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 9, September 2024

DOI: 10.17148/IJARCCE.2024.13914

Diagnosis of Chronic Kidney Disease 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: Chronic Kidney Disease (CKD) is a major noncommunicable disease (NCD) that affects a significant fraction of the millions of people annually who are negatively impacted by NCDs worldwide. This paper presents an artificial intelligence (AI) model trained on publicly available chronic kidney disease data that could be incorporated into a chronic kidney disease diagnosis module within a modular and comprehensive artificial intelligence-driven healthcare system.

Keywords: Chronic Kidney Disease (CKD), Noncommunicable Disease (NCD), Artificial Intelligence (AI), Artificial Neural Network (ANN), Deep Learning (DL), Healthcare System.

I. INTRODUCTION

The World Health Organization (WHO) reports that kidney diseases comprise one of the top ten causes of mortality globally [1] and are among the major noncommunicable diseases that lead to millions of deaths annually around the world [2]. Those adversely affected include people residing in low- and middle-income countries (LMICs) who are among the most vulnerable populations with inadequate healthcare facilities and abysmally low doctor to patient ratios exacerbated by the emigration of qualified medical practitioners to developed countries in search of greener pastures.

In order to help proffer solutions to these issues, Ekpar [3] created a comprehensive artificial intelligence (AI)-powered healthcare system that is scalable and that significantly improves the productivity and performance of medical doctors wielding the tool by dramatically reducing the time and effort required to diagnose, predict and manage a wide variety of health conditions including, but not limited to, heart disease and diabetes. The system also ameliorates medical professional burn-out and associated staff attrition and shortages.

Researchers have utilized algorithms and systems amenable to implementation on computers including artificial intelligence and deep learning in the diagnosis of a wide range of health conditions [4] – [22]. These systems are characterized by diverse strengths and weaknesses.

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

This paper harnesses publicly available chronic kidney disease data to build and AI model that could be incorporated into the comprehensive artificial intelligence-driven healthcare system [3] for utilization in the Kidney Disease Module.

II. MATERIALS AND METHODS

Participant Recruitment

Participants willingly engaged in the studies that contributed to the creation of the advanced AI-driven healthcare system. Each participant provided informed consent before participating in the research.

Ethical Approval

Ethical clearance for the studies was granted by the Health Research Ethics Committee at the Institute of Biomedical Research, University of Uyo. The research adhered to all applicable ethical and regulatory standards, and publicly available data were used in accordance with the licensing terms set by their creators.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⩰ Peer-reviewed & Refereed journal ⩰ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13914

Methodology

To improve publicly accessible healthcare datasets, local experimental and data collection efforts are integrated, enhancing the training of AI models for actionable predictions from new data. Public datasets are sourced from organizations such as the Centers for Disease Control, University of California Irvine Machine Learning Repository, American Epilepsy Society, and Kaggle.

Including local data enhances model robustness, reduces bias, and supports inclusivity and global applicability. This project uniquely combines diagnostic measurements, which may involve electrocardiographic results, with EEG data from both traditional and novel three-dimensional multilayer EEG systems.

For the local data collection efforts, ethical approval has been secured from the relevant research ethics committees in the area where the studies are conducted. Furthermore, the project has obtained the cooperation of licensed medical doctors who have direct patient access and are providing anonymized clinical data for AI model validation.

The developed AI models may be integrated into a comprehensive healthcare system to offer clinical decision support to medical professionals and to generate brain-computer interfaces (BCIs). These systems will use actionable insights and predictions from new clinical data to assist in the early detection, diagnosis, treatment, prediction, and prevention of various conditions such as diabetes, cardiovascular diseases, stroke, autism, and epilepsy.

The project emphasizes open science, reproducibility, and collaboration. Therefore, the generated data will be made publicly available on platforms like GitHub.

System Design and Implementation

The healthcare system described here is designed with a modular architecture, where each module addresses a specific condition (for example, heart disease, diabetes mellitus, stroke, epilepsy, autism). This design allows for the future addition of modules for new conditions and easy updates to existing modules with new data. BCIs, including those based on the motor imagery paradigm, will use EEG data to generate actionable commands and responses.

The system includes instructions for adapting traditional EEG systems to new three-dimensional multilayer EEG systems, as developed by Ekpar [23], [24]. These novel systems leverage approximations of carefully chosen features of bio-signal sources to characterize or manipulate the biological system.

