Turkish Journal of Computer and Mathematics Education *Vol.12 No.14(2021), 2784- 2792*

Research Article

Improving Eligibility Classification on Clinical Trials Document using Bidirectional Long Short Term Memory Recurrent Neural Network

Jasmir Jasmir[a,b], Siti Nurmaini[c] , Reza Firsandaya Malik[d], Bambang Tutuko[c]

[a] Doctoral Program of Engineering Science, Faculty of Engineering Universitas Sriwijaya, Palembang Indonesia
[b]Computer Engineering, Universitas Dinamika Bangsa Jambi Indonesia
[c]Intelligence System Research GroupFaculty of Computer ScienceUniversitas Sriwijaya Palembang Indonesia
[c]Communication Networks and Information Security Research Lab, Faculty of Computer Science Universitas Sriwijaya, Palembang Indonesia

[a]ijay-jasmir@yahoo.com,[c]sitinurmainii@gmail.com,[d]rezafm@unsri.ac.id,[c]bambang_tutoko@unsri.ac.id

Article History: Received: 9 August 2021; Revised: 15 August 2021 ;Accepted:24 August 2021; Published online:

Abstract: Cancer clinical trials intervention are generally too restrictive, and some patients are often excluded on the basis of comorbidity, past or concomitant treatments, or the fact that they are over a certain age. In this research we built a classification model for clinical information using public clinical trial protocols labeled as eligible or not eligible. Text classifications are trained using deep learning to determine the predictive outcome of short free text statements reflectingeligible and not eligible clinical information.then we also performed semantic analysis for the obtained word-embedding representations and were able to identify similar treatments. We have proven that learning outcomes using deep learning methods to extract medical information from clinical trial documents have been successful in assisting health practitioners in prescribing treatments. The evaluation results showed a value with an accuracy value is 77.74%, precision is 76.8%, recall is 80.80%, and F1-score is 78.80%

Keywords:clinical trials, text classification, BiLSTM RNN, deep learning

1. Introduction.

Clinical trial is a test of the efficacy of a new drug in humans, which was previously initiated by animal testing or pre-clinical testing. Basically, clinical trials confirm the effectiveness, safety and description of the side effects that often arise in humans due to administration of a drug. Using healthy or sick humans in experiments is justified in medical science because it will benefit the public at large to understand the effects of these drugs so that they can be used in the wider community with more confidence about their effectiveness and safety[1].

The subject of clinical trials will be discussed in this study including classifying clinical trial documents in the form of free text and ensuring that the documents fall into one or more predetermined categories[2]. These text classifications already have many useful applications in the field of sentiment analysis [3], spam detection[4], subject categorization [5] precise sentence [6] and so on. With the success that deep learning has had in some cases such as visual and speech recognition, many deep architectural models have been implemented to address various issues in NLP, so that better performance can be achieved[7]. Text classification has also attracted the attention of researchers to continue to develop innovations and testing, including those sourced from clinical texts commonly referred to as clinical trials.

Clinical trials is a type of research that studies howsafe it is to help test or care given to patients. [8]. Clinicaltrials play an important role in translating scientificresearch into the practice of medical outcomes[9]. Inclinical trials, there is a term or the most significant partcalled the eligibility criteria that determines the cost,duration and success of the clinical trial process[10]. Study on eligibility criteria in clinical trials isusually written in free text, but it will be difficult ifinterpreted by computer. A popular method for processingeligibility criteria is knowledge representation, whichoften requires extensive knowledge and hard work fromexperts in the sector of medical coding to identifyeligibility criteria[11]. In completing the problem of theeligibility analysis of clinical trials, the optimal method tosolve it is artificial intelligence methods such as machine learning and deep learningarchitecture. [12].

