IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13802

Diagnosis of Diabetes Within a Comprehensive Artificial Intelligence-Driven Healthcare System

Frank Edughom Ekpar[1,2,3]

Founder, Scholars University Ltd, Port Harcourt, Nigeria[1]

Associate Professor, Department of Computer Engineering, Rivers State University, Port Harcourt, Nigeria[2]

Principal Investigator, Department of Computer Engineering, Topfaith University, Mkpatak, Nigeria[3]

Abstract: Diabetes affects millions of people around the world with a significant economic, physical, social and psychological burden leading to avoidable suffering, disability and sometimes mortality. This significant burden is exacerbated in less developed countries by the chronic and seemingly intractable dearth of qualified medical practitioners. Given accurate diagnoses and predictions, efficacious life-saving therapies could be developed. This paper presents a module for the diagnosis of diabetes within the framework of a modular, extensible and comprehensive artificial intelligence-driven healthcare system.

Keywords: Diabetes, Artificial Intelligence, Deep Learning, Healthcare System.

I. INTRODUCTION

According to the World Health Organization (WHO), millions of people worldwide are affected by diabetes [1] which exacts a significant toll in terms of physical and psychological suffering as well as negative economic impact. Furthermore, the number of patients has been on the increase in low- and middle-income countries (LMICs) based on the latest surveys.

Apart from a perfect storm of circumstances leading to an acute shortage of qualified medical practitioners including low levels of funding, weak educational institutions and relatively low patronage, LMICs are currently experiencing a protracted brain drain in the medical and healthcare professions. Many doctors trained in LMICs prefer to practice in developed countries for higher remuneration and improved working conditions. This threatens to reduce the already precipitously low doctor to patient ratios to unacceptable levels.

Ekpar [2] introduced a comprehensive artificial intelligence (AI)-powered healthcare system that addresses the aforementioned issues and helps medical doctors save lives and improve living conditions and enables orders of magnitude improvements in medical doctor productivity and dramatic reductions in healthcare professional burn-out and staff shortages. The system features a modular design and permits the addition of new modules as well as the fine-tuning of existing modules to extract improved performance.

Studies have been carried out involving a wide variety of approaches to the computer-aided diagnosis of diseases including the use of AI and related algorithms and systems with a variety of merits and demerits [3] – [21].

This paper presents the Diabetes Diagnosis Module of the comprehensive artificial intelligence-driven healthcare system [2] – highlighting the adoption of the approach of synthesizing custom adapted artificial neural networks (ANNs) for training, testing, validation and inference in the diagnosis of diabetes mellitus.

II. MATERIALS AND METHODS

Participant Enrollment

Individuals volunteered to participate in the studies aimed at developing the comprehensive AI-driven healthcare system, with each participant providing informed consent for their involvement.

Ethical Oversight

The Health Research Ethics Committee of the Institute of Biomedical Research at the University of Uyo granted ethical clearance and approval for the studies. All research activities adhered strictly to applicable ethical and regulatory standards. Publicly available data were used in accordance with the licensing terms set forth by their creators.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13802

III. METHODOLOGY

Enhancing publicly accessible healthcare datasets involves integrating locally gathered experimental and collected data into AI models to generate actionable predictions from new data. Public sources of healthcare datasets include the Centers for Disease Control, University of California Irvine Machine Learning Repository, American Epilepsy Society, and Kaggle.

Incorporating local data ensures resilience, mitigates bias, and promotes inclusivity and global applicability. A distinctive approach in this project involves combining diagnostic measurements (such as electrocardiographic results) from local experiments with EEG data, including novel three-dimensional multilayer EEG systems.

Ethical clearance for local data acquisition has been obtained from relevant research ethics committees overseeing the geographic regions of experimentation. Collaboration with licensed medical doctors facilitates access to patients and clinicians, enabling the provision of anonymized clinical data for validating AI models.

Trained AI models are integrated into a comprehensive healthcare system designed to provide clinical decision support and Brain-Computer Interfaces (BCIs). These models offer actionable insights derived from new clinical data, aiding in early detection, diagnosis, treatment, prediction, and prevention of various conditions like diabetes, cardiovascular diseases, stroke, autism, and epilepsy.