For each condition-specific module, robust AI models are developed and trained on well-formatted data. These models may incorporate genetic, environmental, lifestyle, and other relevant factors to accurately represent the participants' circumstances.

In Figure 1, the system is depicted with representations of the health conditions (for example, chronic kidney disease, diabetes mellitus, stroke, epilepsy) captured in the modules hosting the trained AI models and high-level representations of their functional relationships.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13914

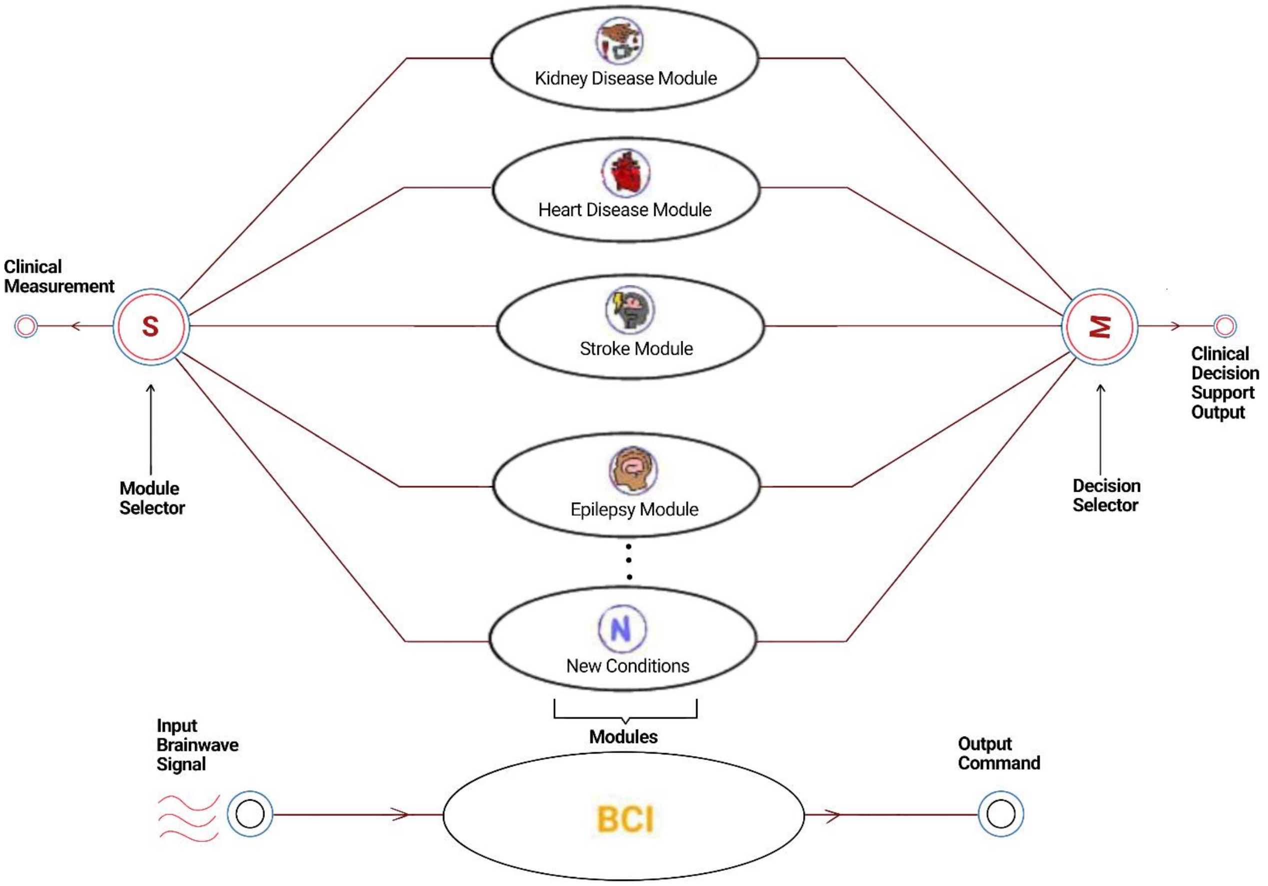

Fig. 1: System Schematic Design Diagram for the Comprehensive AI-Driven Healthcare System and Brain Computer Interface System. The New Conditions component is a placeholder for additional health conditions that can be incorporated into the system by leveraging new modules.

