Journal of Computers and Intelligent Systems
Vol. 4, No. 1, 2026, pages 01 - 13
Journal Homepage:
https://journals.iub.edu.pk/index.php/JCIS/

Information Extraction From Electronic Health Records Using Deep Learning

Fatima Masood[1], Muhammad Summair Raza[1], Muhammad Ramzan[1], Muhammad Sulman Bashir [2], Mahwish Ilyas [3]
[1] Department of Software Engineering, University of Sargodha
[2]Department of Computer Science, Virtual University of Pakistan
[3]Department of Computer Science, University of Raul, Mandi Bahauddin Pakistan

ARTICLE INFO

Article History:

Received:	October	11, 2025
Revised:	January	17, 2026
Accepted:	March	10, 2026
Available Online:	March	25, 2026

Keywords:
Information Extraction
Deep Learning
Electronic Health Records
LSTM
BiLSTM
NLP
Named Entity Recognition

Classification Codes:

Funding:
This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

Information extraction involves the automatic extraction of structured data from its unstructured or semi-structured form. Electronic Health Records contain useful information in an unstructured form, on which we can apply an information extraction strategy to extract structured information. Many traditional methods are available for information extraction, including both manual and rule-based information extraction. The rule-based extraction is simple but inadequate and does not cover all aspects. The manual extraction, on the other hand, is effective but very slow, costly, and not scalable. The deep learning models have recently shown excellent performance when applied to various Natural Language Processing tasks. In this study, we have presented a deep learning framework for multi-label medical entity extraction from Electronic Health Records to overcome the problem in which a medical term may represent numerous clinical roles. Additionally, the proposed approach supports real-time entity prediction, allowing users to input terms and instantly receive comprehensive insights, including diseases, symptoms, diagnostic tests, and treatments, thereby offering a practical tool for clinical decision support. Firstly, it preprocesses the EHR data and then applies deep learning models and the Runtime Context Identifier to extract the information. The results have shown that the proposed model achieved an average F1-score of 91% across five diverse EHR datasets, demonstrating robust overall performance.

Corresponding Author's Email:

Citation:

1. Introduction

The introduction of the social web has resulted in a significant information overload making it difficult to accurately extract information from huge volumes of data. To solve this issue Information Extraction (IE) has been used as one of the prominent techniques to extract and formalize structured information from non-structured documents. Such structured formats are used in various applications, including Information Retrieval (IR), question answering, knowledge bases, reasoning, and decision-making support. Named Entity Recognition (NER) [1], [2], [3], sentiment analysis [4], [5], and text classification [6],[7], [8] are various methodologies used as part of Information Extraction to extract and identify specific clinical entities [9], assess sentiments of patients and classify clinical data.

Electronic Health Records (EHRs) are digital versions of patients' data that contain medical information [10]. These records store a lot of unstructured clinical data, including prescriptions, current diagnosis, medical history, demographic information, etc. However, this unstructured format is too difficult to use in computational analysis. Therefore, the first step for constructing a medical-domain-specific knowledge graph is IE from these unstructured clinical EHRs, which can be beneficial for many fields, i.e., disease diagnosis, decision-making, prognosis prediction, and risk prediction.

Information in EHR may exist in textual, tabular, or image form. However, clinical notes mostly exist in an unstructured textual format, making the IE process extremely important for extracting the structured information from these EHRs. The existing approaches for IE from clinical notes include manual information extraction and rule-based information extraction. Manual extraction is effective but expensive, slow, and also not scalable. It is an insufficient, error-prone, inefficient, labor-intensive, and hectic approach for IE from clinical notes in EHRs. Rule-based approaches, on the other hand, include rules and an interpreter but may extract information with different meanings from the same patterns. Moreover, these rules do not cover all the aspects and give results with low precision. So, Deep Learning (DL) based IE approaches have become prominent these days to capture structured data through learning semantics. DL has been successfully applied to solve complicated problems, as it uses multilayered neural networks. DL models have achieved a high level of performance in different Natural Language Processing (NLP) tasks in recent years. In this research, we have presented a DL method to extract structured data out of unstructured EHRs to minimize manual extraction of clinical data. Additionally, it deals with run-time entity extraction using Runtime Context Identifier (RCI), which uses certain rules and patterns to extract the information that a trained NER model cannot extract. It facilitates more accurate and time-efficient information processing, which can help clinical assistants make better decisions and contribute to the overall increase in healthcare delivery. In several reviews, the application of deep learning in general EHR data analysis has been summarized [11], [12], [13], [14]. The key feature of deep learning is to combine NLP with neural network architecture for structured IE from unstructured medical data.

This research proposed a BiLSTM-based NER model (BNER) that firstly preprocesses the EHR data, including tokenization, stopword removal, text normalization, stemming, and lemmatization. Then, it applies deep learning models to extract information. In addition, RCI is used for runtime prediction of a term by returning its medical category (e.g., symptom, test, condition) along with relevant associated information. In order to justify the proposed approach, five datasets were used for experimentation purposes. Results were analyzed and compared with other approaches using measures like precision, accuracy, and recall. The research contributions are as follows:

- It provides an efficient approach for Information Extraction from unstructured EHRs
- It provides a mechanism to extract the information using RCI, which can further identify the patterns skipped by DL based models.

2. Related Work

Ying et al. [15] proposes a novel Question Answering (QA) pipeline to extract data from EHRs. The approach solves the problem of scaling training data by automatically generating QA pairs from annotated EHRs. The method starts with cleaning up the text and analyzing its dependency tree, and then makes use of a QA model to answer different clinical questions. Finally, the model is refined to increase its practical use in real clinical situations. Even though it is a new approach, the method faced challenges due to the wide range and complexity of the questions handled in extraction. For example, questions about diagnoses and disease classification proved to be the hardest because of the complex language and inadequate training data.