To uphold principles of open science, reproducibility, and collaboration, all generated data will be publicly uploaded to repositories such as GitHub.

System Design and Implementation

The healthcare system described employs a modular architecture, where each module corresponds to a specific condition (for example diabetes mellitus, heart disease, stroke, epilepsy, autism, among others), ensuring adaptability for future conditions and efficient updates with fresh data. Modules designed for BCIs, such as those leveraging the motor imagery paradigm, can utilize EEG data to generate actionable commands and responses.

The system includes guidelines for adapting conventional EEG systems to novel three-dimensional multilayer EEG systems, based on a framework developed by Ekpar [22] – [23]. These systems utilize approximations of representative bio-signal features for characterizing and manipulating biological systems.

Each module incorporates robust AI models trained on appropriately formatted aggregated data, potentially integrating genetic, environmental, lifestyle, and other relevant factors to enhance accuracy in participant circumstances representation.

A schematic representation of the design and functional relationships of the key components of the system appears in Fig. 1.

 This work is licensed under a Creative Commons Attribution 4.0 International License

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024
DOI: 10.17148/IJARCCE.2024.13802

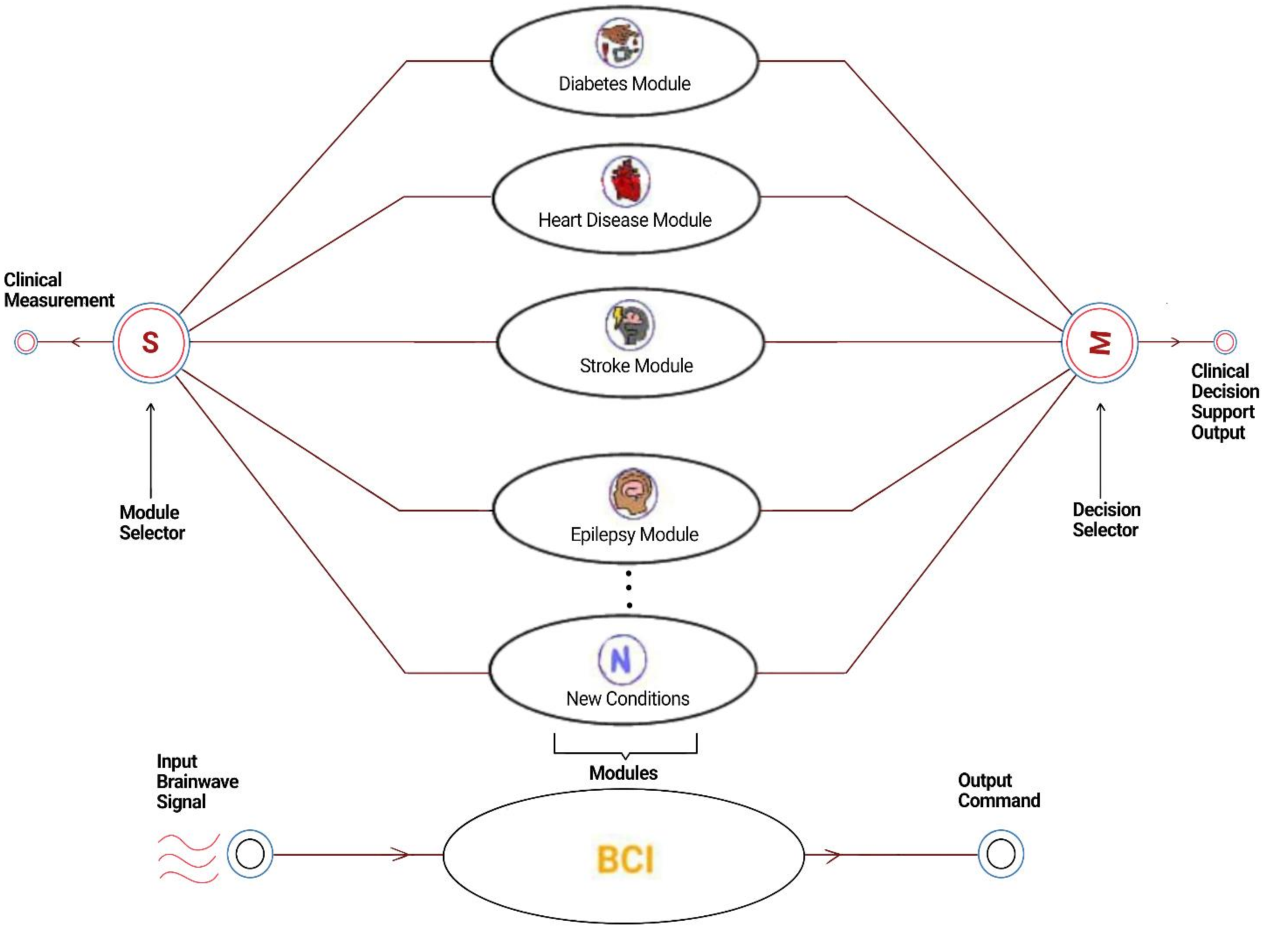

Fig. 1: System Schematic Design Diagram for the Comprehensive AI-Driven Healthcare System and Brain Computer Interface System. The New Conditions component represents additional health conditions that can be incorporated into the system by leveraging new modules.

Approaches Used for Building AI Models

The AI models are developed using four (4) distinct methodologies:

1. **Direct Utilization of Large Language Models (LLMs):**

o Large Language Models (LLMs) like GPT-4 serve as inference engines, utilizing collected data formatted into multidimensional input vectors. This may involve fine-tuning of the LLM for specific tasks.

2. **Application of Prompt Engineering to LLMs:**

o Platforms such as Bard and GPT-4 (including their future iterations) are employed for prompt engineering. This process generates a sequence of steps recommended by the LLM, which are implemented by leveraging extensive knowledge in AI, neural networks, deep learning, Python programming, TensorFlow, Keras, and other machine learning and visualization tools like Scikit-learn and Matplotlib.

