



INFORMATICS  
INSTITUTE OF  
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

## **A transformer neural network for EHR data imputation**

Final Year Project

Mr. Hewa Dewage Emith Dinsara

Supervised by

Mrs. Janani Harischandra

Submitted in partial fulfilment of the requirements for the BEng in Software Engineering degree at the University of Westminster.

**April 2025**

## **ABSTRACT**

Electronic Health Records (EHR) are a very valuable data mine for clinical data analytics. The prevalence of missing data in EHR poses a significant challenge to healthcare analytics and patient care. This affects the reliability of data-driven decisions. It is evident that healthcare professionals can obtain an immense advantage and make educated and efficient decisions in clinical context. To deal with missing data in EHR, apart from other complexities such as sparsity etc. prevalent in EHR, imputation methodologies are introduced. Most traditional imputation methods are not suitable for EHR data imputation due to the suboptimal accuracy and generalizability. This study intends to address the need for a solid and robust EHR imputation methodology, proposing a transformer-based neural network model to effectively handle various missing data mechanisms, such as Missing Completely At Random (MCAR), Missing At Random (MAR), Missing Not At Random (MNAR).

The research adapts the self-attention mechanism in transformers, which enables the model to dynamically weigh features based on relevance, enhancing adaptability to diverse data patterns. A structured approach was applied which included data preprocessing, model training and iterative validation with custom training loops.

The implementation of the transformer neural network was able to outperform the existing EHR data imputation methodologies by a close yet considerable margin and a demonstration platform was built to upload, impute and download datasets.

### **Subject Descriptors:**

- Computing methodologies -> Machine learning -> Neural networks -> Deep learning
- Applied computing -> Life and medical sciences -> Health care information systems
- Data science -> Data processing -> Data imputation

### **Keywords:**

- EHR data imputation
- Transformer neural network
- Missing data mechanisms