Jain et al. [16] worked on the RadGraph Benchmark, which is a deep learning model for finding both entities and relationships inside chest X-ray reports. They worked with the RadGraph dataset, which is made up of structured annotations for both clinical findings and their relationships. There were some issues cited, even though the model did well in relation extraction. The reports included in the dataset were mainly from a limited number of U.S. chest X-ray hospitals, making it hard to say if the model will work for other places and kinds of reports. Also, variations in how findings were annotated, along with having fewer CheXpert patients from insufficient groups, increased challenges for the model's performance.

Dada et al. [17] studied the methods of manipulating the BERT-based models to derive information from unstructured radiology reports, which also have natural language queries. The approach to enhance the F1 score on answering questions involved training on about 850,000 unlabeled reports and then using a medical encyclopedia. The results of the performance were positive; however, the model failed to be very effective when dealing with several or conflicting pieces of information. Moreover, the logic behind the model was not easy to interpret, and it might prevent its implementation in the clinical setting. The research showed that transparency and explainability are key factors for the use of QA models in healthcare.

Isabella C. and colleagues [18] aimed to use the Large Language Model Meta AI-2 (LLaMA-2) to identify clinical features associated with decompensated liver cirrhosis in MIMIC-IV records. They used approaches that relied on zero- or one-shot learning, and their approach did not consider model fine-tuning. They were employed to use prompt engineering techniques and various model sizes, and compared to a gold standard determined by three medical professionals. Although it was good at finding explicit features, the method was not as effective with features that were only in the text. The study also found that LLMs are biased as well (e.g., assuming that a patient is a woman based on statistical prevalence and not based on textual clues), which is a major ethical issue when used in a clinical context.

Lio et al. [19] tested the performance of general-purpose large language models (LLMs) such as LLaMA-2-7B and Generative Pre-trained Transformer-4 (GPT-4) on their capabilities to work with critical care reports. MetaMap 2020a was used to extract concepts based on critical care reports, and the CADA tool was used to annotate the concepts by the team of clinicians. The models were assigned the role of identifying meaningful concepts, the judgment of relevance of each concept, and negated ideas. Low-rank adaptation (LoRA) was used to fine-tune the models, and the 300 GPT-4 responses were reviewed qualitatively. The main positive aspect of this study is that it is done by layering clinical reasoning assessment. Nevertheless, the data were all in a single center, which might not be generalizable. To make matters worse, using third-party LLMs to do this will bring issues of data safety, which are yet to be resolved.

Han et al. [20] worked on extracting Social Determinants of Health (SDOH) from free-text EHR data by applying deep learning-based NLP models. Experts guided the building of an SDOH taxonomy, and three models were evaluated by the study. Researchers used a Convolutional Neural Network (CNN), a BiLSTM, and a BERT classifier. The database consisted of 3,000 clinical records that were labeled by humans. BERT was better than standard machine learning mechanisms for accuracy and recall, according to the results. Still, since it was only conducted in one healthcare system, it is not clear whether the results can be trusted elsewhere. In addition, the biases that arose during manual annotation can affect how well the model learns and how it is used in later applications.

J. Song et al. [21] made use of a local-global memory network to record the development of individual patients alongside population-wide patterns for mortality prediction purposes. The model uses LSTM-based recurrent layers and uses data from four intensive care units in Denmark for training. Patient-specific information was recorded by the local memory, and the global memory picked up on general disease patterns. Even though the model could make accurate predictions, its lack of interpretability makes it hard for doctors to trust and use it.

Park et al. [22] used Gated Recurrent Units (GRUs), a kind of recurrent neural network, to predict bacteremia in surgical patients by analyzing time-stamped records from 36,023 medical records at a tertiary hospital in South Korea. It observed patient records from ICU and general wards all the time to try to identify bloodstream infections as soon as they began. It was accurate, yet the model was not interpretable since the inner workings of the Recurrent Neural Network (RNN) were impossible to understand. Clinicians were unable to check or interpret the prediction of the model, making it less likely the model could be used in real hospital environments.

HiTANet, the Hierarchical Time-Aware Attention Network, was introduced by Luo et al. [23] to help with predicting clinical risks from EHR data. This model made use of embedding that considered time to model how patient information changed over time. When tested on three actual health records datasets, the model achieved F1 metrics and made risk prediction for scalable diseases. Even with its merits, the method had problems with data irregularity that often occurs in EHRs and remained difficult to interpret, which made it difficult for clinicians to evaluate.

Table 1 gives the related papers, methodology, strengths, and gaps used in that paper.

Table. 1. Related Work Representing Methodologies, Strengths, and Gaps of Existing Literature.