3. **Generation of Custom AI Models:**

o Specific AI models are created utilizing the capabilities of LLMs such as Bard and GPT-4, incorporating an automated pipeline for model generation.

4. **Direct Design of AI Architecture:**

o A tailored AI architecture is developed based on comprehensive expertise in AI, neural networks, deep learning, Python programming, TensorFlow, Keras, and additional machine learning and visualization tools like Scikit-learn and Matplotlib.

DOI: 10.17148/IJARCCE.2024.13802

Throughout the development process, all steps taken and tools utilized are meticulously documented to ensure seamless transferability and reusability of the system.

The resulting AI models are rigorously evaluated and compared based on their performance metrics such as specificity, sensitivity, and suitability for addressing specific challenges encountered in the project.

IV. DIABETES MODULE

This section highlights details of the implementation of a Diabetes Diagnosis Module for the comprehensive AI-driven healthcare system.

The Diabetes Diagnosis Module relies on the synthesis of the appropriate ANN architecture as highlighted in the fourth approach listed in the foregoing.

For each participant, the clinical or diagnostic measurements examined include diastolic blood pressure, plasma glucose concentration 2 hours in an oral glucose tolerance test, number of times pregnant (for female participants), triceps skin fold thickness, 2-Hour serum insulin, body mass index, diabetes pedigree function, age and sex.

An appropriate ANN architecture is designed and implemented in the Python programming language by leveraging the TensorFlow framework and the Keras application programming interface (API) package [24] – [25]. The ANN comprised multiple layers including an input layer, hidden layers and an output layer as illustrated schematically in Fig. 2. In Fig. 2, I_1, I_2, ..., I_n represents inputs while O represents the output. Dense sequential layers are utilized. Sigmoid activation is applied to the output layer while rectified linear unit activation is applied to the rest of the ANN. The input layer comprises 8 units, one unit each for the diagnostic or clinical measurements (2-Hour serum insulin, body mass index, diastolic blood pressure, age, and so on) listed earlier while omitting sex or gender and including number of times pregnant for female participants. The output layer has only 1 unit to represent the diagnostic outcome indicating presence or absence of diabetes mellitus. Based on experimentation, 32 units are allocated to each of two hidden layers.