One popular deep learning method is BiLSTM [13]. A Bidirectional LSTM**,** orbiLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction. BiLSTMs effectively increase the amount of information available to the network,

improving the context available to the algorithm (e.g. knowing what words immediately follow *and* precede a word in a sentence).. [14][15]

In this case the BiLSTM RNN Model built has one input layer and one hidden layer.The number of nodes in this model varies depending on the number of inputs and outputs.The number of input neurons depends on the number of markers and the number of hidden neurons.Then this network is trained and validated by k-fold cross validation.

The aim of this study was to construct a classification model of eligibility criteria from clinical trial statements, and use it to train and validate models that predict whether a short free-text statement is considered an eligible or not eligible criteria for this trial. We also analyzed the obtained word-embedding representations and were able to identify similar treatments. The final aim of this work is to assess whether representation learning using a multilayer perceptron can be successfully applied to extract the medical knowledge available in clinical trial protocols, thus opening up opportunities for future major research.

2. Related Work

Some researches related to this theme such asthe classification with BiLSTM [16][17], clinical text classification[18][19] andeligibility criteria[9][20].The other research that discussesthe clinical textclassification of eligibility criteria, Riccardo Miotto [21]which discusses a new pattern to determine the semantic index of automatically controlled vocabulary for clinical trial eligibility documents. Kevin Zhang[22]which discusses the development of a method for the automatic classification of eligibility criteria in facilitating trial matching of patients to specific populations.The proposed method is capable of acting as a filter in search engines to test patients.C. Chuan [23]proposes an active deep learningapproachto automatically classify clinical trial eligibilitycriteria. The experimental results showed thatactivedeep learning performed significantly better than the K-NearestNeighbor method. Thus, Y. Ni et al [24]discussthedevelopment of an automatic screening eligibilityalgorithm to recognize patients who meet the coreeligibilitycharacteristics of oncology clinical trials.

In addition to the research above, there is also research that discusses the features of word embedding.The word embedding feature is also able to improve classification text, including Zeynep H. Kilimci using word embedding with Random Forest and CNN [25]. Lihao Ge uses word embedding with naive bayes and Support vector machine [26] and Duyu Tang uses word embedding with Support vector machine and N-gram [27]

In this study we made a classification model of theeligibility criteria of cancer clinical texts using word embedding and the deep learning.

3.Material and Methods

Figure-1 shows the framework, consisting of clinical text, text preprocessing, word embedding, labeling and deep learning classifier

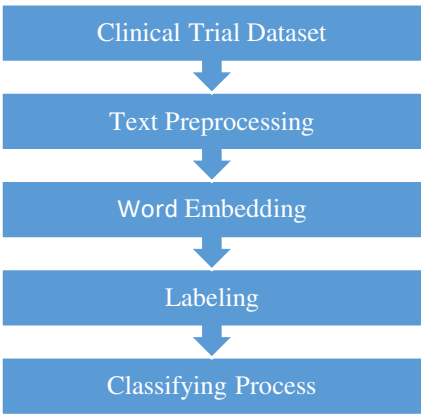

Figure-1 : Research framework

A.Clinical Dataset

Data were taken from clinical statements. Extracted from Clinical TrialProtocols on cancer which can be downloaded fromhttps://clinicaltrials.gov.

B.Text Preprocessing

The Text Preprocessing stage is the stage where the application selects the data to be processed in each document. This preprocessing process includes (1) case folding, (2) tokenizing, (3) filtering, and (4) stemming. [28][29][30]

Case folding is to change all letters in a document to lowercase. Only letters "a" through "z" are accepted. Characters other than letters are removed and are considered delimiters. Case Folding is needed to convert the entire text in a document into a standard form.

Tokenizing stage is the stage of cutting the input string based on each word that compose it. Tokenization broadly breaks down a set of characters in a text into word units, how to distinguish certain characters that can be treated as word separators or not.

The Filtering stage is the stage of taking important words from the token results. Can use a stoplist algorithm (removing less important words) or a wordlist (storing important words). Stoplist / stopwords are non-descriptive words that can be discarded in a bag-of-words approach

Stemming, Indexing is done because a document cannot be recognized directly by an Information Retrieval System (IRS). Therefore, the document first needs to be mapped into a representation using the text contained in it.[31].

