AI Doctor: Disease Prediction Process

1

Patient Symptoms

Input of patient-reported symptoms

2

AI Model Processing

AI algorithms analyze symptoms

3

Disease Prediction

AI provides probable disease predictions

AI Doctor is a machine learning-powered system designed to assist healthcare professionals in predicting diseases based on patient-reported symptoms.

Dataset Description

AI Doctor utilizes two datasets: a medicine dataset and a disease dataset.

The medicine dataset, "Medicine.csv", provides a comprehensive list of medicines and their associated properties. This dataset includes detailed information on various medications, including their predicted tests and their effectiveness in treating different diseases.

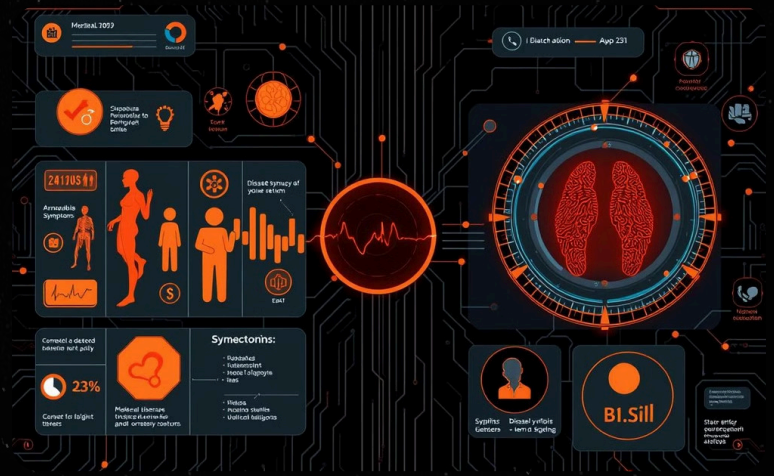
The Model training dataset

We're using raw_data.xlsx to create the cleaned_data.csv dataset for our disease prediction model training. We'll clean the data and fill in any missing values using forward-filling 'ffill'.

Dataset Description

Medicine Dataset: Medicine.csv

This dataset contains detailed information on medicines, including properties like predicted tests and effectiveness in treating diseases.



Disease Dataset: cleaned_data.csv

This dataset is used to train the disease prediction model. It contains cleaned and processed data on symptoms and associated diseases.

Disease Prediction Framework

The framework utilized in this code is **Transformers** from the **Hugging Face Transformers library**.

Key Steps

- **Transformer-Based Embedding Generation**:

The code uses **Sentence-Transformers**, a powerful framework for generating dense vector representations (embeddings) from textual data.

It converts textual descriptions of symptoms into meaningful embeddings that capture semantic relationships between symptoms.

- **Tokenizer and Encoder Initialization**:

Sentence-Transformers' encoder is initialized and used to process input symptom data.

Binary symptom vectors are converted into descriptive strings of active symptoms, which are then passed through the encoder to generate embeddings.

- **Disease Prediction Using Decision Tree**:

The embeddings generated by Sentence-Transformers are input into a **Decision Tree model**.

The Decision Tree learns patterns in the embeddings to classify or predict diseases effectively.

- **Representation Learning**:

Binary symptom vectors are transformed into descriptive strings that highlight active symptoms.

These strings are tokenized and encoded using Sentence-Transformers, ensuring a robust and semantically informed representation of the symptoms.

