future internet

MDPI

Article

Deep Learning Framework for Advanced De-Identification of Protected Health Information

Ahmad Aloqaily [1,*], Emad E. Abdallah [1], Rahaf Al-Zyoud [1], Esraa Abu Elsoud [2], Malak Al-Hassan [3] and Alaa E. Abdallah [4]

[1] Department of Information Technology, Faculty of Prince Al-Hussein Bin Abdullah II for Information Technology, The Hashemite University, P.O. Box 330127, Zarqa 13133, Jordan; emad@hu.edu.jo (E.E.A.); 2070593@std.hu.edu.jo (R.A.-Z.)

[2] Department of Computer Science, Faculty of Information Technology, Zarqa University, P.O. Box 330127, Zarqa 13133, Jordan; eabuelsoud@zu.edu.jo

[3] King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan; m_alhassan@ju.edu.jo

[4] Department of Computer Science, Faculty of Prince Al-Hussein Bin Abdullah II for Information Technology, The Hashemite University, P.O. Box 330127, Zarqa 13133, Jordan; aabdallah@hu.edu.jo

* Correspondence: aloqaily@hu.edu.jo

Abstract: Electronic health records (EHRs) are widely used in healthcare institutions worldwide, containing vast amounts of unstructured textual data. However, the sensitive nature of Protected Health Information (PHI) embedded within these records presents significant privacy challenges, necessitating robust de-identification techniques. This paper introduces a novel approach, leveraging a Bi-LSTM-CRF model to achieve accurate and reliable PHI de-identification, using the i2b2 dataset sourced from Harvard University. Unlike prior studies that often unify Bi-LSTM and CRF layers, our approach focuses on the individual design, optimization, and hyperparameter tuning of both the Bi-LSTM and CRF components, allowing for precise model performance improvements. This rigorous approach to architectural design and hyperparameter tuning, often underexplored in the existing literature, significantly enhances the model's capacity for accurate PHI tag detection while preserving the essential clinical context. Comprehensive evaluations are conducted across 23 PHI categories, as defined by HIPAA, ensuring thorough security across critical domains. The optimized model achieves exceptional performance metrics, with a precision of 99%, recall of 98%, and F1-score of 98%, underscoring its effectiveness in balancing recall and precision. By enabling the de-identification of medical records, this research strengthens patient confidentiality, promotes compliance with privacy regulations, and facilitates safe data sharing for research and analysis.

Keywords: protected health information; electronic health record; deep learning; de-identification; Bi-LSTM-CRF

check for updates

Academic Editor: Petros Patias

Received: 15 December 2024
Revised: 2 January 2025
Accepted: 13 January 2025
Published: 20 January 2025

Citation: Aloqaily, A.; Abdallah, E.E.; Al-Zyoud, R.; Abu Elsoud, E.; Al-Hassan, M.; Abdallah, A.E. Deep Learning Framework for Advanced De-Identification of Protected Health Information. *Future Internet* **2025**, *17*, 47. https://doi.org/10.3390/fi17010047

1. Introduction

The data revolution has experienced a significant transformation of the global environment, a phenomenon particularly underscored by the introduction of electronic chips in the late 1960s. This technological milestone caused a paradigm shift in data management techniques and laid the foundations for the ultimate transition in several industries, including healthcare, from analog documentation to digital records [1]. EHRs have emerged as a vast compendium of health-related data, compiled from many sources, including clinics, hospitals, labs, and other healthcare facilities. This development improved clinical workflows and patient care delivery by radically changing how medical information is stored and retrieved [2].

The Internet and information technology advancements have allowed health records to integrate and become more easily accessible to authorized users, beyond traditional boundaries [3]. This fluid sharing of information facilitates a comprehensive understanding of a patient's medical history, treatment plans, and results, encouraging cooperation amongst medical providers and improving the standard of patient care. In addition, the integration of health records into healthcare systems promotes evidence-based decision-making and improves administrative effectiveness, which in turn raises the standard and effectiveness of medical care [4]. Additionally, this integrated approach to health data has the potential to advance medical research, make customized treatment options possible, and stimulate innovations that could have a significant impact on the efficacy and outcomes of healthcare.

The adoption of EHRs in place of traditional paper-based health records has completely changed the way data are accessed, integrated, and shared among healthcare providers, where it offers many advantages, such as better clinical decision-making, increased care coordination, and reduced administrative responsibilities [5]. These electronic databases include patient demographics, medical histories, the results of diagnostic tests, treatment plans, and progress notes, providing an in-depth understanding of a patient's status [6]. This extensive and interconnected data ecosystem empowers healthcare providers to make informed decisions and deliver tailored, effective care. The ongoing progress of EHR technology and data interoperability provides substantial opportunities for improving research, public health initiatives, and healthcare delivery and enhancement of global healthcare [7].

Health-related data are sensitive and critical; therefore, protecting them is essential. The legal framework, represented by the Health Insurance Portability and Accountability Act (HIPAA), requires the safeguarding of patient privacy to avoid harm and uphold patients' rights [8]. Unauthorized access to or disclosure of health information may have serious repercussions, such as discrimination, identity theft, and reduced access to healthcare services. Securing patient data and maintaining patient–provider confidence requires strong security measures such as encryption, access controls, and regular audits [9]. Patient notes stored in electronic health records may contain vital information for medical investigations. Most medical investigators can only access de-identified notes to protect patients' confidentiality. The Health Insurance Portability and Accountability Act of 1996 in the United States defines 18 types of Protected Health Information that must be removed to de-identify patient notes [10]. De-identification can be performed using automated or manual procedures. Experts must identify and label PHI to perform manual de-identification; however, this approach is restricted by the number of people who are allowed access to identified patient notes and the possibility of human error. Automated de-identification systems, on the other hand, make use of rule-based or machine-learning techniques. Typically, rule-based systems depend on human-defined patterns expressed as lexicons and regular expressions [11].

Our research highlights the importance of sophisticated natural language processing (NLP) techniques in safeguarding health information and offers an effective approach for de-identifying EHR that comply with privacy laws.

The primary aim of this research paper is to provide a strong framework for de-identifying sensitive data in EHRs by utilizing deep learning and machine learning techniques. This research employs various algorithms to guarantee the secure extraction of personal patient data while concurrently upholding the integrity of essential clinical information. The dataset utilized for this research is sourced from Harvard University, specifically designated for academic users, by the Data Use and Confidentiality Agreement established with Partners HealthCare System, Inc. This agreement confers access to de-identified patient discharge summaries for research purposes, with a particular focus on the enhancement of natural language processing (NLP) techniques within the healthcare domain. This research investigates the following pivotal inquiries:

- How much have machine learning and deep learning algorithms contributed to the security of EHR?
- Is the proposed algorithm capable of effectively removing sensitive patient data from a large dataset?
- What is the significance of the Conditional Random Field (CRF) model in safeguarding patient information, particularly concerning the 23 critical categories of Protected Health Information (PHI) present in EHRs?
- How effective is the Bidirectional Long Short-Term Memory (Bi-LSTM) deep learning algorithm at improving de-identification in medical records?

The remainder of this paper is organized as follows: Section 2 encompasses a comprehensive review of related works in E-health records privacy. It critically discusses and analyzes various approaches and experimental findings from these works, while Section 3 outlines the research methodology, Section 4 presents the results and discussion, and Section 5 contains the conclusion.

2. Related Work

This section presents relevant literature on the protection of health records using machine learning and deep learning techniques.

A survey titled "Use of Electronic Health Records in Hospitals in the United States" presented comprehensive research of all American hospitals provided to count the number of records in hospitals, where they studied a proportion of the responses obtained with an average of 63.1 percent [12]. Only 1.5% of hospitals in the United States have a comprehensive electronic records system (i.e., one that is present in all clinical units), whereas 7.6% have a basic system (i.e., one that is present in at least one clinical unit). Only 17% of hospitals have automated provider–order input for drugs in place. Larger hospitals are more likely to have EHRs [13]. Leevy et al. [14] employed CRF and Bi-LSTM layers, where the work was defining the text on the RNN neural network, the Conditional Random Field, and employing RNN-based techniques such as the LSTM algorithm. The study contains vital information that helps in the safe and secure preservation of patient privacy. Qin and Zeng [15] researched clinical named entity recognition using Bi- LSTM-CRF. The goal of their study was to use neural networks to extract clinical concepts, where a two-way Bi-LSTM model was used in conjunction with the CRF model to detect three named medical entities, as they were entered into a training process in CBOW in both the field and non-field. They used the i2b2 dataset on NER and compared to other research. They achieved an F1-score of 0.85307, and they revealed certain specific concerns with clinical concepts to clarify more profound studies.

Tang et al. [16] de-identified clinical text using Bi-LSTM-CRF and neural language models. They focused on identifying the entity named NER, as it was applied to the latest deep learning models Bi-LSTM-CRF to remove the de-identification. They used two datasets: i2b2-2014 and CEGS N GRID-2016. The goal of [17]'s work was to relieve the necessity of manual labeling or feature preparation. However, they proceeded with an automated procedure, and this time it was performed through desensitization and sequence labeling. The data were then fed into the Bi-LSTM model, which applied word embeddings to each word to capture contextual information from both forward and backward directions. An attention layer processed semantic labels from the surrounding text so that the model could pay more attention to those words. A Conditional Random Field (CRF) model was then used to extract entities in the end. The results show the high performance of Bi-LSTM-CRF for entity extraction on Chinese EMRs. Hong et al. [18] proposed a powerful tool called BBC-Radical that combines three of the most important models BERT (Bidirectional Encoder Representations from Transformers), Bi-LSTM, and CRF. The tool is implemented by incorporating features based on Chinese character symbols in the word, and radical root features to show how this approach improves the model performance for the representation of common meanings.

