

# Integrated Deep Learning with Attention Layer Based Approach for Precise Biomedical Named Entity Recognition

Pooja H. and Prabhudev Jagadeesh M. P.\*

Department of Computer Science and Engineering, JSS Academy of Technical Education,  
Visvesvaraya Technological University, Karnataka, India  
Email: pooja.h.28@gmail.com (P.H.); prabhu.jagadeesh@gmail.com (P.J.M.P.)

\*Corresponding author

**Abstract**—Biomedical Named Entity Recognition (BioNER) is a critical task in extracting valuable information from a biomedical corpus. There are different medical terminologies available, therefore it is difficult for scientists and researchers to find the important ones. An efficient integrated deep learning method that combines the advantages of Convolution Neural Network (CNN), Bidirectional-Long Short-Term Memory (Bi-LSTM) with Conditional Random Field (CRF) for precise and effective named entity recognition in the biomedical field is proposed. The proposed integrated model leverages the strengths of the deep learning models to effectively capture contextual information, learn label dependencies, and improve the accuracy of entity recognition. The MIMIC III free-text Electronic Health Record (EHR) dataset is processed using the SpaCy pipeline as a base. Different Entities are trained and tested. The CNN component allows for multi-level feature extraction by capturing local patterns and compositional features, including character-level representations and word embeddings. This allows the model to extract relevant and important features from the input sequence. The Bi-LSTM component further enhances the model's performance by modeling contextual dependencies in both forward and backward directions, enabling a comprehensive understanding of the input sequence. By considering long-range dependencies, the Bi-LSTM component captures intricate relationships between words and the accuracy is improved. To incorporate label dependencies, a CRF layer is used on top of the Bi-LSTM layer. The CRF layer models the global structure of named entity labels and encourages consistent predictions, leading to more coherent and accurate entity recognition. Different performance parameters were considered to be compared with Named Entity Recognition (NER) systems. Comprehensive experimental findings demonstrate the model's improved performance.

**Keywords**—Named Entity Recognition (NER), Bidirectional-Long Short-Term Memory (Bi-LSTM), Conditional Random Field (CRF), Convolutional Neural Networks (CNN), attention model

## I. INTRODUCTION

The cornerstone technique of natural language processing employed in the medical field is medical named entity recognition whose main requirement is extracting named entities. Several approaches for extracting the entities present in large medical databases are found in the literature. Due to the increasing demand for medical data of patients, Electronic Health Record (HER) was created. There is a rise in the utilization of EHR by medical practitioners since it is considered as the primary source of information about the patients [1]. A deep learning-based approach is implemented, as there is a rapid increase and advancement in areas like bioinformatics, disease recognition, medical image analysis, clinical decision-making, and drug discovery to provide advanced performance [2]. Various models are created for data feature learning, mining the information, medical image analysis, and disease or drug prediction. In the biomedical field, the application of deep learning techniques achieves better results in automatic disease coding and integrated analysis of multidata when compared to traditional algorithms.

In recent years, deep learning methods that use neural networks have made remarkable changes in the field of natural language processing. Natural Language Processing (NLP) focuses on the interaction between

Computers and human languages, mainly in applications like text mining and classification. More research is reported where neural network models are used for problems involving classification, recognition, and prediction. Here, the models are tested and trained to learn and extract the higher-level features or retrieve information from the raw data. Deep learning models produce promising results in the area of Named Entity Recognition (NER). With the rapid increase of biomedical texts, there exists a huge amount of biomedical knowledge in unstructured form, and extraction of that information has become a critical requirement for biomedical research. Thus, the use of named entity recognition in the biomedical field has a huge application value by overpassing the different problems, such as large

biomedical data, out-of-vocabulary problems, etc., that the traditional methods could not achieve. A lot of research has been undertaken by applying the Named Entity Recognition (NER) in the field of biomedicine. In the medical field, medical documents such as discharge summaries, clinical cases, hospital records, and inspection records contain a large amount of medical knowledge, which the machine can use to read the medical texts as inputs and extract the different entities that improve the overall performance of clinical research. The recognition of these biomedical entities is a crucial step in information retrieval, information extraction, obtaining knowledge, and answering questions, etc. from the medical data [3].

The process of extracting medical entities from unstructured free texts is named as entity extraction. With

the growing number of biomedical articles and resources [4], the search and extraction of useful data from texts have become a problematic task. To improve efficiency, researchers look to different knowledge sources and convert unstructured text data to distilled information. However, biomedical experts' manual annotations and characteristics are inefficient because the procedure is complicated and not cost-effective. Therefore, useful, and reliable NLP technologies are becoming vital for computational data analytics. Also, up-to-date Text Mining technology is required to automatically analyze biomedical literature and extract useful knowledge from the texts [5, 6]. Fig. 1 illustrates the sample of biomedical NER.



Fig 1. Example of biomedical NER.

NER should be done first as the noted references that plays a vital role in the text mining research, particularly in extracting relationships. It is essential in the biomedical NER to classify named organisms in the biological sector, including chromosomes, disease, chemicals, or medicinal names [7–9]. However, the biomedical NER is an incredibly complicated task, because (1) new entities keep evolving with new discoveries, (2) entities often have a huge number of synonyms, (3) written in long sentences, (4) a mix of a letter, symbol, and punctuation is used.

