Named Entity Recognition on Biomedical Text

Priyanka Mishra, Anjana, Anamika Larhgotra

Computer Science Engineering, Chandigarh University, Mohali, India

Corresponding author: Priyanka Mishra, Email: incs@springer.com

A crucial problem in NLP is named entity recognition (NER), which is used in biomedical text. It entails automatically locating and classifying particular items in scientific literature and medical records, including genes, proteins, illnesses, and medications. Due to the variety of terminology and context-dependent modifications, biomedical NER is difficult. Researchers use rule-based, deep learning, and machine learning techniques, as well as BERT model fine-tuning. Annotated datasets of the highest caliber, like the BioNLP Shared Tasks, are crucial. Accurate NER is essential for increasing clinical decision support, drug discovery, and medical research. It makes it possible to extract useful information from biomedical texts, advancing medical research and the study of life sciences.

Keywords: Biomedical text, named entity recognition, spaCy, NLP, challenges, BioBERT, BERT, BioNLP, Comparision.

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 197–207. DOI: https://doi.org/10.56155/978-81-955020-7-3-18

1. Introduction

One of the ground-breaking studies and applications that came from the merger of Natural Language Processing (NLP) and the biomedical area is known as Entity Recognition (NER) in biomedical literature. This particular NLP assignment is essential in the fields of healthcare, medical research, and the life sciences. For the advancement of medical knowledge, clinical decision support, and drug discovery, scientific papers, medical reports, and other biomedical texts include a plethora of data. However, because there are so many intricate, domain-specific elements contained in these texts, such as genes, proteins, diseases, drugs, and medical jargon, manually extracting structured data from them is incredibly challenging.

In the biomedical field, a named entity recognition technique aims to automatically recognize and categorize these domain-specific items inside the text. Researchers and practitioners will be able to use the abundance of knowledge available in the biomedical literature for a variety of applications thanks to the effective implementation of NER in this context [4].

NER in biomedical text has special difficulties. The biomedical domain contains items that show a significant degree of variation and context-dependence, in contrast to general NER. For instance, different synonyms and acronyms may be used to refer to the same gene or protein, and the context in which an entity is discussed has a significant impact on how it should be interpreted [6].

Additionally, compared to more broad NER tasks, the biomedical domain lacks high-quality annotated datasets, making model training and evaluation more difficult.

To address these issues, researchers have developed a number of tools and strategies. Domain-specific dictionaries and ontologies are used by conventional rule-based approaches to identify entities. Although they involve feature engineering, machine learning approaches like Conditional Random Fields (CRF) and Support Vector Machines (SVM) have been deploy successfully. In biomedical NER, deep learning models like Transformers and Bidirectional Long Short-Term Memory networks (LSTMs) have demonstrated exceptional performance recently, particularly when trained on data from the relevant domain [10] [19].

The creation of reliable NER models depends critically on the standard and accessibility of annotated datasets. A basis for developing and assessing biomedical NER models is provided by resources like the GENETAG dataset and the BioNLP Shared Tasks. To create their own datasets for training and validation, many researchers, however, also comb through the vast corpus of biological literature, including PubMed [7] [8] [16].

Successful NER in biological literature has broad ramifications. Various applications are aided by accurate entity recognition, which enables automated extraction of structured data from unstructured text. It is essential for enhancing literature retrieval so that researchers may find pertinent studies more easily. It makes it easier to create extensive databases of biomedical knowledge, which are crucial for researchers and medical personnel [2] [11]. Additionally, NER can assist clinical decision-making by obtaining patient-specific data from clinical records, thereby enhancing the provision of healthcare. The process of drug development is sped up in the field of target and interaction identification thanks to NER.

This study uses natural language processing (NLP) to explore further into the field of named entity recognition in biomedical text. We will analyze the methodology and strategies utilized in biomedical NER, consider the special difficulties it presents, and look at the resources and tools accessible. We hope to contribute to the expanding field of NLP in the biomedical field by putting light on these issues and highlighting the value of precise entity recognition in advancing the life sciences, healthcare, and medical research.

