Improving Disease Named Entity Recognition for Clinical Trial Matching

Md Abdullah Al Hafiz Khan, Md Shamsuzzaman, Sadid A. Hasan,
Mohammad S Sorower, Joey Liu, Vivek Datla, Mladen Milosevic,
Gabe Mankovich, Rob van Ommering, Nevenka Dimitrova
Philips Research North America, Cambridge, MA, USA
{firstname.lastname}@philips.com

***Abstract*—Disease named entity recognition (NER) is an important enabling technology to develop various downstream biomedical natural language processing applications. This is a challenging task, which requires addressing potential ambiguities due to variable contextual usage of the disease name mentions in clinical texts. In particular, clinical trial texts have unique complexities compared to patient-focused clinical reports or information-rich biomedical research articles, as they typically define drug testing eligibility requirements for patient cohorts via compound contextual and logical relationships. In this paper, we propose a novel disease NER model for clinical trial texts by using deep contextual embeddings with relevant domain-specific features, word embeddings, and character embeddings in a bidirectional long short-term memory network-conditional random field (BiLSTM-CRF) framework. Experiments and analyses on a clinical trial dataset and the benchmark NCBI scientific article dataset show the effectiveness of the proposed model.**

***Index Terms*—Disease Named Entity Recognition, Clinical Trial, Deep Learning, Contextual Embedding, Domain Knowledge Embedding**

I. Introduction

Disease named entity recognition (NER) is an important enabling technology to develop various downstream biomedical natural language processing (NLP) applications such as clinical trial matching, patient profiling, population health management, biomedical article retrieval, pharmacovigilance etc. [1]. This is a challenging task, where potential ambiguities due to variable contextual usage of the disease name mentions need to be addressed. In particular, clinical trial texts have unique complexities compared to patient-centric clinical reports or biomedical research articles, as they typically define patient cohorts for drug testing via communicating relevant clinical characteristics with compound contextual and logical relationships.

Several approaches such as rule-based or traditional machine learning and deep learning techniques have been investigated in the literature for disease named entity recognition over the years [2], [1]. The rule-based systems or traditional machine learning-based techniques heavily depend on handcrafted rules or features and domain knowledge. Such systems are generally not scalable as they are dependent on extensive manual intervention. Therefore, deep learning techniques have gained popularity in the recent past due to their ability of extracting features automatically to learn meaningful representations from clinical text [3]. Word embeddings and character embeddings are the most common techniques used to represent the features from clinical text. Such methods can help recognize disease name mentions in a clinical text, but they fail to consider relevant contextual information. For example, in the sentence "chronic obstructive pulmonary disease (COPD) is a life threatening disease", The term 'COPD' can be treated as both gene mention and disease mention when only word or character-level information is considered. Therefore, models need to understand the overall context to conclude that 'COPD' is a disease in the given sentence. Thus, existing disease NER models face difficulty to recognize disease name mentions due to complex inclusion and/or exclusion criteria describing patient eligibility in clinical trial texts.

To this end, we propose a novel disease NER model for clinical trial texts by incorporating deep contextual embeddings [4] with relevant domain-specific features, word embeddings, and character embeddings in a bidirectional long short-term memory network-conditional random field (BiLSTM-CRF) framework [1]. Our main contributions are summarized as follows:

- We propose a novel disease NER model for clinical trial texts by incorporating sentence-level contextual information via deep context embeddings learned with language models (ELMo) [4] in addition to relevant domain-specific features, word embeddings, and character embeddings within a bidirectional long short-term memory network-conditional random field (BiLSTM-CRF) framework [1].
- We evaluate the performance of our model in comparison to various existing disease NER models using a clinical trial dataset [5] and a biomedical article dataset [6]. To the best of our knowledge, this is the first deep learning based model proposed for extracting clinical concepts specifically from clinical trial text. Experimental results demonstrate that our model outperforms existing state-of-the-art approaches and further qualitative analysis shows the effectiveness of our model for clinical trial texts.

II. Related Work

Recently, deep learning models have gained popularity and have been effectively applied in the biomedical NLP domain. Various deep learning models such as convolutional neural

networks (CNN) and recurrent neural networks (RNN) have been applied to perform the disease NER task [7], [2]. Zhao et al. proposed a disease named entity recognition model that used character embeddings generated via stacking the convolution layers of a character-based CNN model [8]. However, in our proposed model, we use external domain knowledge embeddings to improve the recognition performance.

Bidirectional long-short term memory network (Bi-LSTM) also has been extensively used in the literature due to its ability to consider both forward and backward contexts with respect to the specific mentions for recognizing clinical events such as diseases, and treatments [9], [10]. In addition, Conditional Random Fields (CRF) consider the whole sentence instead of individual word positions and can produce higher accuracy for the entity recognition task by maximizing the probability of observing a sentence given a specific set of mentions. Researchers combine Bi-LSTM and CRF networks to consider both relevant input features and sentence-level annotation information to infer named entities from input text [11]. Si et al. proposed a disease NER model that cascades a CNN model with RNN to obtain character embeddings [3]. By contrast, we use an integrated embedding scheme with both character-based CNN and LSTM models.

Deep learning techniques have been also found to be effective with integrated domain knowledge and contextual information for the biomedical concept extraction task [12], [13], [1], [4], [14]. Word embeddings that accurately reflect the context of the words in a sentence can be obtained using a new language model-based method called, ELMo [4]. Si et al. explored ELMo, but they did not incorporate domain knowledge embeddings and did not evaluate the model on clinical trial texts [14]. The model architecture of [1] is similar to our proposed model, however, they did not consider contextualized embeddings from the given corpus. By contrast, we consider both contextual word embeddings and domain knowledge embeddings to learn better representations of the clinical trial text.

III. Methods

NER can be cast as a sequence-labeling task. In this work, we propose a hybrid LSTM-CRF model to identify disease name mentions. This bidirectional LSTM network enables considering both forward and backward features in a given sentence and sequential conditional random field (CRF) annotates the tags taking the softmax output of Bi-LSTM layer as input. Our model comprises of the following: *(i)* a feature representation layer, and *(ii)* a bidirectional LSTM-CRF network. The overall architecture is shown in figure 1. We discuss the details of the architecture below.

A. Feature Representation Layer

The feature representation layer takes a sequence of input $(x_1, x_2, ..., x_n)$ containing n words and generates a $d-$dimensional feature vector for each word. We consider several types of embeddings to capture the inherent features of the sentences. In particular, the $d-$dimensional feature vector for each sentence is obtained via concatenation of four types of representations: *word embedding*, *context embedding*, *domain knowledge embedding*, and *character embedding*. The detailed feature representation layer is shown in figure 2.