Synthesis of AI Models

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

1. **Direct Application of Large Language Models (LLMs):** Leveraging large language models (LLMs) like GPT-4 as inference engines with data formatted into multidimensional input vectors. This approach may include fine-tuning the LLM to improve its performance.

2. **Prompt Engineering with LLMs:** Using prompt engineering techniques with LLMs such as Bard and GPT-4 (and their future iterations) to outline the steps for developing the AI system. These steps are then executed using the creator's expertise in AI, neural networks, deep learning, Python programming, TensorFlow, Keras, and other machine learning and visualization tools like Scikit-learn and Matplotlib.

3. **Automated AI Model Generation:** Utilizing the capabilities of LLMs like Bard and GPT-4 (and their future versions) through an automated pipeline to generate specific AI models.

4. **Manual AI Architecture Design:** Creating an appropriate AI architecture based on the creator's extensive knowledge and experience with AI, neural networks, deep learning, Python, TensorFlow, Keras, and other machine learning and visualization tools.

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

The performance of the generated AI models is evaluated and compared based on metrics such as specificity and sensitivity, as well as their suitability for addressing the identified challenges.

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 9, September 2024

DOI: 10.17148/IJARCCE.2024.13914

III. CHRONIC KIDNEY DISEASE MODULE

The implementation of the Chronic Kidney Disease Diagnosis sub-system, relying on the generation of an appropriate AI architecture based on the creator's expertise, is expounded in this section. Sequel to development, testing and validation, the sub-system could be integrated into the comprehensive AI-driven healthcare system under consideration.

A wide variety of input variables were solicited from each participant. These variables include patient information such as **demographic particulars and lifestyle factors** (age, gender or sex, ethnicity, education level, socioeconomic status, body mass index, physical activity, smoking, alcohol consumption, diet quality, sleep quality), **medical history** (family history of kidney disease, family history of hypertension, family history of diabetes, previous acute kidney injury, urinary tract infections), **diagnostic measurements** (systolic blood pressure, diastolic blood pressure, fasting blood sugar level, hemoglobin A1c level, serum creatinine level, blood urea nitrogen level, glomerular filtration rate, protein levels in urine, albumin-to-creatinine ratio, serum sodium level, serum potassium level, serum calcium level, serum phosphorus level, hemoglobin level, total cholesterol level, low-density lipoprotein cholesterol level, high-density lipoprotein cholesterol level, triglycerides level), **medications** (use of ACE inhibitors, use of diuretics, frequency of NSAIDs use, use of statins, use of antidiabetic medications), **symptoms and quality of life** (presence of edema, fatigue level, frequency of nausea and vomiting, frequency of muscle cramps, itching severity, quality of life score), **environmental and occupational exposures** (exposure to heavy metals, occupational exposure to harmful chemicals, quality of water), and **health behaviors** (frequency of medical check-ups per year, medication adherence score, health literacy score).

The artificial neural network (ANN) that underpins the AI model for the sub-system was constructed by utilizing the Python programming language and leveraging the TensorFlow framework and the Keras application programming interface (API) package [25] – [26]. After extensive experimentation, two (2) hidden layers with 64 units each were incorporated into the ANN which further comprised an input layer with 51 units (each unit representing one of the 51 clinical measurements embedded in the data and outlined in the foregoing) and an output layer with a single unit standing in for the diagnostic outcome. Sigmoid activation units were used in the output layer while rectified linear units were used in the rest of the ANN. The ANN featured dense sequential layers and connections.

Figure 2 shows the pictorial representation of the ANN where CM_1, CM_2, ..., CM_N depict the clinical measurements (a total of N=51 as listed earlier) and CD depicts the clinical diagnosis or output of the artificial neural network.