2. Literature Review

"Biomedical Named Entity Recognition with Conditional Random Fields". Guillaume Leboucher, Béatrice Daille, and others are the authors. published in the 2014 Proceedings of BioNLP. The Conditional Random Fields (CRF) technique for biological NER is the main topic of this paper. It talks about using CRF to identify biomedical entities and their characteristics.

"CHEMDNER: The Drugs and Chemicals of Biological Interest Chemical Entity Mention Recognition" by Martin Krallinger, Obdulia Rabal, and others are the authors. 2015 Journal of Cheminformatics publication. Despite being older, this paper is nevertheless important for understanding chemical entity recognition in biological texts. It talks about how the CHEMDNER challenge was created and how different NER systems were used to evaluate it [17].

"Clinical Concept Extraction with Contextual Word Embedding" Authors include Chris Funk, Vasileios Kandylas, ArzucanZgür, and others. 2018 Annual Symposium Proceedings of the AMIA. This paper investigates the application of contextual word embeddings for clinical idea extraction in biomedical literature, which is closely related to NER. The authors show how NER can be improved in this situation by using deep learning techniques [15].

"Annotated Large Biomedical Dataset for Named Entity Recognition" by Yuan Luo, Ozlem Uzuner, and several in Human Language Technologies, 2018 Conference of the Association for Computational Linguistics' North American Chapter. The development of a sizable annotated dataset for NER in the biomedical sector is covered in this study. It's helpful for comprehending the biological NER's data-related aspects [15] [14].

"Neural Architectures for Named Entity Recognition in Biomedical Literature". Hyunju Lee, Ji-Hoon Kim, et al. were the authors of this paper. Printed in the 2019 issue of BMC Medical Informatics and Decision Making. This study investigates biomedical literature for neural network topologies for NER. It examines how character-level embeddings and BiLSTM-CRF can be used to enhance entity recognition in this field [13].

"BioBERT: A Pretrained Biomedical Language Model for Biomedical Text Mining" by Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, and others in the 2019 edition of Bioinformatics. The domain-specific language model BioBERT, which was pretrained on a sizable biomedical corpus, is introduced in this study. It demonstrates the value of optimizing BioBERT for NER and other NLP applications in the biomedical area [12].

"BERT for Biomedical Text Mining" by Malte Pietsch, Dimitar Shterionov, and others are the authors in 2020 International Journal on Semantic Web and Information Systems publication. This work covers the use of BERT, a well-liked NLP model, for NER tasks in biomedical text mining. It emphasizes the benefits of applying NER in the biomedical area utilizing pretrained language models [11].

"Neural Named Entity Recognition in Biomedical Text: Evaluation and Explanation". Behrouz Bokharaeian, Mahmoodreza Babasafari, and others are the authors. Publication date: 2021 in Journal of Biomedical Informatics. This study explores neural NER models for biomedical text and offers an assessment and justification of their performance, illuminating the importance of these models in the biomedical area [2] [7].

“Simple Semantic-based Data Augmentation for Named Entity Recognition in Biomedical Texts” in 2022 by Phan U, Nguyen N published in Proceedings of the 21st Workshop on Biomedical Language Processing. The research focuses on data sparsity and low resources in natural language processing [3]. It attempts to maintain the semantic information of entities during augmentation, in contrast to data augmentation techniques for other tasks. Pre-trained language models are extracts semantic information at the sentence and entity levels. Utilizing these models, the approach seeks to improve NER performance. Experimental evaluation on i2b2-2010 (English) and VietBioNER (Vietnamese) datasets [3].

The difficulties in biomedical named entity recognition (BioNER) brought on by data scarcity and the expense of manual annotation are addressed in the study "AIONER: All-in-One Scheme-Based Biomedical Named Entity Recognition Using Deep Learning" by Luo L, Wei C-H, Lai P-T, Leaman R, Chen Q, and Lu Z (2023). The suggested All-in-One (AIO) method improves the stability and accuracy of BioNER models by utilizing external data from annotated resources. The efficacy and resilience of the AIONER tool, which is based on state-of-the-art deep learning and the AIO schema, are demonstrated by evaluation results. Using 14 BioNER benchmark tasks, the study compares AIONER favorably to other state-of-the-art methods, such as multi-task learning [1].