Word Embedding: We obtain word embeddings using the publicly available GloVe[1] representation. It generates word embeddings by considering both local context window and the global matrix factorization. We use 300 dimensional vector representation for each word and denote this embedding as V_{word}.

Context Embedding: GloVe-based word embedding mainly relies on word-level co-occurrence statistics. In order to encode context-level information, we obtain context-aware word representations using the language model-based method, ELMo [4]. It comprises of a character-based Convolutional Neural Network (char-CNN) and two-layer bidirectional-Language Model (bi-LM) to embed contextual information through a highway connection and a low-dimensional projection layer, which are introduced after stacking the char-CNN and bi-LM layers. Unlike traditional word embedding that represents a stable embedding vector for downstream tasks, ELMo captures contextual information dynamically for each word as each word is represented as a function of the given sentence. We denote this embedding as V_{elmo}.

Character Embedding: Following [1], we generate a charMIX embedding vector, V_{mix}, which is formed by concatenating two embedding vectors from charCNN (V_{cnn}) and charLSTM (V_{lstm}). The charCNN model takes words as input, then first looks up in the character embedding matrix $Q \in R^{d \times |C|}$ (where the embedding vector dimension is d) and forms the embedding matrix, C^k. The matrix, C^k is then convoluted with multiple filter/kernel matrices. Thereafter, we apply a pooling operation to get the final fixed-dimensional embedding vector, V_{cnn}. The charLSTM architecture comprises a bi-directional LSTM layer and takes the sequence of characters in a word to generate the character embedding vector V_{lstm}, which consists of both forward and backward hidden states $[h_t^f; h_t^b]$.

Domain Knowledge Embedding: Inspired by [1], the domain knowledge (DK) embedding is obtained from two sources: (i) lexicon/clinical vocabulary, and (ii) a hybrid clinical NLP engine [15]. The domain knowledge embedding is denoted as $V_{dk} = [V_{lex}; V_{tag}]$. We combine multiple disease specific dictionaries such as MEDIC, UMLS etc. to obtain a rich clinical vocabulary to generate the lexicon embeddings. We build a TRIE-like data structure for efficient access of the vocabulary. Generally, TRIE-like data structures are preferable to store words of a dictionary such that adding, modifying, and querying the words become efficient. Such structure also stores tags associated with each word in the vocabulary. Therefore, a given sentence can be easily searched in the TRIE dictionary tree and corresponding BIO sequence tags can be annotated automatically. We transform the BIO tags to generate lexicon embeddings V_{lex}. The clinical NLP engine uses a syntactic parser and clinical ontologies such as SNOMED-CT to provide

[1]https://nlp.stanford.edu/projects/glove/

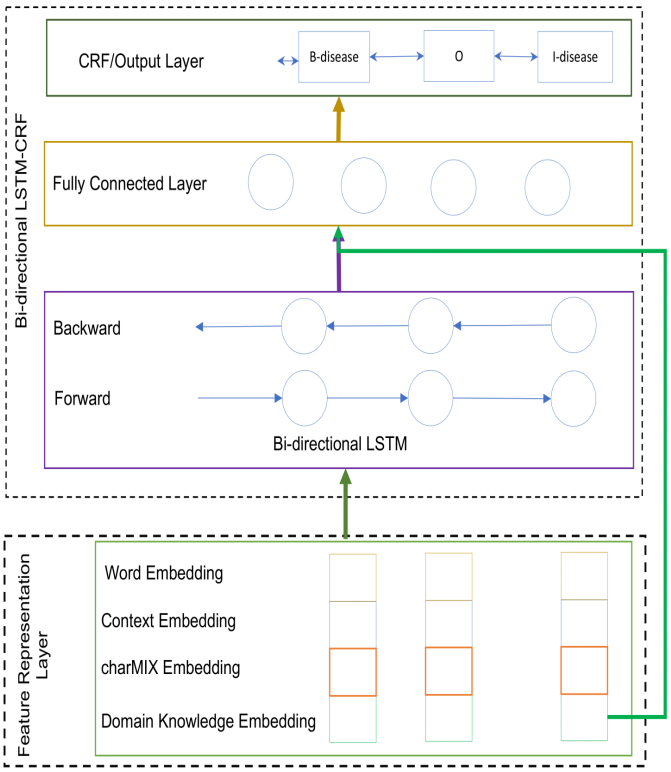


Fig. 1: The proposed architecture for disease name recognition.

tags. Then, we generate the external tagging embeddings V_{tag} based on the sequence of tags provided by the NLP engine.

In summary, the feature representation layer concatenates the above four embeddings to represent a given sentence. We denote the final feature embedding as $V_{fet} = [V_{word}; V_{elmo}; V_{mix}; V_{dk}]$. These features are then fed into the bidirectional LSTM-CRF layer for tagging the sequence of the clinical trial text.

B. Bidirectional LSTM-CRF Layer

We use a bidirectional LSTM-CRF architecture to predict the corresponding tags: O (O=outside), B-disease (B=beginning), and I-disease (I=intermediate) from a given clinical trial text. In particular, Bidirectional LSTM takes a sequence of embedded feature vector, V_{fet} as input and generates the sequence $y = (y_1, y_2, ..., y_n)$ that represents feature encodings from the embedded features. We denote this encoded feature as $V_{bi-lstm} = [h_i^f; h_i^b]$. This feature vector, $V_{bi-lstm}$ is then concatenated with domain knowledge embedding, V_{dk} to generate the input feature vector, $[V_{bi-lstm}; V_{dk}]$ for the fully connected layer. We consider the output of the fully connected layer that is multiplied with corresponding weights, W_f and bias, b_f, as input for the CRF layer and finally, the model predicts the most probable tag sequence.

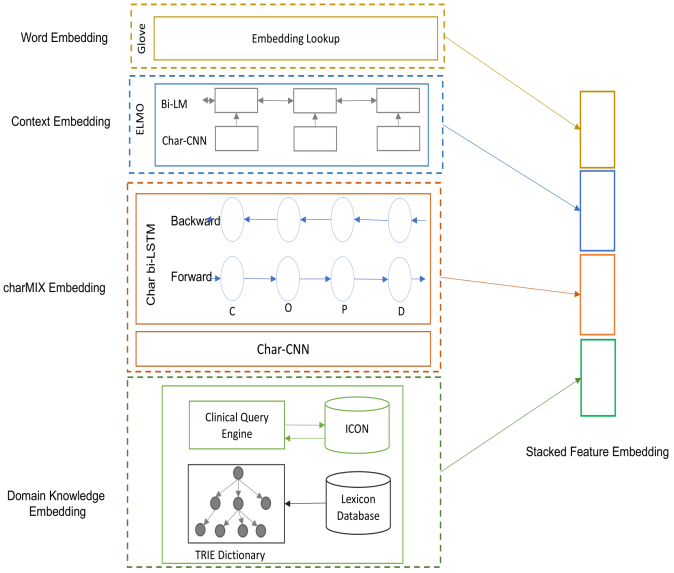


Fig. 2: Feature Representation

IV. Experimental Setup

Our model is implemented using Tensorflow[2]. We evaluate the effectiveness of our models with two datasets. We discuss the implementation details below.