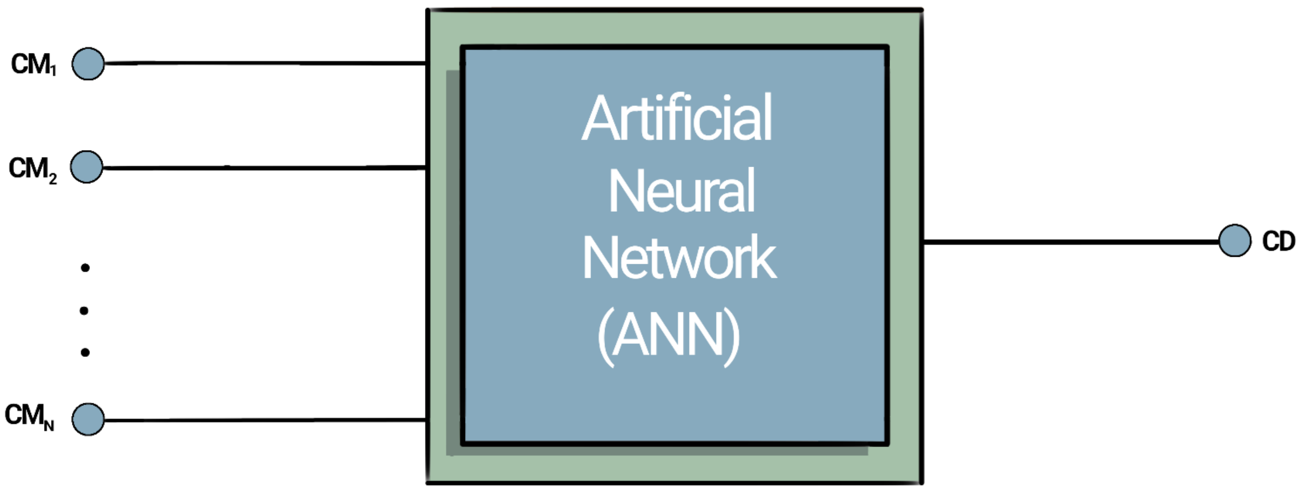

Fig. 2: Schematic Graphical Representation of Artificial Neural Network (ANN) Architecture. CM_1, CM_2, ..., CM_N represent the inputs while CD represents the output indicating the suggested clinical diagnosis.

The Kaggle dataset repository is the source of the chronic kidney disease dataset under consideration. The dataset comprised 1659 rows and 52 columns of usable data. Each row represents a participant or patient. The first usable 51 columns represent the clinical measurements outlined earlier while the last usable column represent the diagnosis indicating the presence (value of 1) or absence (value of 0) of chronic kidney disease.

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 9, September 2024

DOI: 10.17148/IJARCCE.2024.13914

Data Availability

The chronic kidney disease dataset under consideration is available from the Kaggle dataset repository at https://www.kaggle.com/datasets/rabieelkharoua/chronic-kidney-disease-dataset-analysis.

IV. RESULTS

In order to train, test and validate the synthesized ANN, the chronic kidney disease dataset was split into a training dataset with 60% of the data and a testing and validation dataset with 40% of the data. Preprocessing involved shuffling the data at random to ameliorate bias. The Adam Optimizer [27] – [28] was employed for optimization. Training was carried out over 500. Binary cross-entropy loss function was harnessed in the ANN. The learning rate and batch size utilized were the default values of 0.001 and 32, respectively.

The performance of the trained AI model on the validation dataset was characterized by a precision in the range of 92% to 98%, a sensitivity in the range of 92% to 98% and a specificity in the range of 38% to 56%. Measured sensitivity and precision metrics were impressively high but the measured specificity was not as high, possibly owing to the preponderance of positive diagnostic results and paucity of negative diagnostic outcomes in the dataset under consideration.

Figure 3 illustrates a subset of clinical measurements featuring the diagnostic measurements, age and sex of a selected patient and the corresponding inference indicating the clinical decision suggested by the trained AI. The outcome is positive (presence of chronic kidney disease) for this patient.