All eligibility criteria are converted into a simple word sequence.Information about the study intervention and disease type was added to each eligibility criteria by separating text into statements, then removing punctuation marks, space characters, all non-alphanumeric symbols, separators, and single character words from the extracted text.All words are lowercase.

C.Word Embedding

Word embedding is a technique for converting a word into a vector or array consisting of a set of numbers.With the word embedding technique, words can be converted into a vector containing numbers with a size that is small enough to contain more information.The information obtained will be sufficiently large that our vectors will be able to detect meaning.Every one word is charted to one vector.One vector is carried out learning which is similar to a neural network model, then combined in the field of deep learning[32]

In simple terms, word embedding is the process of converting a text into numbers. Why do we need this process? because most machine learning algorithms and deep learning architectures are unable to perform the analysis process on the input data in the form of strings or text, so they require numbers as input. A simple example of converting a word into a number vector. For example given the following sentence: "Word Embedding are Word Converted into numbers." A dictionary will contain a list of all unique words. So that the dictionary that is formed is: ["Word", "Embeddings", "are", "Converted", "into", "numbers"]. Using the one-hot encoding method will generate a vector where 1 represents the position of the word, and 0 for other words. The vector representation of the word "numbers" refers to the dictionary format above is [0,0,0,0,0,1] and the word "Converted" is [0,0,0,1,0,0]. Above is an example of a form of representation of text into numbers.

D.TF-IDF

The thing that needs to be considered in finding information from a heterogeneous collection of documents is term weighting. Terms can be words, phrases or other indexed units in the document that can be used to find out the context of the document. Because each word has different importance in the document, an indicator is given for each word. . However, there may be some words that are not really important, but are not yet on the stopwords list. So another approach to measuring this problem is to look at Inverse Document Frequence. IDF to reduce the dominance of words that often appear in various documents. This is necessary because words that appear in many documents can be considered general terms, so their value is not important.

E.Labeling

The labeling in this clinical statement is described as "eligible" and "not eligible". "Eligible" means the class of patients who will have to undergo a clinical trial, while "not eligible" means the class of patients who will not be subject to clinical trials.

In order to determine the appropriate and unworthy labels of the clinical statements, we labeled them according to:Judging from the inclusion and exclusion criteria contained in each sentence, if the related phrases are not found, then the statement is marked as "eligible". Inclusion criteria that start from the word "not" are transformed into statements that arepositive, labeled "not eligible". Any other possible means of omitting the statement will be processed in detail by the classifier

F. Long Short Term Memory

LSTM takes words from an input sentence i n distributed word representation format. Distributed wordrepresentation is a *n*-dimensional vector of continuous values used to represent a word in the vocabulary. Each word in dictionary ($w \in W$) is embedded into *n*-dimensional space ($L \in \mathbb{R}^{n \times |W|}$). Finally, aword vector can be seen as a single vector in the columnof *L*.The LSTM architecture consists of a set of recurrently connected memory blocks. Each block contains one memory cell *c* and three multiplicative units(gates). These gates help the LSTM memory cell toperform write, read, and reset operation. [33]They enablethe LSTM memory cell to store and access informationover a period of time. These gates are so-called inputgate *i*, forget gate *f*, and output gate *o*. Mathematically one block of LSTM can be viewed as:

$$it = \sigma(Wxixt + Whiht-1 + Wcict-1 + bi), \quad (1)$$
$$ft = \sigma(Wxixt + Whiht-1 + Wcict-1 + bf), \quad (2)$$
$$ct = ftct-1 + it \tanh(Wxcxt + Whcht-1 + bc), \quad (3)$$
$$ot = \sigma(Wxoxt + Whoht-1 + Wcoct + bo), \quad (4)$$
$$ht = ot \tanh(ct), \quad (5)$$

where *xt* is a single distributed vector *L* of word *w*,an input to the LSTM, σ is a logistic sigmoid function,and *h* is a hidden vector. The weight *W* and *b* subscriptrepresent the edge connection matrix and bias vector.