For word embedding, we use 300-dimensional embeddings from publicly available GloVe[3] that is trained on 6 billion tokens from a large corpus text [16].

We embed sentence-level context using the ELMo model. We use the default configuration for ELMO embedding used in [4]. ELMo embedding comprises a charCNN embedding followed by a two layer bi-LSTM model. The charCNN model provides an embedding size of 16. It comprises 7-layers of filters with width $1, 2, 3, 4, 5, 6, 7$ and the filter numbers are $32, 32, 64, 128, 256, 512, 2014$. The bi-LSTM model has 4096 dimensional hidden layers. The output of these two embeddings are then projected to a 512 dimensional space and a highway connection is added at the end.

We keep a fully connected layer after the bi-LSTM layer and the dimension of this layer is set to 500, whereas the dimensions of domain knowledge and bi-LSTM layers are set to 200, and 300, respectively.

We initialize our charCNN and charLSTM models with random values and generate the character embeddings with vector size of 100. This charCNN model has 7 filters with dimensions of 25, 50, 75, 100, 100, 100, 100 and window size of 1,2,3,4,5,6,7. It allows the maximum word length of 40; We padded -1 to keep all the word lengths uniform. We also set the char-LSTM embedding dimension to 100.

As mentioned before, our TRIE dictionary is built with a rich lexicon/vocabulary and it predicts BIO tags. We then use a randomly assigned vector for each tag. Similarly, we form the

[2]https://www.tensorflow.org/
[3]https://nlp.stanford.edu/projects/glove

external embeddings. Both the lexicon and external embedding dimensions are set to 100.

We use mini batch gradient descent methods with batch size 20 to train our proposed model. Initial learning rate is set to 0.001, and in every epoch the learning rate is decayed exponentially, and the decayed rate is set to 0.9. We use Adam Optimizer to train our model. We also set early stopping criteria during training and in this process, we stop training our model if we observe no performance improvement (F1-score) on the validation set for more than 40 epochs.

V. Datasets

In this work, we use two publicly available datasets and a collection of domain knowledge sources for our experiments and evaluation.

NCBI Dataset: Due to unavailability of a disease name annotated clinical trial dataset, we leverage the NCBI disease corpus[4] for our experiments [6]. This dataset has a collection of 793 PubMed abstracts and is fully annotated at the mention and concept level. It has a total of 6,892 disease mentions with 790 unique disease concepts. Table I shows the detailed statistics of the NCBI dataset.

TABLE I: NCBI Dataset statistics.

Split Type	# sentences	sentence length (avg.)	# tokens	# unique tokens	# labels
Train	5,576	23	132,584	9,805	2,911
Test	941	25	24,019	3,679	535
Valid	918	25	23,456	3,580	487

Clinical Trials: We analyze our proposed model performance using an unannotated clinical trial dataset [5] containing a total of 241,006 clinical trial descriptions from the Text REtrieval Conference (TREC), 2017 precision medicine track[5]. We train the ELMo model on this corpus for one epoch achieving the perplexity score of 1.55. Due to lack of annotation, we compare disease name mentions recognized by both our proposed model and the DK-BiLSTM-CRF model [1] by manually evaluating the outputs.

Domain Knowledge Source: Following the work [1], we use the MEDIC[6] disease vocabulary, UMLS clinical terms of 'Disease or Syndrome' semantic types[7] and an in-house cancer specific ontology to generate lexicon features. In addition, we use a hybrid clinical NLP engine [15] to obtain disease tags that are added as external tagging embeddings in our disease NER framework.

VI. Evaluation Results and Discussion

We use accuracy, precision, recall and F1-score metrics to measure the performance of our proposed models. We compare the performance of our method with existing state-of-the-art methods. We compute accuracy based on exact match at the word level and precision, recall and F1-score are computed at phrase level. Table II shows the performance comparison of our proposed methods with state-of-the-art methods. We also show the impact of contextualized embeddings via a detailed analysis on clinical trial texts.

A. Performance comparison on NCBI dataset

In this study, we consider four different scenarios of our proposed model, henceforth termed as the context-BiLSTM-CRF model. For context embedding, we use pretrained ELMo on a general corpus [4], and also fine-tune the ELMo model on clinical trial text [5], and Wiki&MIMIC dataset [17]. We also train ELMo from scratch using the clinical trial text to learn and explore the performance of our proposed Context-BiLSTM-CRF on clinical trial context. We compare these four different context-based scenarios while training our proposed models using the NCBI training dataset and then report the results on NCBI test set in Table II.

TABLE II: Comparative performance measurement

	Accuracy	Precision	Recall	F1-Score
Existing Disease NER systems				
Sahu et al. (2016)[3]	-	0.849	0.741	0.791
Dogan et al. (2012)[18]	-	-	-	0.818
Habibi et al. (2017) [19]	-	0.853	0.836	0.844
Zhao et al. (2017) [8]	-	0.851	0.853	0.852
Amazon Comprehend medical [20]	-	0.430	0.749	0.546
DK-BiLSTM-CRF [1]	0.981	**0.868**	0.839	0.853
Our Proposed Models				
Context-BiLSTM-CRF (clinical)	0.982	0.847	0.864	0.855
Context-BiLSTM-CRF	0.983	0.858	0.865	**0.861**
Context-BiLSTM-CRF (Wiki & MIMIC)	0.983	0.844	0.865	0.854
Context-BiLSTM-CRF (clinical scratch)	**0.984**	0.840	**0.870**	0.855

Table II shows the performance comparison with existing NER systems. We observe that our proposed model, context-BiLSTM-CRF outperforms other existing disease NER systems due to the use of contextual embeddings, which confirms the effectiveness of contextual information in our proposed neural network architecture. Comparison of our models with various existing systems [21], [22], [19], [8] including the recently released Amazon Comprehend Medical NLP tool[8] [23] thus demonstrates the usefulness of our proposed architecture.