Thousands of digital medical records, known as electronic health records, contain unstructured text content. One of the hardest things for researchers is to recognize entities from a wide range of unstructured data. In order to develop a model for identifying medical entities, a large number of works are reported under the medical domain that investigates the integration of NER that provides real-time information extraction for healthcare providers. Identifying and classifying specific entities, like symptoms, drugs, treatments, dosages, frequencies, durations, etc., from the medical data is the main purpose of NER. Different methods were explored by the researchers to solve the NER problems, like rule-based methods, dictionary-based methods, machine learning methods, and deep learning methods. Different NER methods work differently to extract entities. However, by using deep learning models, the performance is high. The paper attempts to solve named entity identification issues in the biomedical field. Finding and extracting items associated with the medical field is a challenging process. New dictionaries are created specifically for medical terms, including ailments, signs, dose units, dosing forms, and species. The first step is training the model on large data, and the second involves applying the pre-trained model to get the results. In NLP, pre-training becomes very useful, but the annotations are difficult to consider [10].

## II. LITERATURE REVIEW

In biomedical named entity recognition, pre-processing methods are utilized for various tasks like information retrieval, hypothesis development, and relationship extraction among objects. Some of the effects of NER include information extraction, search retrieval, text summarization, entity linking, etc. Different approaches were used to extract entities in the biomedical field.

- **Rule-Based Approach:** The data patterns and rules found in sentences serve as the foundation for the rule-based approach. However, rule-based methods can utilize context information to overcome the issues of multiple named entities; however, these rules must be written manually before employing them [11].
- **Dictionary-based Approach:** The dictionary system lists the words used to describe names appearing in the text. String matching is worked out in this method. It determines whether a word or a set of terms from the book corresponds to a dictionary term. These methods follow an easy structure but suffer from the issue of handling unknown items and words with several meanings.
- **Machine learning methods:** The Machine-Learning (ML) techniques and list of dictionaries are commonly used by the NER algorithm. For the NER algorithms, there are two types of ML models: supervised and uncontrolled ML models [12, 13]. For unregulated NER, no training data is required. The purpose of this approach is to create future annotations for data. This method of learning is not common in ML, since this unattended method of learning cannot be adapted without supervised methods. Supervised learning approaches usually employ Hidden Markov Models (HMMs) [14], decision-making boards, support vector methods, and Conditional Random Boards (CRF). Supervised learning approaches usually train based on different linguistic rules with data on several features

and assess results using test data not contained in the training data.

Current advancements in machine learning schemes have led to the development of deep learning methods for biomedical text mining applications. Long Short-Term Memory (LSTM) and Conditional Random Field (CRF) are the two widely adopted methods in this domain. However, training these machine learning methods on general, non-specific domains improves the performance of the system; achieving the performance that is desired for biomedical data will be a challenging task. The rule-based and dictionary-based approaches are easily scalable, but they have to be 'fitted' manually with the dataset. Moreover, these systems fail to achieve the desired performance because they do not include the word in the training set, which causes Out-of-Vocabulary Problems (OOV). This problem affects the performance of medical domain analysis because of the frequent addition of new drugs, medicines, and perhaps new viruses [15]. Currently, deep learning-based approaches have gained more attention in the domain of NLP, which reduces the dependence on feature engineering. In order to overcome problems like feature engineering and out-of-vocabulary that arise in the previous traditional methods, a combination of the Long Short-Term Memory (LSTM), Conditional Random Field (CRF) model and the Recurrent Neural Network (RNN), Conditional Random Field (CRF) model was developed. The combinations achieved good performance in the results by using word vectors and character vectors, with improvements required in high-quality dictionary features. The NER method, using a neural network structure, has many advantages in the medical field. In this work, deep learning-based methods based on BI-LSTM-CRF-CNN with good entity recognition are used, and the results showed that there is an improvement in the performance.

Li *et al.* [16] surveyed various deep-learning schemes for NER. The deep learning based schemes achieve better accuracy but suffer from various challenges such as data annotation which is time consuming task, informal text & unseen entities. From the traditional approach to Deep Learning techniques have gained success in different ways, but recently deep learning has gained more and more consideration for the entity recognition task. The methods used in deep learning techniques diminish the issues of annotated data in Bio-NER.

This section provides a succinct discussion of current biomedical NER models that use deep learning. As mentioned before, the existing techniques such as rule-based NER, dictionary based NER and machine learning based NER systems suffer from various issues in NER. However, the biomedical NER is typically challenging because of the diversity of medical entities. Several deep learning-based techniques have been presented during the last few years.

Cho *et al.* [17] developed a deep learning model for NER. The biomedical word tokens are represented in this model using feature embedding in a combinatorial architecture. The basic idea of this model is obtained from Bi-LSTM with CRF, later it is enhanced by incorporating

two-different level representations of CNN and Bi-LSTM. Along with this, the biomedical word tokens are represented in this model using feature embedding in a combinatorial architecture. This helps to improve the long-term dependency problem in the LSTM model, resulting in improving entity recognition. Kim *et al.* [18] discussed that machine learning techniques require a significant amount of labeled corpora to achieve an efficient performance. Creating a huge corpus manually is time-consuming and not considered as an efficient solution. Thus, in order to construct the biomedical NER corpus and cover all relevant sub-domains, writers developed a novel method. At first, it is thought that only a modest amount of manually annotated data will produce a sizable amount of training data for the corpus. Finally, LSTM and CRF models are trained to enhance the performance.

Li *et al.* [19] proposed an RNN-based method known as WCP-RNN which uses RNN for learning and W, C, and P represent the word embedding, character embedding, and POS tags. Firstly, this model trains word embeddings using Skip-gram and CBOW, character embeddings are trained using CWE then both trained embeddings are concatenated together. Furthermore, linguistic features such as POS tags are also incorporated as priori word information which aims to enhance the final performance.