Chinese EMRs were applied to process a large amount of medical data in [19]; the study proposed a BERT-IDCNN-CRF model. In the methodology, the EMRs were first loaded into BERT, trained, and finally, the model had to be manually annotated for relevant parameters. The experimental stage shows a very high accuracy of 94.1%, recall of 93.8%, and F1-score of 94.5% using the BIOES (Begin, Inside, Outside, End, Single) tagging scheme supervised learning approaches. Ming et al. [20] addressed the drawbacks of the LSTM model, which found it difficult to take full advantage of the GPU parallelism while processing substantial amounts of medical data. Furthermore, it was observed that the word order and semantic information were ignored by ID-CNNs. The authors suggested an ID-CNNs-CRF model with attention methods to address these problems. With F1-scores of 0.9455 and 0.9117, this model was able to effectively capture word order, attentiveness, and local contextual variables.

Using a Bi-LSTM-CRF model, Liu et al. [21] concentrated on detecting named items in biomedical data, using one layer of CRF and two layers of LSTM. The model achieved an F1-score of 0.872. This study used the TASS-2018 dataset and achieved first place in its subtask. Liu et al. [22] focused on identifying clinical quantitative data and connecting them to relevant entities. For this study, 1359 documents from a nearby general hospital's petroleum department were examined. The results indicated the efficacy of the strategy with an F1-score of 0.9477 for recognizing quantitative information, and an accuracy of 0.9460 for associating entities with quantities. Shunli et al. [23] sought to tackle the issue of modern systems lacking the ability to effectively capture low-frequency terms through manual features. The suggested approach involved utilizing a Bi-LSTM-CRF neural network on electronic medical records. The procedure started with using CCNs (Convolutional Neural Networks) for encryption, then proceeded to the Bi-LSTM layer, and concluded with the CRF layer for entity extraction. The research utilized two clinical datasets that are openly accessible: the THYME corpus and the 2012 i2b2 dataset.

The key distinction between our research and previous studies lies in our dedicated approach to the design, optimization, and hyperparameter tuning of both the Bi-LSTM and CRF models. Unlike prior work, which typically combines Bi-LSTM and CRF into a single unified model, our research takes a more granular approach by treating each model individually. This allows for the more precise optimization of each model at every stage of development. We carefully designed the architecture of both deep learning models, placing particular emphasis on fine-tuning their hyperparameters to achieve optimal accuracy. This focused process of hyperparameter optimization, a critical yet under-explored aspect in the literature, plays a pivotal role in enhancing model performance.

Additionally, by conducting comprehensive evaluations across 23 distinct categories, we provide a detailed and independent analysis of each model's performance. This thorough approach not only offers a more nuanced understanding of the models' behaviors but also ensures that the achieved accuracy surpasses that of existing studies. Our rigorous and systematic optimization process, coupled with independent testing across diverse categories, addresses a significant gap in prior research where such meticulous architectural and performance comparisons are frequently overlooked.

3. Methodology

Electronic health records have completely revolutionized how medical data are recorded, accessed, and shared in the modern digital health environment. They enable more efficient data management, which enhances the information that healthcare professionals have access to and supports the integration of patient care in different healthcare settings. However, there are significant concerns regarding patient data security and privacy as EHRs expand. Deep learning, a field of artificial intelligence, has emerged as a effective technique for strengthening the security of sensitive EHRs. Bi-LSTM is a novel deep learning paradigm that builds upon the traditional LSTM architecture to make it highly effective in sequence modeling tasks. Bi-LSTM analyzes sequences bidirectionally, which allows it to identify contextual dependencies from the past and future more efficiently than regular LSTM, which processes information in a single direction [24]. This bidirectional feature leverages the complete context that is included in the data to enable the creation of more accurate predictions. The structure consists of interconnected layers of neurons, or cells, with the ability to store, change, and reset internal states [25]. The network acquires knowledge sequence patterns and correlations as data flow through these cells, making Bi-LSTM especially appropriate for time-series data, such EHRs. The advantages of this technology include enhanced contextual sensitivity, real-time monitoring, and adaptability to changing patterns, among other attributes that raise its significance in enhancing security standards in healthcare systems [26]. Healthcare institutions can effectively protect patient data from new threats by incorporating Bi-LSTM into security frameworks. This ensures patient privacy and preserves confidentiality in digital healthcare infrastructures.

Our proposed methodology integrates CRF and Bi-LSTM, as illustrated in Figure 1, to accurately identify categories of Protected Health Information (PHI). CRF, a statistical model renowned for its ability to predict label sequences, is leveraged to enhance the performance of PHI identification. Central to our contribution is the careful design, optimization, and hyperparameter tuning of both models, which forms the backbone of our deep learning approach. Unlike conventional methods, we optimize each model individually, fine-tuning the hyperparameters to achieve the best possible performance. By combining these models and emphasizing rigorous model design and hyperparameter optimization, our approach aims to surpass traditional methods in accuracy, recall, precision, and overall effectiveness, offering a more robust solution for PHI categorization.

3.1. Data Extraction

In this research, we utilize the i2b2 (Informatics for Integrating Biology and the Bedside) dataset, obtained from the Department of Biomedical Informatics at Harvard Medical School [27–29]. The dataset is publicly accessible at https://n2c2.dbmi.hms.harvard.edu/data-sets (accessed on 10 December 2024), subject to a Data Use Agreement (DUA). The i2b2 dataset is a valuable resource for healthcare researchers. It is a collection of de-identified clinical records that include a variety of medical data, such as clinical notes, discharge summaries, and radiology results. The dataset is well known for its diversity and breadth, making it an excellent candidate for training NLP models, particularly in the context of clinical text processing and information extraction. The i2b2 dataset includes PHI categories, enabling researchers to address the difficulty of PHI identification while respecting patient privacy through data anonymization. The Data Usage Agreement (DUA) is used to access and retrieve the dataset, and registration was required. The i2b2 dataset consists of 21 compressed files, and all medical records are stored in the (XML) file format.

3.2. Data Preprocessing

The data preprocessing phase includes three key phases to prepare the dataset for feature extraction.

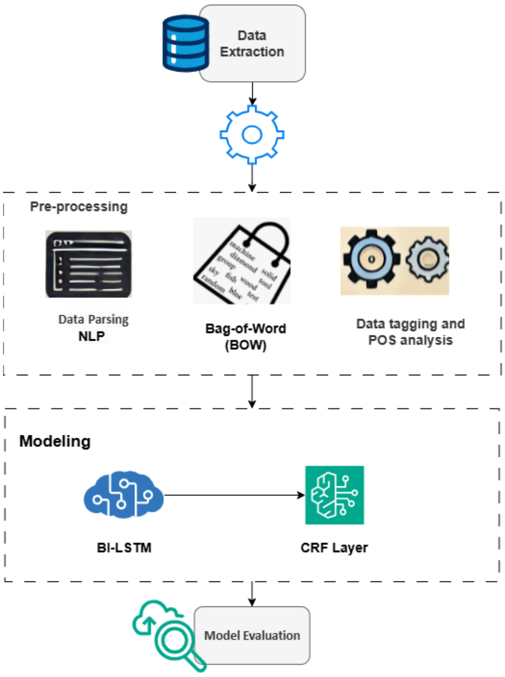


Figure 1. The proposed methodology.

3.2.1. Data Parsing

The first step in data preprocessing focuses on transforming the incomprehensible, unreadable, and unstructured text into a more coherent and interpretable format. This critical step facilitated the extraction of relevant information without the time-consuming and manual preparation that is typically required. Advanced NLP techniques are used in the process, which effectively transformed raw data into more understandable texts, allowing for efficient information retrieval and analysis.

3.2.2. Bag-of-Word (BOW)

The data are tokenized, in which all the words within each record are segmented using a delimiter (,). This tokenization method converts the text into an arbitrary "bag of words" (BOW) model, which is defined by fixed and specific length vectors. Each vector represents the frequency with which individual words appear in the text [30]. In the context of BOW models, this method is commonly referred to as a "directed" approach. The BOW model is based primarily on the identification of a vocabulary, which includes a set of words found in the record. Individual words are isolated and distinguished using the delimiter (,) to form the vocabulary. The number of times each word appears in the record is then counted and recorded. This process produces a numerical representation of the text's content, effectively capturing word frequencies for further analysis and modeling.