Paper	Methodology	Strengths	Gaps
[15]	QA techniques for extraction of information from EHRs	Excellent performance in information extraction, capable of handling few-shot or zero-shot settings, emphasizes the importance of transfer learning	It cannot addresses diverse question types and constrained answer options in EHR extraction.

		and each component in the proposed pipeline.	
[16]	Information extraction schema for entities and relationships from radiology reports using the development of RedGraph.	1- High coverage of clinically relevant information in reports. 2- Achieved a micro F1 of 0.82 for relation extraction on the MIMIC-CXR test set.	1- The schema employed does not reflect information on clinical context in radiology reports, except for some parameters. 2- Cases of ambiguity in reports.
[17]	Use of BERT models on unexplained radiology reports and labeled reading comprehension datasets, and proceeding with pre-training on radiology data sets and a medical encyclopedia.	1- 96.01 percent accuracy of identification of unanswerable questions. 2- An Effective method of retrieving data out of radiology reports by eliminating the use of manual searches.	It was challenging to compare reports from various radiologists due to their different styles and word choices.
[18]	Created an open-source pipeline based on a locally deployed Large Language Model (LLM), named it Llama 2, to extract quantitative data from clinical text.	Shown high success in finding major clinical variables such as liver cirrhosis, ascites, abdominal pain, dyspnea, and confusion in the medical history of the patient.	LLMs' decisions may lead to certain potential biases due to a probabilistic approach.
[19]	MetaMap and clinician annotations were applied on MIMIC-III nursing notes, evaluated by LLMs and fine-tuned models, with performance assessed through qualitative review and statistical analysis.	A comprehensive approach was used to assess the proficiency of LLMs in processing real-world clinical notes from adult critical care settings.	Using single-center data may limit generalizability to other centers.
[20]	Developing a framework to create a set of SDOH categories based on existing taxonomies and expert input then Evaluating three deep learning models: CNN, BiLSTM, and BERT-based Trained models on a manually annotated dataset of over 3,000 clinical notes.	Demonstrated improved performance of deep learning models over traditional machine learning baselines in classifying SDOH mentions BERT-based model achieved the best overall performance.	1- Limited to a single healthcare system. 2- Validation across different clinical settings is needed to assess the generalizability of the approach.
[21]	Utilizes local-global memory neural networks (based on LSTM) for mortality prediction by learning both individual patient patterns (local memory) and global disease evidence (global memory) from ICU data.	Captures both personalized and generalized information for better predictive performance; designed for longitudinal patient data.	1- Lack of model interpretability (black-box nature). 2- Limited insight into decision rationale reduces clinical trust.
[22]	Applies RNNs (specifically GRUs) on time-series EHR data from over 36,000 surgical inpatients to detect bacteremia; data sourced from ICU and general wards.	Real-time, continuous monitoring; effective for infection prediction in dynamic clinical environments.	1- Model opacity makes it difficult to understand prediction logic. 2- Limited to data from a single hospital, affecting generalizability.

[23]	Implements a transformer-based hierarchical attention network incorporating temporal information from EHRs to enhance risk prediction accuracy across three real-world datasets.	Captures temporal dependencies; performs well on irregular clinical data; scalable for risk prediction tasks.	1- Struggles with irregularities in EHR documentation. 2- Lack of model explainability; decision-making process remains opaque.

3. Proposed Methodology

This section discusses the datasets and methodology used in our research.

3.1 Datasets

The datasets that are utilized in this study are publicly available clinical text datasets based on a variety of medical cases and EHRs. The data can be found on Kaggle [24], under the name of Medical Text for Text Classification. We divided the data set into five sections and named them Fold_1, Fold_2, Fold_3, Fold_4 and Fold_5. Each dataset consists of annotated entities, with several clinical categories (TEST, TREATMENT, SYMPTOM, CONDITION, and PROBLEM), allowing multi-label entity extraction and evaluation in diverse medical settings.

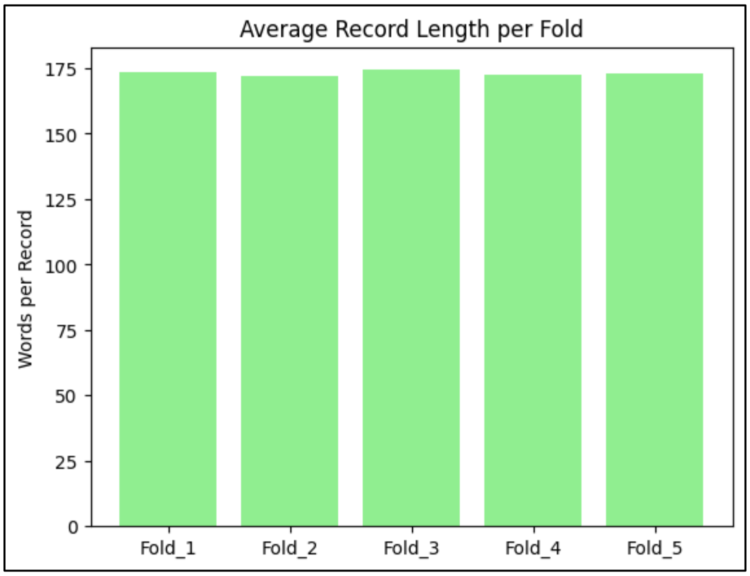

Figure. 1 (a) Average Record Length per Fold.

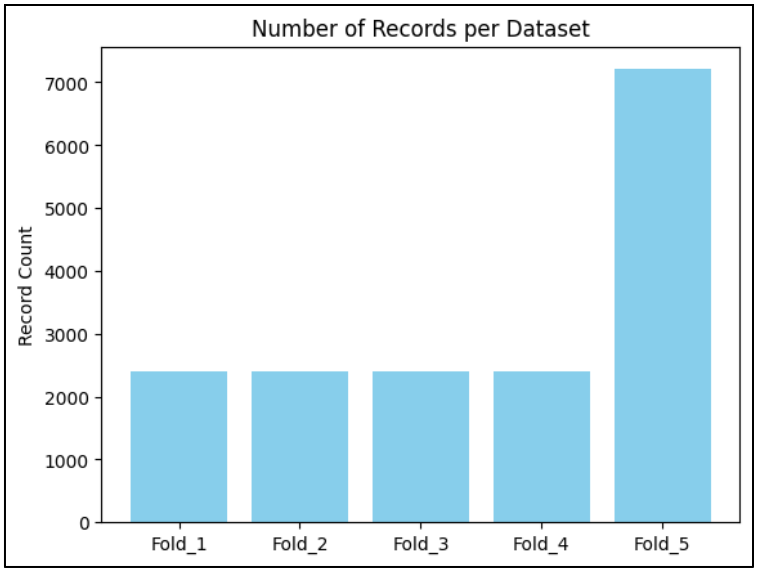

Figure. 1 (b) Number of Records per Fold.