Crichton *et al.* [20] developed a supervised multi-task CNN model for biomedical NER. In this work, the authors introduced a single-task model, multi-output, and dependent multi-task model. These models are tested on 15 different types of medical data which contain various entities such as Anatomy, Chemical, Disease, Gene/Protein, and Species. Zhao *et al.* [21] presented Disease Named Entity Recognition (DNER) in Bio-NLP. The current deep learning systems employ the CRF layer to gather the label correlation data. A Multiple-label Convolutional Neural Network (MCNN) for NER was proposed by the researchers. This method substitutes a multiple-label scheme for the CRF layer (MLS). During the first stage, the embedding at the character, word, and lexical feature levels are concatenated. Later, several convolutional layers are applied to the concatenated embedding. The MLS model is then put into action to gather the correlation data. Yoon *et al.* [22] introduced CollaboNet which is designed by combining multiple types of NER models. In this model, the trained models are connected to each other which helps to reduce the false positives. This model uses Bi-LSTM-CRF, Word Embedding, and Character Level Word Embedding models to design CollaboNet. Neumann *et al.* [23] presented a fast and robust model known as ScispaCy for biomedical NLP. The ScispaCy is a python-based library which uses the spacy module. In Scispacy, the spacy library is retrained and two new functions are defined as `en_core_sci_md` and `en_core_sci_sm`. The `en_core_sci_md` module uses a large vocabulary and includes word vector for NER whereas the `en_core_sci_sm` model does not consider word vectors and has a smaller vocabulary.

Naseem *et al.* [24] reported several challenges in biomedical NER such as limited amount of training data,

one entity can refer to multiple objects depending on the context, and dependence on acronyms. The models fail to address these challenges, thus authors introduced a new approach which is called bioALBERT. This model focus on domain specific information extraction while training on large-scale biomedical corpora. Moreover, this model uses self-supervised loss function to extract the coherence among attributes and learn the context-dependent features with low memory requirement and faster training speed.

An *et al.* [25] reported that existing deep learning schemes fail to fully utilize global and multi-level semantic information in the data. To overcome with the issues of deep learning schemes, authors introduced a character level representation approach for clinical NER which embeds the character labels to enhance the feature representation. Furthermore, multi-head self-attention based bidirectional LSTM conditional random field is presented where multi-head self-attention is mixed with the medical dictionary for weight calculation which helps to capture the weights effectively. Thus, this model is trained on rich feature information resulting in better accuracy.

Ning *et al.* [26] addresses the issues of traditional machine learning algorithms in NER based Natural language processing. According to this study, these algorithms are highly dependent on the manual attributes. The current deep learning has been considered a promising approach in this field. Thus, authors presented a novel deep learning scheme called as Glove-BLSTM-CRF. Here, the Glove model is used for training where word vectors with their semantic attributes are trained, the Bi-LSTM model is used for training the word vectors with their morphological features. Finally, these models are combined to represent the final word.

Liu [27] proposed a Named Entity Recognition in equipment support field using Tri-Training algorithm and Text information extraction technology that supports word segmentation, classification, annotation, and information extraction.

Che *et al.* [28] proposed an effective biomedical named entity recognition using a temporal convolutional network with a conditional random field.

Motivation for the choice of specific architectures for the proposed model:

- The main challenge in entity recognition is that Biomedical texts are often complex and contain domain-specific terminology.
- The use of CNN and Bi-LSTM is motivated by the need to capture complex linguistic structures, domain-specific word formations, and contextual dependencies in biomedical documents.
- Local patterns include the character-level and word-level features that are important for recognizing entities in BioNER.
- Bi-LSTMs capture long-range dependencies by processing text in both forward and backward directions, which is crucial in biomedical NER.
- In BioNER, entities often have constraints related to their labels, CRF ensures that these dependencies are captured perfectly.

### III. MATERIALS AND METHODS

#### A. Data Sets

The MIMIC III free-text EHR dataset was processed using the SpaCy pipeline. This dataset contains a collection of discharge letters from the Intensive Care Unit (ICU), which consist of rich and detailed information regarding treatment. To fine-tune the NER task, we used a self-supervised, pre-trained model on the MIMIC III dataset of 2 million free-text patients' records. We have taken the word count, trained, and tested it, and the entity recognition techniques are applied and the performance is calculated.

A neural network method is proposed for biomedical NER named Att-CNN-Bi-LSTM-CRF (Attention-based Convolution Neural Networks, Bidirectional Long Short-Term Memory with a Conditional Random Field Layer). The first phase uses a word and character embedding scheme. Later, a combination of CNN, Bi-LSTM, and CRF models with an attention layer is used to augment the efficacy of the NER system. This paper proposes a biomedical named entity recognition model based on Convolutional Neural Networks (CNN), bidirectional long-short-term memory networks (Bi-LSTM), and Conditional Random Fields (CRF), along with character embedding, word embedding, and additional feature extraction.

#### B. Models Used

Convolutional Neural Network (CNN): The CNN component is often used for feature extraction. It can capture local patterns and structures in the text, making it suitable for detecting word-level features and patterns within a window of words.

Bidirectional Long-Short Term Memory Network (Bi-LSTM): Bi-LSTM models are excellent at capturing long-range dependencies and context information in sequential data. They can understand the relationships between words that are farther apart in the text.

Conditional Random Field (CRF): The CRF layer helps in modeling label dependencies and sequence-level constraints. It is beneficial for ensuring that the predicted entity labels form coherent sequences.

Additional Feature Extraction: This can include a wide range of features, such as part-of-speech tags, gazetteers, or domain-specific knowledge. These features can further enhance the model's ability to identify entities.

The inclusion of an attention layer in a Biomedical Named Entity Recognition (NER) model that already combines Convolutional Neural Network (CNN), Bidirectional Long-Short Term Memory Network (Bi-LSTM), and Conditional Random Field (CRF) is a powerful addition. An attention mechanism is designed to capture and emphasize relevant information within the input sequence while suppressing irrelevant or noisy information. In the NER, an attention layer can be used to focus on specific parts of the input text that are most informative for making entity recognition decisions.