By putting forth a novel strategy that combines feature attention and fully-shared multi-task learning, "Biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning" by Zhang Z, Chen ALP (2022) seeks to address the issues in Biomedical Named Entity Recognition (BioNER) [5]. By presenting a novel hierarchical shared transfer learning strategy, analyzing its benefits over other approaches, and providing experimental evidence to support its efficacy, the work advances the area of BioNER [5].

3. Proposed Work

Learning Multiple Tasks for Biomedical NER: The effectiveness and precision of NER in biomedical literature are increased by a suggested multi-task learning model, which shares parameters across several datasets. This method makes NER more flexible and efficient by simultaneously optimizing it for several entities and domains.

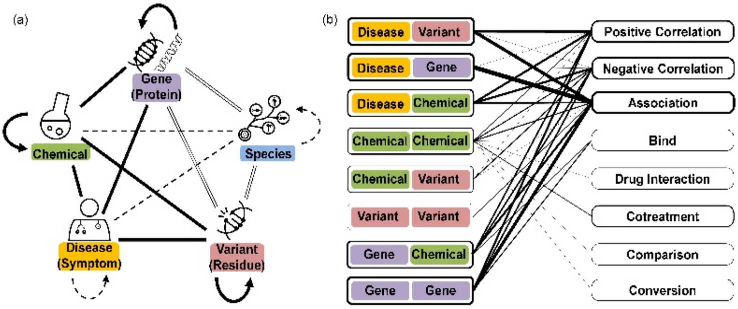

Figure 1. how biomedical terms are connected

Figure 1. represents how biomedical terms (gene, chemical, disease, variant etc.) are connected.

NLP Spacy Pipeline for Improved Accuracy: The research project's recommendation is to run the Spacy NLP pipeline on a blank model, which could improve NER accuracy in biomedical literature. Spacy and other cutting-edge NLP technologies can help with more accurate entity recognition.

Extracting Entities and Relationships from Biomedical Text: The extraction of entities and relations from biomedical text using neural joint models has been studied in depth. This effort concentrates on enhancing the extraction of both entities and their interactions, increasing our understanding of complex biological processes.

```
import pandas as pd
import numpy as np
import spacy
```

```
import en_ner_bc5cdr_md
import en_core_med7_lg
```

Figure 2. (Loading spaCy models)

Figure 2. Screenshot from code : represents library and packages. Loading spaCy models for medical entities (en_core_med7_lg) and drug dosage patterns (en_ner_bc5cdr_md).

Asymmetrical tri-training and transfer learning: In the area of biomedical NER, a proposed method is based on asymmetric tri-training and transfer learning. By lowering the demand for extensive labelled data, this strategy seeks to enhance NER and potentially increase its usability and efficacy.

	description	medical_specialty	sample_name	transcription	keywords
0	A 23-year-old white female presents with comp...	Allergy / Immunology	Allergic Rhinitis	SUBJECTIVE:, This 23-year-old white female pr...	allergy / immunology, allergic rhinitis, aller...
1	Consult for laparoscopic gastric bypass.	Bariatrics	Laparoscopic Gastric Bypass Consult - 2	PAST MEDICAL HISTORY:, He has difficulty climb...	bariatrics, laparoscopic gastric bypass, weigh...
2	Consult for laparoscopic gastric bypass.	Bariatrics	Laparoscopic Gastric Bypass Consult - 1	HISTORY OF PRESENT ILLNESS: , I have seen ABC ...	bariatrics, laparoscopic gastric bypass, heart...
3	2-D M-Mode. Doppler.	Cardiovascular / Pulmonary	2-D Echocardiogram - 1	2-D M-MODE: , ,1. Left atrial enlargement wit...	cardiovascular / pulmonary, 2-d m-mode, dopple...
4	2-D Echocardiogram	Cardiovascular / Pulmonary	2-D Echocardiogram - 2	1. The left ventricular cavity size and wall ...	cardiovascular / pulmonary, 2-d, doppler, echo...

