Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 4 (2022) 505 - 510 ISSN Media Electronic: 2580-0760

Bidirectional Long Short-Term Memory and Word Embedding Feature for Improvement Classification of Cancer Clinical Trial Document

Jasmir Jasmir[1], Willy Riyadi[2], Silvia Rianti Agustini[3], Yulia Arvita[4], Despita Meisak[5], Lies Aryani[6]
[1,2]Computer Engineering, Universitas Dinamika Bangsa Jambi Indonesia
[3,4,5,6]Information System, Universitas Dinamika Bangsa Jambi Indonesia
[1]ijay_jasmir@yahoo.com, [2]wriyadi5@gmail.com, [3]silviarianti7@gmail.com, [4]yulia_arvita@yahoo.co.id,
[5]liesaryani6@gmail.com, [6]despitam88@gmail.com

Abstract

In recent years, the application of deep learning methods has become increasingly popular, especially for big data, because big data has a very large data size and needs to be predicted accurately. One of the big data is the document text data of cancer clinical trials. Clinical trials are studies of human participation in helping people's safety and health. The aim of this paper is to classify cancer clinical texts from a public data set. The proposed algorithms are Bidirectional Long Short Term Memory (BiLSTM) and Word Embedding Features (WE). This study has contributed to a new classification model for documenting clinical trials and increasing the classification performance evaluation. In this study, two experiments work are conducted, namely experimental work BiLSTM without WE, and experimental work BiLSTM using WE. The experimental results for BiLSTM without WE were accuracy = 86.2; precision = 85.5; recall = 87.3; and F-1 score = 86.4. meanwhile the experiment results for BiLSTM using WE stated that the evaluation score showed outstanding performance in text classification, especially in clinical trial texts with accuracy = 92,3; precision = 92.2; recall = 92.9; and F-1 score = 92.5.

Keywords: Deep Learning, BiLSTM, Text classification, Word Embedding, Clinical Trials

1. Introduction

Clinical trials (CT) are very important activities to determine the safety and effectiveness of medical treatment. These trials form the basis for conducting clinical practice guidelines [1] and greatly assist medical teams in conducting health practice and when making treatment-related decisions. Clinical documents containing clinical statements require a review of the feasibility of clinical trial results (feasible and unfit), the eligibility criteria used in clinical trials are very limited and require a computerized process or text classification [2][3].

Text classification is considered a fundamental problem in Information Sciences. Text classification, known as an effective method for text information organization and management, is widely employed in the fields of information sorting[4] sentiment analysis[5][6], spam filtering [7][8], clinical text [9] etc. The method of deep learning is deemed as an effective method for classification.[10] Moreover, an increasing number of scholars have applied commonly used neural networks, including the conditional random filed[11] and the recurrent neural network (RNN), to text classification[12],

Recently, neural networks have had tremendous success for text classification, and are showing more significant advances than other models.One of the challenges of building a text classification model is how to capture features for different units of text, such as phrases, sentences and documents. Making use of its iterative structure, Recurrent Neural Network (RNN), as a type of alternative neural network, is particularly suitable for processing text of variable length.[13][14] Because RNN is equipped with recurrent network structure which can be used to maintain information, it can better integrate information in certain contexts[15]. For the purpose of avoiding the problem of gradient exploding or vanishing in a standard RNN, long short-term memory (LSTM) [16][17] have been designed for the improvement of remembering and memory accesses. Living up to expectations, LSTM does show a remarkable ability in the processing of natural language.

Accepted by editor: 12-04-2022 | Final Revision: 20-07-2022 | Online Publication: 22-08-2022

In improving performance and computing, Zuheros[18] in a study explained the challenges of one of the methods of deep learning. One of the challenges of deep learning problems as revealed in the research of Vincent Menger [19] which states that in some cases the approach with deep learning techniques applied to clinical text classification can produce conclusions that are in line with expectations, but will be different if tested on datasets. different clinical settings and with different domains and different sizes. They propose a neural network architecture to improve the system performance and computing efficiency. While Zhipeng Jiang [20] utilized LSTM and the result produced good accuracy, however the result indicated with low computing performance. The LSTM development is BiLSTM. Beakcheol Jang [21] has also used BiLSTM to improve accuracy in text classification. In another study, there is also research that discusses the features of word embedding [22][23]. The word embedding feature is also able to improve classification text. For instance including Zeynep H. Kilimci used the word embedding with Random Forest and CNN [24]. Lihao Ge uses word embedding with naive bayes and Support vector machine [25] and Duyu Tang uses word embedding with Support vector machine and N-gram [26], Jasmir etc using Bidirectional Long Short Term Memory Recurrent Neural Network to improve eligibility classification clinical trials document [27]

There are several studies related to BiLSTM for instance Guixian Xu [28] discussed the use of BiLSTM in sentiment analysis, meanwhile Annisa Darmawahyuni [28] conducted Myocardial Infarction Classification with BiLSTM, this paper is a preliminary study so that contains only brief analysis and plan. However, it can present other point-of-view to process cardiac rhythm that associated in timesteps based on deep learning approach. Beakcheol Jang[21] used BiLSTM to improve text classification.