The proposed architecture for biomedical NER is shown in Fig. 2. In this architecture, a sentence is given as input

to the network. In the phase1 of architecture, word embedding and character-level embedding are done using the CNN and Bi-LSTM models. The word-level embedding uses the pre-trained model, which contains several word vectors from the biomedical corpora. The word and character level weights are concatenated and processed as input to the simple, fully connected network. The outcome of this fully connected layer is used as the input to the Bi-LSTM model. The Bi-LSTM model uses forward and backward learning. The LSTM model helps to acquire comprehensive context information for the given sentence. The outcome of this is fed into the attention layer. This layer regulates the strength of other characters to the considered character of interest. In phase2, the tag scores that are obtained are processed through the CRF layer, and finally, the entities are obtained. The CRF layer finds the relationship between neighbor labels.

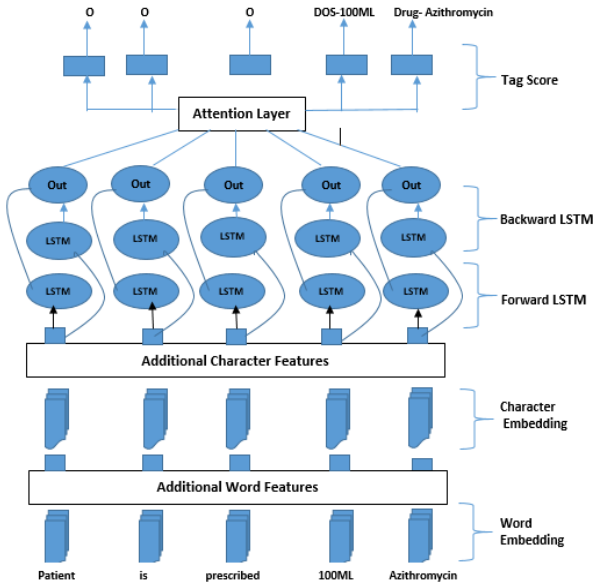


Fig. 2. Proposed architecture of biomedical NER.

### C. Different Layers

#### 1) Input layer

In layer one of the proposed architecture of biomedical NER, input is provided in the form of a sentence. The input sequence given is segmented into words. For example, we have taken input as a sentence as “Patient is prescribed 100 ML Azithromycin” which is segmented and given as input as depicted in Fig. 3. In general, let us consider that  $s$  denotes a clinical sentence which is expressed as:

$$s = w_0, w_1, \dots, w_n \quad (1)$$

where  $w$  denotes the word in a sentence which consists of total  $n$  number of words,  $w_t (1 \leq t \leq n)$  is the  $t^{th}$  character which is represented as  $x_t = [c_{wt}; r_{wt}]$  where  $c_{wt}$  denotes the word embedding,  $r_{wt}$  denotes the radical Convolutional neural networks (CNN) layer.

The CNN is an important stage of this model which is employed to obtain the local context information of input character of interest by applying four steps which are input

matrix, convolution operation, pooling and full connection shown in Fig. 3.

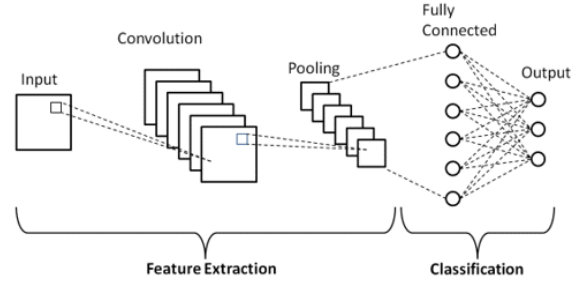


Fig. 3. CNN architecture.

The input layer has the input matrix which contains the entire information of current word and character by using word and character embedding’s. Moreover, it contains the position embedding of words.

In CNN, the operation is performed on the given input data by using filter or kernels of different size to extract the features are generate the feature map of input data. In this work, we considered  $M$  number of kernels with different size that are employed for feature extraction.

Let us consider that total  $L$  filters or feature maps are preset for each size of  $M$  the  $v^{th}$  filter of size  $u$  is expressed as  $W^{(u,v)}$  ( $1 \leq u \leq M, 1 \leq v \leq L$ ), then the convolution operation on input matrix  $Q$  is applied as:

$$F_i^{(u,v)} = \sigma(W^{(u,v)} Q_{[i:i+k-1]} + b^{(u,v)}) \quad (1 \leq i \leq m - k +) \quad (2)$$

where  $F_i^{(u,v)}$  represents the  $i^{th}$  feature obtained from the context matrix  $Q$  by using filter  $W^{(u,v)}$ ,  $\sigma$  denotes the component wise sigmoid function, denotes the element wise product and  $b^{(u,v)}$  is the bias vector. The extracted for each word by using this  $W^{(u,v)}$  can be denoted as:

$$F_i^{(u,v)} = (F_1^{(u,v)}, F_2^{(u,v)}, \dots, F_{m-k+1}^{(u,v)}) \quad (3)$$

The aim of pooling layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. In the work Max-pooling is used, it identifies the maximum element of feature map which are covered by the applied filtered. Hence, it generates the output in the form of feature map which contains the most significant features. This feature map is considered more robust when compared with previous used maps. We apply this operation of the generated feature set, to obtain the most significant features. It can be given as:

$$F_{max}^{(u,v)} = (F_1^{(u,v)}, F_2^{(u,v)} - F_{m-k+1}^{(u,v)}) \quad (4)$$

Fully Connected (FC) layer is acts as feed forward neural network which is the last layer of the CNN. The output of pooling layer or convolution layer is fed to the FC layer as input. All outcome of max pooling layer are concatenated to generate the local context of wt. This concatenation is given as:

$$g_t = (F_1^{(1)}, F_2^{(2)} Ft^{(M)}) \quad (5)$$

After processing through the CNN module, the original input data is redefined as  $g = (g_1, g_2, \dots, g_n)$  which is now considered as input to the LSTM model.