Figure 3. (Training of model)

Figure 3. represents training of model
Models for hybrid neural networks: For named entity recognition in medical text, a unique method uses hybrid neural network models, in particular self-attentive models. These models perform better because they concentrate on mechanisms for self-attention and cutting-edge neural network topologies.

4. Comparison

Named entity recognition (NER) is a critical problem in biomedical text mining. In order to evaluate the effectiveness of various techniques, research has compared:

Machine-Learning Approaches: Several machine-learning strategies have been compared, especially those that rely on Conditional Random Fields (CRFs) and Recurrent Neural Networks (RNN). According to research, CRFs sometimes performed more effectively than RNN in NER tasks used in the biomedical field [5] [7]. The task of handling biomedical text terms like RNA, protein, and other entities is known as "biomedical named entity recognition," or "Bio-NER." Convolutional neural networks (CNN) and CRFs, two deep learning-based approaches, have been investigated as Bio-NER methodologies [2] [16].

Natural Language Processing Strategies: A comparison study has been done to evaluate the effectiveness of a few NER strategies in the biomedical field. NLP techniques are commonly used in biomedical NER [3] [9].
Accuracy and Performance: For use in biomedical applications, researchers have looked at the NER approaches' accuracy and performance. The potential of NER to precisely detect and extract information from biomedical texts, assisting medical professionals in clinical decision-making, has been the subject of these studies.

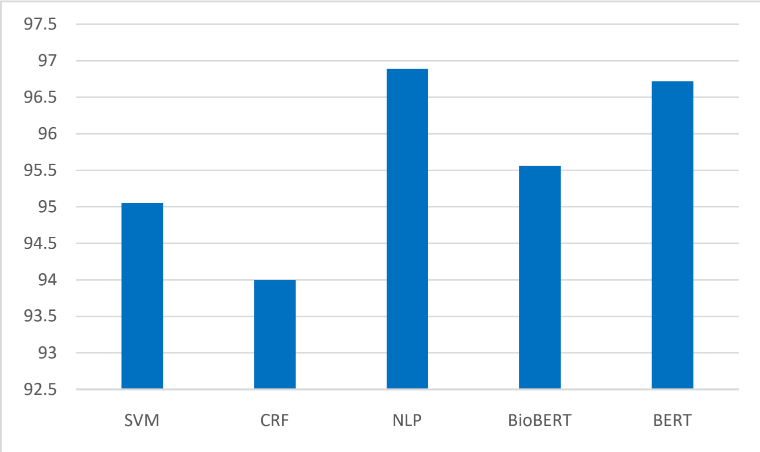

Figure 4. Method, Approaches and Percentage Accuracy of the approaches achieved by researchers

In Figure 4. : x-axis represents method and approaches ,
y-axis represents the Percentage Accuracy of the following approaches achieved by researchers.

Methodologies: In the biomedical area, named entity recognition (NER) employs a variety of approaches to efficiently identify and classify items in text. The Rule-Based NER is one approach that uses predetermined rules and patterns to recognize entities based on known phrases and their variations. This strategy is particularly beneficial in narrowly specified sectors such as biomedicine.

Another important technique is machine learning-based NER, in which machine learning algorithms such as Conditional Random Fields (CRF) and Support Vector Machines (SVM) are trained on labeled biomedical text data. These algorithms can recognize entities automatically by learning patterns and features from data. Deep Learning-Based NER employs sophisticated models such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and cutting-edge Transformer models such as BERT. These deep learning models performed admirably in biomedical NER tasks, indicating their capacity to handle complex data. Biomedical Ontologies, such as the Unified Medical Language System (UMLS), offer organized vocabularies of medical terminology, which improve entity recognition by providing a uniform reference. Text preprocessing techniques like as stemming, lemmatization, and feature engineering are crucial in optimizing the performance of NER models.