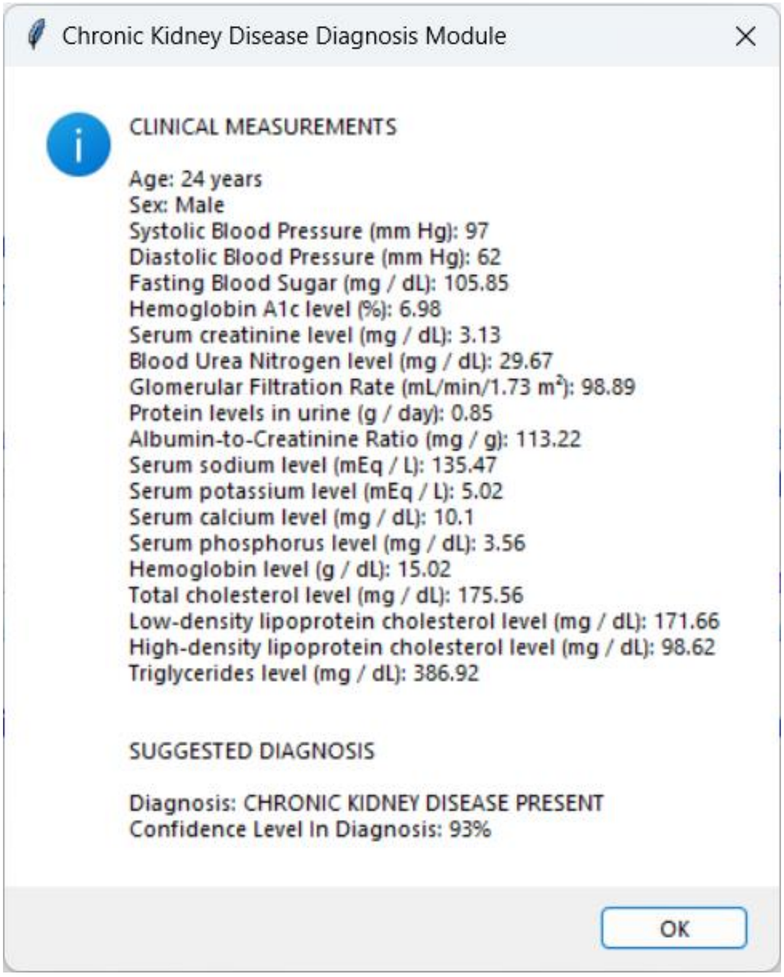

Fig. 3: Positive Chronic Kidney Diagnosis Output of Trained AI Model and Underlying Clinical Measurements.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13914

Computation of the performance metrics of sensitivity, precision and specificity for the trained AI model was carried out in accordance with the equations below.

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

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

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

In the equations specified, TN represents true positives, FP represents false positives, FN represents false negatives, and TN refers to true negatives. Negative here indicates the normal or chronic kidney disease-free condition while positive refers to the presence of chronic kidney disease.

Implementing the comprehensive AI system described here will provide actionable insights for clinical decision support, potentially saving lives and enhancing quality of life. It will reduce the economic, social, psychological, and physical burdens associated with predicted conditions, leading to more efficient prevention, early detection, diagnosis, treatment, and management.

Electronic Health Records (EHR), including clinical diagnostic measurements and EEG data, will be created by participating medical doctors and their affiliated colleagues. EEG data may also be collected through experiments involving brain-computer interfaces (BCIs). These datasets will be gathered in line with ethical approvals and will be anonymized before being published in publicly accessible repositories, along with related scholarly research articles.

V. CONCLUSION

This study introduced an AI-based model for chronic kidney disease diagnosis. This model could be part of a broader AI healthcare system that assists doctors in making clinical decisions. The system is flexible enough to leverage various data points like lifestyle, genetics, and environment to provide more accurate predictions and diagnoses. By permitting the integration of local data, the system aims to be more effective globally and to reduce bias. The system is designed to be flexible and can be expanded to diagnose other diseases or improve existing features based on fresh data and insights.

REFERENCES

[1]. World Health Organization (WHO) - Top 10 Causes of Death: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Retrieved (2024).
[2]. World Health Organization (WHO) – Noncommunicable Diseases: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. Retrieved (2024).
[3]. Ekpar, F. E. A Comprehensive Artificial Intelligence-Driven Healthcare System, European Journal of Electrical Engineering and Computer Science, 8(3), Article 617. (2024).
[4]. 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).
[5]. 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).
[6]. 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).
[7]. 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).
[8]. 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).
[9]. Chaikijurajai, T., Laffin, L., Tang, W. Artificial intelligence and hypertension: recent advances and future outlook, *Am J Hypertens* 33:967–974 (2020).

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13914

[10]. 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).

[11]. 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).

[12]. 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).

[13]. 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).

[14]. 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).

[15]. 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).

[16]. 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).

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

[18]. 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).

[19]. 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).

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

[21]. 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).

[22]. 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).

[23]. 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).

[24]. Ekpar, F. E. A Novel Three-dimensional Multilayer Electroencephalography Paradigm, *Fortune Journal of Health Sciences*, 7(3): 466-480. (2024).

[25]. 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).

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

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

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