Figure. 1 (c) Word Clouds of five Datasets.

In order to explain the features of the datasets better, Figure 1 (a) shows the average length of the records of each of the five datasets and reveals fluctuations in the textual density within different sources. Figure 1 (b) shows the number of words in each dataset, which provides information about the size and distribution of the datasets. Such visualizations helps in evaluating the quantity of clinical entities utilized in the research. Moreover, Figure 1 (c) presents word clouds of each dataset, which includes the most common terms, highlighting important medical concepts. These statistics combined in a comprehensive picture of the linguistic and contextual diversity within the data.

3.2 Model Design

The proposed methodology follows an iterative approach for real-time medical entity prediction from clinical text. The entire process is divided into four main phases, i.e., Data Preparation, Information Extraction, Prediction, and Results. Figure 2 shows the framework of the proposed method with details of all four phases.

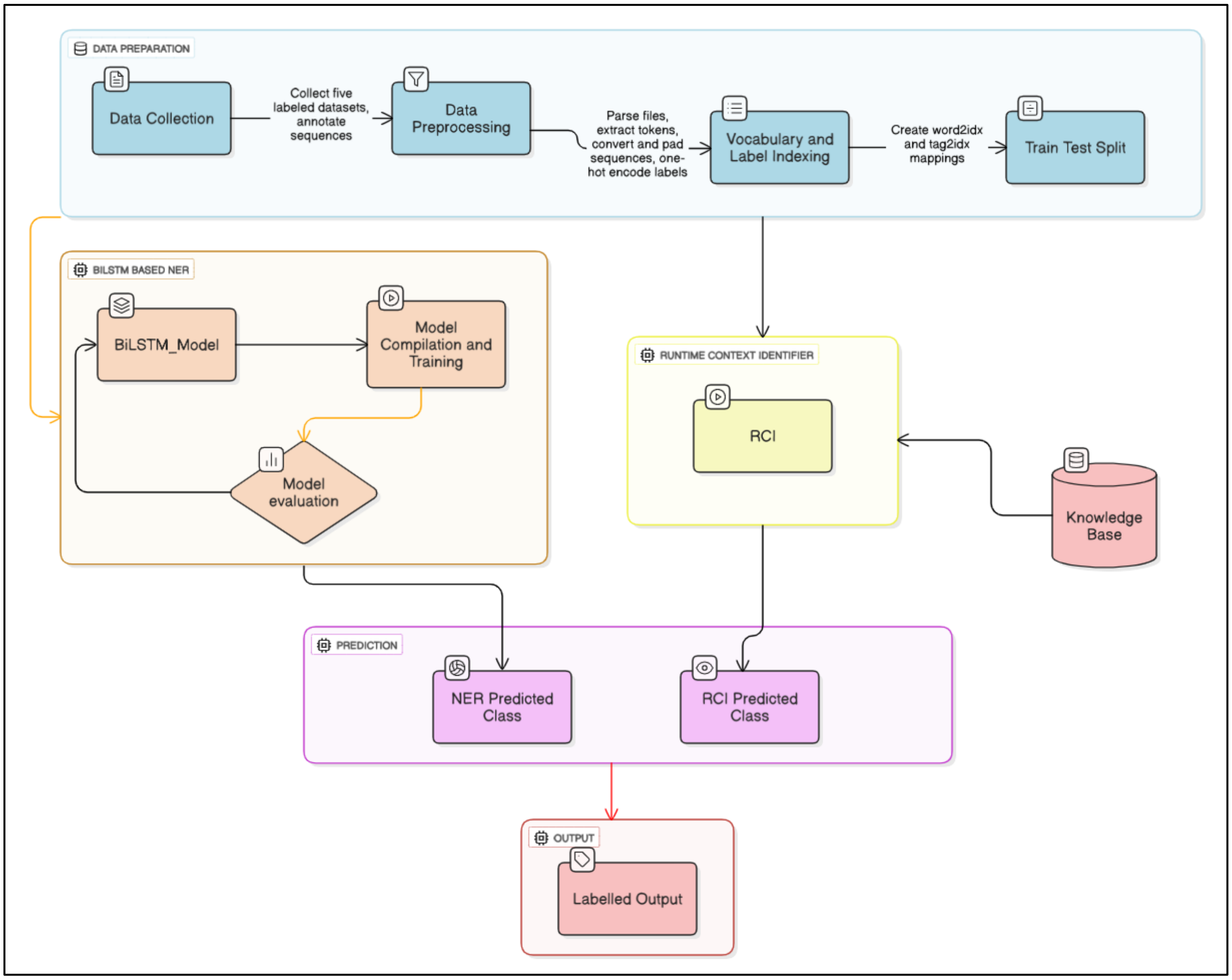

Figure. 2. Framework Diagram of Proposed Methodology.

