Automated Prediction of Preeclampsia Using Artificial Intelligence

FRANK EDUGHOM EKPAR
Founder, Scholars University Ltd
Associate Professor, Department of Computer Engineering, Rivers State University
Principal Investigator, Department of Computer Engineering, Topfaith University

Abstract- This paper presents a system that builds artificial intelligence models for the automated prediction of the likelihood of occurrence of preeclampsia in pregnant women based on a suite of clinical measurements during the course of the pregnancy including proteinuria, amniotic fluid levels, fetal weight, gravida, parity, body mass index, systolic blood pressure, diastolic blood pressure, gestational age, presence or absence of diabetes, hemoglobin, history of hypertension and the age of the pregnant woman. The system is trained on publicly accessible preeclampsia datasets that could be augmented with locally sourced data for mitigation of bias, balance and robustness. The trained artificial intelligence models could be fine-tuned and integrated into preeclampsia prediction modules in a comprehensive artificial intelligence-driven healthcare system, saving lives and improving outcomes in pregnancy.

Indexed Terms- Preeclampsia, Automated Disease Prediction, Artificial Intelligence (AI), Deep Learning (DL), Artificial Neural Network (ANN), TensorFlow, Healthcare System.

I. INTRODUCTION

Preeclampsia is a health condition that affects a relatively small fraction of pregnant women and can sometimes present soon after delivery of the baby and is one of the leading causes of maternal morbidity. In some cases, and especially if not treated, it can lead to severe complications and adverse outcomes for both the pregnant woman and fetus [1] – [2].

As with many other health conditions, the burden of preeclampsia has the greatest negative impact on low- and middle-income countries (LMICs) with their often inadequate and severely resource-limited healthcare systems.

The prediction of the risk of preeclampsia in pregnancy and attendant effective treatment could lead to significantly improved outcomes. The earlier and more reliably this prediction of risk could be made, the more effective the resulting interventions could be.

Numerous studies have highlighted the utilization of artificial intelligence and machine learning systems to the prediction, detection or diagnosis of myriads of diseases with a tilt in focus towards data gathered in developed countries [3] – [21].

In this study, artificial intelligence models are constructed and trained, tested and validated on clinical measurements comprising diagnostic measurements and biomarkers such as proteinuria, amniotic fluid levels, gravida, parity, body mass index, systolic blood pressure, diastolic blood pressure, gestational age, presence or absence of diabetes, history of hypertension, fetal weight, and so on, coupled with the associated risk of occurrence of preeclampsia (low, medium, high) with a view to automatically generating early predictions of the likelihood of preeclampsia given a relevant set of clinical measurements.

With refinement aimed at improving performance and robustness, including the integration of data acquired through local data gathering drives, the system developed here could be incorporated into Scholar Medic, the comprehensive artificial intelligence-driven healthcare system created by Ekpar [22] – [25] with a modular architecture to accommodate a wide range of health conditions and permit efficient improvements to existing modules on the basis of fresh data as well as other unique features including the ability to harness novel three-dimensional multilayer electroencephalography (Ekpar EEG) systems [26] – [28] and permit adaptations of traditional EEEG systems to the innovative, advanced three-dimensional multilayer EEG (Ekpar EEG) paradigm for game-changing applications in a wide range of domains.

II. MATERIALS AND METHODS

Participant Recruitment

Individuals willingly chose to participate in the research studies that helped develop the AI-driven healthcare system. All participants gave informed consent before engaging in the studies.

Ethical Approval

The studies were approved by the Health Research Ethics Committee of the Rivers State University Teaching Hospital at Rivers State University. They followed all relevant ethical and regulatory guidelines. Publicly accessible data were used in accordance with the licensing terms set by their respective providers.

III. METHODOLOGY

Healthcare datasets available in the public domain can be improved by integrating data from local experiments and data collection initiatives. This combined dataset can then be used to train AI models that provide actionable predictions based on new input. Examples of such publicly accessible healthcare datasets include those from the Centers for Disease Control, the University of California Irvine Machine Learning Repository, the American Epilepsy Society, and Kaggle.

Incorporating local data enhances model accuracy, reduces potential biases, and promotes inclusivity and global applicability. A key aspect of this project is combining diagnostic measurements, such as electrocardiographic data, from local experiments with EEG data, where appropriate, including both traditional and advanced three-dimensional multilayer EEG systems.

