BSCI- ECG Arrhythmia Diagnosis AI Model

Team TacoBell Coding Inc

Team Members:
Fatima Hassan
Hoang Tran
Kien Le
Premal Patel
Oanh Dinh
Ariya Ansari

Final Report

Table of Contents:

Introduction:

Boston Scientific, a global medical device company, tasked our team with creating an internal tool to analyze electrocardiogram (ECG) data and diagnose arrhythmias using machine learning. ECG signals represent the heart's electrical activity, and clinicians use the waveforms to diagnose heart conditions. However, manual ECG interpretation is time-consuming and requires expertise. Automated ECG analysis using machine learning can improve diagnosis speed and accuracy.

Our initial goal was to train a model to predict rhythms from ECG images, but this approach required extensive preprocessing. We pivoted to developing a tool that reads raw ECG data in CSV format and uses classification or time series models to predict the diagnosis. We utilized public ECG databases and designed a data processing pipeline with steps like noise removal, heartbeat detection, feature extraction, and model training/evaluation.

Project Goals:

The original objective of this project was train an AI model that reads ECG images and outputs a predicted rhythm. After completing this task with a 57% accuracy and creating a functional webapp, the sponsor introduced us to an AI professor, Dr. Elham Buxton, who adjusted our implementation.

Dr. Buxton told us the complexity and inefficiency with trying to train a model on the images and using the model to classify them. Instead, she transitioned us into using raw lead data from the ECG. We found the following dataset:
https://figshare.com/collections/ChapmanECG/4560497/1

The above dataset is formatted into 12 columns with over 5000 rows.

She also transitioned us to using a new state of art model called Timesnet which is a time series models which are designed to capture temporal dependencies in sequential data. These recurrent architectures have shown promise in ECG classification tasks, as we can effectively model the dynamic patterns and long-term dependencies in ECG signals. In addition to the Timesnet, we also continued using the XGBoost model, a classification. Classification models are well-suited for ECG analysis due to their ability to automatically learn discriminative features from the raw signal.

In addition to the core machine learning models, we also planned to develop a user-friendly front-end interface. This interface would allow users to input their own ECG data in CSV format and receive the model's predicted rhythm as output. By providing a seamless and intuitive user experience, the tool could be easily adopted by healthcare professionals and researchers for rapid arrhythmia screening and analysis.

Project Accomplishments:

We achieved several significant milestones in developing an AI-based ECG arrhythmia diagnosis system. A key accomplishment was training and testing an XGBoost classifier model to predict ECG rhythms from raw ECG data. XGBoost is a powerful gradient-boost algorithm known for its strong performance and interpretabilit. Our XGBoost model achieved an impressive accuracy of around 91% in classifying arrhythmias.

This high accuracy demonstrates the model's ability to reliably detect abnormal heart rhythms, which is crucial for timely diagnosis and treatment. We also explored a more advanced deep learning approach by training a Timesnet model. Timesnet is a neural network architecture designed for time series data like ECGs. While the current Timesnet model achieved 11% accuracy, we recognized that it requires further training and optimization to reach its full potential. This highlights the iterative nature of machine learning projects and the importance of continuous improvement.

To make the arrhythmia detection system accessible and user-friendly, we built a web application. The app allows users to input their own ECG data in CSV format and receive the model's predicted rhythm as output. This streamlines the diagnosis process and enables healthcare professionals to leverage the power of AI without needing technical expertise.

Lastly, we implemented a database to store the history of all previously processed ECGs. This database serves as a valuable resource for future model improvements, research, and auditing purposes. It ensures that the system can learn from past predictions and facilitates long-term data tracking.

Project Challenges:

We encountered several significant challenges throughout the project that tested our skills and resilience. One major hurdle was the steep learning curve associated with the complex domains of ECG analysis and AI modeling. We had to dedicate substantial time to self-study and research to gain the necessary background knowledge. This included understanding ECG waveforms, arrhythmia types, signal processing techniques, and training models such as Timesnet & XGBoost.

Training the machine learning models, especially the deep learning Timesnet architecture, also required long computation times. The computational cost of training models, particularly Timesnet, imposed a financial burden on us. We ended up $100s on google cloud to train this model, which achieved a much lower accuracy of 11%. Balancing model complexity, training time, and cost was a difficult trade-off.

Lastly, we were forced to adapt to changing project goals and requirements later in the semester. This forced us to re-evaluate our approach, discard some work, and quickly adjust our plans. Effectively managing these moving targets added to the overall challenge.

General Team Experiences and Challenges:

Our journey throughout the project was marked by significant growth and learning. Initially, we struggled with disorganization and lack of focus, hindering our progress. However, a turning point came when the sponsor introduced an AI professor, Dr. Elham Buxton, to provide guidance and support. This addition, coupled with an adjustment of project goals, catalyzed a dramatic improvement in our productivity and collaboration. The AI professor brought valuable expertise and insights, helping us refine our approach and align our efforts more effectively. By clarifying project objectives and providing a clearer roadmap, we were able to channel our energy and skills more productively.

Conclusion:

The complexity of this project speaks for itself. We were tasked to make a reality out of theoretical research of AI reading ECG data while given minimal resources. However, this project was lead by Hoang and Fatima who took the greatest initiative in the self-learning and implementation process. They kept us accountable and lead by example. (Premal)

The increasing vitality of AI in fields like medicine underscores the transformative potential of this technology. This project provided valuable exposure to the cutting-edge approaches used by innovative companies to tackle complex, real-world problems. It required the team to rapidly learn new domains, master emerging technologies, and collaborate effectively to deliver a functional product - essential skills for successfully transitioning from academia to industry.