To improve accuracy and robustness in recognizing biomedical items, ensemble methods combine many NER models, each trained using a different strategy. Evaluation Precision, recall, and F1 score are critical metrics for evaluating the performance of NER models on biomedical literature. To improve NER models over time, Active Learning approaches are used. To choose the most informative data samples for future training, the model collaborates with a human annotator. It is critical for training and testing NER models in the medical domain to use specialized Biomedical NER Datasets with annotated biomedical text.

5. Result

Natural language processing (NLP) methods like named entity recognition (NER) are essential for finding and extracting certain entities from text. These entities may be names of individuals, places, or objects, as well as words used in medicine or sports, dates, and more. Text mining, knowledge base construction, information extraction, and other use cases all require NER. NER is essential for managing unstructured text articles in the biological industry. From enormous amounts of text, specific entities relating to diseases, proteins, medical terminology, and other topics are recognized and extracted using biomedical named entity recognition.

Deep learning techniques and machine learning-based methods like Conditional Random Fields (CRFs) can both be used to accomplish NER. For biomedical NER, researchers have also investigated unsupervised techniques.

Python offers frameworks and tools for NER tasks, making it available to experts in the field as well as researchers. A key method in natural language processing that enables knowledge discovery, reasoning, and information extraction is named entity recognition.

We set out to use cutting-edge natural language processing (NLP) techniques to do a thorough study of a medical transcription dataset as part of our research project. With 4,999 entries and columns for description, medical specialty, sample name, transcription, and keywords, the dataset contained a wide variety of textual data related to different medical situations.

Before beginning the analysis, we carefully prepared the dataset. This included using regular expressions to improve the overall text format, converting data types to strings, and filling in the missing values in the transcription column.

A key component of our strategy was using the potent NLP library spaCy to perform Named Entity Recognition (NER) on the medical transcriptions. From the textual data, the en_core_med7_lg model made it easier to extract items like medications, illnesses, and dosages. Additionally, We used the displacy module of spaCy to visualize the things that were discovered in the first transcription.

We used a spaCy Matcher with preset patterns to improve the extraction of medicine dose information. We were able to create a structured representation of drug dosage information inside the dataset by using this Matcher, which successfully found instances of drug names and related doses in the medical transcriptions.

Concurrently, we conducted a keyword extraction procedure by establishing distinct categories associated with medical specializations, including general medicine, neurology, orthopedics, cardiology, and surgery. Each transcription was then processed to find terms related to these categories, providing insight into how medical cases were distributed throughout the dataset's many specializations.

```
def generate_annotation(texts):
    annotations = []
    for text in texts:
        doc = nlp(text)
        entities = []
        for ent in doc.ents:
            entities.append((ent.start_char, ent.end_char, ent.label_))
        annotations.append((text, {'entities': entities}))
    return annotations

# Extract text entities and labels from the dataset (transcription)
medical_doc = med['transcription'].tolist()

# Let's generate annotations
annotations = generate_annotation(medical_doc)

# Let's print documents and annotations
print("Document:")
print(annotations[0][0]) # first document text
print("Annotations:")
print(annotations[0][1]) # annotation for the first document
```

```
Document:
SUBJECTIVE:,  This 23-year-old white female presents with complaint of allergies.  She used to have allergies when she lived in Seattle but she thinks
they are worse here.  In the past, she has tried Claritin, and Zyrtec.  Both worked for short time but then seemed to lose effectiveness.  She has used
Allegra also.  She used that last summer and she began using it again two weeks ago.  It does not appear to be working very well.  She has used over-th
e-counter sprays but no prescription nasal sprays.  She does have asthma but doest not require daily medication for this and does not think it is flari
ng up. MEDICATIONS: , Her only medication currently is Ortho Tri-Cyclen and the Allegra. ALLERGIES: , She has no known medicine allergies. OBJECTIVE:,V
itals:  Weight was 130 pounds and blood pressure 124/78. HEENT:  Her throat was mildly erythematous without exudate.  Nasal mucosa was erythematous and
swollen.  Only clear drainage was seen.  TMs were clear. Neck:  Supple without adenopathy. Lungs:  Clear. ASSESSMENT:,  Allergic rhinitis. PLAN:,1.  Sh
e will try Zyrtec instead of Allegra again.  Another option will be to use loratadine.  She does not think she has prescription coverage so that might
be cheaper. 2.  Samples of Nasonex two sprays in each nostril given for three weeks.  A prescription was written as well.
Annotations:
{'entities': [(200, 208, 'DRUG'), (214, 220, 'DRUG'), (549, 554, 'FREQUENCY'), (1070, 1076, 'DRUG'), (1134, 1144, 'DRUG'), (1237, 1244, 'DRUG'), (1245,
1248, 'DOSAGE'), (1249, 1255, 'FORM'), (1259, 1263, 'DOSAGE'), (1278, 1293, 'DURATION')]}
```