For data collection, ethical approval has been obtained from the appropriate research ethics committees in the regions where the experiments take place. The project has also partnered with licensed medical professionals who have direct access to patients and clinical teams. These medical experts are providing anonymized clinical data for validating the AI models.

After the models are trained, they will be integrated into a comprehensive healthcare system aimed at offering clinical decision support to healthcare providers and facilitating brain-computer interfaces (BCIs).

The system will deliver actionable insights and predictions based on new clinical data, helping with the early detection, diagnosis, treatment, prediction, and prevention of various conditions, including preeclampsia, tuberculosis, chronic kidney disease, liver disease, diabetes, heart disease, stroke, autism, and epilepsy.

This project is committed to promoting open science, reproducibility, and collaboration, with all generated data made publicly available on platforms like GitHub.

IV. SYSTEM DESIGN AND IMPLEMENTATION

This paper presents a healthcare system with a modular design, where each health condition (e.g., chronic kidney disease, heart disease, liver disease, stroke, epilepsy, autism, etc.) is assigned to a specific module. This structure not only enables flexibility in diagnosing and predicting future conditions but also allows for easy updates to modules through the integration of new data. Additionally, Brain-Computer Interface (BCI) modules, such as those using the motor imagery paradigm, can analyze EEG data to generate actionable commands and appropriate responses.

The system also outlines a process for upgrading traditional EEG systems to advanced three-dimensional, multilayer EEG systems. These systems, developed by Ekpar and called Ekpar EEG systems [26] - [28], are based on a conceptual framework that uses approximations of carefully selected bio-signal features to model or influence the biological systems involved.

For each module, AI models are developed and trained with appropriately structured data, as described in the paper. These models take into account factors such as genetic, environmental, and lifestyle information to provide more precise assessments of the participants' conditions.

Figure 1 offers a conceptual overview of the system's design.