3.2.3. Data Tagging and POS Analysis

In this phase, the text is segmented into eight main parts of speech, collectively known as Part of Speech (POS) categories. These categories comprise nouns, pronouns, verbs, adjectives, adverbs, prepositions, and interjections. Data tagging facilitates data annotation for subsequent data tagging tasks. It is carried out using a three-letter notation consisting of "B", "I", and "O". Each letter corresponds to a specific meaning; "B" indicates the "Beginning" of a tagged entity, and "I" refers to the "Intermediate" entity, while

"O" refers "Other", indicating that the word does not belong to any tagged entity [31]. Within the electronic medical records, a total of 23 basic categories are identified for data tagging, including Date, Patient, Age, Profession, Medical Record, Hospital, Doctor, Street, City, State, ZIP, Phone, Id_Num, Username, Organization, FAX, Country, BIOID, Location, Device, Email, Health Plan, And URL. These categories help in organizing specific information within medical records, laying the groundwork for future analysis and model development. Each tag in the record is associated with the corresponding text above it during this process. To denote their positions and roles, the existing tags are modified with additional letters. The letter "B" is appended to the first word in a category, while the letter "I" is appended to subsequent words in the same category. The letter "O" on the other hand, denotes unimportant words that do not fall into any of the required 23 categories specified in the tags. The tags are associated with several entity types, including Person (PER), Location (LOC), Organization (ORG), and Miscellaneous (MISC). For example, the sentence "EU rejects German call to boycott British lamb", is tagged as (B-ORG O B-MISC O O O B-MISC O O), where we have the following:

- B-ORG: "EU" is identified as the beginning of an organization.
- O: "rejects" does not fit within any recognized category.
- B-MISC: "German" is tagged as the beginning of a miscellaneous entity.
- The sequence continues with more O tags that are not part of any recognized category.
- B-MISC: "British" is tagged as the beginning of a miscellaneous entity.

The number of occurrences for each of the 23 main PHI categories is summarized in Table 1. The number of occurrences for the "Other" (O) category after data reprocessing is 775,812.

Table 1. Number of B-labels and I-labels for PHI categories.

#	PHI Categories	Number of B-Label	Number of I-Label
1	Date	12,437	1362
2	Patient	2184	1184
3	Age	1993	10
4	Profession	411	346
5	Medical Record	1028	47
6	Hospital	2305	1819
7	Doctor	4782	3466
8	Street	350	713
9	City	651	170
10	State	502	18
11	ZIP	350	0
12	Phone	523	100
13	Id_Num	455	30
14	Username	356	0
15	Organization	205	173
16	FAX	10	2
17	Country	183	21
18	BIOID	1	0
19	Location-Other	17	15
20	Device	15	2
21	EMAIL	5	0
22	Health Plan	1	1
23	URL	2	4

3.3. Bi-LSTM-CRF Model

The idea behind the sequence labeling challenge is to convert an input sequence into a corresponding, identically sized output sequence. As a result, constructing a Recurrent

Neural Network with a sufficient number of hidden layers has produced outstanding outcomes for sequence-to-sequence conversions. Here, recurrent connections serve as a kind of memory that allows the network's internal state to hold on to insights from the previous inputs. The label applied to the input data is represented by the final output, which is established by taking previous inputs into account as well. There have been several versions of Recurrent Neural Nets developed, but the LSTM architecture stands out since it solves the problem of disappearing gradients effectively. Therefore, the LSTM network is frequently preferred for labeling words inside possibly long sentences. However, since the whole phrase is known ahead of time, the label assignment of a given word can be carried out more precisely if it takes into account the words that come before and after it. It is more efficient to represent these dependencies by using the Bidirectional LSTM [32]. To analyze data from both directions of the input sequence, the model's network design (as illustrated in Figure 2) includes forward and backward LSTM units. The model is more capable of capturing word dependencies, regardless of where they are in the sentence, due to this bidirectional arrangement. The model is able to comprehend each word's meaning and grammatical structure better as a result of the concatenation of the PoS vectors and the embedding layer output. As each word's label can be predicted individually by the Bi-LSTM, the relationships between labels in the output sequence cannot be captured by the model. One word that is labeled as the beginning of an entity (B) in PoS tagging, for instance, increases the possibility that the following word will be marked as the inner (I) of the same entity. CRF comes in. To guarantee that the predicted labels adhere to correct sequential patterns, the CRF layer is layered on top of the Bi-LSTM. The possibility of label transitions (e.g., from B to I, or from I to O) is taken into consideration to effectively model the relations between the expected labels.

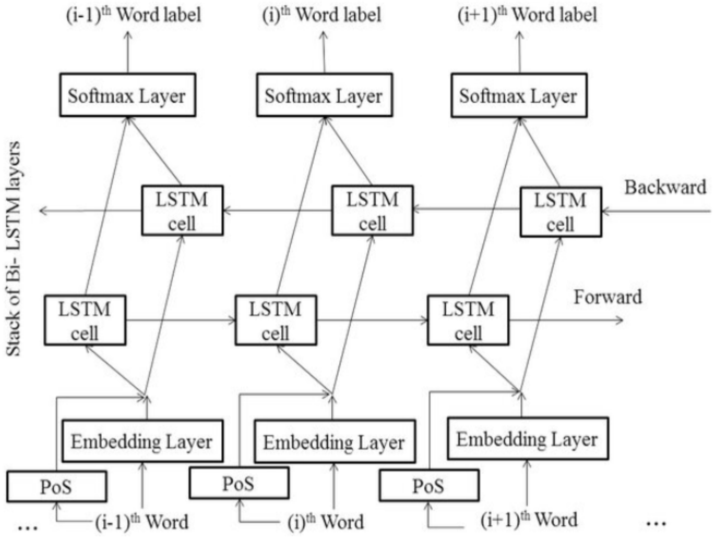


Figure 2. Bi-LSTM model.

3.4. Bi-LSTM-CRF Model Architecture, Hyperparameter Tuning, and Optimization

Model Architecture

The Bi-LSTM-CRF model developed for this research integrates a Bi-LSTM network with a Conditional Random Field (CRF) layer to enhance sequential tagging tasks by capturing both contextual dependencies and transition constraints between output tags. The model's key components are as follows:

- **Embedding Layer**: The input words are transformed into dense vector representations using an embedding layer with a dimensionality range of 50 to 300.

- **Bi-LSTM Layer**: A multi-layer bidirectional LSTM network is employed with a hidden dimension ranging from 50 to 512, to capture both forward and backward contexts for each token.
- **CRF Layer**: A transition matrix, initialized randomly, scores transitions between tags, including constraints to prevent invalid transitions to and from designated start and stop tags.

Hyperparameter Tuning

Key hyperparameters were fine-tuned to balance performance and computational efficiency:

- **Learning Rate**: Ranges from 0.001 to 0.05, optimizing model convergence while maintaining stability.
- **Weight Decay**: Regularization is applied with a weight decay parameter ranging from 1×10^{-4} to 1×10^{-2}, effectively controlling overfitting.
- **Batch Size**: The model processes samples in batches, with a batch size ranging from 8 to 64, enhancing computational efficiency and convergence.
- **Epochs**: Training is performed over 300 epochs to ensure sufficient learning while monitoring for convergence.

Optimization

The model training utilizes Stochastic Gradient Descent (SGD), selected for its straightforward implementation and effective balance of convergence speed and stability. The design allows for the potential inclusion of gradient clipping to manage exploding gradients, a common consideration in Bi-LSTM models.

The Stochastic Gradient Descent (SGD) optimizer in this model is configured with the following key parameters:

- **Learning Rate**: This is set within the range of 0.001 to 0.05 to balance the learning speed with convergence stability. This range is selected based on empirical observations, providing consistent convergence without overshooting.
- **Weight Decay**: This is set within the range of 1×10^{-5} to 1×10^{-2} as a regularization term, which helps control overfitting by penalizing large weights during training.

This Bi-LSTM-CRF architecture thus provides a robust framework for sequence labeling, leveraging the ability of Bi-LSTM to model contextual dependencies and the structured output layer of the CRF to enforce tag transition rules as summarized in Table 2.

Model Training and Evaluation Metrics

The dataset consists of 1304 records, which are split into 790 training records and 514 testing records, equivalent to a 60% to 40% split, as presented in Table 3. Each record represents specific medical information about individual patients. In total, there are 269 unique patients, and each patient's data are represented by 3 to 5 files of XML type. The file naming convention follows a pattern, denoted as (XXX-YY.xml), where XXX indicates the patient number and YY represents the record number within that patient's dataset. The dataset structure is designed to efficiently manage and differentiate the medical records associated with each patient, coherently facilitating comprehensive analysis and research.

Table 2. Summary of the Bi-LSTM-CRF model architecture.

Component	Details
Input Representation	Concatenation of word embeddings and PoS vectors
Embedding Layer	Embedding dimensionality ranges from 50 to 300
Bi-LSTM Layers	Multi-layer bidirectional LSTM with hidden dimensions ranging from 50 to 512
CRF Layer	Transition matrix with constraints for valid tag transitions
Hyperparameters	Learning rate: 0.001 to 0.05 Weight decay: 1×10^{-4} to 1×10^{-2} Batch size: 8 to 64
Epochs	300
Optimizer	Stochastic Gradient Descent (SGD) with gradient clipping
Learning Rate	Tuned between 0.001 to 0.05 for stability and convergence
Regularization	Weight decay: 1×10^{-5} to 1×10^{-2} to control overfitting
Sequence Constraints	CRF ensures valid tag transitions between output labels
Objective	Enhances sequential tagging by capturing contextual dependencies and transition constraints

Table 3. Dataset split into training and testing records.