We notice that Context-BiLSTM-CRF (clinical scratch) performs better than existing methods and achieves accuracy and recall values of 0.984 and 0.870 respectively. Context-BiLSTM-CRF (clinical scratch) performs better due to the contextual knowledge learned exclusively from clinical trial texts and the classifier encountering less ambiguity with other contextual knowledge such as general text from Wiki. Context-BiLSTM-CRF model is trained with general text, which also contains clinical concepts and achieves better F1 score due to the smaller difference on both precision and recall values.

[4]https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE/

[5]http://www.trec-cds.org/2017.html\#documents

[6]http://ctd.mdibl.org/voc.go?type=disease

[7]https://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html

[8]https://aws.amazon.com/comprehend/medical/

This model identifies disease concepts more correctly out of the predicted disease name mentions. Although our proposed model achieves a better recall score, the DK-BiLSTM-CRF model achieves a better precision score (0.868), which indicates that our proposed model predicts more disease relevant text. Therefore, our best model with context and domain knowledge embeddings has the best accuracy of 0.984, and recall of 0.870 establishing the new state-of-the-art performance compared with existing methods on the NCBI dataset.

To further analyze the performance of our model, we report normalized confusion matrices for both DK-BiLSTM-CRF and Context-BiLSTM-CRF models in figure 3. Gold labels are denoted on the y-axis and predicted labels are denoted on the x-axis. From the figure, we note that the DK-BiLSTM-CRF (left confusion matrix) model predicts 0.84, 0.85, and 0.99 fractions of the gold labels as B-disease, I-disease, and O tags, respectively. By contrast, our proposed Context-BiLSTM-CRF (right confusion matrix) predicts 0.89, 0.89 and 0.99 fractions of the gold labels as B-disease, I-disease and O tags, respectively. We conclude that our model outperforms DK-BiLSTM-CRF for tagging both the B-disease and I-disease name mentions considerably.

B. Importance of learning contextual embeddings from clinical trial text

In this section, we investigate the importance of learning contextual embeddings from clinical trial text. We evaluate the performance of our models on clinical trial text. We present a representative result of our evaluation in figure 4, and 5. We highlight the annotated text using different colors. The "green" colored text represents text that is appropriately annotated as disease concepts by the context-biLSTM-CRF model and are missed by the DK-biLSTM-CRF model. The red highlights represent concepts which are incorrectly annotated as disease concepts by the context-biLSTM-CRF model but not by the other model. Finally, the texts highlighted with yellow are the ones that are correctly identified as "non-disease" concepts by the context-biLSTM-CRF model, but are falsely labeled as disease concepts by the DK-biLSTM-CRF model.

Table III, and IV show the comparative disease detection performance for both DK-biLSTM-CRF, and our proposed context-biLSTM-CRF models for two clinical trial text documents. These tables show appropriate detection of disease names by the individual models. For this experiment, we train these models using NCBI dataset and evaluated them on the clinical trial text. Manual evaluation of the outputs from these two models show how our proposed model performs better than the other model.

From table III, and IV, it is observed that DK-BiLSTM-CRF often fails to recognize disease name mentions, while our proposed context-biLSTM-CRF model is able to recognize disease names effectively. In several occasions, DK-BiLSTM-CRF falsely identified various clinical and non-clinical terms as disease concepts. For example, DK-BiLSTM-CRF detects "Criteria", "vascular access" and "toxicity" as disease named entities, whereas our proposed model successfully regards these as non-disease concepts. Another aspect where the context-biLSTM-CRF model outperforms the DK-BiLSTM-CRF model is the disambiguation of acronyms. Various acronyms found in these two representative clinical trials are inappropriately identified as disease concepts by the DK-BiLSTM-CRF model, e.g., "ULN, SC, MTD, KRAS, DLT" etc. By contrast, our proposed model understands the surrounding contexts of the concepts better, and can correctly identify them as non-disease concepts. Finally, our proposed model can accurately identify the extended disease names along with associated qualifiers. For example, our model perfectly recognizes the context and appropriately detects the full disease name as opposed to the short uninformative form such as, "allergy to egg protiens" vs "allergy" only, "HER2-ve advanced breast cancers" vs "breast cancers" only, "soft tissue sarcoma" vs "sarcoma" only etc.

Overall, our in-depth analyses on clinical trial texts suggest that the proposed context-BiLSTM-CRF model would more likely fit various downstream clinical concept recognition use cases and applications that require more interpretability of

TABLE III: Model comparison for trial #NCT02432963

Predicted Disease Concepts by either models	**Is Disease**	**Cases where Context-BiLSTM-CRF is Correct/Better**	**Cases where DK-BiLSTM-CRF is better**	**Reason for Not Correct**
anti-programmed cell death	No	✓	✗	Not Disease
allergy to egg proteins	Yes	✓	✗	Partial Match
ipilimumab	No	✗	✓	Not Disease
HER2-ve advanced breast	Yes	✓	✗	Partial Match
Vitiligo	Yes	✓	✗	Not Annotated
SC	No	✓	✗	Not Disease
bladder, soft tissue sarcoma, triple-negative breast cancer	Yes	✓	✗	Partial Match
Vaccinica	No	✓	✗	Not Disease
IRB	No	✓	✗	Not Disease
Criteria	No	✓	✗	Not Disease
Toxicity	No	✓	✗	Not Disease
Unresectable Solid Neoplasm	Yes	✓	✗	Not Annotated
immune-mediated adverse reactions	Yes	✓	✗	Not Annotated

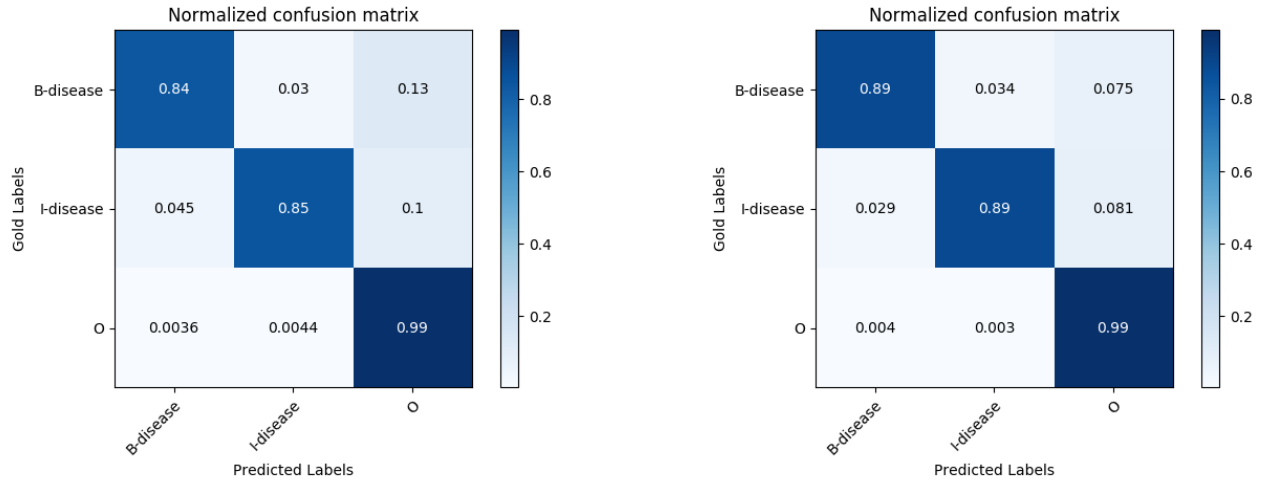


Fig. 3: Normalized confusion matrix for DK-BiLSTM-CRF (left) and Context-BiLSTM-CRF (right) on NCBI dataset