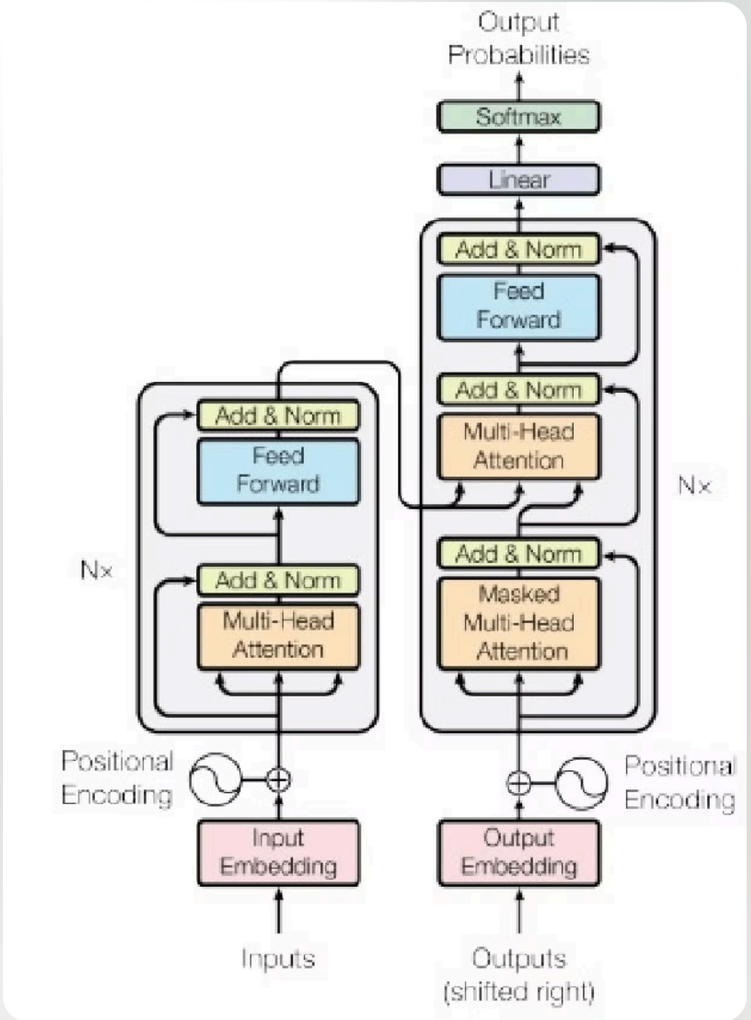
Decision Tree

Key Details

- **Type**: Machine Learning Model for Classification and Regression
- **Model Name**: Decision Tree
- **Source**: Scikit-learn Library (or similar ML frameworks)
- **Architecture**: Hierarchical structure with nodes representing decisions based on feature values
- **Interpretability**: Provides clear and interpretable decision paths for predictions

Key Features

- **Feature-Based Decisions**: Makes predictions by splitting data based on feature values at each node.
- **No Pre-training Required**: Learns patterns directly from the provided dataset without requiring pre-trained knowledge.
- **Handles Non-linear Relationships**: Capable of capturing complex decision boundaries.
- **Fine-tuning**: Hyperparameters such as tree depth, splitting criteria (e.g., Gini Index or Entropy), and minimum samples per leaf can be adjusted for optimal performance.
- **Visualization**: Easily interpretable as a flowchart-like structure, making it simple to understand the decision-making process.



