



## ABSTRAK

Rekam Medis Elektronik (RME), yang berformat teks bebas (*free-text*), menghadirkan tantangan dalam pemahaman dokumen, sehingga proses pengodean diagnosis ICD-10 menjadi sulit dan rentan terhadap kesalahan jika dilakukan secara manual oleh koder. Kesalahan pengodean ini dapat merugikan berbagai pihak. Untuk mengatasi masalah tersebut, beberapa penelitian terdahulu telah mengembangkan sistem prediksi kode diagnosis ICD menggunakan model *deep learning* seperti *Convolutional Neural Network* (CNN), *Long-Short Term Memory* (LSTM), dan *Bidirectional Long-Short Term Memory* (BILSTM) untuk memberikan hasil yang lebih tepat. Akan tetapi, penelitian-penelitian tersebut masih terbatas pada penggunaan data RME dalam bahasa selain Bahasa Indonesia.

Penelitian ini bertujuan untuk menganalisis performa berbagai model *deep learning* dalam sistem prediksi kode ICD-10 dengan menggunakan data rekam medis elektronik berbahasa Indonesia. Selain itu, penelitian ini akan menganalisis bagaimana penggunaan data rekam medis elektronik berbahasa Indonesia untuk membangun sebuah sistem prediksi kode ICD-10.

Penelitian ini menunjukkan bagaimana proses *preprocessing* seperti penghapusan karakter "enter", *case folding*, *replace abbreviation*, dan *remove punctuation* dalam rekam medis elektronik berbahasa Indonesia. Selanjutnya, data tersebut diubah dalam bentuk vektor dengan menggunakan *word embedding*, *word2vec* untuk digunakan sebagai *embedding layer* pada model *deep learning*. Data yang dalam bentuk *embedding matrix* digunakan sebagai input untuk melakukan pelatihan beberapa model prediksi. Penelitian ini mendapatkan hasil LSTM memiliki akurasi rata-rata tertinggi dengan nilai 0,7666. Penelitian ini menemukan hasil yang didapatkan masih belum optimal dikarenakan adanya nilai *loss* dan *val\_loss* yang tinggi. Selain itu, ditemukan indikasi *overfitting* pada model CNN *Parallel Layer* dan LSTM serta adanya indikasi *underfitting* pada CNN *vanilla* dan BILSTM.

**Kata kunci** – Rekam Medis Elektronik, *International Classification of Disease*, Sistem Prediksi Kode ICD-10, *Deep Learning*, CNN, LSTM, BILSTM



## ABSTRACT

Electronic Health Records (EHR), which are in a free-text format, present challenges in document comprehension, making the ICD-10 diagnosis coding process difficult and prone to errors when performed manually by coders. These coding errors can be detrimental to multiple parties. To address this issue, several previous studies have developed ICD diagnosis code prediction systems using deep learning models such as Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and Bidirectional Long-Short Term Memory (BILSTM) to provide more accurate results. However, these studies were still limited to using EHR data in languages other than Indonesian.

This study aims to analyze the performance of various deep learning models in an ICD-10 code prediction system using Indonesian-language electronic health record data. Additionally, this research will analyze the construction of an ICD-10 code prediction system using Indonesian-language electronic health records.

This study demonstrates the preprocessing process in Indonesian-language electronic health records, such as removing "enter" characters, case folding, abbreviation replacement, and punctuation removal. Subsequently, the data is converted into vector form using word embedding (word2vec) to be used as an embedding layer in the deep learning models. The data, in the form of an embedding matrix, is used as input to train several prediction models. This study found that the LSTM model achieved the highest average accuracy with a value of 0.7666. The research concludes that the obtained results are still suboptimal due to high loss and val\_loss values. Furthermore, indications of overfitting were found in the CNN Parallel Layer and LSTM models, as well as indications of underfitting in the vanilla CNN and BILSTM models.

**Keywords** : Electronic Medical Records, International Classification of Disease, ICD-10 Coding Prediction System, Deep Learning, CNN, LSTM, BILSTM