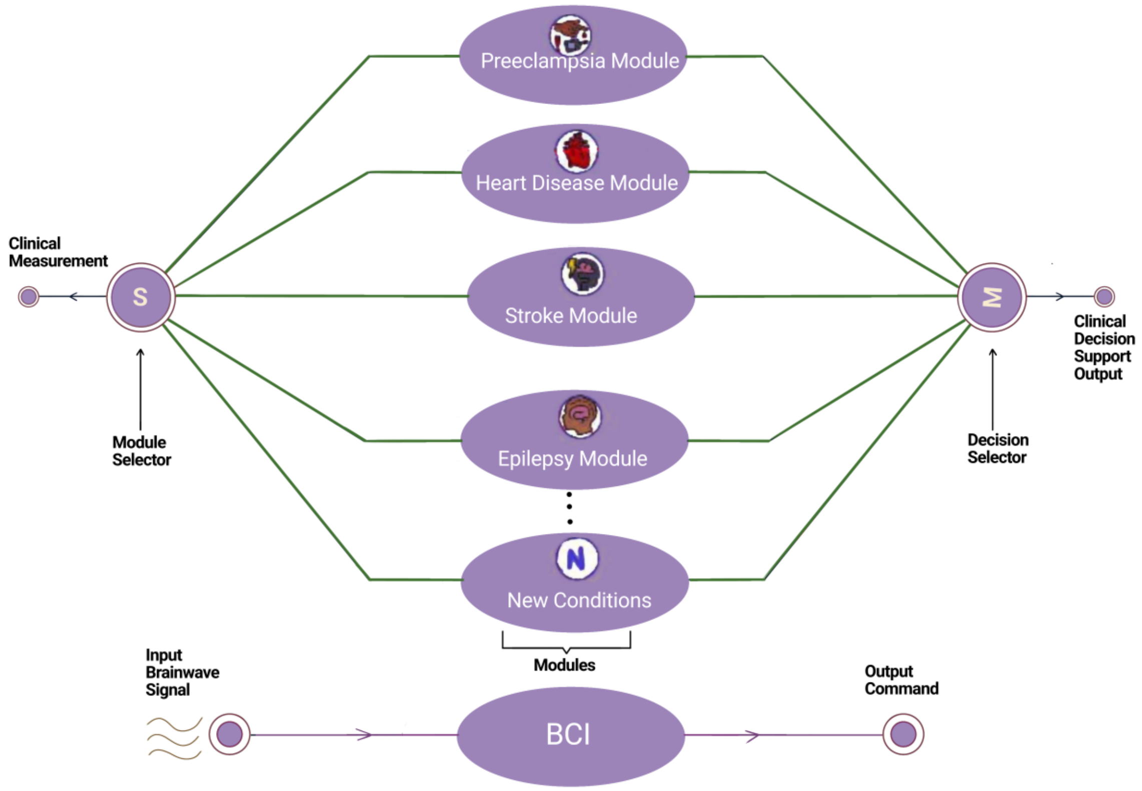

Fig. 1: System Schematic Design Diagram for the Comprehensive AI-Driven Healthcare Solution and Brain Computer Interface System. The New Conditions component represents additional health conditions that can be incorporated into the solution via new modules.

The AI models are developed using four different approaches, outlined below:

1. Direct Use of LLMs: Large Language Models (LLMs), such as GPT-4, serve as inference engines that process collected data formatted into multidimensional input vectors. Fine-tuning the LLMs may also be part of the process.
2. Prompt Engineering with LLMs: LLMs like Bard and GPT-4 (including their future versions) are used in prompt engineering to define the steps necessary for developing the AI system. Developers then apply their expertise in AI, neural networks, deep learning, and tools such as Python, TensorFlow, Keras, Scikit-learn, and Matplotlib to implement these steps.
3. Automated Model Generation: Specific AI models are generated automatically through a pipeline leveraging the capabilities of LLMs, such as Bard and GPT-4 (and future versions).
4. Manual AI Architecture Design: The AI architecture is designed directly based on the developer's in-depth knowledge of AI, neural networks, deep learning, and programming frameworks like Python, TensorFlow, Keras, Scikit-learn, and Matplotlib.

All tools and processes involved in the development are thoroughly documented to ensure easy transfer and reuse of the solution.

The resulting AI models are evaluated using performance metrics like accuracy, specificity, and sensitivity to determine how well they solve the intended problems.

V. AUTOMATED PREECLAMPSIA PREDICTION MODULE

The fourth approach is adopted from the list of approaches listed earlier and involves building custom AI models based on AI expertise and experimentation.

In order to actually build the AI models or artificial neural networks (ANNs), the characteristics of the dataset need to be considered.

VI. DATASET

A publicly accessible preeclampsia dataset was obtained from the Kaggle dataset repository.

The dataset comprises a total of 203 rows of data representing clinical measurements captured from 203 individual participants or pregnant women.

Diagnostic measurements for each pregnant woman and these measurements are the basis for the columns in each row of data. There are 14 columns with the first 13 columns representing factors such as gravida (number of pregnancies), parity (number of pregnancies carried past the threshold of viability), gestational age, age of participant, body mass index, presence or absence of diabetes, history of hypertension, systolic blood pressure, diastolic blood pressure, hemoglobin, fetal weight, proteinuria and amniotic fluid levels while the last and fourteenth column represents the level of preeclampsia risk – low, medium or high.

One-hot encoding was utilized for the outputs (found in the last or fourteenth column of the original dataset indicating the risk of preeclampsia), assigning codes for each of the three (3) output classes representing the risk or likelihood of the occurrence of preeclampsia, namely, LOW, MEDIUM, HIGH, as shown in Table 1.

Table 1: One-hot Encoding for Three Output Classes (Risk or Likelihood of Preeclampsia)

OUTPUT CLASS	ENCODING
LOW	0 0 1
MEDIUM	0 1 0
HIGH	1 0 0

VII. DATA AVAILABILITY

The preeclampsia dataset harnessed in this study is available from Kaggle at https://www.kaggle.com/datasets/muhammadasifwazir/preeclampsia-in-pregnant-women-dataset.

VIII. ARTIFICIAL NEURAL NETWORK (ANN)

The artificial neural network designed in this study had 13 input units for the 13 clinical measurements, three hidden layers with 64, 32 and 32 rectified linear (ReLU) units respectively while the output layer had 3 units with sigmoid activation.

Figure 2 graphically illustrates the generalized representation of the artificial neural network. In Fig. 2, CM_1, CM_2, ..., CM_N are inputs and stand in for clinical measurements such as the aforementioned 13 factors and biomarkers, yielding N=13 as the total number of inputs or clinical measurements while CD stands in for the clinical diagnosis and is the output of the artificial neural network comprising 3 output units.