3.2.1. Data Preparation and Preprocessing

During the initial stage of data preparation, five labeled datasets were tagged with the medical entities that were relevant. These datasets were preprocessed with a preprocessing pipeline that involved the text preprocessing steps such as token extraction, sequence padding, and feature encoding and labeling. The step in this stage was to divide the dataset into training and testing in an 80/20 proportion and set up the data to develop the model.

3.2.2. Information Extraction

During this step, we work on the recognition and classification of medical objects in the processed clinical text. One of the most important tasks in any clinical NLP pipeline is information extraction because it bridges raw textual data and structured, machine-readable representations of medical knowledge. In order to accomplish this, there are two complementary strategies used: BNER and RCI. Whereas BNER concentrates on labeling sequences on a token level, RCI extends it by providing an inference of relationships among recognized entities to provide a level of semantic richness that is especially useful in complicated clinical stories. These two approaches combine into a powerful extraction framework that can accommodate linguistic diversity and the domain-specific vocabulary of medical text.

3.2.2.1. BiLSTM-based NER (BNER)

During this phase, the BNER is trained using preprocessed data to predict the classes of different medical entities. The model is compiled and trained to predict the class of an entity. The BNER model is trained on the preprocessed dataset described in earlier stages of the pipeline. During training, the model learns to map input token sequences to a predefined set of entity class labels using the BIO (Beginning, Inside, Outside) tagging scheme, which allows the model to capture multi-token entity spans with precision. The bidirectional nature of the LSTM means that for each token, the model considers not only the left-to-right context but also the right-to-left context, which is particularly important in medical language, where the meaning of a term can shift significantly depending on what follows it. Figure 3 represents the NER Pipeline used in our study and its several components as a structure: User, System, DataLoader, Preprocessor, Splitter, Model, and Evaluator.

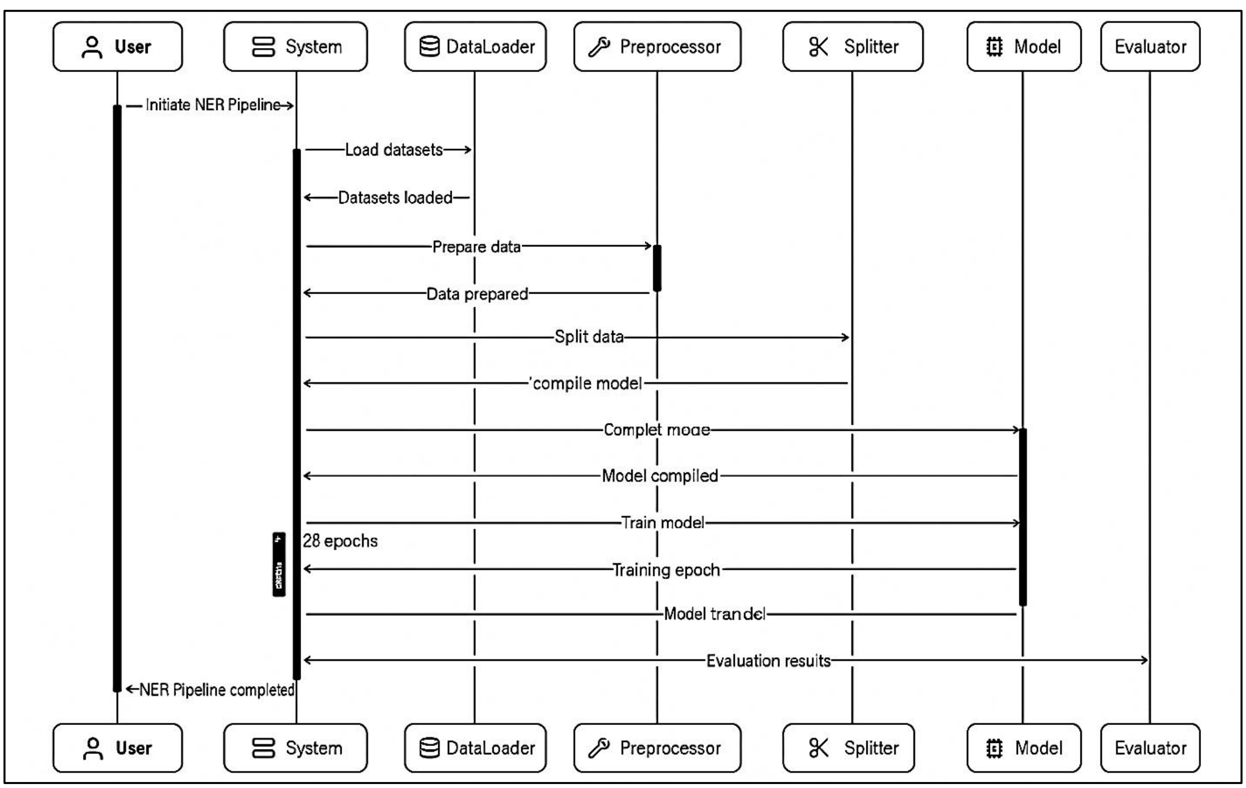

Figure. 3. Detailed High-Level Architecture by using NER Pipeline with BiLSTM.

Following model training, the evaluation was performed using four performance metrics. Table 2 represents the performance evaluation measures along with their mathematical formulas.

Table. 2. Performance Evaluation Measures.

Sr#	Metrics	Formula
1	Precision	$Precision = \frac{TP}{TP+FP}$
2	Recall	$Recall = \frac{TP}{TP+FN}$
3	F1-Score	$F - Score = \frac{2*precision*recall}{precision+recall}$

The output was then post-processed to generate the presentable results. The entire training process was repeated a number of times to enhance the performance and accuracy by optimizing different parameters.