### 2) Long Short-Term Memory (LSTM Layer)

The LSTM layer considers the inputs from CNN as  $g = (g_1, g_2, \dots, g_n)$  and generates a new representation of the sequence as  $L = (l_1, l_2, \dots, l_n)$  where  $l_t = [l_{ft}; l_{bt}] (1 \leq t \leq n)$  considers the output from for backward and forward LSTM as  $l_{ft}$  and  $l_{bt}$ , respectively. The LSTM model contains one memory unit ( $c_t$ ) and three gates as input gate ( $i_t$ ), forget gate ( $f_t$ ) and output gate ( $o_t$ ) for each step  $t$ . Let us consider  $g_t, l_{t-1}, c_{t-1}$  as the initial inputs to LSTM at step  $t$ , then the steps produced by LSTM, can be represented as:

$$i_t = \sigma(W_{gi}g_t + W_{li}l_{t-1} + W_{ci}c_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{gf}g_t + W_{lf}l_{t-1} + W_{cf}c_{t-1} + b_f) \quad (7)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{gc}g_t + W_{tc}l_{t-1} + b_c) \quad (8)$$

$$h_t = \sigma(W_{go}g_t + W_{lo}l_{t-1} + W_{co}c_t + b_o) \quad (9)$$

where  $\sigma$  denotes the component wise sigmoid function, Denotes the element wise product,  $W_i, W_f, W_c$ , and  $W_o$  denotes the weight matrices,  $b_i, b_f, b_c$ , and  $b_o$  denotes the bias vector.

### 3) Attention layer

The attention layer is applied in order to generate the relativity of sentence with other biomedical words. For example, in our case we have considered the input sentence as ‘‘Patient is prescribed 100 ML Azithromycin’’ where using ‘‘100ML’’ separately cannot be considered as biomedical entity but combining with ‘‘Azithromycin’’ with ‘‘100ML’’ is used as biomedical entity. This task is accomplished using attention layer. Fig. 4 shows the model of attention layer.

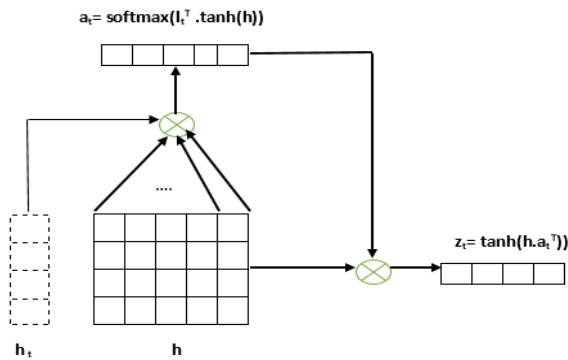


Fig. 4. Attention layer architecture.

Let us consider the exit state of LSTM layer  $L$  is considered as the entry state to the attention layer which generates a new sequence as  $z = (z_1, z_2, \dots, z_n)$  where  $z_t$  at step  $t$  can be computed as:

$$z_t = \tanh(L.a_t^T) \quad (10)$$

where  $\tanh$  denotes the activation function,  $L$  is the output matrix of LSTM layer,  $a_t$  denotes the weight vector of each word present in the sentence. The weight of each word is computed as:

$$a_t = \text{softmax}(l_t^T . \tanh(h)) \quad (11)$$

where SoftMax denotes the normalization function,  $l_t$  denotes the LSTM output at step  $t$ . The generated new sequence is used for label prediction in the CRF layer.

### 4) Conditional random field

Conditional Random Field (CRF) is widely adopted technique of machine learning. The main idea is to predict the sequence labels from the sequences of measurement by considering the sequentially of data. The output sequence of attention layer  $z = (z_1, z_2, z_3, \dots, z_n)$  is sent to the CRF layer as input to forecast the label sequence.  $y = (y_1, y_2, y_3, \dots, y_n)$ . Let us consider that we have a training data  $D$  and the parameters of CRF are denoted by  $\theta$  which are estimated by using log-likelihood as follows:

$$L(\theta) = \sum_{(s,y) \in D} \log p(y|vz, \theta) \quad (12)$$

where  $y$  represents the corresponding labels for input sentence  $s$ ,  $p$  denotes the conditional probability. We assume that the  $S_\theta(z, y)$  denotes the score of label sequence for given sentence. For this, the conditional probability ( $p$ ) can be computed by normalizing the sequence scores. Further, we incorporate a transition matrix  $T$  with an emission matrix  $E$  to compute the label score  $S_\theta(z, y)$ , as follows:

$$S_\theta(z, y) = \sum_{t=1}^n (E_{y_t, t}) + T_{(y_{t-1}, y_t)} \quad (13)$$

$E_{y_t}$  is the probability of that word belongs to label  $y_t$ ,  $T_{y_{t-1}, y_t}$  is the probability that word  $z_{t-1}$  belongs to label  $y_{t-1}$ .

Bidirectional Long Short-Term Memory-Conditional Random Field model

The unidirectional LSTM considers inputs from one direction i.e. data is fed from left to right. However, in Bi-LSTM model the data can be processed in both forward and backward direction which helps to store the information from both direction.

For given input sentence  $= w_0 w_1 \dots w_n$ , the hidden states are computed for forward and backward direction as  $h_t^1$  and  $h_t^2$ , respectively. These hidden states are concatenated as  $h_t = [h_t^1; h_t^2]$ . This helps to represent word token in each step  $t$ . The CRF layer is added above the Bi-LSTM layer. The CRF layer decodes the best tag with the help of state transition matrix. In aforementioned section, we have presented the model to compute the label score. After computing the final score, we apply softmax function to find the conditional probability of  $i^{th}$  word in the given sentence.