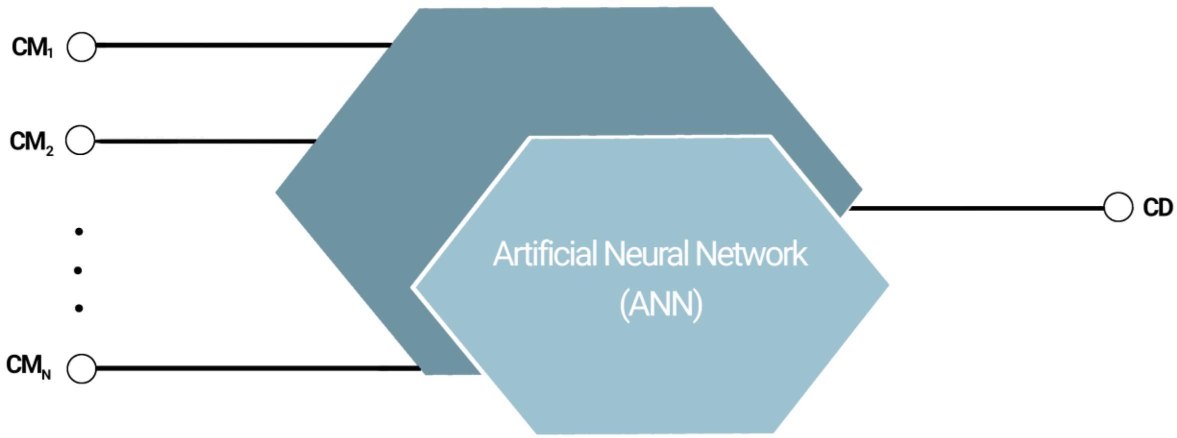

Fig. 2: Schematic Graphical Representation of Artificial Neural Network (ANN) Architecture. CM_1, CM_2, ..., CM_N represent the inputs while CD represents the output indicating the suggested clinical diagnosis.

IX. RESULTS

Preprocessing involved shuffling the data randomly to mitigate bias. The data was split into training and testing/validation sets with the training dataset taking up 80% of the original dataset while the testing/validation dataset took up 20% of the original dataset. The artificial neural network was implemented by leveraging the TensorFlow platform and the Keras Application Programming Interface (API) in the Python programming language [29] – [30].

The neural network was trained over 300 epochs using categorical cross entropy loss function and the Adam Optimizer [31] – [32] with the default batch size and learning rate.

An accuracy of approximately 95% was achieved for the training and validation datasets.

Figure 3 demonstrates the history of the accuracy and loss performance metrics over the training epochs.

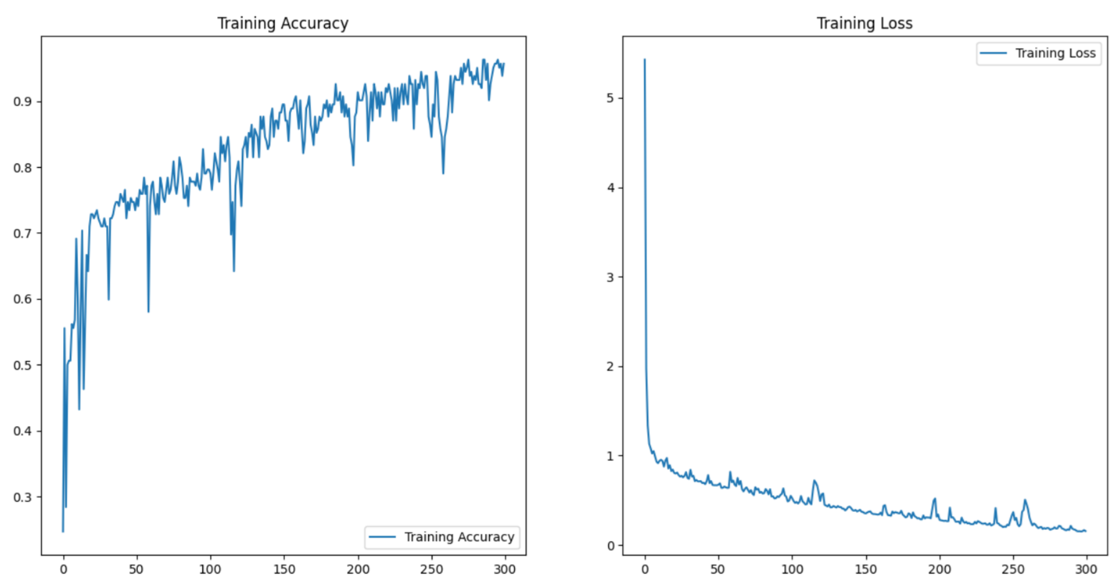

Fig 3: History of training accuracy and loss.