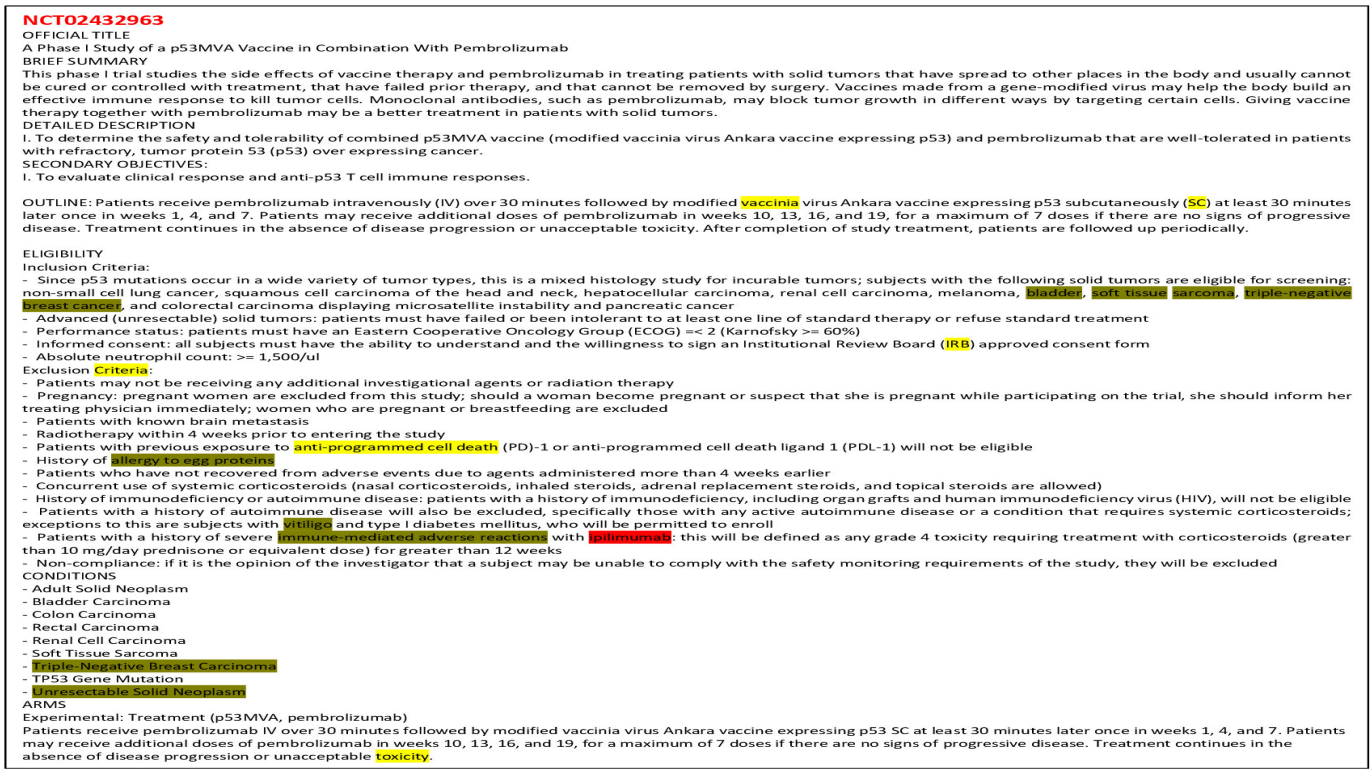
TABLE IV: Model comparison for trial #NCT02389842

Predicted Disease Concepts by either models	**Is Disease**	**Cases where Context-BiLSTM-CRF is correct /Better**	**Cases where DK-BiLSTM-CRF is better**	**Reason for Not Correct**
Palbociclib	No	✗	✓	Not Disease
advanced solid tumours	Yes	✓	✗	Partial Match
MTD	No	✓	✗	Not Disease
HER2-ve advanced breast cancers	Yes	✓	✗	Partial Match
KRAS	No	✓	✗	Not Disease
DLT	No	✓	✗	Not Disease
PIK3CA	No	✗	✓	Not Disease
ER+ve and HER2-ve post-menopausal breast cancer	Yes	✓	✗	Partial Match
Mitomycin-C	No	✓	✗	Not Disease
NB	No	✓	✗	Not Disease
ULN	No	✓	✗	Not Disease
DLCO	No	✗	✓	Not Disease
central nervous system	No	✓	✗	Not Disease
vascular access	No	✓	✗	Not Disease

the input text. Therefore, our model is more suitable to the applications where it requires understanding of detailed context to satisfy end users. Furthermore, our proposed model can be fine-tuned on new clinical datasets towards better generalization of the models.

VII. Conclusion and Future Work

In this paper, we proposed a novel disease NER model for clinical trial texts by using deep contextual embeddings with relevant domain-specific features, word embeddings, and character embeddings in a bidirectional long short-term memory network-conditional random field (BiLSTM-CRF) framework. Extensive experiments and analyses on a clinical trial dataset and the benchmark NCBI scientific article dataset show the effectiveness of the proposed model. In the future, we will experiment with deep bidirectional transformer-based language models (BERT) [24] to generate deep contextualized embeddings of clinical trial texts.