It can be expressed as:

$$p(y | x) = \frac{\exp(S([w]_1^T, [y]_1^T))}{\sum_{y \in Y(w)} \exp(S([w]_1^T, [y]_1^T))} \quad (14)$$

The training model optimizes the log probability during training in order to reduce training error. After finishing

the training, the best output is decoded with the final score. The final output is presented as:

$$\hat{Y} = \arg \max_{y \in \mathcal{Y}(w)} ( ([w]_1^T, [y]_1^T ) ) \quad (15)$$

Below is a simplified pseudocode to help you understand the basic structure of the proposed model.

Steps:

- Start by installing SciSpaCy and loading the necessary components.

```
import spacy
import scispacy
```

- Create a custom NER model that incorporates CNN, Bi-LSTM, CRF, and attention components. This involves implementing a custom pipeline component for SciSpaCy.
- Define custom layers for CNN, Bi-LSTM, CRF, and the attention mechanism.

```
nlp = spacy.load("en_core_sci_md")
```

```
ner = EntityRecognizer(nlp.vocab)
```

```
nlp.add_pipe(ner, name="biomed_ner", after="ner")
```

```
biomedical_ner = BiomedicalNER(nlp)
```

- Define the layers for CNN, Bi-LSTM, CRF, and the attention mechanism.

#### IV. RESULT AND DISCUSSION

The results compared with the existing models are shown in Tables I and II. The datasets and pre-processing of these corpora were the first steps in the experimental setup. The MIMIC III free-text EHR dataset is processed using the SpaCy pipeline. This dataset contains the collection of discharge letters from the Intensive Care Unit (ICU), which contains rich and detailed information regarding treatment. 100 samples were considered. Below given Table I shows the dataset details in terms of training and testing for considered entities.

TABLE I. DATASET DETAILS FOR TRAINING AND TESTING

Entity	Total	Trained	Tested
Dosage	422	6,908	2,681
Drug	162,578	26,832	10,575
Duration	593	970	378
Form	6,658	11,016	4,359
Frequency	6,280	10,293	4,012
Route	546	8,973	3,513
Strength Documents	6,695	10,924	4,230
Count	303	505	202
Words Count	957,972	1,585,743	627,771
Unique Words Count	27,602	35,763	21,729

An example of entities present in this dataset is given below:

- Dosage: It is the quantity or amount of medicine prescribed. For example, “15 ml” or “2”.

- Drug: It is the brand name of the medicine or its generic name, example “paracetamol” or “Avil”.
- Duration: It is the course time of the medication. Example, “5 days” or “for five days”.
- Form: It is the form of medicine that is prescribed, example, “capsule” or “tablet”.
- Frequency: It is the regularity of the intake of medicine, example, and “every 3 hours”, “daily” or “bid”.
- Route: It is the way of medicine intake in the body, example, and “injection”, “by mouth or PO”.
- Strength: It is the amount or hardness of the dose, example, “50 mg” or “0.02%”.

The complete implementation of the proposed workflow is done in Python language Python 3.7 and the SciSpaCy library is used for model training and development. Word2Vec embeddings pre-trained on EHR texts are used for the base Bi-LSTM-CRF model. The architecture of the Bi-LSTM-CRF comprises two bidirectional LSTM layers trailed by a CRF layer.

To fine-tune the NER task, a self-supervised, pre-trained model on the MIMIC III dataset of 2 million free-text patients’ records is used. 100 samples were considered for testing. The model predicts the center word given left-to-right in one phase and right-to-left in the next phase for context representations, i.e., instead of predicting an exact word matched from the vocabulary, the GloVe word’s vector is predicted using a static embedding table with a cosine loss function. Also, to bootstrap patterns and keywords in the MIMIC III dataset, we utilized a method similar to sense2vec. sense2vec has an advantage over word2vec in that it also performs learnings through words’ language annotations to avoid context ambiguity. The initialization of weights for CNN layers is done through a pre-trained language model instead of initializing any random weights. Various combinations of language model parameters are tested, such as row count and embedding tables’ width and depth of CNN layers, to check the best result. Also, the parameters of the CNN architecture, such as regularization type, batch size, dropout rate, and learning rate, are adjusted to get better results.

In order to evaluate the outcome of proposed approach, we measure precision, recall, and F1-Score which are computed based on the true positive, false positive and false negative as given below:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Here, TP denotes the true positive which shows the model correctly predicted the test class to its corresponding positive class, True Negative (TN) is shows that the model has correctly predicted the corresponding negative class. False Positive (FP) denotes that model has incorrectly predicted the positive class and False Negative

(FN) denotes the number of negative classes are predicted incorrectly.

Based on these parameters, we compared the performance of proposed model with state of art machine learning based techniques of NER like Conditional Random Field, Smooth Support Vector Machine, Baseline method, Convolution Neural Network, Recurrent Neural Network, Semi Markov method with the same dataset as mentioned in [28].

Conditional Random Field (CRF): This is the standard model for predicting the most alike sequence of labels for the input. Structural Support Vector Machine (SSVM): A classification technique for pattern recognition, Baseline Method: This method contextualize the results of the trained models. Semi-Markov: It easily constructs entity level features. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Deep learning architecture recognize patterns.

Table II shows the comparative analysis of different existing NER techniques with the proposed model in terms of precision, recall, and F1-Score.

