Reference ID : STS2024-05062

Title- Fine-tuning AI on Coronary Artery Disease (CAD) Dataset for Improved Diagnosis as LLM in Physiologically Complex Patients.

Introduction-Artificial intelligence (AI) has emerged as a powerful tool in healthcare, with the potential to revolutionize medical research and clinical practice. Large language models (LLMs) are a type of AI that can understand and generate human-like text, making them well-suited for tasks such as medical literature review, clinical decision support, and patient communication.
Need for an LLM Specifically Designed for Researchers and Doctors
While existing LLMs have shown promise in healthcare applications, they often lack the specialized knowledge and context required for medical research and clinical practice. A tailored LLM, developed specifically for researchers and doctors, could address this gap by:
Understanding complex medical terminology and concepts: LLMs can be trained on vast amounts of medical literature, enabling them to comprehend the nuances of medical language and the relationships between different medical concepts.
Generating high-quality medical text: LLMs can assist researchers in writing scientific papers, grant proposals, and other medical documents, ensuring clarity, accuracy, and adherence to medical conventions.
Facilitating efficient literature review: LLMs can rapidly summarize and synthesize large volumes of medical literature, helping researchers keep up with the latest advances and identify relevant studies for their work.
Providing clinical decision support: LLMs can be integrated into electronic health records and clinical decision support systems, offering real-time guidance to doctors based on the latest medical knowledge and evidence.
Improving patient communication: LLMs can help doctors communicate complex medical information to patients in a clear and understandable manner, enhancing patient engagement and adherence to treatment plans.

Our Approach

We propose to develop an LLM specifically designed for researchers and doctors. Our approach involves:

Collecting a comprehensive medical dataset: We will gather a large and diverse dataset of medical literature, including research papers, textbooks, and clinical guidelines.

Training a transformer-based LLM: We will train a transformer-based LLM on the medical dataset, using advanced natural language processing techniques.

Fine-tuning for specific tasks: We will fine-tune the LLM on specific tasks relevant to medical research and clinical practice, such as literature review, medical writing, and clinical decision support.

User interface and integration: We will develop a user-friendly interface for researchers and doctors to access the LLM and integrate it into their workflows.

Expected Impact

The development of an LLM specifically designed for researchers and doctors has the potential to:

Accelerate medical research: By providing researchers with tools for efficient literature review and high-quality medical writing, we can expedite the discovery of new knowledge and the development of new treatments.

Improve clinical practice: By providing doctors with real-time clinical decision support and enhanced patient communication tools, we can improve patient outcomes and reduce healthcare costs.

Empower patients: By making medical information more accessible and understandable, we can empower patients to make informed decisions about their health and participate actively in their own care.

Objectives of the Research -

To develop an artificial intelligence (AI)-powered tool for the diagnosis of coronary artery disease (CAD) using electrocardiogram (ECG) data.

To train and validate the AI model on a large and diverse dataset of ECG recordings from patients with and without CAD.

To evaluate the performance of the AI tool in a clinical setting and compare its accuracy to that of traditional CAD diagnostic methods.

To develop a user-friendly interface for the AI tool, making it accessible to healthcare professionals and researchers.

To explore the potential of the AI tool to improve the early detection and management of CAD, leading to better patient outcomes.

Methodology-

Data Collection and Preprocessing

Collect a large and diverse dataset of ECG recordings from patients with and without CAD.

Preprocess the ECG recordings to remove noise and artifacts, and segment the ECG signals into individual heartbeats.

Label the heartbeats with their corresponding CAD status (e.g., normal, CAD present).

AI Model Selection and Training

Select a pre-trained language model (LM) from Google AI Studio that is suitable for medical text analysis.

Fine-tune the LM on the labeled ECG dataset using transfer learning.

Optimize the LM's hyperparameters using automated hyperparameter optimization techniques.

Model Evaluation

Split the dataset into training, validation, and test sets.

Evaluate the performance of the fine-tuned LM on the test set using standard metrics for medical diagnosis, such as accuracy, sensitivity, specificity, and F1 score.

Compare the performance of the LM to that of traditional CAD diagnostic methods, such as cardiologists' interpretations of ECGs.

User Interface and Integration

Develop a user-friendly interface for the AI tool, allowing healthcare professionals and researchers to easily access and use the LM for CAD diagnosis.

Integrate the AI tool with electronic health records (EHRs) and other clinical systems to facilitate its use in real-world clinical settings.

Clinical Validation

Conduct a clinical study to evaluate the performance of the AI tool in a real-world clinical setting.

Recruit a cohort of patients with suspected CAD and collect their ECG recordings.Compare the AI tool's diagnoses to the diagnoses of experienced cardiologists.Assess the accuracy, sensitivity, specificity, and F1 score of the AI tool in this clinical setting.

Additional Considerations

Data Security and Privacy: Implement appropriate security measures to protect patient data privacy and confidentiality.

Explainability and Interpretability: Develop methods to explain the AI tool's predictions and make them interpretable to healthcare professionals.

Continuous Improvement: Establish a process for monitoring the performance of the AI tool over time and making adjustments as needed to maintain its accuracy and effectiveness.

Implications and Outcomes -

The AI tool has the potential to:

Improve the accuracy and efficiency of CAD diagnosis.

Reduce healthcare costs by enabling early detection and prevention of complications.

Increase access to care by providing remote access to expert-level CAD diagnosis.

The API output of the model will provide a probability score for the presence of CAD, which can be used by healthcare professionals to inform their clinical decision-making.

Future directions include:

Integrating the AI tool with other diagnostic modalities.

Developing personalized CAD risk assessment tools.

Implementing the AI tool in clinical practice through large-scale clinical trials.

References-

Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-118. doi:10.1038/nature21056

Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine. 2017;23(5):540-546. doi:10.1038/nm.4305

Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Mendis B, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine. 2019;25(6):954-961. doi:10.1038/s41591-019-0447-0

Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer-aided diagnosis: A brief survey. Journal of Medical Imaging and Health Informatics. 2018;8(2):131-141. doi:10.1186/s13623-018-0620-9