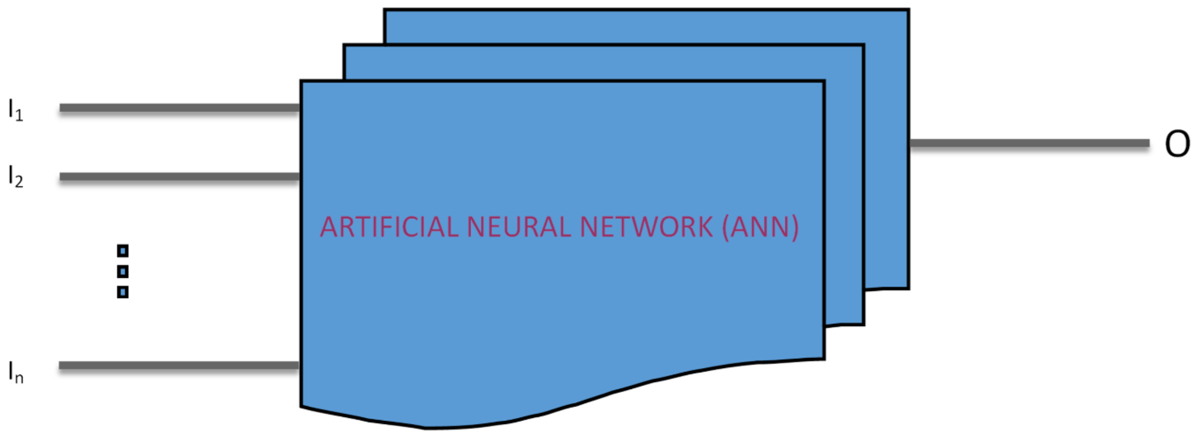

Fig. 2: Schematic Graphical Representation of Artificial Neural Network (ANN) Architecture. I_1, I_2, ..., I_n represent the inputs while O represents the output.

The Pima Indians Diabetes database was retrieved from the Kaggle dataset repository and employed for the training, testing and validation of the AI models. Altogether, the dataset contains 768 rows of data. For each row, there are 9 columns. The first 8 columns represent the clinical or diagnostic measurements such as diastolic blood pressure, plasma glucose concentration 2 hours in an oral glucose tolerance test, number of times pregnant, triceps skin fold thickness, 2-Hour serum insulin, body mass index, diabetes pedigree function and age of the participant. The ninth and final column represents the diagnosis with a value of 1 indicating the presence of diabetes mellitus and a value of 0 indicating a normal outcome.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⋈ Peer-reviewed & Refereed journal ⋈ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13802

Data Availability

The data utilized in this study are available from **GitHub** at https://github.com/frankekpar/pima_indians_diabetes_dataset/blob/main/dataset.zip and derived from the Pima Indians Diabetes Database publicly accessible from Kaggle at https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

V. RESULTS

The Pima Indians Diabetes dataset was split into two parts, namely, a training dataset comprising 80% (614 rows) of the total dataset and a testing and validation dataset comprising 20% (154 rows) of the total dataset. The data is shuffled at random to eliminate bias before processing. Optimized using the Adam optimizer [26] – [27], the neural network was trained on the training dataset for 300 epochs. The loss function employed was the binary cross-entropy loss function. Training was carried out at a learning rate of 0.001 relying on the TensorFlow default batch size of 32.

With respect to performance of the trained AI model on the validation dataset, the specificity obtained was approximately 88% coupled with a sensitivity of 70% and a precision of 72%.

After training, the AI model is integrated into the Diabetes Diagnosis Module of the comprehensive AI-driven healthcare system -Scholar Medic - created by Ekpar [2].

Fig. 3 depicts an instance of the Diabetes Diagnosis Module in Scholar Medic featuring a set of diagnostic or clinical measurements derived from a patient as well as the inference of the trained AI model in the form of the suggested diagnosis of the system.