TABLE II. COMPARATIVE ANALYSIS OF TECHNIQUES

Technique	Precision	Recall	F-Score
CRFs	82%	73%	77%
Baselines	87%	81%	84%
SSVMs	87%	84%	86%
Semi-Markov	87%	84%	85%
CNN	85%	81%	83%
RNN	85%	87%	86%
Att-CNN-Bi-LSTM-CRF	89%	87%	87%

In this experiment, the performance of the suggested technique in terms of recall, precision, and F1-Score is improved. In existing schemes, Smooth Support Vector Machine achieved better performance as it achieved a precision score of 87.38%; better recall was obtained using Recurrent Neural Network as 86.56%; and it achieved a better F-Score performance as 85.94%. Similarly, we measured the performance of each entity and compared the performance of the proposed model with the recurrent neural network-based model and the Bi-LSTM model. The comparative performance is presented in Table III, Figs. 5 and 6. For this comparison, we have considered drug, frequency, dose, duration, and route as entities.

TABLE III. ENTITY RECOGNITION RESULTS

Entity	Precision			Recall			F1-Score		
	RNN	Bi-LSTM	Proposed model	RNN	Bi-LSTM	Proposed model	RNN	Bi-LSTM	Proposed model
Drug	86%	87%	88%	90%	84%	92%	88%	86%	88%
Frequency	85%	82%	86%	86%	81%	89%	85%	82%	87%
Dose	88%	88%	90%	84%	83%	84%	86%	86%	88%
Duration	65%	74%	81%	79%	82%	83%	71%	78%	81%
Route	88%	89%	93%	92%	91%	93%	90%	90%	92%

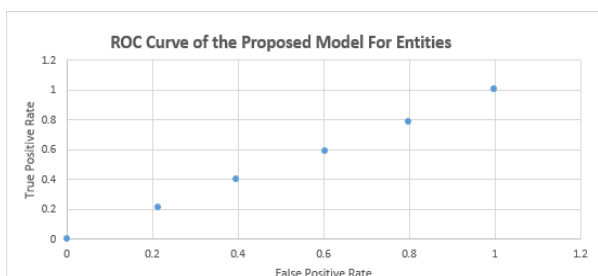


Fig. 5. ROC curve of the proposed model for entities.

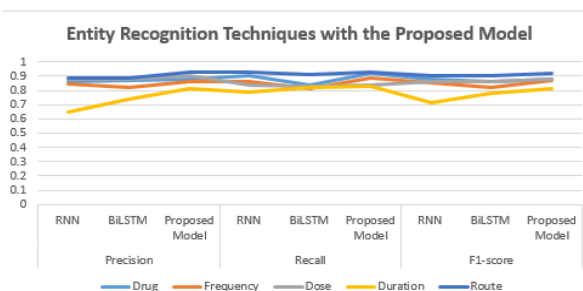


Fig. 6. Different entity recognition techniques with the proposed model.

The investigational results proved that the performance of a developed method is better in terms of F1-Score, recall, and precision when compared with the prevalent schemes.

The proposed model in Table III out performs RNN and Bi-LSTM by different factors. In BioNER the entities have different constraints and labels are dependent, in which the

Conditional Random Field ensures that the entity labels are coherent and is a sequence labeling model. Bi-LSTM. Bi-LSTM can capture long range dependencies and the context in which entities appear should be understandable. In biomedical texts, where information is scattered across lengthy documents, the attention mechanism helps the model identify and give more weight to words that are relevant to entities.

## V. CONCLUSION

Our proposed integrated Biomedical Named Entity Recognition Model based on Convolution Neural Network (CNN), Bidirectional Long-Short Term Memory Network (Bi-LSTM), and Conditional Random Field (CRF) provides significant advancements in BioNER.

The integrated model leverages the strengths of each component to effectively capture contextual information, learn label dependencies, and improve the accuracy of entity recognition. Here, character embedding, word embedding, and additional feature extraction tasks are considered.

In summary, adding an attention layer to the biomedical NER model, which already includes CNN, Bi-LSTM, and CRF, is a promising approach. It enhances the model's ability to capture and emphasize contextually relevant information, improving entity recognition accuracy, especially in the biomedical texts.



Careful model training, hyper parameter tuning, and evaluation on benchmark datasets are key steps in realizing the potential benefits of this comprehensive NER model.

The model is trained, validated, and tested to get better performance in terms of recall, precision, and F1-Score. A CNN, Bi-LSTM, and CRF model are trained using 1000 samples. Word embedding and character-level embedding using the CNN and Bi-LSTM models are done. The word-level embedding uses the pre-trained model that contains several word vectors from the biomedical corpora. The word and character level weights are concatenated and processed as input to the simple, fully connected network.

The outcome of this fully connected layer is used as the input to the Bi-LSTM model. The Bi-LSTM model uses forward and backward learning. The LSTM model helps to acquire comprehensive context information for the given sentence. The outcome of this is fed into the attention layer. This layer regulates the strength of other characters to the considered character of interest. In the next step, tag scores are obtained, which are processed through the CRF layer, and finally, the outcome is obtained in the form of entities.

Experiments are conducted and analyzed. The results of the proposed model show its performance is boosted compared to other NER approaches. The experimental analysis demonstrates that the suggested technique performs better in terms of recall, precision, and F1-Score. This performance improvement have significant practical implications across the biomedical and healthcare sector.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTIONS

Dr. Prabhudev Jagadeesh M. P. supervised the research development. Mrs. Pooja H. implemented the model and lead in writing the manuscript. All authors have read and agreed to the published version of the manuscript.