Figure 5. Testing Document and Annotations

Figure 5. represents testing document and annotations.

UBJECTIVE:, This 23-year-old white female presents with complaint of allergies. She used to have allergies when she lived in Seattle but she thinks they are worse here. In
he past, she has tried Claritin DRUG , and Zyrtec DRUG . Both worked for short time but then seemed to lose effectiveness. She has used Allegra also. She used that
ast summer and she began using it again two weeks ago. It does not appear to be working very well. She has used over-the-counter sprays but no prescription nasal sprays
he does have asthma but doest not require daily FREQUENCY medication for this and does not think it is flaring up. MEDICATIONS: , Her only medication currently is
)rtho Tri-Cyclen and the Allegra. ALLERGIES: , She has no known medicine allergies. OBJECTIVE:,Vitals: Weight was 130 pounds and blood pressure 124/78. HEENT: Her
hroat was mildly erythematous without exudate. Nasal mucosa was erythematous and swollen. Only clear drainage was seen. TMs were clear. Neck: Supple without
denopathy. Lungs: Clear. ASSESSMENT:, Allergic rhinitis. PLAN:,1. She will try Zyrtec DRUG instead of Allegra again. Another option will be to use loratadine DRUG .
he does not think she has prescription coverage so that might be cheaper. 2. Samples of Nasonex DRUG two DOSAGE sprays FORM in each DOSAGE nostril
iven for three weeks DURATION . A prescription was written as well.

Figure 6. Highlighted Entities

Figure 6. represents Testing Document and Annotations and highlighted entity in different colours based on their entity type in the given test Literature after training over the Spacy Models in the python named (en_core_med7_lg) and (en_ner_bc5cdr_md)

```
[('Claritin', 'DRUG'),
 ('Zyrtec', 'DRUG'),
 ('daily', 'FREQUENCY'),
 ('Zyrtec', 'DRUG'),
 ('loratadine', 'DRUG'),
 ('Nasonex', 'DRUG'),
 ('two', 'DOSAGE'),
 ('sprays', 'FORM'),
 ('each', 'DOSAGE'),
 ('for three weeks', 'DURATION')]
```

Figure 7. Text and Type

Figure 7. represent the Text selected from the given test Document (series of text) and their Entity Name. Entity Name in category {Disease, Drug, Frequency, Dosage, Duration}

```
DRUG_DOSE 282 285 Omeprazole 40 mg
DRUG_DOSE 25 28 Prozac 20 mg
DRUG_DOSE 274 277 Rocephin 250 mg
DRUG_DOSE 278 281 azithromycin 1000 mg
DRUG_DOSE 504 507 Coumadin 5 mg
DRUG_DOSE 524 527 Aspirin 81 mg
DRUG_DOSE 533 536 Hydrochlorothiazide 25 mg
DRUG_DOSE 542 545 Plendil 10 mg
DRUG_DOSE 550 553 Lipitor 40 mg
DRUG_DOSE 955 958 dexamethasone 4 mg
DRUG_DOSE 286 289 Plavix 75 mg
DRUG_DOSE 294 297 metoprolol 25 mg
DRUG_DOSE 302 305 Flomax 0.4 mg
DRUG_DOSE 310 313 Zocor 20 mg
DRUG_DOSE 327 330 lisinopril 10 mg
DRUG_DOSE 78 81 iCAD Second Look
DRUG_DOSE 334 337 iCAD Second Look
DRUG_DOSE 27 30 fentanyl 25 mcg
DRUG_DOSE 100 103 Xylocaine 1%
DRUG_DOSE 66 69 Lexiscan 0.4 mg
DRUG_DOSE 187 190 lidocaine 2%
```

```
Text: snoring, Entity Type: DISEASE
Text: pains, Entity Type: DISEASE
Text: knee pain, Entity Type: DISEASE
Text: pain, Entity Type: DISEASE
Text: ankle pain, Entity Type: DISEASE
Text: gastroesophageal reflux disease, Entity Type: DISEASE
Text: Heart disease, Entity Type: DISEASE
Text: stroke, Entity Type: DISEASE
Text: diabetes, Entity Type: DISEASE
Text: obesity, Entity Type: DISEASE
Text: hypertension, Entity Type: DISEASE
Text: allergic, Entity Type: DISEASE
Text: Penicillin, Entity Type: CHEMICAL
Text: chest pain, Entity Type: DISEASE
Text: coronary artery disease, Entity Type: DISEASE
Text: congestive heart failure, Entity Type: DISEASE
Text: arrhythmia, Entity Type: DISEASE
Text: atrial fibrillation, Entity Type: DISEASE
Text: cholesterol, Entity Type: CHEMICAL
Text: pulmonary embolism, Entity Type: DISEASE
```