G. Bidirectional Long Short-Term Memory

One drawback of LSTM architecture is that they areonly considering the previous context. In order to makeLSTM aware of both previous and subsequent context,it needs to process data in both directions with twoseparate hidden layers [19]. Later these two hiddenlayers are combined to the same output layer. Thisarchitecture is so called BiLSTM.By computing both a forward and backward layer, BiLSTM allows us to exploit future and history context together at once. Recently, BiLSTM has been used intensively for real-world applications, ranging from signalprocessing tasks[16][14] to text processing tasks.

Bi-directional Long Short Term Memory (BiLSTM)structure allows networks to have information about thesequence both backward and forward at all times. Usingbidirectional LSTM, your inputs will run in two ways: onefrom past to future and the other from future to past andwhat varies from unidirectional is that in the LSTM thatoperates backwards, you are able to preserve knowledgefrom the future and use the two hidden states together tomaintain details from the past and the future at any point intime.The idea underlying Bidirectional Recurrent NeuralNetworks (RNNs) is very simple. It entails replicating thefirst repeated layer in the network and supplying the inputsequence as it is sent to the first layer, then presenting thereplicated layer with a reversed copy of the input sequence.This overcomes the conventional RNN limitations.Bidirectional recurrent neural network (BRNN) can beeducated in the past and future of a particular time-stepusing all available input data. Splitting state neurons inregular RNN is responsible for forward states (positivedirection of time) and backward states (negative direction oftime)[17].

H. Recurrent Neural Network

Recurrent Neural Network (RNN) is a Neural Network,which uses backward propagation, where the result obtainedfrom the step before is fed to the current step as input. Intraditional neural network all the input and outputs areindependent, but in cases such as when it is important topredict the next word of a sentence, the preceding words arerequired and, therefore, it is necessary to remember theprevious words. This led to the existence of RNN, henceRNN with help of hidden layer was able to solve this issue.RNN's principal and most significant feature is the hiddenstate, the hidden state has some memory of the sequence.The most important feature of RNN is hidden state, whichhas the ability to remember information about a sequence. Ithas a memory that recalls all information which is required.Unlike other neural networks this reduces the difficulty ofparameters. A recurrent neural network uses time t torepresent the forms that the input sequence takes to reachthe final output sequence that is necessary. There is a hiddenstate ht to represent the status of the input processing neuralnetwork system at a particular time t. RNN accepts input xtat time t, and a non-linear function helps predict the systemstatus at time t using the status at time t-1, ht = f (ht-1 , xt)[6]

E.Precision Recall and F1-Score

To determine the classification performance [34]**,**it can be seen in table 1

Table-1 : Confusion Matrix

	PredictedClass		
ActualClass		Class = Yes	Class = No
	Class = Yes	TP	FN
	Class = No	FP	TN

TP: True Positive

FP: False Positive

FN: False Negative

TN: True Negative

Precision is a representation of uniformity and repetition of measurements. Precision is the degree of excellence, on the performance of an operation or technique used to get results.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

Recall is a measure of the success of a system in finding and retrieving information.Furthermore, F-Measure is a process of calculating evaluation by combining precision and recall calculations.recall and precission in a situation can have different weights.The measure that displays the reciprocity between Recall and Precission is F-Measure which is the average harmonic weight and realization and precision..