#### REFERENCES

- [1] D. Fushman, N. Elhadad, and C. Friedman, "Natural language processing for health-related texts," *Biomedical Informatics*, pp. 241–272, 2021. doi:10.1007/978-3-030-58721-58
- [2] X. Zhou, L. Yue, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 912–921, 2020. doi: 10.1109/TCBB.2020.2994780
- [3] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Named entity recognition and relation extraction: State-of-the-art," *ACM Computing Surveys (CSUR)*, vol. 54, no. 1, pp. 1–39, 2021. <https://doi.org/10.1145/3445965>
- [4] V. Kocaman and D. Talby, "Biomedical named entity recognition at scale," in *Proc. International Conference on Pattern Recognition*, 2021, pp. 635–646.
- [5] A. Kishwar and K. Batool, "Comparison of named entity recognition models based on neural network in biomedical," in *Proc. 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021, pp. 426–431. doi: 10.1109/IBCAST51254.2021.9393197
- [6] S. Ali, K. Masood, A. Riaz, and A. Saud, "Named entity recognition using deep learning: A review," in *Proc. 2022 International*

- Conference on Business Analytics for Technology and Security (ICBATS)*, 2022.
- [7] E. Karatzas *et al.*, "Darling: A web application for detecting disease-related biomedical entity associations with literature Mining," *Biomolecules*, vol. 12, no. 4, 520, 2022. <https://doi.org/10.3390/biom12040520>
- [8] R. M. Ghoniem and D. L. Elshewikh, "A novel genetic artificial bee inspired neural network model for drug name recognition," *Procedia Computer Science*, vol. 189, pp. 48–60, 2021. doi: 10.1016/j.procs.2021.05.069
- [9] N. Zhang *et al.*, "CBLUE: A Chinese biomedical language understanding evaluation benchmark," arXiv preprint, arXiv:2106.08087, 2021.
- [10] X. Li, H. Zhang, and X. H. Zhou, "Chinese clinical named entity recognition with variant neural structures based on BERT methods," *Journal of Biomedical Informatics*, vol. 107, 2020. <https://doi.org/10.1016/j.jbi.2020.103422>
- [11] A. F. Zipper *et al.*, "GoMi-A new gold standard corpus for miRNA named entity recognition to test dictionary, rule-based and machine-learning approaches," *Biorxiv*, pp. 1–13, 2021. <https://doi.org/10.1101/2021.10.18.464801>
- [12] Y. Xiong *et al.*, "Improving deep learning method for biomedical named entity recognition by using entity definition information," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–13, 2021. <https://doi.org/10.1186/s12859-021-04236-y>
- [13] H. Pooja and M. P. Jagadeesh, "A collective study of data mining techniques for the big health data available from the electronic health records," in *Proc. the 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, Bangalore, India, 2019, pp. 51–55. doi: 10.1109/ICATIECE45860.2019.9063623
- [14] Y. Li *et al.*, "BERTifying the hidden Markov model for multi-source weakly supervised named entity recognition," arXiv preprint, arXiv:2105.12848, 2021.
- [15] U. Naseem, K. Musial, P. Eklund, and M. Prasad, "Biomedical named-entity recognition by hierarchically fusing bioBERT representations and deep contextual-level word-embedding," in *Proc. 2020 International Joint Conference on Neural Networks*, 2020. doi: 10.1109/IJCNN48605.2020.9206808
- [16] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, pp. 50–70, 2020. doi: 10.1109/TKDE.2020.2981314
- [17] M. Cho, J. Ha, C. Park, and S. Park, "Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition," *Journal of Biomedical Informatics*, vol. 103, 2020. <https://doi.org/10.1016/j.jbi.2020.103381>
- [18] J. Kim, Y. Ko, and J. Seo, "A bootstrapping approach with CRF and deep learning models for improving the biomedical named entity recognition in multi-domains," *IEEE Access*, vol. 7, 2019.
- [19] J. Li *et al.*, "WCP-RNN: A novel RNN-based approach for BioNER in Chinese EMRs," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 1450–1467, 2020.
- [20] G. Crichton, S. Pyysalo, B. Chiu, and A. Korhonen, "A neural network multi-task learning approach to biomedical named entity recognition," *BMC Bioinformatics*, vol. 18, no. 1, 368, 2017.
- [21] Z. Zhao *et al.*, "Disease named entity recognition from biomedical literature using a novel convolutional neural network," *BMC Medical Genomics*, vol. 10, no. 5, 73, 2017. <https://doi.org/10.1186/s12920-017-0316-8>
- [22] W. Yoon, C. H. So, J. Lee, and J. Kang, "Collabonet: Collaboration of deep neural networks for biomedical named entity recognition," *BMC Bioinformatics*, vol. 20, no. 10, 2019.
- [23] M. Neumann, D. King, I. Beltagy, and W. Ammar, "Scispacy: Fast and robust models for biomedical natural language processing," arXiv preprint, arXiv:1902.07669, 2019. <https://doi.org/10.48550/arXiv.1902.07669>
- [24] U. Naseem, M. Khushi, V. Reddy, S. Rajendran, I. Razzak, and J. Kim, "Bioalbert: A simple and effective pre-trained language model for biomedical named entity recognition," in *Proc. 2021 International Joint Conference on Neural Networks (IJCNN)*, 2021.
- [25] Y. An, X. Xia, X. Chen, F. X. Wu, and J. Wang, "Chinese clinical named entity recognition via multi-head self-attention based Bi-

- LSTM-CRF,” *Artificial Intelligence in Medicine*, vol. 127, 2022. doi: 10.1016/j.artmed.2022.102282
- [26] G. Ning and Y. Bai, “Biomedical named entity recognition based on Glove-BLSTM-CRF model,” *Journal of Computational Methods in Sciences and Engineering*, vol. 21, no. 1, pp. 125–133, 2021.
- [27] C. Liu, Y. Yu, X. Li, and P. Wang, “Named entity recognition in equipment support field using tri-training algorithm and text information extraction technology,” *IEEE Access*, vol. 9, pp. 126728–126734, 2021.
- [28] C. Che *et al.*, “Fast and effective biomedical named entity recognition using temporal convolutional network with conditional random field,” *Mathematical Biosciences and Engineering*, vol. 17, no. 4, 2020.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.