STATEMENT OF PURPOSE

The human body, in all of its complexity, has become less of a black box thanks to recent advances in imaging technology such as magnetic resonance imaging and computed tomography. For a long time, these imaging data could only be analyzed by trained clinicians with an experienced eye and a deep knowledge base. However, artificial models have been developed that outperform human experts on certain metrics. These models are not perfect however, and face several challenges when deployed in a clinical setting. These include issues of robustness, explainability, evaluation, and uncertainty. As a result I am pursuing a PhD to tackle these open problems in medical imaging and I believe that the Paul G. Allen school of computer science and engineering at the University of Washington (UW) would be the perfect environment for me to do that.

My initial experience in analyzing medical data and probing the robustness of clinically deployable models began in my third year of bachelors where I worked on extracting actionable information from EEG data in the Biosignal Processing, Instrumentation, and Control Lab (BSIC) at Obafemi Awolowo University (OAU). I was focused on tracking artifacts caused by forced eyelid movement as an indicator to pass instructions to a wheelchair controller from scalp-EEG data acquired from an Emotiv Epoc+ device. This project evolved to become my undergraduate thesis, where I developed a technique to integrate EEG dataset from multiple disparate sources with different acquisition pipelines and features. This is important for increasing generalizability of deep convolutional neural network (CNN)-based techniques for epilepsy detection. To combine disparate datasets for training, we transform the data into sequences of feature vectors which form images that retain the temporal and spatial information of the EEG time series signal. Images from different datasets are then combined using a domain generalization approach. Our trained recurrent-CNN (RCNN) model was tested on a privately collected dataset obtained from collaborators at the University's Teaching Hospital. We achieved results comparable with previous large-scale validation studies and show that the distributional shift between disparate datasets can be minimized to enable merging datasets available for automatic seizure detectors. My thesis, supervised by Dr K.P. Ayodele, contributed to our manuscript which was published in *Computers in Biology and Medicine*[1]. Through this research, I learned to form scientific questions, scour the literature for ideas and information, apply techniques to test our hypotheses, and discuss the results and its implications.

After graduating from my Bachelor's programme, I returned to the lab as a research assistant, working on dimensionality reduction (DR). With insights from my previous project, we looked into developing a method for extracting increasingly more useful information than was previously done from the spectral bands of each EEG electrode signal. This pursuit required that I first address an intermittent question: How do we, empirically, characterize the temporal dynamics of spectral components? This became relevant, because the research direction of the DR problem was predicated on knowing the dynamic evolution of the data in the frequency domain. I investigated the temporal evolution of each spectral component of the EEG data after spectral decomposition with the Fourier transform, empirically determining the stationarity, complexity, and chaoticity properties. We observed that all the spectral components contained information and no evidence that the evolution of the individual spectral components were chaotic while interpreting the estimated largest Lyapunov exponent. This work actively contributed to a manuscript in collaboration with Dr. A. Osuntuyi accepted now at the *"international journal of online and biomedical engineering"* [2] and taught me how to think scientifically, in terms of rigorous and reproducible testing. Building on this empirical study, I am currently developing a DR technique for pre-processing epileptic EEG data prior to modeling.

Traditionally, DR methods like frequency banding and principal component analysis (PCA) have only been applied to lower spectral components (LSC) after cutting off higher spectral components (HSC) in a bid to increase the signal-to-noise ratio of the data since the majority of EEG activity was found in these lower bands. However, our work on characterizing the temporal dynamics of the spectral components alongside other literature affirmed that there is usable spectral information in HSC's. Consequently, I explored several DR methods on all components(both LSC and HSC), the most promising - the dynamical component analysis (DyCA) - this method corresponds to a generalized eigenvalue problem exploiting the empirical observation that the deterministic nature of epilepsy signals closely resembles the trajectories of a known and describable chaos type. While we had success adjusting this technique to work in the spectral domain, the RCNN models trained on the reduced data did not outperform the traditional methods in both patient-specific and non-patient-specific downstream epilepsy detection tasks. Nonetheless, we compared the DyCA method empirically against classical methods like PCA and are poised to submit a manuscript describing our findings and future directions. With its many moving parts and the independence I received from working on this project, I gained valuable experience in software engineering and data science, including working with High-Performance Computing node clusters and open source code. This project motivated me to pursue further training in computational neuroscience.

I received further training on CNNs and modeling at the IBRO-Simons computational neuroscience (ISICNI) summer school in Cape Town, South Africa. At the school, I worked on two projects. The first was on applying biological visual attention to deep convolutional neural networks, supervised by Dr. Grace Lindsay, and for my second project, with Dr. Chris Currin, I explored CNN classification of non-image data and asked, "Can we create statistically correct natural images from non-image data points?" The culture the organizers had established at the summer school provided an enabling environment to gain insights quickly. The projects and lectures required me to actively learn, engage with my peers and the teaching assistants, apply learned concepts, and exchange ideas with the faculty at the school while piquing my interest in a broad array of related topics such as reinforcement learning and generative adversarial networks (GANs).

Ultimately, my career goal is to become a professor and principal investigator at an academic institution. Just as my aspirations to become a scientist have been fostered by professors, I want to train the next generation of scientists and engineers. To reach my career goals, the collaborative environment, excellent faculty, and motivated students at UW's school of computer science and engineering would be ideal for my training. With my interest in medical image analysis (MIA), I am curious about Professor Sun-In Lee's research on explainable artificial intelligence models in medicine, and Professor Linda Shapiro's research in applying attention to the task of classification and detection of breast cancer. Furthermore, since I have been in a highly collaborative environment, I can easily identify with the school's broad range of researchers/faculty and clinicians and understand that they will be instrumental to my development as a researcher. The field of MIA in general can benefit from promoting the robust and ethical use of machine learning methods. During my doctoral program, I would investigate ways to improve our current understanding of complex artificial intelligence models in MIA and provide crucial guidance to the clinical use of these technologies.

Thank you for taking the time to consider my application.

Citations

[1] Ayodele, K. P., W. O. Ikezogwo, M. A. Komolafe, and Philip Ogunbona. 2020. "Supervised Domain Generalization for Integration of Disparate Scalp EEG Datasets for Automatic Epileptic Seizure Detection." Computers in Biology and Medicine 120: 103757.

[2] Ayodele, K. P., W. O. Ikezogwo, and A. A. Osuntuyi. (in press) "Empirical Characterization of the Temporal Dynamics of EEG Spectral Components." International Journal of Online and Biomedical Engineering (IJOE).