2. Research Methods

Dataset containing clinical texts that have been labeled was intervened by inserting words or data containing several other types of cancer for each criterion. The insertion of words or data for other types of cancer in this case is called the word embedding process. The result of the Word Embedding process is integrated into the IDF TF algorithm and produces a weighted word vector. The weighted word vector was entered into BiLSTM and then generated a BiLSTM model from this clinic dataset. The next stage is to evaluate the classification performance with a configuration matrix. In this paper the authors compare the evaluation results with BiLSTM without Word Embedding and BiLSTM with Word Embedding.

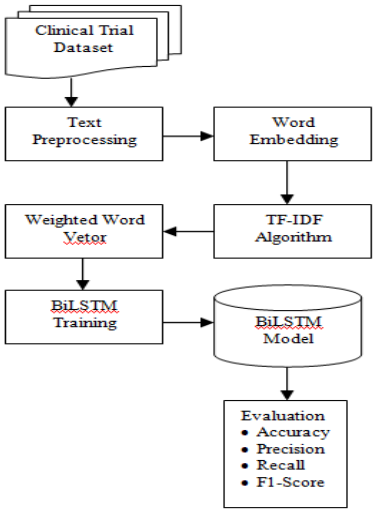

Figure 1. Conceptual Model

A. Construction of Weighted Word Vector

In this paper, the Word2vec model is used to achieve a distributed word representation. The Word2vec model is part of the CBOW model and the Skip-gram model. The advantage of this word2vec model is the ability to predict words based on context distribution [29]. This model also contains an input layer, a projection layer and an output layer. For the word w_k, the context is stated in formula (1)

$$Context(W_k) = \{w_{k-t}, w_{k-(t-1)}, ., w_{k+(t-1)}, w_{k+t}\} \quad (1)$$

In contrast, the Skip-gram model predicts the context based on the word target wk.

TF-IDF is the most commonly used weight calculation method in text categorization. This method considers the frequency of words and distribution of words in the document, so that the features in the classification are reflected in this method. The TF-IDF formula can be seen in the formula (2,3)

$$w(ti,d) = \frac{tf(ti,d)xidf\ (ti)}{\sqrt{\sum_{ti\epsilon d}[tf(ti,d)xidf(ti)]^2}} \quad (2)$$

$$idf(ti) = \log\left(\frac{N}{Nti}\right) + 1 \quad (3)$$

w (ti, d) is the weight of the word *ti* in the document *d*, *tf (ti, d)* is the frequency of the word *ti* in the document d, *N* is the number of documents and *ti* is the number of documents where the word *ti* appears

The weight calculation method for word vectors is as follows:

$$wi = tf - idfi\ .e \quad (4)$$

DOI: https://doi.org/10.29207/resti.v6i4.4005

$$e = \{^{\alpha,\ tiisnoteligibilityclass}_{1,\ tiiseligibilityclass} \quad (5)$$

B. 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[30].

In simple terms, word embedding is the process of converting a text into numbers, 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.

C. BiLSTM Layer

Long Short-term Memory (LSTM)[31] evolved from the Recurrent Neural Network (RNN). The main idea used is to add a "gateway" to the Recurring Neural Network for the purpose of controlling the passing data. Generally, LSTM architecture consists of memory cells, input gates, output gates, and forget gates. The LSTM is presented in the form of a chain constructed with repeated modules of neural networks. With the information stored inside, memory cells travel along the chain. In addition, the other three gates are mainly designed to control whether to add or block information to the memory cell.

The LSTM transition functions are defined as follows:

$$It = \sigma(WxiXt + Whiht - 1 + WCict - 1 + bi) \quad (6)$$

$$Ft = \sigma(WxfXt + Whfht - 1 + WCfct - 1 + bf) \quad (7)$$

$$Ct = FtCt - 1 + Ittanh(WxctXt + Whcht - 1 + bc) \quad (8)$$

$$Ot = \sigma(WxoXt + Whoht - 1 + WCoct + bo) \quad (9)$$

$$ht = Ottanh(ct) \quad (10)$$

σ refers to the logistic sigmoid function that has an output in [0, 1], tanh indicates the hyperbolic tangent function that has an output in [-1, 1], and ● denotes the element wise multiplication. At the current time t, h_t refers to the hidden state, f_t represents the forget gate, i_t indicates the input gate, and o_t denotes the output gate. W_t, W_o, and W_f represent the weight of these three gates, respectively, while b_t, b_o, b_f refers to the biases of the gates. As for BLSTM, it is regarded as an extension of the unidirectional LSTM, and it not only adds another hidden layer but also connects with the first hidden layer in the opposite temporal order. Because of its structure, BLSTM can process the information from both the past and the future.

Therefore, BLSTM is adopted to capture the information of the text input in this paper. In general, the BiLSTM architecture can be seen in the figure 2.