Dataset	Train	Test
Records	790	514
Total	1304	

The Bi-LSTM-CRF model is trained using the i2b2 dataset over iterative cycles, referred to as epochs. Each epoch represents a full pass through the training dataset, during which the model's parameters, including weights, are updated based on the input data. After one epoch, the model processes each sample in the training set once, allowing for gradual improvement in its performance through repeated evaluations. This iterative training process enables the model to enhance its data processing capabilities with each cycle, progressively refining its ability to handle complex inputs. To ensure optimal performance, key hyperparameters, such as the learning rate, batch size, and number of epochs, were tuned based on empirical testing. A learning rate of 0.001 is selected to ensure stable convergence, while a batch size of 32 allowed for efficient processing without overloading the system's memory. The number of epochs is set to 50, providing sufficient training time for the model to learn complex patterns in the data without overfitting. By executing the model over multiple epochs with carefully chosen hyperparameters, sufficient computational resources are allocated to improve its predictive accuracy, thereby enhancing its capacity to identify patterns and generate more reliable predictions.

To better understand the dataset, the model improves its "knowledge base" after each cycle. The iterative approach helps the model to get to know the subtleties of the dataset and make more accurate predictions based on the knowledge it has gained. The model performance is evaluated through three critical metrics: accuracy, recall, and the F1-score. However, relying only on accuracy is insufficient in evaluating an imbalanced dataset (where certain kinds of data may be disproportionately represented). As a result, we also include recall, which assesses the model's ability to recognize important data, and the F1-score, which combines recall and accuracy to produce a more complex evaluation of the model's performance.

4. Results and Discussion

In this research, we use the Bi-LSTM-CRF and CRF models, where the CRF model is an easier-to-understand architecture and helps in reducing overfitting, especially on smaller datasets. However, the Bi-LSTM-CRF model makes use of deep learning's advantages for sequential data processing by combining a Bi-LSTM network with a CRF layer. The Bi-LSTM component is potentially appropriate for detecting hidden trends and contextual relations since it can capture intricate correlations within sequences. Bi-LSTM and other deep learning models demonstrate high accuracy on larger and more complex datasets, but their performance is highly sensitive to dataset size and requires careful hyperparameter tuning. These models are particularly advantageous for identifying intricate patterns that may be overlooked by simpler models, making them well suited for complex data analysis tasks.

4.1. Optimized Parameter Settings and Model Configuration for Bi-LSTM-CRF

As outlined in Section 3.4, this section presents the optimized parameter settings for the Bi-LSTM-CRF model, tailored for effective PHI tag detections. The optimization of the Bi-LSTM-CRF model is crucial, as it enhances its ability to accurately capture sequential dependencies and improve tagging performance, especially in complex datasets. Table 4 provides a detailed summary of the optimal configuration achieved for the model's architecture, hyperparameter tuning, and optimization strategies. As the experimental results show, the architecture leverages an embedding layer with a dimensionality of 60 to convert input words into dense vector representations, which enhances the model's ability to learn meaningful relationships between words. A single-layer Bi-LSTM network with a hidden dimension of 80 is employed to capture the bidirectional context for each token, ensuring that both past and future contexts contribute to the model's understanding of sequential dependencies. A CRF layer, initialized with a transition matrix, scores transitions between output tags and enforces constraints to prevent invalid tag transitions, contributing to improved tagging accuracy. Hyperparameter tuning is performed to achieve an optimal balance between model performance and computational efficiency. The learning rate is set to 0.01, which provides a stable convergence. A weight decay parameter of 1×10^{-4} is used to regularize the model, helping control overfitting by penalizing large weights. Additionally, a batch size of 32 is selected, which allows the model to process data efficiently while benefiting from gradient updates. The model is trained over 300 epochs, with the best performance observed at epoch 31, indicating the need for early stopping to prevent overfitting. Finally, the optimization is conducted using the SGD algorithm, chosen for its straightforward implementation and effectiveness in balancing learning speed and stability. The table summarizes these configurations, highlighting the model's robust framework for sequence labeling tasks.

Table 4. Best parameter settings for Bi-LSTM-CRF model architecture, hyperparameter tuning, and optimization.

Component	Parameter	Setting
Model Architecture	Embedding Dimension	60
	Bi-LSTM Hidden Dimension	80
	CRF Transition Matrix	Randomly initialized

Table 4. *Cont.*

Component	Parameter	Setting
Hyperparameter Tuning	Learning Rate	0.01
	Weight Decay	1×10^{-4}
	Batch Size	32
	Epochs	300 (best: 31)
Optimization	Optimizer	Stochastic Gradient Descent (SGD)
	Learning Rate	0.01
	Weight Decay	1×10^{-4}

4.2. CRF Model

The CRF model works with conditional distributions instead of joint distributions, where the conditional distribution does not enforce independent labels, making it easier to interpret natural language. Moreover, it defines the phrase's order by integrating the probabilistic evaluation of the unique symbols contained in a sentence [33]. The CRF model is independently applied as part of the initial evaluation phase. Table 5 presents the outcomes of these evaluations in an organized manner. This table provides a comprehensive overview of the model's effectiveness, covering a variety of assessment criteria such as accuracy, precision, recall, and F1-score.

Table 5. Conditional Random Fields (CRFs) results.

Evaluation Method	Results
Precision	98%
Recall	99%
F1-score	98%
Accuracy	99%

The CRF performs exceptionally well in terms of accuracy metrics, with an accuracy rate of 99%, an F1-score of 98%, and a recall of 99% as shown in Table 5. These analyses provide additional knowledge about the CRF model's performance on the i2b2 dataset. The enhanced performance of the HIPAA-level sets has the potential to significantly improve data security and privacy in EHR, which might have a significant positive impact on real-world healthcare applications.

The results of the comprehensive assessment for each of the 23 basic categories of PHI as defined by HIPAA are shown in Table 6. B-DATE performs best out of all of these categories, achieving the maximum level of precision and efficiency. Within this evaluation framework, the B-DATE category achieves an exceptional precision score of 99%, indicating the model's ability to correctly identify actual positive instances. Moreover, the model's 98% recall score demonstrates its ability to identify real positive cases. A crucial measure for balancing recall and precision, the F1-score, is 98%, indicating that the B-DATE model is robust in achieving an equitable balance between these two crucial performance aspects.

Table 6. Evaluation findings for each of the 23 PHIAA categories using the CRF model.

Categories	Precision (P%)	Recall (R%)	F1-Score (F1%)	Support
B-AGE	0.56	0.48	0.52	764
B-City	0.76	0.51	0.61	260
B-Country	0.68	0.11	0.19	117
B-Date	0.99	0.98	0.98	4980
B-Device	0	0	0	8
B-Doctor	0.65	0.57	0.60	1912
B-Email	0	0	0	1
B-FAX	0	0	0	2
B-Hospital	0.87	0.59	0.71	875
B-ID_Num,	0.98	0.76	0.86	195
B-Location Other	0	0	0	13
B-Medical Record	0.97	0.96	0.96	422
B-Organization	0.50	0.07	0.13	82
B-Patient	0.53	0.25	0.34	879
B-Phone	0.96	0.93	0.94	215
B-Profession	0.59	0.21	0.31	179
B-State	0.91	0.76	0.83	190
B-Street	0.98	0.92	0.95	136
B-Username	0.99	0.90	0.94	92
B-ZIP	0.99	0.96	0.97	140
I-Age	0	0	0	1
I-City	0.83	0.36	0.50	81
I-Country	0	0	0	13
I-Date	0.95	0.88	0.91	509
I-Device	0	0	0	2
I-Doctor	0.73	0.74	0.74	1350
I-FAX	0	0	0	1
I-Health Plan	0	0	0	0
I-Hospital	0.93	0.78	0.85	700
I-ID_Num	1.00	0.57	0.73	14
I-Location Other	0	0	0	7
I-Medical Record	0	0	0	22
I-Organization	0.44	0.18	0.26	61
I-Patient	0.54	0.30	0.38	479
I-Phone	0.96	0.94	0.95	48
I-Profession	0.77	0.38	0.51	156
I-State	0	0	0	15
I-Street	0.91	0.92	0.92	280
Other "O"	0.99	1.00	0.99	301,937

According to Table 6 and Figure 3, we can note that the numbers of OTHER "O" category represents 95% of all records with 4980 records, and it is achieved the best classifier results with precision and an F1-score of 99%, as well as the other categories with a high number of support records such as B-DATE with 4980 records, B-DOCTOR with 1912, and I-DOCTOR with 1350. All of these categories achieve good metrics results as presented in Figure 4.

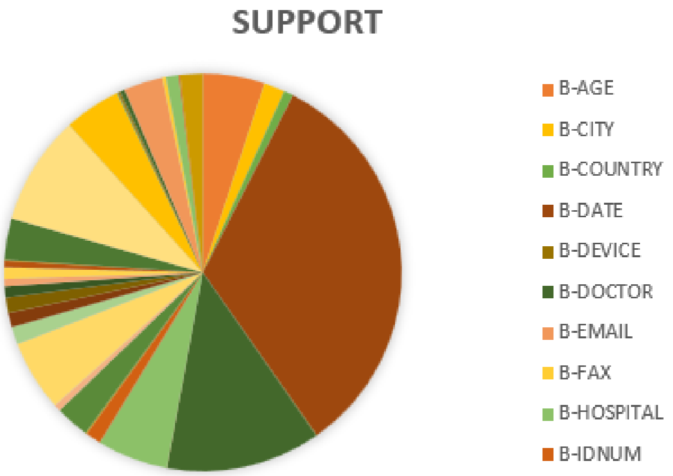


Figure 3. Support in each category without other in CRF model.