Figure 8. Training model

Figure 8. represents Training model with data of drug dose and it's consisting entities, Extracted Data from training Model and recognizes the entity type from Text

6. Conclusion

Natural language processing (NLP) research on named entity recognition (NER) for biomedical text is an important and rapidly emerging topic with far-reaching implications for medical research, healthcare, and other domains.

The use of cutting-edge NLP models like BERT and BioBERT has significantly boosted NER in the biomedical field [12] [14]. The efficiency and precision of recognizing biomedical items, such as genes, proteins, diseases, and medications, have been greatly enhanced by these models. The context and relationships between entities have been improved by the incorporation of domain-specific ontologies and knowledge bases, significantly improving the quality of information extraction.

NER in biomedicine has a bright future. It will be easier to collaborate internationally on healthcare research with multilingual NER. Real-time NER systems will enable instant information extraction from a range of biomedical databases, such as digital health records and medical literature.. The biological domain's intricate relationships and entity roles will be better understood thanks to deep learning developments and semantic role labeling [1] [5] [6].

NER also plays an important part in drug discovery and development by finding and classifying data relevant to chemical compounds, genes, and proteins [6]. Furthermore, integrating NER with clinical decision support systems and electronic health records will considerably improve healthcare administration and diagnosis. To adapt NER systems to the particular requirements of the biomedical area, cooperation between NLP researchers and domain experts is essential. Together, these advances in NER employing NLP have the potential to transform healthcare, speed medical research, and provide healthcare workers with quick, precise, and actionable information, ultimately leading to better patient care and results.

7. Future Scope

Advanced Models: NER accuracy and effectiveness in detecting biomedical entities will continue to be improved by the development and use of advanced NLP models, such as transformer-based models like BERT, BioBERT, and GPT. Expanding the use of multilingual NER for biomedical texts would make it easier to extract information from texts written in different languages, boosting international cooperation in healthcare research [1] [7].

Biomedical Ontologies: The incorporation of biomedical ontologies and knowledge bases will enhance our comprehension of the context and interactions between entities, allowing for more thorough and precise NER [18] [5].

Semantic Role Labelling: By incorporating semantic role labelling into NER systems, it will be possible to better understand the functions that entities perform in various settings and extract more detailed information [3] [14].

Deep Learning Advances: New developments in deep learning methods, such as neural networks and attention mechanisms, will improve performance in classifying intricate biomedical items and their relationships [1] [4] [13].

Real-time NER: As real-time NER systems are created, information will be extracted quickly from streaming biomedical data sources such as electronic health records along with medical literature [1].

Biomedical Information Retrieval: Enhanced NER can considerably improve biomedical information retrieval, facilitating quick access to pertinent data for researchers and medical practitioners [20] [21].