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

F-Measure or F1-score is one of the evaluation calculations in information retrieval that combines recall and precission. The recall value and Precission in a situation can have different weights. The size that displays reciprocity between Recall and Precission is F-Measure which is the mean harmonic weight and reall and precission

$$\text{F1 Score} = \frac{2*(Recall*Precision)}{Recall+Precision} \quad (8)$$

3.Result

A.Evaluation using K-fold Cross Validation

Cross validation is a technique for performing statistical analysis of independent data.This technique is used to create prediction and classification models that are able to estimate the level of accuracy of the predictions and classifications that are carried out.In a predictive model, the model will be given a set of training data and testing data.The purpose of cross validation is to describe the data set to "test" the model in the training phase (i.e., data validation).in order to restrict issue such as overfitting, providing concept into how the model will generalize independent of the dataset (i.e., unknown datasets, for example from real issue), etc.[35]. One round of cross-validation involves dividing the sample data into complementary subsets, performing analysis on one subset (named training data), and validating the analysis on another subset (named validation or testing data). To degrade variability, multiple rounds of cross validation are implemented using disparate partitions, and the validation results are averaged over the rounds

One of the primaryargument for using cross validation instead of using conventional validation (e.g. partitioning a data set into two sets, i.e. 70% for training and 30% for testing) is that there is not adequate data available to partition it into separate training and lossless testing data. significant modeling or testing capabilities. In these cases, a fair way to predict the prediction model with accuracy is to use cross validation as a powerful common technique [36]. In short, cross validation combines (averaged) the fit measure (prediction error) by looking at the average value for each round to obtain a more accurate estimate of the prediction model's performance.

The table-2 shows a snippet of the dataset used which consists of document and class attributes. We used a short free text of the cancer clinical trial protocol available on the public registry to practice embedding words, which were labeled as eligible or not eligible.

Tabel-2 Snippet Of Dataset

Document	Class
first or second line hertwo negative breast cancer diagnosis and pregnant or breast feeding females	1
first or second line hertwo negative breast cancer diagnosis and patients with active brain metastases c	1
first or second line hertwo negative breast cancer diagnosis and patients who have undergone major su	1
first or second line hertwo negative breast cancer diagnosis and no prior therapy with an antiresorptive	1
first or second line hertwo negative breast cancer diagnosis and life expectancy of greater_than six mo	0
first or second line hertwo negative breast cancer diagnosis and history of non compliance to medical r	1
first or second line hertwo negative breast cancer diagnosis and history of bone metabolism diseases	1
first or second line hertwo negative breast cancer diagnosis and current active dental problems includir	1
ewing sarcoma diagnosis and wbc greater_than two thousand Âµl	0
ewing sarcoma diagnosis and onest complete remission but very high risk features less_than twenty sui	0
ewing sarcoma diagnosis and more than one cycle of other chemotherapy prior to registration	1
ewing sarcoma diagnosis and informed consent from patient or legal guardian if patient is minor	0
ewing sarcoma diagnosis and haematological parameters	0
ewing sarcoma diagnosis and forty-five days after diagnostic biopsy surgery	0
ewing sarcoma diagnosis and age and sex	0
evidence of inflammatory breast cancer or distant metastasis	1
estrogen receptor positive breast cancer diagnosis and regular use of anti inflammatory agents with the	1
estrogen receptor positive breast cancer diagnosis and postmenopausal women	0
estrogen receptor positive breast cancer diagnosis and postmenopausal status defined as	0
estrogen receptor positive breast cancer diagnosis and patients who will receive neoadjuvant therapy p	1
ductal carcinoma in situ diagnosis and moderate to severe renal impairment serum creatinine greater th	0
ductal carcinoma in situ diagnosis and exclusion	0
ductal carcinoma in situ diagnosis and ecog performance status zero two	0
ds stage plasma cell myeloma diagnosis and no prior treatment with denosumab	0
ds stage plasma cell myeloma diagnosis and no current treatment with investigational agent s	0
ds stage plasma cell myeloma diagnosis and corrected serum calcium greater_than equal_than doc m	0
ds stage plasma cell myeloma diagnosis and calculated creatinine clearance greater_than equal_thar	0
ds stage ii plasma cell myeloma diagnosis and at least one site of bone metastasis or bone involvement	0
disease free period of other malignant tumor is less than five years except cured basal cell skin cancer a	1
diagnosis of at least one cancer related bone lesion	0
current or previously documented diagnosis of at least one bone metastasis due to prostate cancer pat	0
confirmed multiple myeloma stage ii iii as per salmon and durie one thousand	0
concurrent treatment including chemotherapy hormonal therapy and or biologic therapy for metastatic	0
concurrent participation in another clinical treatment trial for this cancer unless the patient is longer rec	1
concurrent aromatase inhibitors	1