NCT02432963
OFFICIAL TITLE
A Phase I Study of a p53MVA Vaccine in Combination With Pembrolizumab
BRIEF SUMMARY
This phase I trial studies the side effects of vaccine therapy and pembrolizumab in treating patients with solid tumors that have spread to other places in the body and usually cannot be cured or controlled with treatment, that have failed prior therapy, and that cannot be removed by surgery. Vaccines made from a gene-modified virus may help the body build an effective immune response to kill tumor cells. Monoclonal antibodies, such as pembrolizumab, may block tumor growth in different ways by targeting certain cells. Giving vaccine therapy together with pembrolizumab may be a better treatment in patients with solid tumors.
DETAILED DESCRIPTION
I. To determine the safety and tolerability of combined p53MVA vaccine (modified vaccinia virus Ankara vaccine expressing p53) and pembrolizumab that are well-tolerated in patients with refractory, tumor protein 53 (p53) over expressing cancer.
SECONDARY OBJECTIVES:
I. To evaluate clinical response and anti-p53 T cell immune responses.

OUTLINE: Patients receive pembrolizumab intravenously (IV) over 30 minutes followed by modified vaccinia virus Ankara vaccine expressing p53 subcutaneously (SC) at least 30 minutes later once in weeks 1, 4, and 7. Patients may receive additional doses of pembrolizumab in weeks 10, 13, 16, and 19, for a maximum of 7 doses if there are no signs of progressive disease. Treatment continues in the absence of disease progression or unacceptable toxicity. After completion of study treatment, patients are followed up periodically.

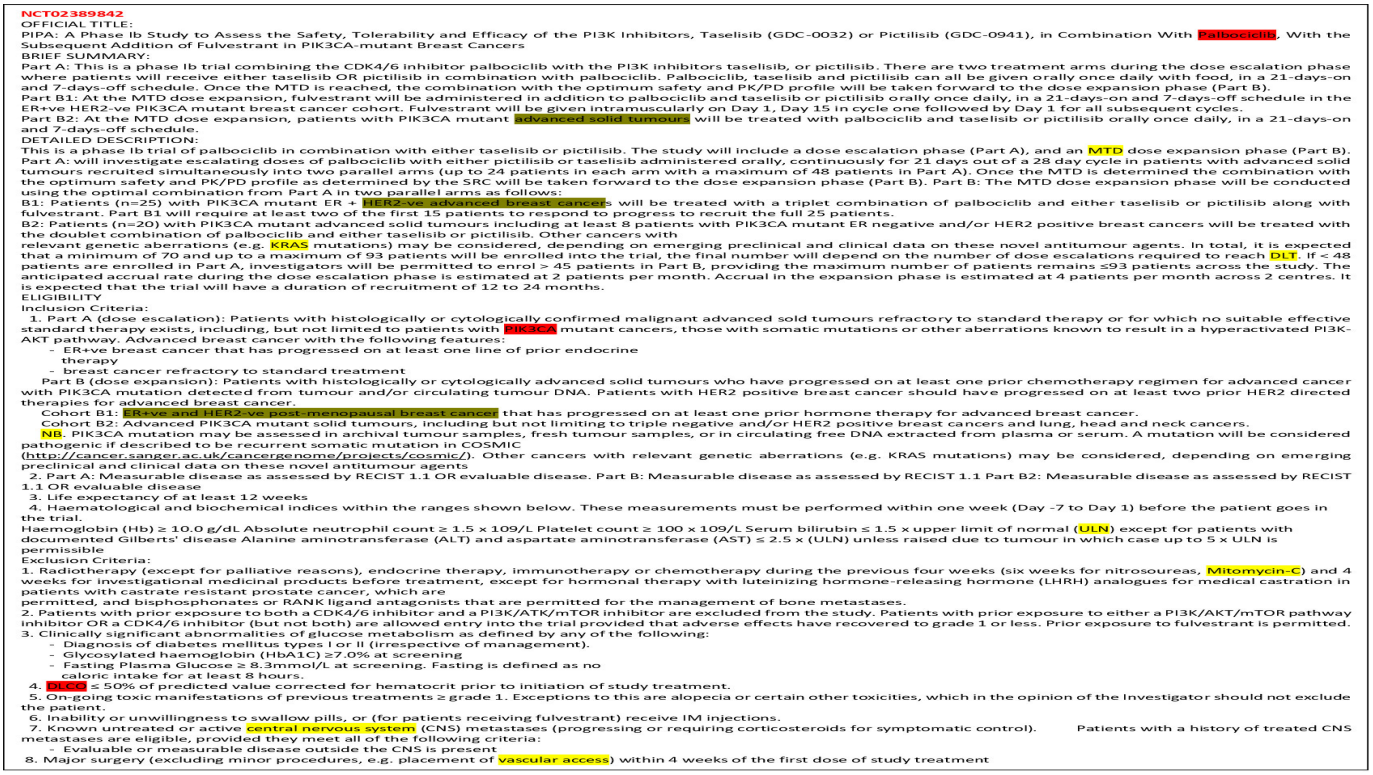
ELIGIBILITY
Inclusion Criteria:
- Since p53 mutations occur in a wide variety of tumor types, this is a mixed histology study for incurable tumors; subjects with the following solid tumors are eligible for screening: non-small cell lung cancer, squamous cell carcinoma of the head and neck, hepatocellular carcinoma, renal cell carcinoma, melanoma, bladder, soft tissue sarcoma, triple-negative breast cancer, and colorectal carcinoma displaying microsatellite instability and pancreatic cancer
- Advanced (unresectable) solid tumors: patients must have failed or been intolerant to at least one line of standard therapy or refuse standard treatment
- Performance status: patients must have an Eastern Cooperative Oncology Group (ECOG) =< 2 (Karnofsky >= 60%)
- Informed consent: all subjects must have the ability to understand and the willingness to sign an Institutional Review Board (IRB) approved consent form
- Absolute neutrophil count: >= 1,500/ul

Exclusion Criteria:
- Patients may not be receiving any additional investigational agents or radiation therapy
- Pregnancy: pregnant women are excluded from this study; should a woman become pregnant or suspect that she is pregnant while participating on the trial, she should inform her treating physician immediately; women who are pregnant or breastfeeding are excluded
- Patients with known brain metastasis
- Radiotherapy within 4 weeks prior to entering the study
- Patients with previous exposure to anti-programmed cell death (PD)-1 or anti-programmed cell death ligand 1 (PDL-1) will not be eligible
- History of allergy to egg proteins
- Patients who have not recovered from adverse events due to agents administered more than 4 weeks earlier
- Concurrent use of systemic corticosteroids (nasal corticosteroids, inhaled steroids, adrenal replacement steroids, and topical steroids are allowed)
- History of immunodeficiency or autoimmune disease: patients with a history of immunodeficiency, including organ grafts and human immunodeficiency virus (HIV), will not be eligible
- Patients with a history of autoimmune disease will also be excluded, specifically those with any active autoimmune disease or a condition that requires systemic corticosteroids; exceptions to this are subjects with vitiligo and type I diabetes mellitus, who will be permitted to enroll
- Patients with a history of severe immune-mediated adverse reactions with ipilimumab: this will be defined as any grade 4 toxicity requiring treatment with corticosteroids (greater than 10 mg/day prednisone or equivalent dose) for greater than 12 weeks
- Non-compliance: if it is the opinion of the investigator that a subject may be unable to comply with the safety monitoring requirements of the study, they will be excluded