3.2.2.2. Runtime Context Identifier

A key feature integrated into this stage is the RCI, a rule-based medical knowledge engine designed to complement the NER model by providing real-time, context-aware predictions of medical entities. It is a dynamic engine to deal with the runtime entities, and is capable of taking new rules at runtime and then extracting the information related to the medical entity. The RCI system operates independently of the training limitations of the NER model. It uses a comprehensive, custom-built dictionary that maps a wide range of medical terms, including diseases, symptoms, diagnostic tests, and treatments, to their associated clinical attributes.

3.2.2.2.1. RCI Functional Overview

The RCI is an advanced post-processing and decision-refinements layer that is automatic and can be generated once the BiLSTM model has generated its raw token predictions. RCI uses explicit medical knowledge, dictionary evidence, and context-sensitive rules in addition to the learned patterns of the neural model to prove or modify or refine predicted entity labels. This provides RCI, especially useful with the ambiguous, infrequent, or invisible medical terms that cannot lead to the correct inference. The RCI does three major functions: Entity Confirmation, in which the model is validated or overridden with dictionary matches; Context Resolution, in which the sentence around the target is analyzed to resolve semantic ambiguities; and Entity Enrichment, in which related medical information, such as diseases related to symptoms, medications related to indications, or tests related to common results is attached.

3.2.2.2.2. RCI Architecture

The RCI architecture is comprised of four closely related modules that process the refinement of entity predictions in a logical sequence. The first one is the Lexical Matcher, which matches after the exact matching, stem-based matching, and fuzzy matching, a single word expression and multi-word expression. It incorporates a large custom dictionary containing synonyms, acronyms, misspellings, and morphological variants and allows one to recognize medical entities well even when they appear in other forms. The second one, the Context Analyzer, functions on a user-configurable context window and identifies clinically significant co-occurrence patterns, such as symptoms appearing around diseases, medications appearing around dosage keywords, or diagnostic tests appearing around typical result terms. The analyzer makes sure that the tokens are interpreted appropriately according the surrounding context, by using a set of predefined templates of contextual rules. The third module is the Rule Engine which is a dynamic and programmable system that enables the addition of new rules, removal of rules and modification of rules with no model retraining. It holds rules of priority, conditional, override and weighted rules with confidence with the flexibility of controlling decisions. The Knowledge Mapper is the last module, which relates recognized entities to medical relationships including symptom disease relationship, disease treatment relationship, medication indication relationship, and test procedure relationship, and thus adds more semantic knowledge to extracted entities and meets the needs of clinical application.

```
# Example usage:
sentence = input("Enter a sentence: ")
medical_RCI(sentence)

Enter a sentence: John is having constant headache

🧠 'headache' is classified as: Symptom
🩺 Possible Diseases: ['migraine', 'tension headache', 'brain tumor']
🧪 Recommended Tests: ['CT scan', 'MRI']
```

Figure. 4. Working of the Runtime Context Identifier (RCI).

The functionality of the RCI is shown in Figure 4 where the system is run in real time, and the user inputs any clinical sentence at run time. After a sentence is typed, e.g., John is having constant headache, the RCI is working on the input and automatically recognizes the medically relevant words, including symptoms, diseases, medications, or tests. This example shows that the RCI recognizes the concept headache, and classifies it as a Symptom according to its dictionary and rule-based logic. Then it adds value to this identified entity by showing other clinical data, including potential diseases related to headache (e.g., migraine, tension headache, brain tumor) and diagnostic tests (e.g., CT scan, MRI) that are suggested. This shows how the RCI is capable of identifying medical entities in real-time as well as offering contextual information that complements clinical reasoning so it is an efficient and dynamic element in the information extraction pipeline.

3.3. Predictions

Finally, during the Prediction phase, the trained model will be used to predict medical entities live. The model, aided by the RCI, can deliver relevant medical entities by having the users input text in real time. Result dashboards visualize the predictions and provide the charts and performance plots to be interpreted and analyzed by the users.

4. Results and Discussions

This section provides the details of the experimental setup, results, figures, and discussions. Table 3 presents the performance metrics, F1-Score, Precision, and recall, of the BNER across five different medical datasets. Overall, the model achieved high performance with average scores of 0.918 for F1-Score, 0.922 for Precision, and 0.918 for Recall, indicating that it is effective in identifying medical named entities across varying data sources. Table 4 presents the comparison of the performance metrics of a few representative approaches.

Table. 3. Comparison of F1-Score, Precision, and Recall of 5 Datasets and their Average.

Dataset	F1-Score	Precision	Recall
1	0.91	0.92	0.91
2	0.91	0.91	0.91
3	0.94	0.94	0.94

4	0.88	0.89	0.88
5	0.95	0.95	0.95
Average(1-5)	0.898	0.932	0.898

Table. 4. Comparison of F1-Score, Precision, and Recall of 5 Datasets of Existing Three Approaches.