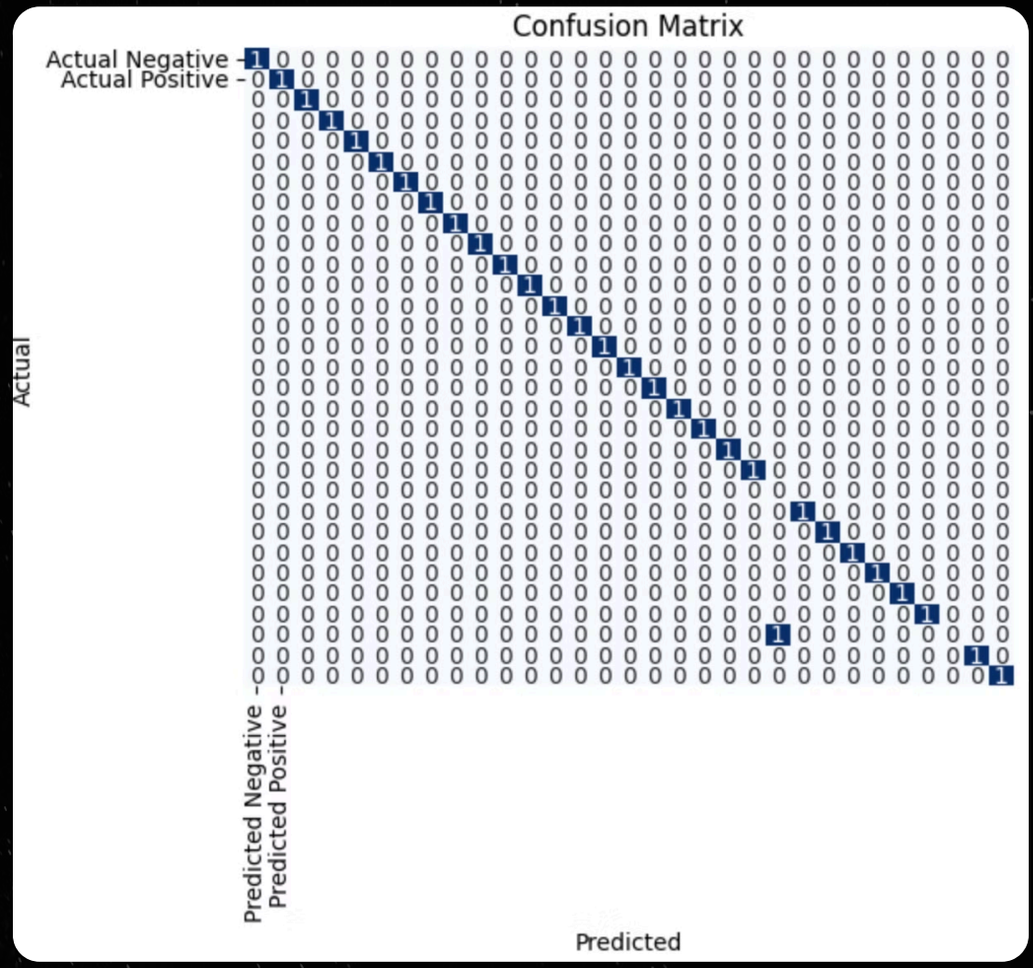
Results: Quantitative Metrics

These metrics will provide a comprehensive assessment of the model's ability to correctly identify diseases based on the provided symptom data.

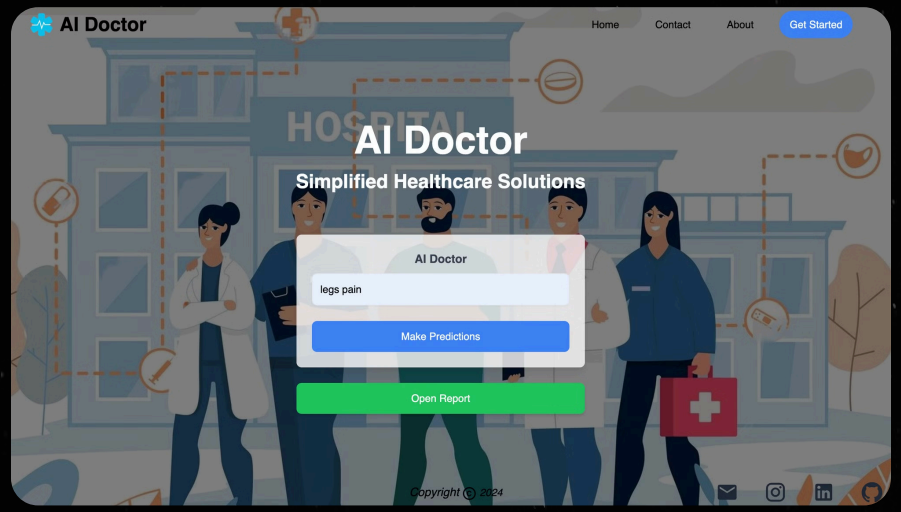
Accuracy: 96.67%
Classification Report:

