Assignment – 01

Medical Diagnosis with Attention Mechanism

Name: Mahima Tejeshwini R

Roll No: 22AD046

Class: AI&DS [‘A’]

1. Introduction

In modern healthcare, Natural Language Processing (NLP) is widely used to analyze unstructured medical text, such as patient records, clinical notes, and discharge summaries. One of the key challenges in automated medical diagnosis is extracting critical symptoms that contribute to disease identification. Traditional rule-based and statistical methods struggle to capture contextual dependencies, making deep learning approaches—especially Transformer-based models—more effective.

This study focuses on using a Transformer model with an attention mechanism to highlight important symptoms in patient records. The model assigns attention scores to words in a medical report, allowing us to visualize which symptoms are deemed most relevant for diagnosis.

2. Transformer Model with Attention Mechanism

2.1 Understanding Transformers in Medical NLP

The Transformer architecture, introduced in Vaswani et al.'s *Attention Is All You Need* (2017), has revolutionized NLP by leveraging the self-attention mechanism. Unlike recurrent models (RNNs, LSTMs), Transformers capture long-range dependencies efficiently using multi-head self-attention, enabling them to understand complex clinical narratives.

For our task, we use a pre-trained Transformer model (e.g., BERT, BioBERT, or ClinicalBERT), fine-tuned on medical texts to extract symptom-related information.

2.2 Attention Mechanism in Medical Texts

The self-attention mechanism computes attention scores based on three vectors:

- Query (Q): Represents the word currently being processed.
- Key (K): Represents all words in the sequence.
- Value (V): Represents the word embeddings used for contextual representation.

The attention scores are computed using the scaled dot-product attention formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

The softmax function ensures that the scores sum to 1, highlighting the most relevant words.

3. Attention-Based Symptom Highlighting

To analyse how the model prioritizes symptoms, we visualize attention heatmaps for different patient records.

3.1 Case 1: Initial Medical Note

Patient-Record:
"Patient reports persistent cough, high fever, and difficulty breathing for the past three days."

Attention Heatmap Analysis: The model assigns high attention scores to:

- "cough"
- "high fever"
- "difficulty breathing"

These symptoms are critical indicators of respiratory infections (e.g., pneumonia, COVID-19, bronchitis). The attention mechanism correctly prioritizes them due to their association with severe diseases in training data.

3.2 Case 2: Modified Medical Note

Patient-Record:
"Mild headache and occasional dizziness, but no fever or cough."

Attention Heatmap Analysis:

- The focus shifts to "headache" and "dizziness", but the scores are more evenly distributed.
- The absence of fever and cough results in lower confidence in diagnosing severe illness.
- The attention map shows a broader spread of attention, indicating lower symptom specificity.

Comparison with Case 1:

- In Case 1, attention was highly concentrated on severe symptoms.
- In Case 2, attention is dispersed, reflecting the lower clinical urgency of mild symptoms.

3.3 Case 3: Introducing Rare but Critical Symptoms

Patient-Record:
"Sudden vision loss and severe chest pain experienced since this morning."

Attention Heatmap Analysis: The model assigns very high attention scores to:

- "vision loss" (potential indicator of stroke, neurological disorders)
- "severe chest pain" (possible heart attack, cardiovascular emergency)

Since these symptoms are medically urgent, the model correctly prioritizes them, aligning with clinical best practices.

4. Model Behaviour Insights

From the above cases, we derive the following observations:

- Symptoms with high clinical significance receive stronger attention scores.
- Non-specific symptoms lead to more distributed attention patterns.
- Rare but serious symptoms trigger high model confidence, indicating their critical nature for diagnosis.

5. Conclusion

The study demonstrates the effectiveness of attention mechanisms in highlighting key symptoms for medical diagnosis. By visualizing attention heatmaps, we observe how the model prioritizes symptoms, which can assist healthcare professionals in identifying critical cases. The future enhancements can be done in

1. Domain-Specific Pretraining: Fine-tune the Transformer model on larger medical datasets (MIMIC-III, PubMed).
2. Multi-Modal Integration: Combine text analysis with medical imaging for more comprehensive diagnostics.
3. Explainability Enhancements: Use Layer-wise Relevance Propagation (LRP) to improve interpretability beyond attention heatmaps.