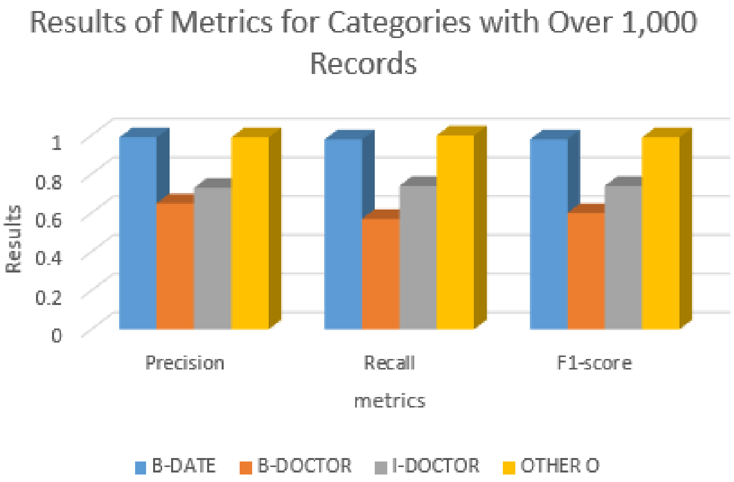


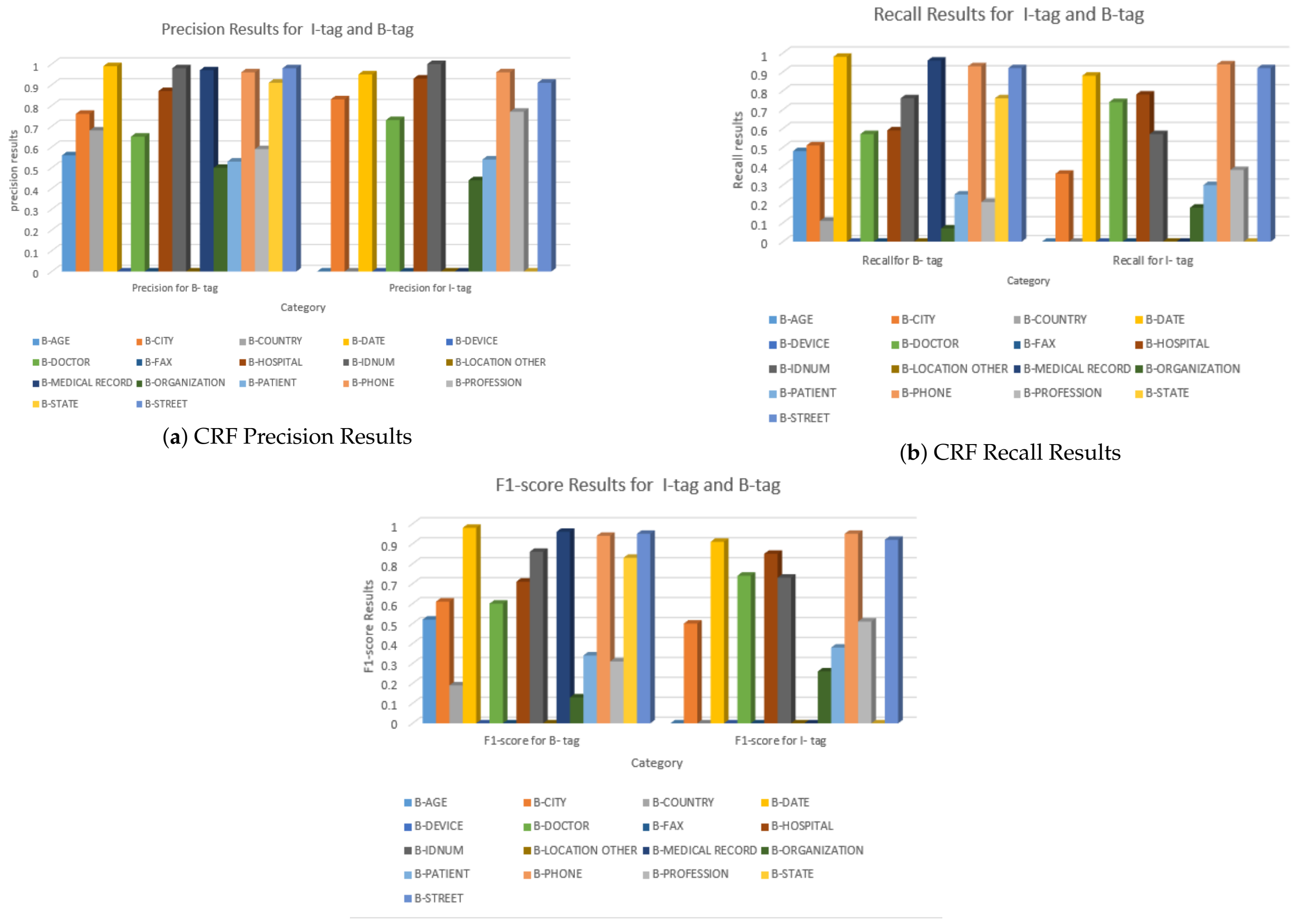
Figure 4. Results of metrics for categories with over 1000 records.

When analyzing the results according to B-tags and I-tags as presented in Table 7 and Figure 5, the B-tags indicate the beginning of entities; hence, it is critical to precisely define these boundary errors in determining the beginning of an entity, as they may result in a series of I-tag mistakes, which reduces its overall effectiveness. The I-tags often consist of fragments of longer or more complex items. Misclassifications are more common when an entity is extended (as shown by I-tags), especially when the model has difficulty recognizing contextual details. Moreover, several I-tags show noticeably lower frequencies of occurrence (that is, fewer instances in the dataset). As an illustration, categories like I-AGE and I-COUNTRY show shallow support (with only 1 or 13 instances), which reduces the model's performance. However, compared to the extension of the entity with I-tags, the identification of the initiation of an entity with B-tags typically performs better since it is frequently less complicated and dependent on contextual clues. Furthermore, the dataset's restricted frequency of I-tag categories might exacerbate I-tags' poor results compared to B-tags.

Table 7. Comparison of B-tag and I-tag Results using the CRF model.

Category	B-Tag				I-Tag			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Age	0.56	0.48	0.52	764	0	0	0	1
City	0.76	0.51	0.61	260	0.83	0.36	0.50	81
Country	0.68	0.11	0.19	117	0	0	0	13
Date	0.99	0.98	0.98	4980	0.95	0.88	0.91	509
Device	0	0	0	8	0	0	0	2
Doctor	0.65	0.57	0.60	1912	0.73	0.74	0.74	1350
Email	0	0	0	1				
FAX	0	0	0	2	0	0	0	1
Hospital	0.87	0.59	0.71	875	0.93	0.78	0.85	700
ID_Num	0.98	0.76	0.86	195	1	0.57	0.73	14
Location Other	0	0	0	13	0	0	0	7
Medical Record	0.97	0.96	0.96	422	0	0	0	22
Organization	0.5	0.07	0.13	82	0.44	0.18	0.26	61
Patient	0.53	0.25	0.34	879	0.54	0.30	0.38	479
Phone	0.96	0.93	0.94	215	0.96	0.94	0.95	48
Profession	0.59	0.21	0.31	179	0.77	0.38	0.51	156
State	0.91	0.76	0.83	190	0	0	0	15
Street	0.98	0.92	0.95	136	0.91	0.92	0.92	280
Username	0.99	0.90	0.94	92				
ZIP	0.99	0.96	0.97	140				
Other “O”					0.99	1	0.99	301,937

The model’s performance, however, shows a strong bias towards high-frequency labels, particularly the “Other” (O) label, which constitutes 95% of the dataset records as shown in Table 5 and Figure 3. This label achieves high precision and an F1-score of 99%, indicating that the model is highly accurate for the “Other” class. Similarly, categories with higher support counts, such as B-DATE, B-DOCTOR, and I-DOCTOR, also display strong metrics as presented in Figure 4. However, the model struggles with lower-frequency categories, particularly those labeled with I-tags, which represent internal segments of entities and often require more nuanced contextual understanding. The infrequent occurrence of certain I-tags, such as I-AGE and I-COUNTRY with only 1 and 13 instances, respectively, hinders the model’s ability to generalize effectively for these classes. As a result, the model’s overall performance across PHI categories is uneven, showing marked accuracy for frequent labels but decreased precision and recall for less common, more complex labels. This highlights a limitation in the model’s generalizability, emphasizing the need for more balanced dataset representation to improve classification across all PHI categories.



(**a**) CRF Precision Results

(**b**) CRF Recall Results

(**c**) CRF F1-score Results

Figure 5. Comparison of CRF results: precision, recall, and F1-score.

4.3. Bi-LSTM-CRF Model

In the Bi-LSTM-CRF design, the CRF component makes post-prediction processing easier and produces sense results. The CRF algorithm uses the relationships between adjacent tokens to correct for any important data missed in the first prediction, improving the accuracy of the result. By implementing the Bi-LSTM-CRF model, we can guarantee enhanced workflow efficiency for medical professionals and improve data protection accuracy. Healthcare organizations can enhance patient trust, improve data-breach defenses, and preserve the ethical responsibility to secure sensitive health records. The results of the Bi-LSTM-CRF model are presented in Table 8. Regarding the accuracy metrics, the Bi-LSTM-CRF model performs admirably, with an accuracy rate of 98%, a recall score of 99%, and an F1-score of 99%.