CONCLUSION

Suitably complex artificial intelligence models were constructed and then trained, tested and validated on preeclampsia datasets to automatically predict the likelihood of the presentation of preeclampsia during pregnancy by monitoring diagnostic measurements such as proteinuria, amniotic fluid levels, hemoglobin, gravida, parity, body mass index, systolic blood pressure, diastolic blood pressure, gestational age, presence or absence of diabetes, history of hypertension and fetal weight. Refinement of the resulting artificial intelligence models could permit their inclusion in a comprehensive artificial intelligence-driven healthcare system as preeclampsia prediction modules. This system could enable accurate early prediction of preeclampsia and lead to life-saving interventions and improved health outcomes for pregnant women.

REFERENCES

[1] World Health Organization (WHO) – Recommendations for Prevention and Treatment of Pre-eclampsia and Eclampsia: https://www.who.int/publications/i/item/9789241548335. Retrieved (2025).

[2] NHS – Pre-eclampsia: https://www.nhs.uk/conditions/pre-eclampsia/. Retrieved (2025).

[3] Nomura, A., Noguchi, M., Kometani, M., Furukawa, K., Yoneda, T. Artificial Intelligence in Current Diabetes Management and Prediction, *Curr Diab Rep.* 21(12):61 (2021).

[4] Kumar, Y., Koul, A., Singla, R., Ijaz, M. F. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda, *Journal of Ambient Intelligence and Humanized Computing* 14:8459–8486 (2023).

[5] Ansari, S., Shafi, I., Ansari, A., Ahmad, J., Shah, S. I. Diagnosis of liver disease induced by hepatitis virus using artificial neural network, *IEEE Int Multitopic.* https://doi.org/10.1109/INMIC.2011.6151515 (2011).

[6] Battineni, G., Sagaro, G. G., Chinatalapudi, N., Amenta, F. Applications of machine learning predictive models in the chronic disease diagnosis, *J Personal Med.* https://doi.org/10.3390/jpm10020021 (2020).

[7] Abdar, M., Yen, N., Hung, J. Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision tree, *J Med Biol Eng* 38:953–965 (2018).

[8] Chaikijurajai, T., Laffin, L., Tang, W. Artificial intelligence and hypertension: recent advances and future outlook, *Am J Hypertens* 33:967–974 (2020).

[9] Fujita, S., Hagiwara, A., Otsuka, Y., Hori, M., Takei, N., Hwang, K. P., Irie, R., Andica, C., Kamagata, K., Akashi, T., Kumamaru, K. K., Suzuki, M., Wada, A., Abe, O., Aoki, S. Deep Learning Approach for Generating MRA Images From 3D Quantitative Synthetic MRI Without Additional Scans, *Invest Radiol* 55:249–256 (2020).

[10] Juarez-Chambi, R. M., Kut, C., Rico-Jimenez, J. J., Chaichana, L. K., Xi, J., Campos-Delgado, D. U., Rodriguez, F. J., Quinones-Hinojosa, A., Li, X., Jo, J. A. AI-Assisted *In Situ* Detection of Human Glioma Infiltration Using a Novel Computational Method for Optical Coherence Tomography, *Clin Cancer Res* 25(21):6329–6338 (2019).

[11] Nashif, S., Raihan, R., Islam, R., Imam, M. H. Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System, *World Journal of Engineering and Technology* Vol 6, No. 4 (2018).

[12] Chen, P. H. C., Gadepalli, K., MacDonald, R., Liu, Kadowaki, S., Nagpal, K., Kohlberger, T., Dean, J., Corrado, G. S., Hipp, J. D., Mermel, C. H., Stumpe, M. C. An augmented reality microscope with real time artificial intelligence integration for cancer diagnosis, *Nat Med* 25:1453–1457 (2019).

[13] Gouda, W., Yasin, R. COVID-19 disease: CT Pneumonia Analysis prototype by using artificial intelligence, predicting the disease severity, *Egypt J Radiol Nucl Med* 51(1):196 (2020).