Class 0 : not eligible
Class 1 : eligible

B.Result using K-fold Cross Validation

K-fold Cross Validation (CV) works by dividing the data into several folds and confirm that each fold is used as a testing set at some point. Table-3 below shows the results of the accuracy, precision, recall and f1-score for testing data using 10-fold cross validation.

Table-3 Result Of Testing Data 10-Fold

```
Correctly Classified Instances        1554               77.7389 %
Incorrectly Classified Instances       445               22.2611 %
Kappa statistic                          0.5541
Mean absolute error                      0.2307
Root mean squared error                  0.4482
Relative absolute error                 46.1524 %
Root relative squared error             89.6515 %

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.808    0.255    0.768      0.808    0.788      0.555    0.808     0.768     1
               0.745    0.192    0.788      0.745    0.766      0.555    0.809     0.754     0
Weighted Avg.  0.777    0.224    0.778      0.777    0.777      0.555    0.809     0.761
```

Training set is a part of the dataset that we train to make predictions or run the functions of an ML algorithm. Table-4 below shows the results of the training data process

Table-4 : Result Of Training Data

```
Correctly Classified Instances        1519               75.988  %
Incorrectly Classified Instances       480               24.012  %
Kappa statistic                          0.5191
Mean absolute error                      0.251
Root mean squared error                  0.4229
Relative absolute error                 50.2246 %
Root relative squared error             84.5962 %


=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.785    0.266    0.755      0.785    0.769      0.520    0.850     0.839     1
                 0.734    0.215    0.765      0.734    0.749      0.520    0.850     0.852     0
Weighted Avg.    0.760    0.241    0.760      0.760    0.760      0.520    0.850     0.845
```

The graph below is a comparison of the evaluation results from training data and testing data. As seen in the graph below, the difference between the training data and the 10-fold testing data is not so significant, meaning that the dataset we use has good validity.

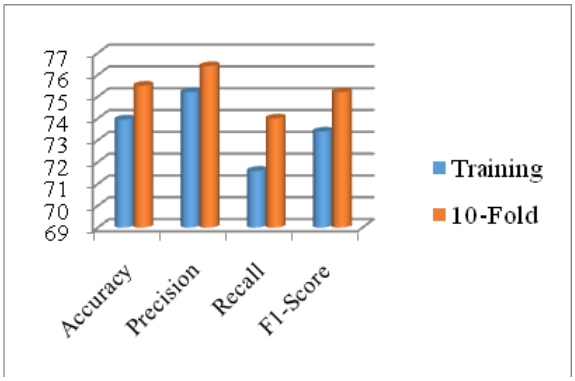

Figure-3 Graph of comparison between training data and testing data

Table-5 describes the experimental results from the dataset without the word embedding feature and the results after using the word embedding feature.

Table-5 Result Of Without And Using Word Embedding Feature

	Accuracy %	Precision %	Recall %	F1-Score %
Without WE	73.94	73.70	76.10	74.90
After WE	77.74	76.80	80.80	78.80