Table 8. Bi-LSTM-CRF model overall results.

Evaluation Method	Results
Precision	99%
Recall	99%
F1-score	99%
Accuracy	98%

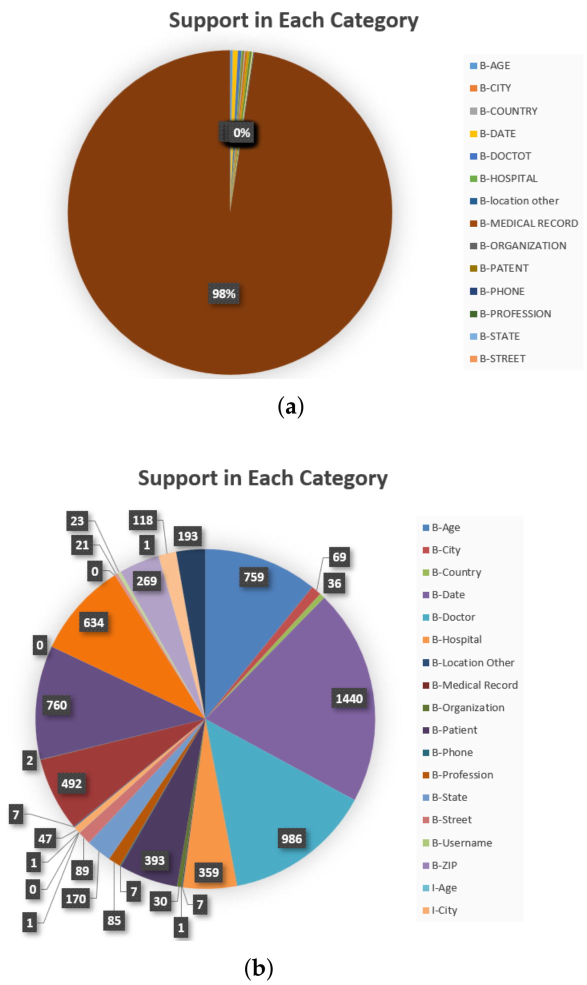
The evaluation results for all 23 core categories of PHI as defined by the HIAA using the Bi-LSTM-CRF model are presented in Table 9. When analyzing the results in the table and by looking into support values, we can see the difference between support in each category as presented in Figure 6, where the other "O" tags represent 98% of the instances, which will affect the evaluation results as shown in Figure 7.

Table 9. Evaluation findings for each of the 23 PHIAA categories using the Bi-LSTM-CRF model.

Categories	Precision P%	Recall R%	F1-Score F1%	Support
B-Age	0.91	0.83	0.87	759
B-City	0.27	0.36	0.31	69
B-Country	0.33	0.14	0.20	36
B-Date	0.46	0.71	0.56	1440
B-Doctor	0.51	0.69	0.59	986
B-Hospital	0.41	0.50	0.45	359
B-Location Other	0.00	0.00	0.00	7
B-Medical Record	0.00	0.00	0.00	1
B-Organization	0.00	0.00	0.00	30
B-Patient	0.53	0.43	0.47	393
B-Phone	0.00	0.00	0.00	7
B-Profession	0.28	0.22	0.25	85
B-State	0.35	0.10	0.16	170
B-Street	0.62	0.67	0.65	89
B-Username	0.00	0.00	0.00	1
B-ZIP	0.00	0.00	0.00	0
I-Age	0.00	0.00	0.00	1
I-City	0.33	0.09	0.14	47
I-Country	0.00	0.00	0.00	7
I-Date	0.85	0.73	0.78	492
I-Device	0.00	0.00	0.00	2
I-Doctor	0.62	0.63	0.62	760
I-Health Plan	0.00	0.00	0.00	0
I-Hospital	0.83	0.45	0.59	634
I-ID_Num	0.00	0.00	0.00	0
I-Medical Record	0.00	0.00	0.00	21
I-Organization	0.00	0.00	0.00	23
I-Patient	0.47	0.24	0.32	269
I-Phone	0.00	0.00	0.00	1
I-Profession	0.65	0.24	0.35	118
I-Street	0.82	0.61	0.70	193
Other	1.00	0.99	0.99	287,426

According to Figure 7, the evolution results for tags with support greater than 400 are considered almost good, compared to those with support less than 400 as illustrated in Figure 8. So, we can conclude that the amount of support will affect the results of the evaluation metrics in some way, where smaller tags might have lower precision or recall because the model struggles to learn from fewer examples, and tags with more support provide more data, which usually leads to more reliable evaluation metrics.

On the other hand, when analyzing the results according to the B tag and I tag, as presented in Table 10, we can conclude that the B-tags exhibit superior performance across various categories. The model demonstrates a heightened level of confidence in recognizing the initiation of entities in contrast to their internal components. I-tag performance is low in categories like "AGE", "CITY", and "COUNTRY", indicating that the model struggles to maintain entity recognition beyond the first token. This could be the result of the model leaning towards making simpler, single-token predictions or a deficiency of sufficient training data for multi-token entities. I-tags perform well in categories like "DATE", "HOSPITAL", and "STREET". These entities often have more structured patterns and contain more tokens, which makes it easier for the model to capture internal tokens efficiently.



(**a**)

(**b**)

Figure 6. Support in each category using Bi-LSTM-CRF. (**a**) Support in each category using Bi-LSTM-CRF. (**b**) Support in each category using Bi-LSTM-CRF except the other category.

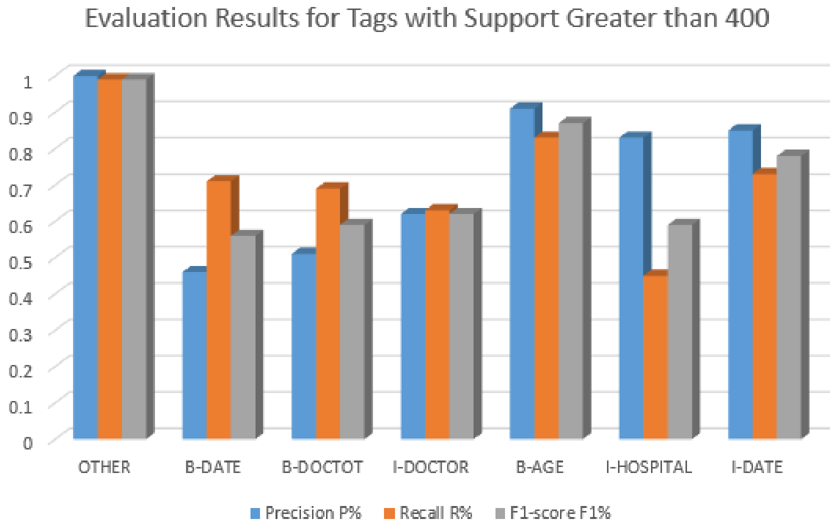


Figure 7. Evaluation results for tags with support greater than 400.

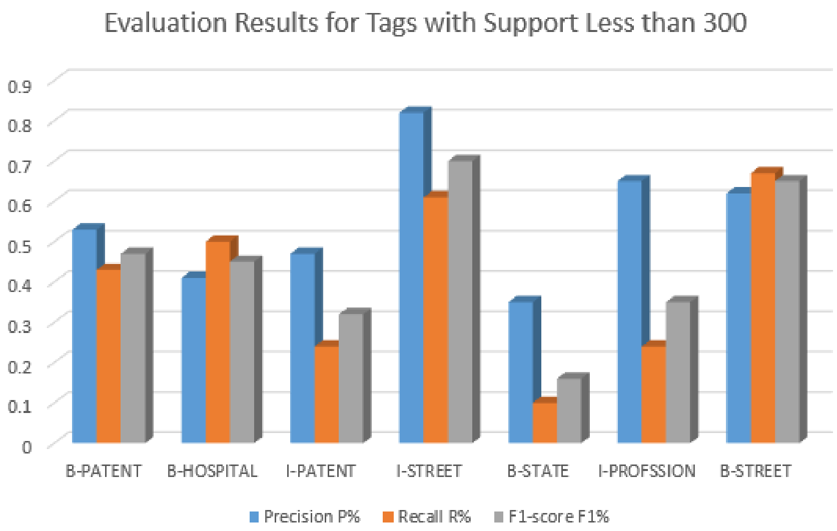


Figure 8. Evaluation results for tags with support less than 300.

Table 10. Comparison of B-tag and I-tag results using Bi-LSTM-CRF model.

Category	**B-Tag**			**I-Tag**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Age	0.91	0.83	0.87	0	0	0
City	0.27	0.36	0.31	0.33	0.09	0.14
Country	0.33	0.14	0.20	0	0	0
Date	0.46	0.71	0.56	0.85	0.73	0.78
Doctor	0.51	0.69	0.59	0.62	0.63	0.62
Hospital	0.41	0.50	0.45	0.83	0.45	0.59
Patient	0.53	0.43	0.47	0.47	0.24	0.32
Profession	0.28	0.22	0.25	0.65	0.24	0.35
Street	0.62	0.67	0.65	0.82	0.61	0.70

For most of the 23 evaluated categories, the CRF model outperforms the Bi-LSTM-CRF model, which has shown promising performance across most of the 23 evaluated categories. However, its effectiveness is influenced by the need for a larger dataset to leverage its complex architecture and hyperparameter tuning fully. While the CRF model has a simpler architecture and is generally less prone to overfitting, the Bi-LSTM-CRF model requires a larger volume of data to maximize its potential. Although the Bi-LSTM-CRF has been carefully optimized and tuned, expanding the dataset would allow for even more effective performance improvements and better utilization of the model's capabilities.

