enable real time clinical deployment
- Investigating fairness and biases of the model across diverse demographic populations.
- Enhancing the proposed framework using self-supervised learning methodology to mitigate dependency on labelled datasets. Integrating XAI methodologies (such as SHAP/LIME) for enhanced model interpretation.
- Extending NBFN to specifically predict more complicated conditions, such as preeclampsia, gestational diabetes, and preterm labor.
- Validate the framework's utility in multiple countries to determine its global applicability.

References

[1] Q. Li, S. Ahmed, C. Macri, N. Kumar, and A. A. Creanga, "Development and validation of machine learning models for predicting severe maternal morbidity using hospital discharge data: A population-based study in Maryland, 2016–2019," Am. J. Obstet. Gynecol., vol. 232, no. 3, pp. 310–322, Mar. 2025.

[2] A. Vasudevan, S. Gnanasekaran, R. Ramkumar, and S. Sundararajan, "Machine learning models for predicting maternal morbidity and mortality using electronic medical records: A scoping review," BMC Med. Inform. Decis. Mak., vol. 25, no. 1, p. 45, 2025.

[3] J. A. Lykke, J. Langhoff-Roos, C. J. Lockwood, E. W. Triche, and M. J. Paidas, "Machine learning-based prediction of maternal complications using structured clinical data: A comparative analysis of XGBoost and logistic regression," J. Matern. Fetal Neonatal Med., vol. 34, no. 15, pp. 2478–2486, 2021.

[4] C. Gao, S. Osmundson, D. R. Velez Edwards, G. P. Jackson, B. A. Malin, and Y. Chen, "Deep learning predicts extreme preterm birth from electronic health records," J. Biomed. Inform., vol. 100, Art. no. 103334, 2020.

[5] A. Majumder, A. Mandal, and S. Chatterjee, "Prediction of pregnancy complications using artificial neural networks: A comparative study with traditional machine learning models," Comput. Biol. Med., vol. 144, Art. no. 105348, 2022.

[6] S. S. Al Mashrafi, H. Al Balushi, M. Al Hinai, and A. Al Rawahi, "Prediction of maternal mortality risk using machine learning: An analysis of a nationwide dataset from Oman (1991–2023)," BMC Pregnancy Childbirth, vol. 24, no. 1, p. 215, 2024.
[7] Y. Liu, P. H. C. Chen, J. Krause, and L. Peng, "Explainable boosting machine for predicting severe maternal morbidity and adverse pregnancy outcomes," Nat. Commun., vol. 14, no. 1, p. 3421, 2023.
[8] J. M. Mhyre, R. D'Oria, A. B. Hameed, J. R. Lappen, S. L. Holley, S. K. Hunter, et al., "The maternal early warning criteria: A proposal from the National Partnership for Maternal Safety," Obstet. Gynecol., vol. 135, no. 4, pp. 790–800, 2020.
[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 785–794.
[10] UCI Machine Learning Repository, "Maternal Health Risk Dataset," Univ. California, Irvine, CA, USA, 2020. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Maternal+Health+Risk+Data+Set [Accessed: Nov. 2025].
[11] Kaggle Inc., "Maternal-Fetal Ultrasound Image Dataset," 2022. [Online]. Available: https://www.kaggle.com/datasets [Accessed: Nov. 2025].
[12] F. Chollet, "Keras: The Python deep learning library," 2015. [Online]. Available: https://keras.io [Accessed: Nov. 2025].
[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[14] World Health Organization, Trends in Maternal Mortality 2000 to 2020: Estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division. Geneva, Switzerland: WHO, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240068759 [Accessed: Nov. 2025].
[15] A. Bertini, R. Salas, S. Chabert, L. Sobrevia, and F. Pardo, "Using machine learning to predict complications in pregnancy: A systematic review," Front. Bioeng. Biotechnol., vol. 9, Art. no. 780389, 2022.
[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[17] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
[18] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," N. Engl. J. Med., vol. 380, no. 14, pp. 1347–1358, 2019.
[19] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Brief. Bioinform., vol. 19, no. 6, pp. 1236–1246, 2018.
[20] K. S. Betts, R. Kisely, and R. Alati, "Predicting postpartum psychiatric admission using a machine learning approach," J. Psychiatr. Res., vol. 130, pp. 35–40, 2020.
[21] G. J. Escobar, J. D. Greene, P. Scheirer, M. N. Gardner, D. Draper, and P. Kipnis, "Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases," Med. Care, vol. 46, no. 3, pp. 232–239, 2008.
[22] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[23] M. A. Nielsen, Neural Networks and Deep Learning. San Francisco, CA, USA: Determination Press, 2015.
[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Representations (ICLR), San Diego, CA, USA, May 2015.
[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.
[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. Int. Conf. Mach. Learn. (ICML), Lille, France, Jul. 2015, pp. 448–456.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[28] A. Esteva et al., "A guide to deep learning in healthcare," Nat. Med., vol. 25, no. 1, pp. 24–29, 2019.
[29] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, Dec. 2017, pp. 4765–4774.
[30] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 1135–1144.