CONDITIONS
- Adult Solid Neoplasm
- Bladder Carcinoma
- Colon Carcinoma
- Rectal Carcinoma
- Renal Cell Carcinoma
- Soft Tissue Sarcoma
- Triple-Negative Breast Carcinoma
- TP53 Gene Mutation
- Unresectable Solid Neoplasm

ARMS
Experimental: Treatment (p53MVA, pembrolizumab)
Patients receive pembrolizumab IV over 30 minutes followed by modified vaccinia virus Ankara vaccine expressing p53 SC at least 30 minutes later once in weeks 1, 4, and 7. Patients may receive additional doses of pembrolizumab in weeks 10, 13, 16, and 19, for a maximum of 7 doses if there are no signs of progressive disease. Treatment continues in the absence of disease progression or unacceptable toxicity.

Fig. 4: Clinical Trial Text (Trial #NCT02432963)



NCT02389842
OFFICIAL TITLE:
PIPA: A Phase Ib Study to Assess the Safety, Tolerability and Efficacy of the PI3K Inhibitors, Taselisib (GDC-0032) or Pictilisib (GDC-0941), in Combination With Palbociclib, With the Subsequent Addition of Fulvestrant in PIK3CA-mutant Breast Cancers
BRIEF SUMMARY:
Part A: This is a phase Ib trial combining the CDK4/6 inhibitor palbociclib with the PI3K inhibitors taselisib, or pictilisib. There are two treatment arms during the dose escalation phase where patients will receive either taselisib OR pictilisib in combination with palbociclib. Palbociclib, taselisib and pictilisib can all be given orally once daily with food, in a 21-days-on and 7-days-off schedule. Once the MTD is reached, the combination with the optimum safety and PK/PD profile will be taken forward to the dose expansion phase (Part B).
Part B1: At the MTD dose expansion, fulvestrant will be administered in addition to palbociclib and taselisib or pictilisib orally once daily, in a 21-days-on and 7-days-off schedule in the ER+ve HER2-ve PIK3CA mutant breast cancer cohort. Fulvestrant will be given intramuscularly on Day 1, Day 15 in cycle one followed by Day 1 for all subsequent cycles.
Part B2: At the MTD dose expansion, patients with PIK3CA mutant advanced solid tumours will be treated with palbociclib and taselisib or pictilisib orally once daily, in a 21-days-on and 7-days-off schedule.
DETAILED DESCRIPTION:
This is a phase Ib trial of palbociclib in combination with either taselisib or pictilisib. The study will include a dose escalation phase (Part A), and an MTD dose expansion phase (Part B).
Part A: will investigate escalating doses of palbociclib with either pictilisib or taselisib administered orally, continuously for 21 days out of a 28 day cycle in patients with advanced solid tumours recruited simultaneously into two parallel arms (up to 24 patients in each arm with a maximum of 48 patients in Part A). Once the MTD is determined the combination with the optimum safety and PK/PD profile as determined by the SRC will be taken forward to the dose expansion phase (Part B). Part B: The MTD dose expansion phase will be conducted using the optimal combination from Part A in two parallel arms as follows:
B1: Patients (n=25) with PIK3CA mutant ER + HER2-ve advanced breast cancers will be treated with a triplet combination of palbociclib and either taselisib or pictilisib along with fulvestrant. Part B1 will require at least two of the first 15 patients to respond to progress to recruit the full 25 patients.
B2: Patients (n=20) with PIK3CA mutant advanced solid tumours including at least 8 patients with PIK3CA mutant ER negative and/or HER2 positive breast cancers will be treated with the doublet combination of palbociclib and either taselisib or pictilisib. Other cancers with
relevant genetic aberrations (e.g. KRAS mutations) may be considered, depending on emerging preclinical and clinical data on these novel antitumour agents. In total, it is expected that a minimum of 70 and up to a maximum of 93 patients will be enrolled into the trial, the final number will depend on the number of dose escalations required to reach DLT. If < 48 patients are enrolled in Part A, investigators will be permitted to enrol > 45 patients in Part B, providing the maximum number of patients remains ≤93 patients across the study. The anticipated accrual rate during the dose escalation phase is estimated at 2 patients per month. Accrual in the expansion phase is estimated at 4 patients per month across 2 centres. It is expected that the trial will have a duration of recruitment of 12 to 24 months.
ELIGIBILITY
Inclusion Criteria:
1. Part A (dose escalation): Patients with histologically or cytologically confirmed malignant advanced solid tumours refractory to standard therapy or for which no suitable effective standard therapy exists, including, but not limited to patients with PIK3CA mutant cancers, those with somatic mutations or other aberrations known to result in a hyperactivated PI3K-AKT pathway. Advanced breast cancer with the following features:
   - ER+ve breast cancer that has progressed on at least one line of prior endocrine therapy
   - breast cancer refractory to standard treatment