4.4. Comparison Between Our Findings with Other Studies

Clinical data processing has been improved using Bi-LSTM-CRF techniques for PHI de-identification in EHRs. This method, which combines the advantages of sequence labeling and deep learning, improves the accuracy and efficiency of identifying sensitive information.

In [34], the authors used the Rennes University Hospital's EHRs to create a French de-identification dataset. The distribution of entities in the training and test sets was consistent, and the dataset included extensive personal information. A Bi-LSTM + CRF model was assessed using manually annotated clinical reports in conjunction with Flair and FastText word embeddings. The model outperformed other tested models and demonstrated the efficacy of this strategy for de-identifying sensitive information with a high F1-score of 96.96%. To identify text that should be classified as sensitive due to privacy concerns, the authors in [35] used three different models: Bi-LSTM, Bi-LSTM with attention, and BERT base models. The words used can disclose important details about a patient's character and personal life, even though it may not at first appear to be sensitive to privacy concerns. Clinical data include hidden features that may cause privacy problems in addition to demographic data. Approximately 92% accuracy was achieved by enhancing the Bi-LSTM model with an attention layer that highlights crucial words that are important for categorization. The dataset comprised 206,926 phrases, of which 80% was used for training and the remaining 20% for testing. Using the Bi-LSTM model alone, the dataset yielded an accuracy of about 90%. In [36], the authors presented a novel multi-medical entity recognition approach combining BART, Bi-LSTM, and CRF using a fusion strategy. EHRs were first cleaned, encoded, and segmented. Then, semantic representations were dynamically merged using the BART model. Subsequently, sequential data were captured by the Bi-LSTM network, and CRF was utilized to decode and produce multi-task entity recognition outcomes. The approach produced an average precision, recall, and F1-score of 0.880, 0.887, and 0.883, respectively. Based on the BiLSTM-CRF deep learning model, the authors in [37] presented the Bi-RNN-LSTM-RNN-CRF electronic medical record named entity recognition model. A bidirectional RNN-LSTM-RNN layer was trained by first gathering an electronic medical record dataset and then utilizing a word vector tool to turn the characters into vectors. After that, a CRF layer received the training data to compute the loss function and produce predictions. To compare the two models, identical procedures were carried out using the conventional BiLSTM-CRF model. According to the experimental results, the Bi-RNN-LSTM-RNN-CRF model outperformed the BiLSTM-CRF model in recognition, with an F1-score of 97.80%. Xiaocheng et al. [19] used a manually annotated corpus to refine the bidirectional transformer pre-training model BERT by the BIOES (Begin Inside Outside End Single) standard. Word vectors, which successfully capture the context in electronic medical records, represent the semantic content of words in the text, which is learned unsupervised. Character sequences are sent into the BERT model, which learns their state properties. The CRF layer uses this information to optimize sequence transition constraints. The BERT-IDCNN-CRF model obtained an average accuracy of 94.5%, recall of 93.8%, and F1-score of 94.1%, respectively.

Our research demonstrates that CRF performs better than Bi-LSTM-CRF in the majority of the 23 assessed categories. This is probably because of the size of the dataset and the significant chance of overfitting in deep learning models on smaller datasets. Our findings show that while Bi-LSTM-CRF shows potential for larger and more diverse datasets, CRF offers a solid baseline with robust generalization. The dual-model strategy used in our work demonstrates the significance of choosing a model based on the properties of the data and the necessity of giving similar importance to deep learning and simpler architectures to attain the best results across a variety of datasets. By doing this, we offer a thorough analysis that clarifies how our methodology successfully manages the trade-offs between model complexity and performance, resulting in a more sophisticated knowledge of PHI detection challenges. Table 11 compares the methods for de-identification used in other recent research.

Table 11. Comparison of methods for medical text de-identification and entity recognition.

Reference	Methods	Dataset	Results	Limitations
[34]	Created a method for automatically de-identifying clinical documents.	Rennes University Hospital's EHRs	F1-score of 96.96%	Lack of annotated data for medical de-identification.
[35]	Three different models: Bi-LSTM, Bi-LSTM with attention, and BERT base models	206,926 sentences used for classification.	Bi-LSTM achieved accuracy of 90%, BERT accuracy of 93%	High computational cost
[36]	A novel multi-medical entity recognition approach combining BART, Bi-LSTM, and CRF using a fusion strategy.	CCKS2019 dataset	Precision 0.880, Recall 0.887, F1-score 0.883	Lack of comparative analysis with other models
[37]	Based on BiLSTM-CRF, the Bi-RNN-LSTM-RNN-CRF medical record named entity recognition model.	No specific dataset details were provided.	F1 97.80%	Slightly inferior recognition effect compared to BiLSTM-CRF model.
[19]	Refines the bidirectional transformer pre-training model BERT using BIOES standard.	No specific dataset details were provided.	Accuracy: 94.5%, Recall: 93.8%, F1: 94.1%	Bi-LSTM-CRF cannot utilize GPU parallelism.
Our work	Using Bi-LSTM-CRF model and deep learning approach	i2b2 dataset	Precision: 99%, Recall: 98%, F1-score: 98%	Limited dataset size impacts model optimization

5. Conclusions and Potential Future Works

The privacy protection mechanism in healthcare has become critical due to the rapid global adoption of EHRs, which contain vast amounts of unstructured textual data. The sensitive nature of Protected Health Information (PHI) within these records presents significant privacy challenges, necessitating robust de-identification techniques. This paper introduces a novel Bi-LSTM-CRF model approach, utilizing the i2b2-2014 dataset from Harvard University to achieve accurate and reliable PHI de-identification. Unlike prior studies that often unify Bi-LSTM and CRF layers, our approach emphasizes the individual design, optimization, and hyperparameter tuning of both components, resulting in precise model performance improvements. This rigorous approach to architecture and tuning, typically underexplored in the existing literature, enhances the model's capacity to effectively detect PHI tags while retaining the essential clinical context.

Comprehensive evaluations were conducted across the 23 PHI categories defined by HIPAA, ensuring robust security across vital domains. The optimized model demonstrates exceptional performance, achieving a precision of 99%, recall of 98%, and an F1-score of 98%, which underscores its effectiveness in balancing recall and precision. By enabling the de-identification of medical records, this research strengthens patient confidentiality, promotes compliance with privacy regulations, and facilitates safe data sharing for research and analysis. Interestingly, certain PHI categories displayed outstanding accuracy, confirming the model's efficacy in identifying and safeguarding PHI within medical records.

The performance differences observed between the CRF and Bi-LSTM-CRF models suggest a range of influences, including dataset characteristics, specific tasks, and model architecture.

6. Future Work

In future research, we plan to involve integrating distributed machine learning techniques, such as federated learning, into the Bi-LSTM-CRF model. Federated learning enables training machine learning models across different devices without centralizing the data, thus preserving user privacy [38]. This approach guarantees privacy by keeping data on local devices and improves model robustness by leveraging diverse data sources. In future work, our objective is to explore the application of federated and split learning-based methods, incorporating robust aggregation techniques to enhance the privacy, security, and performance of the Bi-LSTM-CRF model. By combining these methods, we can ensure that the model maintains privacy while improving its predictive accuracy in the de-identification of Protected Health Information (PHI).

Author Contributions: Conceptualization, A.A., E.E.A. and R.A.-Z.; Methodology, A.A. and E.E.A.; Software, E.E.A., E.A.E. and M.A.-H.; Validation, A.A., E.E.A. and R.A.-Z.; Formal analysis, E.E.A. and E.A.E.; Investigation, R.A.-Z., E.E.A. and M.A.-H.; Resources, E.E.A. and E.A.E.; Data curation, A.A.; Writing—original draft preparation, A.A., E.E.A. and R.A.-Z.; Writing—review and editing, E.E.A., E.A.E. and M.A.-H.; Visualization, M.A.-H.; Supervision, A.A. and E.E.A.; Project administration, A.E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hashemite University grant number H6574.