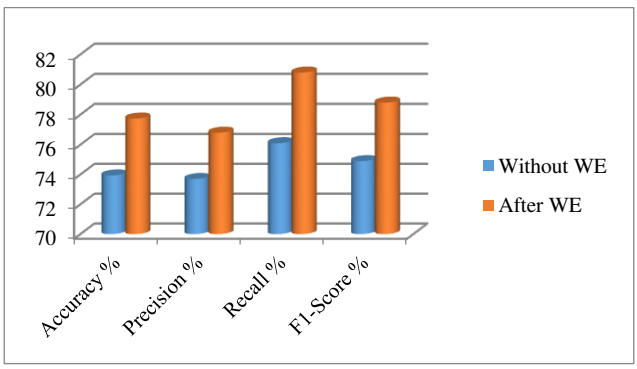

Figure-4 result of without and using word embedding feature

Table-5 and figure-4 show that the results of testing data processing using the word embedding feature and without using the word embedding feature are not much different. There is no visible significant effect of using the word embedding feature. This means that we have proven that representational studies using the multilayer perceptron are successfully utilized to extract medical knowledge from clinical trial protocols to potentially help practitioners when recipe treatments.

Word embedding is important in this research.Adding a trained "word embedding" to the classification adds value to the evaluation.We also use word vectors to generate word clusters. The word vectors generated were also useful for accurately resolving analogy problems.

5. Conclusion

We have constructed a classification model to improve the classification value of clinical trial texts using a multilayer perceptron. The accuracy of the eligibility classification is 77.74%, 76.80% precision, 80.80% recall and F1 score 78.80%. This increase in value occurs because of the effect of word embedding. We use circularity measurements in the dataset and from statistics. Our analysis finds that these quantitative measures lead to transparency in our conclusions.

The next research is a continuation of this research. Variations in research can be created by using a larger dataset or similar dataset and by using other deep learning methods, making it possible to conduct research in multi-label classification.

REFERENCES

[1] K. A. Kessel *et al.*, "Cancer clinical trials – Survey evaluating patient participation and acceptance in a university-based Comprehensive Cancer Center (CCC)," *Clin. Transl. Radiat. Oncol.*, vol. 13, pp. 44–49, 2018.

[2] Y. Wu, M. Jiang, J. Lei, and H. Xu, "Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network," *Stud. Health Technol. Inform.*, vol. 216, pp. 624–628, 2015.

[3] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "ScienceDirect ScienceDirect CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Comput. Sci.*, vol. 111, no. 2015, pp. 376–381, 2017.

[4] A. H. Wang, "DON ' T FOLLOW ME Spam Detection in Twitter," 2010.

[5] Y. Yang and X. Liu, "A re-examination of text categorization methods."

[6] W. K. SARI, D. P. RINI, R. F. MALIK, and I. S. B. AZHAR, "Sequential Models for Text Classification Using Recurrent Neural Network," vol. 172, no. Siconian 2019, pp. 333–340, 2020.

[7] H. Zhou, Y. Zhang, S. Huang, and J. Chen, "A neural probabilistic structured-prediction model for transition-based dependency parsing," *ACL-IJCNLP 2015 - 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. Asian Fed. Nat. Lang. Process. Proc. Conf.*, vol. 1, no. 2014, pp. 1213–1222, 2015.

[8] B. Melinda, "Clinical Trials," *Bill Melinda Gates Found.*

[9] C. Shivade, C. Hebert, M. Lopetegui, M. De Marneffe, E. Fosler-lussier, and A. M. Lai, "Textual inference for eligibility criteria resolution in clinical trials," vol. 58, 2015.

[10] A. Srikanthan *et al.*, "Evolution in the eligibility criteria of randomized controlled trials for systemic cancer therapies," *Cancer Treat. Rev.*, vol. 43, pp. 67–73, 2016.

[11] E. Chondrogiannis, V. Andronikou, A. Tagaris, E. Karanastasis, T. Varvarigou, and M. Tsuji, "A novel semantic representation for eligibility criteria in clinical trials," *J. Biomed. Inform.*, vol. 69, pp. 10–23, 2017.

[12] B. Mackellar and C. Schweikert, "Analyzing Conflicts between Clinical Trials from a Patient Perspective," pp. 479–482, 2015.

[13] S. Liu, K. Lee, and I. Lee, "Knowledge-Based Systems Document-level multi-topic sentiment classification of Email data with BiLSTM and data augmentation," vol. 197, 2020.

[14] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, 2020.