Paper	Dataset	F1-Score	Precision	Recall
RadGraph: Extracting Clinical Entities and Relations from Radiology Reports [16]	Dataset1	0.701	0.696	0.411
	Dataset2	0.712	0.752	0.454
	Dataset3	0.704 0.711	0.723	0.432
	Dataset4	0.732	0.771	0.443
	Dataset5	0.692	0.676	0.410
Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing [20]	Dataset1	0.711	0.657	0.662
	Dataset2	0.732	0.725	0.729
	Dataset3	0.752 0.723	0.712	0.719
	Dataset4	0.733	0.735	0.742
	Dataset5	0.745	0.713	0.718
Local–Global Memory Neural Network for Medication Prediction [21]	Dataset1	0.686	0.657	0.662
	Dataset2	0.676	0.725	0.729
	Dataset3	0.652	0.712	0.719
	Dataset4	0.677	0.735	0.712
	Dataset5	0.645	0.725	0.729

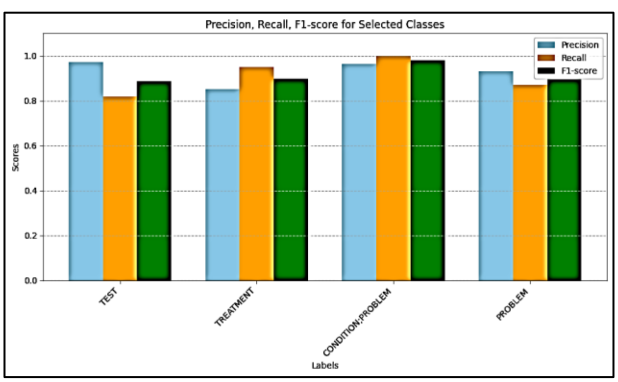

Figure 5 (a) Evaluation Metrics of Dataset 1

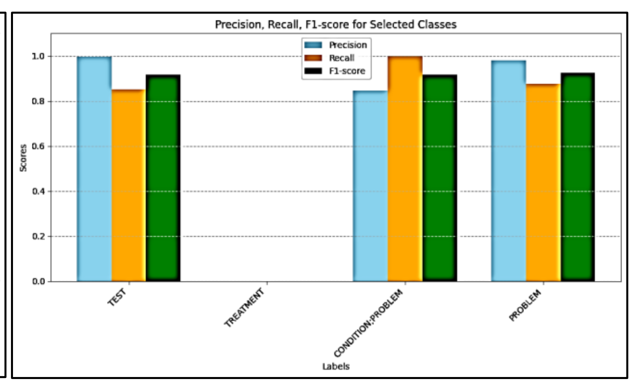

Figure 5 (b) Evaluation Metrics of Dataset 2

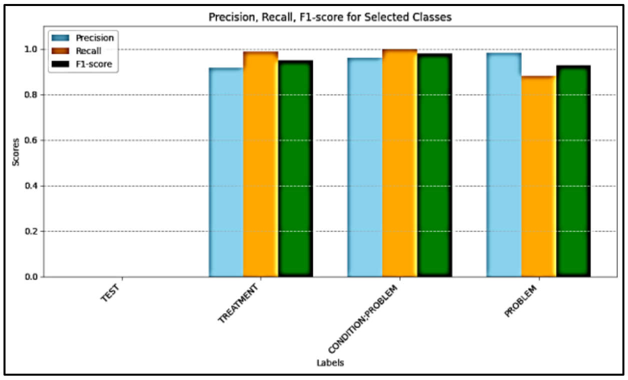

Figure 5 (c) Evaluation Metrics of Dataset 3

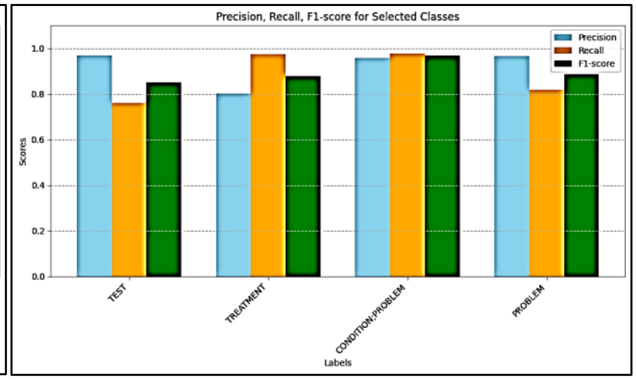

Figure 5 (d) Evaluation Metrics of Dataset 4

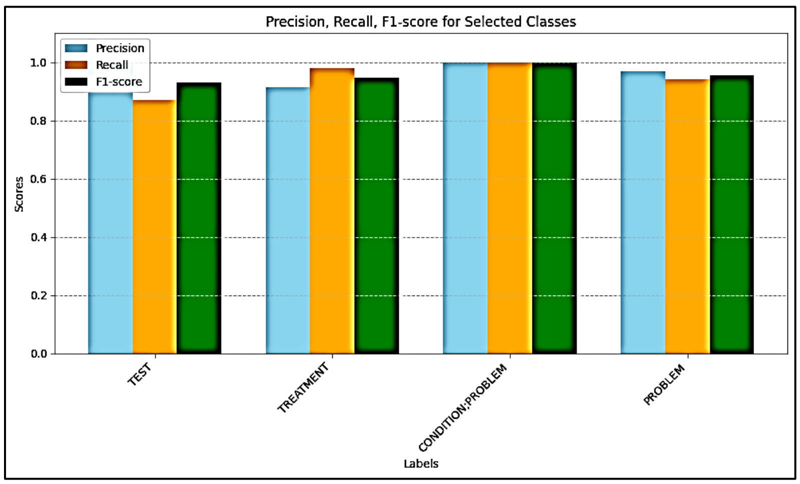

Figure 5 (e) Evaluation Metrics of Dataset

Figure 5 demonstrates the precision, recall, and F1-score of chosen entity classes in selected five datasets with different configurations of models. These findings support the performance and quality of the proposed approach to accurately extract clinically relevant entities.