Data Availability Statement: The dataset used in this research is not publicly available due to privacy, legal, and ethical restrictions. It was obtained from Harvard University under a Data Use and Confidentiality Agreement with Partners HealthCare System, Inc., which provides access to de-identified patient discharge summaries for academic research purposes. Researchers can request access to this dataset by completing the required registration and adhering to the terms of the Data Usage Agreement (DUA).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ahmed, T.; Aziz, M.M.A.; Mohammed, N.; Jiang, X. Privacy preserving neural networks for electronic health records de-identification. In Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, Gainesville, FL, USA, 1–4 August 2021; pp. 1–6.
2. Javaid, M.; Haleem, A.; Singh, R.P. Health informatics to enhance the healthcare industry's culture: An extensive analysis of its features, contributions, applications and limitations. *Inform. Health* **2024**, *1*, 123–148. [CrossRef]
3. Oladele, J.K.; Ojugo, A.A.; Odiakaose, C.C.; Emordi, F.U.; Abere, R.A.; Nwozor, B.; Ejeh, P.O.; Geteloma, V.O. BEHedas: A blockchain electronic health data system for secure medical records exchange. *J. Comput. Theor. Appl.* **2024**, *1*, 231–242. [CrossRef]
4. Okolo, C.A.; Ijeh, S.; Arowoogun, J.O.; Adeniyi, A.O.; Omotayo, O. Reviewing the impact of health information technology on healthcare management efficiency. *Int. Med. Sci. Res. J.* **2024**, *4*, 420–440. [CrossRef]
5. Adeniyi, A.O.; Arowoogun, J.O.; Chidi, R.; Okolo, C.A.; Babawarun, O. The impact of electronic health records on patient care and outcomes: A comprehensive review. *World J. Adv. Res. Rev.* **2024**, *21*, 1446–1455. [CrossRef]
6. Kehl, K.L.; Jee, J.; Pichotta, K.; Trukhanov, P.; Fong, C.; Waters, M.; Nichols, C.; Cerami, E.; Schrag, D.; Schultz, N.; et al. Shareable artificial intelligence to extract cancer outcomes from electronic health records. *Nat. Commun.* **2024**, *15*, 9787. [CrossRef]
7. Ajegbile, M.D.; Olaboye, J.A.; Maha, C.C.; Igwama, G.T.; Abdul, S. The role of data-driven initiatives in enhancing healthcare delivery and patient retention. *World J. Biol. Pharm. Health Sci.* **2024**, *19*, 234–242. [CrossRef]
8. Corte-Real, A.; Nunes, T.; da Cunha, P.R. Reflections about Blockchain in Health Data Sharing: Navigating a Disruptive Technology. *Int. J. Environ. Res. Public Health* **2024**, *21*, 230. [CrossRef]
9. Isibor, E. Regulation of Healthcare Data Security: Legal Obligations in A Digital Age. *SSRN* **2024**. . [CrossRef]
10. Alves, V.M.R.G. De-Identification of Clinical Text Using Sentence Embeddings. Master's Thesis, Universidade do Porto, Porto, Portugal, 2024.
11. Negash, B.; Katz, A.; Neilson, C.J.; Moni, M.; Nesca, M.; Singer, A.; Enns, J.E. De-identification of free text data containing personal health information: A scoping review of reviews. *Int. J. Popul. Data Sci.* **2023**, *8*, 2153. [CrossRef]
12. Jha, A.K.; DesRoches, C.M.; Campbell, E.G.; Donelan, K.; Rao, S.R.; Ferris, T.G.; Shields, A.; Rosenbaum, S.; Blumenthal, D. Use of electronic health records in US hospitals. *N. Engl. J. Med.* **2009**, *360*, 1628–1638. [CrossRef]
13. van der Linden, H.; Kalra, D.; Hasman, A.; Talmon, J. Inter-organizational future proof EHR systems: A review of the security and privacy related issues. *Int. J. Med. Inform.* **2009**, *78*, 141–160. [CrossRef] [PubMed]
14. Leevy, J.L.; Khoshgoftaar, T.M.; Villanustre, F. Survey on RNN and CRF models for de-identification of medical free text. *J. Big Data* **2020**, *7*, 73. [CrossRef]
15. Liu, Z.; Tang, B.; Wang, X.; Chen, Q. De-identification of clinical notes via recurrent neural network and conditional random field. *J. Biomed. Inform.* **2017**, *75*, S34–S42. [CrossRef] [PubMed]
16. Tang, B.; Jiang, D.; Chen, Q.; Wang, X.; Yan, J.; Shen, Y. De-identification of clinical text via Bi-LSTM-CRF with neural language models. In Proceedings of the AMIA Annual Symposium Proceedings, Washington, DC, USA, 16–20 November 2019; American Medical Informatics Association: Bethesda, MD, USA, 2019; Volume 2019, p. 857.
17. Zhang, H.; Kang, X.; Li, B.; Wang, Y.; Liu, H.; Bai, F. Medical name entity recognition based on Bi-LSTM-CRF and attention mechanism. *J. Comput. Appl.* **2020**, *40*, 98-012.
18. Chinese-named entity recognition from adverse drug event records: Radical embedding-combined dynamic embedding–based BERT in a bidirectional long short-term conditional random field (Bi-LSTM-CRF) model. *JMIR Med. Inform.* **2021**, *9*, e26407. [CrossRef] [PubMed]
19. Cai, X.; Sun, E.; Lei, J. Research on application of named entity recognition of electronic medical records based on BERT-IDCNN-CRF model. In Proceedings of the 6th International Conference on Graphics and Signal Processing, Chiba, Japan, 1–3 July 2022; pp. 80–85.
20. Gao, M.; Xiao, Q.; Wu, S.; Deng, K. An attention-based ID-CNNs-CRF model for named entity recognition on clinical electronic medical records. In Proceedings of the International Conference on Artificial Neural Networks, Munich, Germany, 17–19 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 231–242.
21. Zavala, R.M.R.; Martínez, P.; Segura-Bedmar, I. A Hybrid Bi-LSTM-CRF model for Knowledge Recognition from eHealth documents. In Proceedings of the TASS@ SEPLN, Seville, Spain, 18 September 2018; pp. 65–70.
22. Liu, S.; Nie, W.; Gao, D.; Yang, H.; Yan, J.; Hao, T. Clinical quantitative information recognition and entity-quantity association from Chinese electronic medical records. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 117–130. [CrossRef]
23. Zhang, S.; Li, Y.; Li, S.; Yan, F. Bi-LSTM-CRF network for clinical event extraction with medical knowledge features. *IEEE Access* **2022**, *10*, 110100–110109. [CrossRef]

24. Khullar, S.; Singh, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. *Environ. Sci. Pollut. Res.* **2022**, *29*, 12875–12889. [CrossRef]
25. Liu, H.; Qin, Y.; Chen, H.Y.; Wu, J.; Ma, J.; Du, Z.; Wang, N.; Zou, J.; Lin, S.; Zhang, X.; et al. Artificial neuronal devices based on emerging materials: Neuronal dynamics and applications. *Adv. Mater.* **2023**, *35*, 2205047. [CrossRef]
26. Dhoot, A.; Deva, R.; Shukla, V. A Novel Security Model for Healthcare Prediction by Using DL. In Proceedings of the International Conference on Cryptology & Network Security with Machine Learning, Kanpur, India, 27–29 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 787–799.
27. Kumar, V.; Stubbs, A.; Shaw, S.; Uzuner, Ö. Creation of a new longitudinal corpus of clinical narratives. *J. Biomed. Inform.* **2015**, *58*, S6–S10. [CrossRef]
28. Stubbs, A.; Uzuner, Ö. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. *J. Biomed. Inform.* **2015**, *58*, S20–S29. [CrossRef]
29. Stubbs, A.; Kotfila, C.; Uzuner, Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J. Biomed. Inform.* **2015**, *58*, S11–S19. [CrossRef]
30. De Santis, E.; Martino, A.; Ronci, F.; Rizzi, A. From Bag-of-Words to Transformers: A Comparative Study for Text Classification in Healthcare Discussions in Social Media. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *early access*. . [CrossRef]
31. Sumukh, S. Better Understanding of Code-Mixed Social Media Data via Information Extraction. Ph.D. Thesis, International Institute of Information Technology Hyderabad, Hyderabad, India, 2023.
32. Gandhi, H.; Attar, V. Extracting aspect terms using CRF and bi-LSTM models. *Procedia Comput. Sci.* **2020**, *167*, 2486–2495. [CrossRef]
33. Ullah, S.; Ali, N.I.; Chandio, S.M.; Brohi, I.A.; Laghari, B.A. Part-Of-Speech Tagging for Balochi Language: A Data driven application of Conditional Random Fields. *Asian Bull. Big Data Manag.* **2024**, *4*, 229–236. [CrossRef]
34. Azzouzi, M.E.; Coatrieux, G.; Bellafqira, R.; Delamarre, D.; Riou, C.; Oubenali, N.; Cabon, S.; Cuggia, M.; Bouzillé, G. Automatic de-identification of French electronic health records: A cost-effective approach exploiting distant supervision and deep learning models. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 54. [CrossRef]
35. Zalte, J.; Shah, H. Contextual classification of clinical records with bidirectional long short-term memory (Bi-LSTM) and bidirectional encoder representations from transformers (BERT) model. *Comput. Intell.* **2024**, *40*, e12692. [CrossRef]
36. Du, H.; Xu, J.; Du, Z.; Chen, L.; Ma, S.; Wei, D.; Wang, X. MF-MNER: Multi-models Fusion for MNER in Chinese Clinical Electronic Medical Records. *Interdiscip. Sci. Comput. Life Sci.* **2024**, *16*, 489–502. [CrossRef] [PubMed]
37. Dai, C.; Zhuang, X.; Cai, J. Chinese Electronic Medical Record Named Entity Recognition Based on Bi-RNN-LSTM-RNN-CRF. In Proceedings of the 2022 11th International Conference on Computing and Pattern Recognition, Beijing, China, 17–19 November 2022; pp. 577–583.
38. Taheri, R.; Arabikhan, F.; Gegov, A.; Akbari, N. Robust Aggregation Function in Federated Learning. In Proceedings of the International Conference on Information and Knowledge Systems, Portsmouth, UK, 22–23 June 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 168–175.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.