NAME: MOHAMMED MUBAREK

ID: ADS21B00260Y

CAPSTONE PROPOSAL TOPICS

TOPIC 1:

DETECTION AND ANALYSIS OF BRAIN TUMOR FROM MAGNETIC RESONANCE IMAGING (MRI) USING WITH TRANSFER LEARNING WITH MOBILENETV2: A HYBRID DEEP LEARNING APPROACH IN MEDICAL HEALTH DIAGNOSIS

Making effective analysis from MRI scans is quite burdensome to healthcare professionals as it is a complex and delicate process that requires extensive evaluation and, in some cases, further consultation with more experienced specialists. Such, when coupled with increased patient size, makes the entire healthcare process subject to stagnation and consequently, loss of lives due to late diagnosis and treatment. Further, in dealing with brain tumor diagnoses and classification, radiologists often make diagnoses as per the histological subtype (the growth pattern of tumors), this is however highly subjective and prone to inaccuracies in diagnosing the type of tumor.

The extenuating circumstance thus necessitates the need to incorporate Artificial intelligence in the radiology sector. This study aims to apply a hybrid approach by combining a custom Convolutional neural network model with a state of the art pre-trained image classifier model (MobilenetV2), to build a more powerful that would be trained and tested with a publicly available dataset of MRI image samples, which would then be deployed in a web application interface via an application programming interface (API), so new MRI scans of patients can be parsed into, for automated brain tumor detection and analysis.

TOPIC 2:

A RANDOM FOREST ALGORITHM (RF) APPROACH TO IDENTIFY LARGE LANGUAGE MODEL (LLM) GENERATED TEXT: COMBATING A.I PLAGIARISM IN ACADEMIA

Artificial intelligence has been of extreme help to humans, seeing its pervasive essence in health, education, industry sectors, among others. As a result of continuous research in an attempt to achieve optimum use of A. I, there have been significant results with transformer neural networks, which has seen the birth of large language models. Large language models are transformer neural networks that have been trained with billions of data and as a result they are able to understand and generate text of wide domains just like humans. While this seems auspicious, having seen the essence of LLM like Chat-gpt in present day academia, making life of students across the globe, there ought to be a way to ensure the responsible and ethical use of them in academia. Trying to identify plagiarized works presented by students is increasingly becoming difficult as the works they are LLM generated and would fail to be identified by the existing conventional software that is used to detect plagiarism.

Thus, this study aims to use a machine learning algorithm, specifically; the random forest algorithm, which is of the supervised learning descent and uses ensemble techniques, to build a model that would be able to classify human written text from an A.I generated text, by training it with publicly available Kaggle dataset of labeled classes, human written and A.I generated. This model will then be optimized to achieve higher accuracy and deployed in a web application interface to serve as a detection tool that would predict if an inputted batch of text is human written, or AI generated.

TOPIC 3:

ANALYSIS OF THE KERNEL METHODS USED IN MACHINE LEARNING ALGORITHMS: MATHEMATICS OF MACHINE LEARNING

Machine learning draws from the language of mathematics to postulate concepts that seem intuitively obvious but seem to be quite difficult to formalize. With the current growth of the field and the development of pre-built machine learning packages and frameworks, the low-level details are abstracted from the practitioner. While this seems a plausible approach to make the field accessible and easier to use by everyone, it limits practitioners to the conventional approaches in usage and implementation of the algorithms. There is no room for further customization of these algorithms for optimal performance.

Kernel methods in machine learning are used to map a nonlinear N-dimensional space classifier into a higher N- dimensional, where we don't need to understand or visit that region; they have wide applications in bioinformatics, 3D reconstruction, handwriting recognition, among others. In order to fully appreciate what is going on behind the kernel and explore ways to tweak them for optimal performance, we ought to consider the convex optimization that these kernel methods operate on

This paper, hence, aims to dive into the rigorous underlying statistical theory following these kernels.