[14] Han, Y., Han, Z., Wu, J., Yu, Y., Gao, S., Hua, D., Yang, A. Artificial Intelligence

Recommendation System of Cancer Rehabilitation Scheme Based on IoT Technology, *IEEE Access* 8:44924–44935 (2020).

[15] Chui, C. S., Lee, N. P., Adeoye, J., Thomson, P., Choi, S. W. Machine learning and treatment outcome prediction for oral cancer, *J Oral Pathol Med* 49(10):977–985 (2020).

[16] Koshimizu, H., Kojima, R., Okuno, Y. Future possibilities for artificial intelligence in the practical management of hypertension, *Hypertens Res* 43:1327–1337 (2020).

[17] Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., Marx, A., Boor, P., Tacke, F., Neumann, U. P., Grabsch, H. I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., Luedde, T. Deep learning microsatellite instability directly from histology in gastrointestinal cancer, *Nat Med* 25:1054–1056 (2019).

[18] Kwon, J. M., Jeon, K. H., Kim, H. M., Kim, M. J., Lim, S. M., Kim, K. H., Song, P. S., Park, J., Choi, R. K., Oh, B. H. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography, *EP Europace* 22(3):412–419 (2020).

[19] Khan, M. A. An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier, *IEEE Access* 8:34717–34727 (2020).

[20] Oikonomou, E. K., Williams, M. C., Kotanidis, C. P., Desai, M. Y., Marwan, M., Antonopoulos, A. S., Thomas, K. E., Thomas, S., Akoumianakis, I., Fan, L. M., Kesavan, S., Herdman, L., Alashi, A., Centeno, E. H., Lyasheva, M., Griffin, B. P., Flamm, S. D., Shirodaria, C. Sabharwal, N., Kelion, A., Dweck, M. R., Van Beek, E. J. R., Deanfield, J., Hopewell, J. C., Neubauer, S., Channon, K. M., Achenbach, S., Newby, D. E., Antoniades, C. A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography, *Eur Heart J* 40(43):3529–3543 (2019).

[21] Sabottke, C. F., Spieler, B. M. The Effect of Image Resolution on Deep Learning in Radiography, *Radiology: Artificial Intelligence* Vol. 2. No. 1, 2:e190015 (2020).

[22] Ekpar, F. E. A Comprehensive Artificial Intelligence-Driven Healthcare System, European Journal of Electrical Engineering and Computer Science, 8(3), Article 617. (2024).

[23] Ekpar, F. E. Diagnosis of Chronic Kidney Disease Within a Comprehensive Artificial Intelligence-Driven Healthcare System, *International Journal of Advanced Research in Computer and Communication Engineering*, 13(9). (2024).

[24] Ekpar, F. E. Image-based Chronic Disease Diagnosis Using 2D Convolutional Neural Networks in the Context of a Comprehensive Artificial Intelligence-Driven Healthcare System, *Molecular Sciences and Applications*, 4(13). (2024).

[25] Ekpar, F. E. Leveraging Generative Artificial Intelligence Recommendations for Image-based Chronic Kidney Disease Diagnosis, *International Journal of Advanced Research in Computer and Communication Engineering*, 14(1). (2025).

[26] Ekpar, F. E. A Novel Three-dimensional Multilayer Electroencephalography Paradigm, *Fortune Journal of Health Sciences,* 7(3). (2024).

[27] Ekpar, F. E. System for Nature-Inspired Signal Processing: Principles and Practice, *European Journal of Electrical Engineering and Computer Science*, 3(6), pp. 1-10, (2019).

[28] Ekpar, F. E. Nature-inspired Signal Processing, *United States Patent and Trademark Office*, US Patent Application Number: 13/674,035 (Filed: November 11, 2012, Priority Date: December 24, 2011), Document ID: US 20140135642 A1: Published: (2014).

[29] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X. TensorFlow: A System for Large Scale Machine Learning, *Proceedings of the 12th USENIX Symposium on*

Operating Systems Design and Implementation (OSDI '16). (2016).

[30] Pang, B., Nijkamp, E., Wu, Y. N. Deep Learning with TensorFlow: A Review, *Journal of Educational and Behavioral Statistics*. Vol. 45, Iss. 2. (2019).

[31] Kingma, D. P., Ba, J. L. Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations (ICLR)* (2015).

[32] Zhang, Z. Improved Adam Optimizer for Deep Neural Networks, *IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (2018).