   Part B (dose expansion): Patients with histologically or cytologically advanced solid tumours who have progressed on at least one prior chemotherapy regimen for advanced cancer with PIK3CA mutation detected from tumour and/or circulating tumour DNA. Patients with HER2 positive breast cancer should have progressed on at least two prior HER2 directed therapies for advanced breast cancer.
   Cohort B1: ER+ve and HER2-ve post-menopausal breast cancer that has progressed on at least one prior hormone therapy for advanced breast cancer.
   Cohort B2: Advanced PIK3CA mutant solid tumours, including but not limiting to triple negative and/or HER2 positive breast cancers and lung, head and neck cancers.
   NB. PIK3CA mutation may be assessed in archival tumour samples, fresh tumour samples, or in circulating free DNA extracted from plasma or serum. A mutation will be considered pathogenic if described to be recurrent somatic mutation in COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/). Other cancers with relevant genetic aberrations (e.g. KRAS mutations) may be considered, depending on emerging preclinical and clinical data on these novel antitumour agents
2. Part A: Measurable disease as assessed by RECIST 1.1 OR evaluable disease. Part B: Measurable disease as assessed by RECIST 1.1 OR evaluable disease
3. Life expectancy of at least 12 weeks
4. Haematological and biochemical indices within the ranges shown below. These measurements must be performed within one week (Day -7 to Day 1) before the patient goes in the trial.
Haemoglobin (Hb) ≥ 10.0 g/dL Absolute neutrophil count ≥ 1.5 x 109/L Platelet count ≥ 100 x 109/L Serum bilirubin ≤ 1.5 x upper limit of normal (ULN) except for patients with documented Gilberts' disease Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) ≤ 2.5 x (ULN) unless raised due to tumour in which case up to 5 x ULN is permissible
Exclusion Criteria:
1. Radiotherapy (except for palliative reasons), endocrine therapy, immunotherapy or chemotherapy during the previous four weeks (six weeks for nitrosoureas, Mitomycin-C) and 4 weeks for investigational medicinal products before treatment, except for hormonal therapy with luteinizing hormone-releasing hormone (LHRH) analogues for medical castration in patients with castrate resistant prostate cancer, which are
permitted, and bisphosphonates or RANK ligand antagonists that are permitted for the management of bone metastases.
2. Patients with prior exposure to both a CDK4/6 inhibitor and a PI3K/ATK/mTOR inhibitor are excluded from the study. Patients with prior exposure to either a PI3K/AKT/mTOR pathway inhibitor OR a CDK4/6 inhibitor (but not both) are allowed entry into the trial provided that adverse effects have recovered to grade 1 or less. Prior exposure to fulvestrant is permitted.
3. Clinically significant abnormalities of glucose metabolism as defined by any of the following:
   - Diagnosis of diabetes mellitus types I or II (irrespective of management).
   - Glycosylated haemoglobin (HbA1C) ≥7.0% at screening
   - Fasting Plasma Glucose ≥ 8.3mmol/L at screening. Fasting is defined as no caloric intake for at least 8 hours.
4. DLCO < 50% of predicted value corrected for hematocrit prior to initiation of study treatment.
5. On-going toxic manifestations of previous treatments ≥ grade 1. Exceptions to this are alopecia or certain other toxicities, which in the opinion of the Investigator should not exclude the patient.
6. Inability or unwillingness to swallow pills, or (for patients receiving fulvestrant) receive IM injections.
7. Known untreated or active central nervous system (CNS) metastases (progressing or requiring corticosteroids for symptomatic control). Patients with a history of treated CNS metastases are eligible, provided they meet all of the following criteria:
   - Evaluable or measurable disease outside the CNS is present
8. Major surgery (excluding minor procedures, e.g. placement of vascular access) within 4 weeks of the first dose of study treatment

Fig. 5: Clinical Trial Text (Trial #NCT02389842))

REFERENCES

1. Yuan Ling, Sadid A Hasan, Oladimeji Farri, Zheng Chen, Rob van Ommering, Charles Yee, and Nevenka Dimitrova. A domain knowledge-enhanced lstm-crf model for disease named entity recognition. *AMIA Summits on Translational Science Proceedings*, 2019:761, 2019.
2. Gamal Crichton, Sampo Pyysalo, Billy Chiu, and Anna Korhonen. A neural network multi-task learning approach to biomedical named entity recognition. *BMC Bioinformatics*, 18, 12 2017.
3. Sunil Kumar Sahu and Ashish Anand. Recurrent neural network models for disease name recognition using domain invariant features. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*, 2016.
4. Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.
5. Kirk Roberts, Dina Demner-Fushman, Ellen M Voorhees, William R Hersh, Steven Bedrick, Alexander J Lazar, and Shubham Pant. Overview of the trec 2017 precision medicine track.
6. Rezarta Islamaj Doğan, Robert Leaman, and Zhiyong Lu. Ncbi disease corpus: a resource for disease name recognition and concept normalization. *Journal of biomedical informatics*, 47:1–10, 2014.
7. Abhyuday N Jagannatha and Hong Yu. Bidirectional rnn for medical event detection in electronic health records. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2016, page 473. NIH Public Access, 2016.
8. Zhehuan Zhao, Zhihao Yang, Ling Luo, Lei Wang, Yin Zhang, Hongfei Lin, and Jian Wang. Disease named entity recognition from biomedical literature using a novel convolutional neural network. *BMC medical genomics*, 10(5):73, 2017.
9. Willie Boag, Elena Sergeeva, Saurabh Kulshreshtha, Peter Szolovits, Anna Rumshisky, and Tristan Naumann. Cliner 2.0: Accessible and accurate clinical concept extraction. *ArXiv*, abs/1803.02245, 2018.
10. Raghavendra Chalapathy, Ehsan Zare Borzeshi, and Massimo Piccardi. Bidirectional lstm-crf for clinical concept extraction. *arXiv preprint arXiv:1611.08373*, 2016.
11. Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.
12. Yuan Ling, Yuan An, and Sadid Hasan. Improving clinical diagnosis inference through integration of structured and unstructured knowledge. In *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*, pages 31–36, 2017.
13. Yuan Ling, Yuan An, Mengwen Liu, Sadid A Hasan, Yetian Fan, and Xiaohua Hu. Integrating extra knowledge into word embedding models for biomedical nlp tasks. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 968–975. IEEE, 2017.
14. Yuqi Si, Jingqi Wang, Hua Xu, and Kirk Roberts. Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, 07 2019.
15. Vivek V. Datla, Sadid A. Hasan, Ashequl Qadir, Kathy Lee, Yuan Ling, Joey Liu, and Oladimeji Farri. Automated clinical diagnosis: The role of content in various sections of a clinical document. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM*, pages 1004–1011, 2017.
16. Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
17. Henghui Zhu, Ioannis Ch. Paschalidis, and Amir Tahmasebi. Clinical concept extraction with contextual word embedding, 2018.
18. Rezarta Islamaj Doğan and Zhiyong Lu. An improved corpus of disease mentions in pubmed citations. In *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, BioNLP '12, pages 91–99, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
19. Maryam Habibi, Leon Weber, Mariana Neves, David Luis Wiegandt, and Ulf Leser. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics*, 33(14):i37–i48, 2017.
20. Parminder Bhatia, Busra Celikkaya, and Mohammad Khalilia. End-to-end joint entity extraction and negation detection for clinical text. *CoRR*, abs/1812.05270, 2018.
21. Sunil Kumar Sahu and Ashish Anand. Recurrent neural network models for disease name recognition using domain invariant features. In *ACL*, 2016.
22. Rezarta Islamaj Doğan and Zhiyong Lu. An improved corpus of disease mentions in pubmed citations. In *Proceedings of the 2012 workshop on biomedical natural language processing*, pages 91–99. Association for Computational Linguistics, 2012.
23. Parminder Bhatia, Busra Celikkaya, and Mohammed Khalilia. End-to-end joint entity extraction and negation detection for clinical text. *CoRR*, abs/1812.05270, 2018.
24. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.