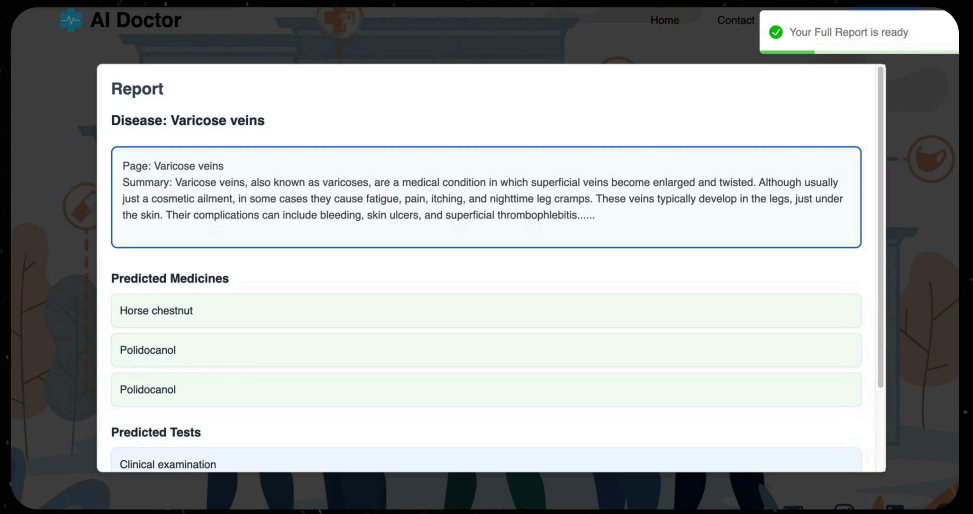
	precision	recall	f1-score	support
HIV	1.00	1.00	1.00	1
bipolar disorder	1.00	1.00	1.00	1
cellulitis	1.00	1.00	1.00	1
cirrhosis	1.00	1.00	1.00	1
colitis	1.00	1.00	1.00	1
confusion	1.00	1.00	1.00	1
delirium	1.00	1.00	1.00	1
delusion	1.00	1.00	1.00	1
dementia	1.00	1.00	1.00	1
endocarditis	1.00	1.00	1.00	1
gastroesophageal reflux disease	1.00	1.00	1.00	1
glaucoma	1.00	1.00	1.00	1
hemiparesis	1.00	1.00	1.00	1
hepatitis	1.00	1.00	1.00	1
hiv infections	1.00	1.00	1.00	1
hypertensive disease	1.00	1.00	1.00	1
insufficiency renal	1.00	1.00	1.00	1
ischemia	1.00	1.00	1.00	1
kidney disease	1.00	1.00	1.00	1
kidney failure acute	1.00	1.00	1.00	1
lymphoma	1.00	1.00	1.00	1

...



Results: visual outputs from the model





- Symptom input triggers disease prediction
- Advanced ML analyzes input to identify condition
- FAISS model recommends relevant treatments
- Delivers robust AI healthcare solution

Conclusion

The key findings of this project are that patients can prompt their symptoms, and the model will accurately predict the underlying disease. Once the disease is identified, the Wikipedia API wrapper will generate a detailed summary to provide more context for the patient.

The implications of these results are significant. By empowering patients to self-diagnose with a high degree of accuracy, this system can improve healthcare accessibility and lead to earlier detection of diseases. Additionally, the FAISS (Facebook AI Similarity Search) model will predict the 3 most relevant medicines and tests based on the identified disease, further streamlining the healthcare process.

Overall, this AI-powered disease prediction and treatment recommendation system has the potential to revolutionize how patients interact with the healthcare system. By putting more control in the hands of individuals, it can drive better outcomes and reduce the burden on overburdened medical facilities.

Project Contributions

The key contributions of this project were made by the two team members, Ansh and Aakash.

Ansh Sharma - developing the Flask backend that powers the core functionality of the disease prediction system. also played a crucial role in implementing the FAISS (Facebook AI Similarity Search) model, which is used to recommend the most relevant medicines and diagnostic tests based on the identified disease.

Aakash - focused on the training and development of the primary disease prediction model. Additionally, Aakash led the development of the Node.js and React-based frontend, ensuring a seamless and intuitive user experience.

Together, we have culminated in a comprehensive AI-powered healthcare solution that empowers patients to take a more active role in their own well-being. By combining disease prediction, treatment recommendations, and informative summaries, this project has the potential to revolutionize how individuals interact with the healthcare system.

References

The key references for this project include the dataset sources from Kaggle, as well as the documentation and resources used for the FAISS (Facebook AI Similarity Search) model.

Specifically, the project team utilized the "Medicine.csv" and "Raw_data.xlsx" datasets from Kaggle to train and evaluate the machine learning models. These datasets provided the necessary information on diseases and their associated symptoms, which formed the foundation of the disease prediction system.

In addition, the team referred extensively to the FAISS documentation, available on the Facebook Research GitHub wiki. This resource was instrumental in implementing the FAISS model, which is used to recommend the most relevant medicines and diagnostic tests based on the identified disease.

Finally, the team also consulted various online resources and GitHub repositories related to machine learning, model training, and backend/frontend development. These supplementary references helped the team refine their approach and address any technical challenges that arose during the project implementation.

By leveraging a combination of curated datasets and well-documented frameworks, the project team was able to develop a comprehensive AI-powered healthcare solution that delivers accurate disease prediction and personalized treatment recommendations.