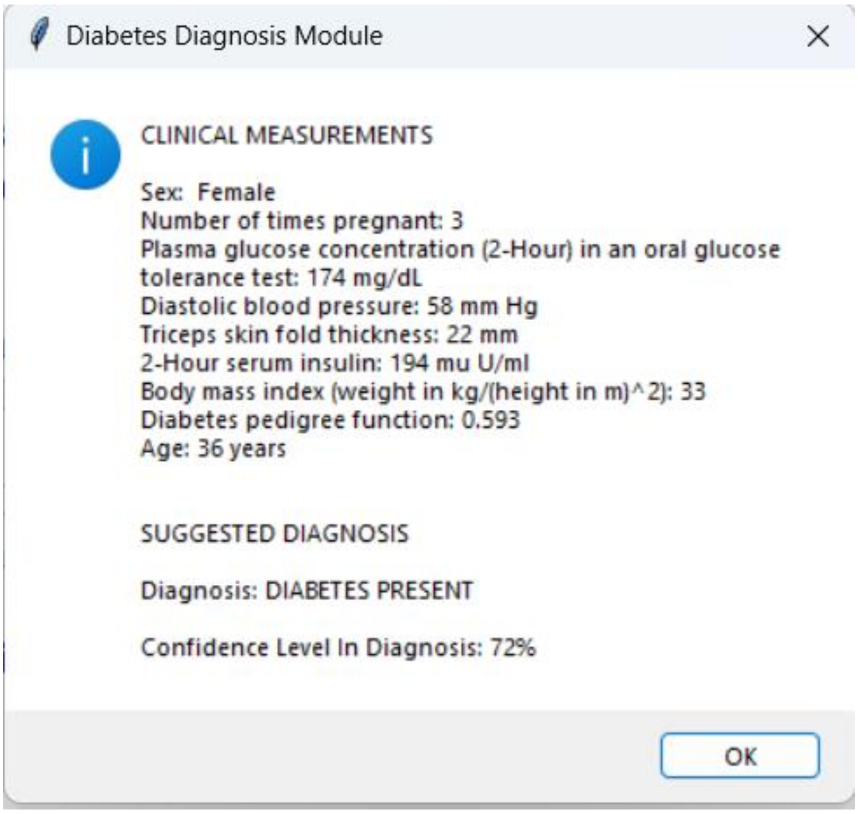

Fig. 3: Diabetes Diagnosis Module of Scholar Medic Showing Clinical Measurements and Corresponding Suggested Diagnosis.

The following equations specify the manner of computation of the precision, sensitivity or recall and specificity performance metrics for the AI model.

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2024.13802

$$\boldsymbol{Precision} \quad = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}}$$

$$\boldsymbol{Sensitivity} \quad = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}$$

$$\boldsymbol{Specificity} \quad = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}}$$

In the equations above, TN refers to true positives, FP refers to false positives, FN refers to false negatives, and TN refers to true negatives. Negative here indicates the diabetes-free condition while positive refers to the presence of diabetes.

Adopting the comprehensive AI system outlined here will provide actionable insights for clinical decision support, leading to life-saving outcomes and improved living conditions. This improvement will come from reducing the economic, social, psychological, and physical burdens associated with the conditions that are predicted, prevented, detected early, diagnosed, treated, and managed more effectively.

Electronic Health Records (EHRs), including clinical diagnostic measurements and EEG data, can be generated by participating medical professionals and their teams. Additionally, EEG data from experiments involving Brain-Computer Interfaces (BCIs) will be collected. All data will be gathered in compliance with ethical approvals, anonymized, and then published in publicly accessible repositories along with related scholarly research.

VI. CONCLUSION

This paper introduced a diabetes diagnosis module designed to be integrated into a comprehensive AI-powered healthcare system. This system aims to save lives and enhance living conditions by supporting medical practitioners with clinical decision-making in predicting and diagnosing various conditions. By incorporating lifestyle, genetic, and environmental data, the system can provide more accurate insights into a patient's context, potentially leading to better health outcomes. Additionally, local data integration helps reduce bias and improves the global applicability of the system's results. The system's modular design allows for future expansion to include modules for diagnosing other conditions and implementing improvements to existing modules.

REFERENCES

[1]. World Health Organization (WHO) Diabetes Statistics: https://www.who.int/news-room/fact-sheets/detail/diabetes. Retrieved (2024).

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

[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,

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13802

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

[23]. Ekpar, F. E. A Novel Three-dimensional Multilayer Electroencephalography Paradigm, *Digital Preprint*, DOI: https://doi.org/10.21203/rs.3.rs-3003621/v1. (2023).

[24]. 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).

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

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

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