[15] S. Sathyanarayana, "A Gentle Introduction to Backpropagation," *Numer. Insight, Inc Whitepaper*, no. July 2014, pp. 1–15, 2014.

[16] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[17] G. Xu, "Sentiment Analysis of Comment Texts Based on BiLSTM," vol. 7, 2019.

[18] J. Jasmir, S. Nurmaini, R. F. Malik, and B. Tutuko, "Bigram feature extraction and conditional random fields model to improve text classification clinical trial document," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 3, pp. 886–892, 2021.

[19] J. Jasmir, S. Nurmaini, R. F. Malik, and D. Zaenal, "Text Classification of Cancer Clinical Trials

Documents Using Deep Neural Network and Fine Grained Document Clustering," vol. 172, no. Siconian 2019, 2020.
[20] B. Mackellar and C. Schweikert, "Patterns for Conflict Identification in Clinical Trial Eligibility Criteria," 2016.
[21] R. M. and C. Weng, "Unsupervised Mining of Frequent Tags for Clinical Eligibility Text Indexing," *Bone*, vol. 23, no. 1, pp. 1–7, 2012.
[22] K. Zhang and D. Demner-Fushman, "Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations," *J. Am. Med. Informatics Assoc.*, vol. 24, no. 4, pp. 781–787, 2017.
[23] C. H. Chuan, "Classifying Eligibility Criteria in Clinical Trials Using Active Deep Learning," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 305–310, 2019.
[24] Y. Ni *et al.*, "Increasing the efficiency of trial-patient matching: Automated clinical trial eligibility Pre-screening for pediatric oncology patients Clinical decision-making, knowledge support systems, and theory," *BMC Med. Inform. Decis. Mak.*, vol. 15, no. 1, pp. 1–10, 2015.
[25] Z. H. Kilimci and S. Akyokus, "Deep learning- and word embedding-based heterogeneous classifier ensembles for text classification," *Complexity*, vol. 2018, 2018.
[26] L. Ge and T. S. Moh, "Improving text classification with word embedding," *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-Janua, pp. 1796–1805, 2017.
[27] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning Sentiment-Specific Word Embedding," pp. 1555–1565, 2014.
[28] L. Zhou *et al.*, "Text preprocessing for improving hypoglycemia detection from clinical notes – A case study of patients with diabetes," *Int. J. Med. Inform.*, vol. 129, no. January, pp. 374–380, 2019.
[29] J. C. Ferrão, M. D. Oliveira, F. Janela, and H. M. G. Martins, "Preprocessing structured clinical data for predictive modeling and decision support: A roadmap to tackle the challenges," *Appl. Clin. Inform.*, vol. 7, no. 4, pp. 1135–1153, 2016.
[30] D. Torunoğlu, E. Çakirman, M. C. Ganiz, S. Akyokuş, and M. Z. Gürbüz, "Analysis of preprocessing methods on classification of Turkish texts," *INISTA 2011 - 2011 Int. Symp. Innov. Intell. Syst. Appl.*, pp. 112–117, 2011.
[31] H. Fan and Y. Qin, "Research on Text Classification Based on Improved TF-IDF Algorithm," vol. 147, no. Ncce, pp. 501–506, 2018.
[32] Z. Yin and Y. Shen, "On the dimensionality of word embedding," *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. NeurIPS, pp. 887–898, 2018.
[33] A. Darmawahyuni, "Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification," *2019 6th Int. Conf. Instrumentation, Control. Autom.*, no. August 2019, pp. 19–23, 2020.
[34] A. Darmawahyuni, S. Nurmaini, and F. Firdaus, "Coronary Heart Disease Interpretation Based on Deep Neural Network," *Comput. Eng. Appl. J.*, vol. 8, no. 1, pp. 1–12, 2019.
[35] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, no. Cv, pp. 78–83, 2016.
[36] G. Seni and J. F. Elder, *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*, vol. 2, no. 1. 2010.