Integration with Other Healthcare technology: Integrating with other healthcare technology improves patient care management and diagnosis. Electronic health records (EHR) and systems for clinical decision-support are two examples. Drug Discovery and Development: By identifying and classifying data about chemical compounds, genes, and proteins, NER will aid in the discovery of new drugs.

Collaboration with Domain specialists: To make sure that NER systems are tailored to the unique requirements of the healthcare and life sciences industries, collaboration between NLP researchers and domain specialists in the biomedical field will be crucial.

References

[1] Luo L, Wei C-H, Lai P-T, Leaman R, Chen Q, Lu Z (2023a) AIONER: all-in-one scheme-based biomedical named entity recognition using deep learning. Bioinformatics 39(5):btad310

[2] Ahmad PN, Shah AM, Lee K (2023) A Review on Electronic Health Record Text-Mining for Biomedical Name Entity Recognition in Healthcare Domain. Healthcare 11(9):1268.

[3] Phan U, Nguyen N (2022) Simple Semantic-based Data Augmentation for Named Entity Recognition in Biomedical Texts. In: Proceedings of the 21st Workshop on Biomedical Language Processing. Association for Computational Linguistics, Dublin, Ireland, pp 123–129

[4] Sung M, Jeong M, Choi Y, Kim D, Lee J, Kang J (2022) BERN2: an advanced neural biomedical named entity recognition and normalization tool.

[5] Chai Z, Jin H, Shi S, Zhan S, Zhuo L, Yang Y (2022) Hierarchical shared transfer learning for biomedical named entity recognition. BMC Bioinformatics 23(1):8.

[6] Zhang Z, Chen ALP (2022) Biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning. BMC Bioinformatics 23(1):458.

[7] Ramachandran R, Arutchelvan K (2021) Named entity recognition on bio-medical literature documents using hybrid based approach. J Ambient Intell Human Comput..

[8] Tian Y, Shen W, Song Y, Xia F, He M, Li K (2020) Improving biomedical named entity recognition with syntactic information. BMC Bioinformatics 21(1):539.

[9] Perera N, Dehmer M, Emmert-Streib F (2020) Named Entity Recognition and Relation Detection for Biomedical Information Extraction. Front Cell Dev Biol 8:673.

[10] Liu, Y., Peng, Y., Wang, Q., Lu, Z., & Zhang, S. (2020). Biomedical named entity recognition with ensemble learning. BMC Bioinformatics.

[11] Li, J., Sun, Y., Jimeno, A., Yang, Z., & Lin, B. (2019). Active learning for biomedical named entity recognition with deep neural networks. PLOS Computational Biology, 15(2).

[12] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J (2019) BioBERT: a pre-trained biomedical language representation model for biomedical text mining.

[13] Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019). Scispacy: Fast and robust models for biomedical natural language processing. .

[14] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding.

[15] Zhu H, Paschalidis IC, Tahmasebi A (2018) Clinical Concept Extraction with Contextual Word Embedding.arXiv

[16] Zhao, Z. (2017). Disease named entity recognition from biomedical literature. Springer.

[17] Lample, G., Denoyer, L., & Lavergne, T. (2016). Neural architectures for named entity recognition.

[18] Krallinger M, Rabal O, Leitner F, Vazquez M, Salgado D, Lu Z, Leaman R, Lu Y, Ji D, Lowe DM, Sayle RA, Batista-Navarro RT, Rak R, Huber T, Rocktäschel T, Matos S, Campos D, Tang B, Xu H, Munkhdalai T, Ryu KH, Ramanan S, Nathan S, Žitnik S, Bajec M, Weber L, and others(2015) The CHEMDNER corpus of chemicals and drugs and its annotation principles. J Cheminform 7(S1):S2.

[19] Zhang, S. (2013). Unsupervised Biomedical Named Entity Recognition. ScienceDirect.

[20] Rodrigo Rafael Villarreal Goulart,Vera Lúcia Strube de Lima, Clarissa Xavier. (2011). A systematic review of named entity recognition in biomedical texts. ResearchGate.

[21] Liu, F. (2008). Named Entity Recognition in Biomedical Literature. Springer Open.