5. Conclusion

In this research we proposed BNER to extract information from unstructured EHRs. It helps to extract the information using RCI, which can further identify the patterns skipped by BNER. Firstly the dataset was loaded, preprocessed, and split for training the embedded layers of BNER for information extraction from EHRs. Then the model was tested using different evaluation metrics, we repeated the process and changed the parameters to achieve accurate results. Additionally RCI is capable of run time entity prediction. The model performed well across multiple datasets with an average with average scores of 0.918 for F1-Score, 0.922 for Precision, and 0.918 for Recall. Such results confirm the potential of deep learning in automated clinical information extraction, and also highlight the future opportunities to improve it, including improved preprocessing, data augmentation, and the incorporation of domain knowledge. Furthermore, the imbalanced datasets may be reviewed to check the results alterations. As it is further developed and validated, this system can be helpful in supporting clinical decision-making and developing healthcare data analysis.

6. References

[1] N. Abdelmageed and others, "BiodivNERE: Gold standard corpora for named entity recognition and relation extraction in the biodiversity domain," *Biodivers Data J*, vol. 10, 2022, doi: 10.3897/BDJ.10.e89481.

[2] K. Jeon, G. Lee, S. Yang, and H. D. Jeong, "Named entity recognition of building construction defect information from text with linguistic noise," *Autom Constr*, vol. 143, 2022, doi: 10.1016/j.autcon.2022.104543.

[3] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artif Intell Rev*, vol. 55, no. 7, 2022, doi: 10.1007/s10462-021-10134-9.

[4] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, 2022, doi: 10.1007/s10462-022-10144-1.

[5] N. Nath, S. H. Lee, and I. Lee, "NEAR: Named entity and attribute recognition of clinical concepts," *J Biomed Inform*, vol. 130, 2022, doi: 10.1016/j.jbi.2022.104092.

[6] A. Mohammed and R. Kora, "An effective ensemble deep learning framework for text classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, 2022, doi: 10.1016/j.jksuci.2021.11.001.

[7] Q. Li and others, "A Survey on Text Classification: From Traditional to Deep Learning," *ACM Comput Surv*, 2022, doi: 10.1145/3495162.

[8] A. Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, "Survey on Text Classification Algorithms: From Text to Predictions," *Information (Switzerland)*, vol. 13, no. 2, 2022, doi: 10.3390/info13020083.

[9] Y. Okamura and others, "Obsessive-Compulsive Disorder with Psychotic Features: Is It a Clinical Entity?," *Healthcare (Switzerland)*, vol. 10, no. 10, 2022, doi: 10.3390/healthcare10101910.

[10] Y. P. Chen, Y. H. Lo, F. Lai, and C. H. Huang, "Disease concept-embedding based on the self-supervised method for medical information extraction from electronic health records and disease retrieval: Algorithm development and validation study," *J Med Internet Res*, vol. 23, no. 1, p. e25113, 2021, doi: 10.2196/25113.

[11] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1419–1428, 2018.

[12] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," *IEEE J Biomed Health Inform*, vol. 22, no. 5, pp. 1589–1604, 2018.

[13] J. R. Ayala Solares *et al.*, "Deep learning for electronic health records: A comparative review of multiple deep neural architectures," *J Biomed Inform*, vol. 101, p. 103337, 2020, doi: 10.1016/j.jbi.2019.103337.

[14] Y. Si *et al.*, "Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review," *J Biomed Inform*, vol. 115, p. 103671, 2021, doi: 10.1016/j.jbi.2020.103671.

[15] H. Ying and S. Yu, "A Question Answering Based Pipeline for Comprehensive Chinese EHR Information Extraction," *arXiv preprint arXiv:2402.11177*, 2024, [Online]. Available: http://arxiv.org/abs/2402.11177

[16] S. Jain *et al.*, "RadGraph: Extracting Clinical Entities and Relations from Radiology Reports," *CoRR*, vol. abs/2106.14463, 2021, [Online]. Available: https://arxiv.org/abs/2106.14463

[17] A. Dada, T. L. Ufer, M. Kim, and others, "Information extraction from weakly structured radiological reports with natural language queries," *Eur Radiol*, vol. 34, pp. 330–337, 2024, doi: 10.1007/s00330-023-09977-3.

[18] I. C. Wiest *et al.*, "From Text to Tables: A Local Privacy Preserving Large Language Model for Structured Information Retrieval from Medical Documents," *medRxiv*, 2023, doi: 10.1101/2023.12.07.23299648.

[19] D. Liu *et al.*, "Evaluation of General Large Language Models in Contextually Assessing Semantic Concepts Extracted from Adult Critical Care Electronic Health Record Notes," *arXiv preprint arXiv:2401.13588*, 2024, [Online]. Available: http://arxiv.org/abs/2401.13588

[20] S. Han *et al.*, "Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing," *J Biomed Inform*, vol. 127, p. 103984, 2022, doi: 10.1016/j.jbi.2021.103984.

[21] J. Song and others, "Local–Global Memory Neural Network for Medication Prediction," *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 4, pp. 1723–1736, 2021, doi: 10.1109/TNNLS.2020.2989364.

[22] H. J. Park, D. Y. Jung, W. Ji, and C.-M. Choi, "Detection of Bacteremia in Surgical InPatients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study," *J Med Internet Res*, vol. 22, no. 8, p. e19512, 2020, doi: 10.2196/19512.

[23] J. Luo, M. Ye, C. Xiao, and F. Ma, "HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

[24] Chaitanya Krishna Kasaraneni, "Medical Text," Kaggle, 2017. [Online]. Available: https://www.kaggle.com/datasets/chaitanyakck/medical-text.