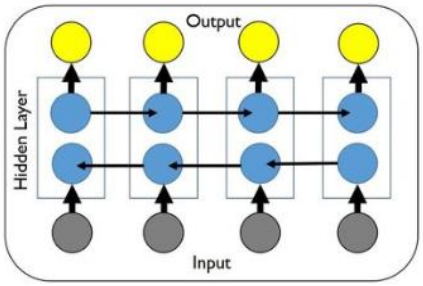

Figure 2. BiLSTM Architecture

3. Result and Discussion

3.1. Accuracy Precision Recall and F-1 Score

Measures of classification performance can be defined based on the confusion matrix [32] as seen in Table1. The confusion matrix provides information on the comparison of the classification results carried out by the system (model) with the actual classification results. The confusion matrix is in the form of a matrix table that describes the performance of the classification model on a series of test data whose true values are known.

Table 1. Measures of classification performance

		Actual Class	
Predicted Class	Class = Yes	Class = Yes True Positive = TP	Class = No False Positive = FP
	Class = No	False Negative = FN	True Negative = TN

Accuracy is how close or far off a given set of measurements (observations or readings) are to their true value.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

DOI: https://doi.org/10.29207/resti.v6i4.4005

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.

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

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 Precision in a situation can have different weights. The measure that displays the reciprocity between Recall and Precision is F-Measure which is the average harmonic weight and realization and precision.

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

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

$$\text{F1 Score} = \frac{2*(recall*Precision)}{(Recall+Precision)} \quad (14)$$

3.2. Model Evaluation and Validation

This study uses public data from the National Library of Medicine about cancer in free-text language. Data taken from Clinical Trial.gov with a total of 500,000 records.

In this paper, two experiments are carried out for text classification. The first experiment used the BiLSTM model without the Word Embedding (WE). The second experiment was to combine BiLSTM using WE.

Table-2 and figure-3 are the results of BiLSTM without WE. The evaluation results from BiLSTM stated Accuracy = 86.2; Precision = 85.5; Recall = 87.3; and F-1 Score = 86.4.

Table 2. Table of BILSTM without WE

Evaluation	BiLSTM (%)
Accuracy	86.2
Precision	85.5
Recall	87.3
F-1 Score	86.4

The propose Bidirectional LSTM (BiLSTM) with Word Embedding (WE). Table-3 and figure-4 are the results of BiLSTM with WE. The evaluation results from BiLSTM stated Accuracy = 92.3; Precision = 92.2; Recall = 92.9; and F-1 Score = 92.5;

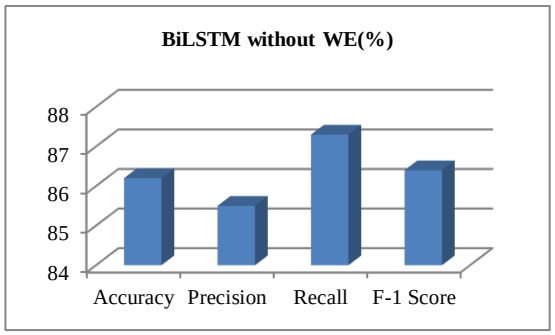

Figure 3. Graph of BiLSTM without WE

Table 3. Table of BILSTM With WE

Evaluation	BiLSTM + WE (%)
Accuracy	92.3
Precision	92.2
Recall	92.9
F-1 Score	92.5

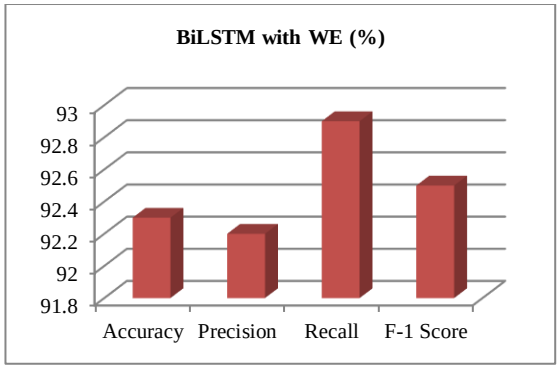

Figure 4. Graph of BiLSTM with WE

To clarify the difference in evaluation results between the use of BiLSTM without WE and the use of BiLSTM using WE in the clinical trial text dataset presented in Table 4 and Figure 5.

Table 4. Table of Compare BILSTM Without WE And BILSTM using WE

Evaluation	BiLSTM (%)	BiLSTM + WE (%)
Accuracy	86.2	92.3
Precision	85.5	92.2
Recall	87.3	92.9
F-1 Score	86.4	92.5

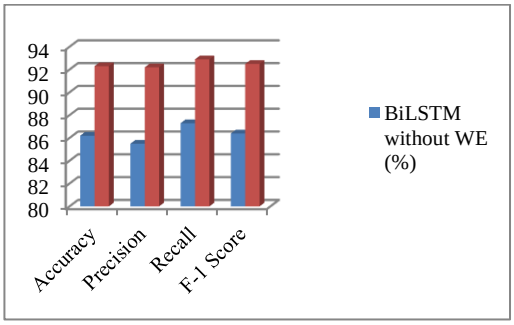

Figure 5. Graph of Compare BiLSTM without WE and BiLSTM with WE

DOI: https://doi.org/10.29207/resti.v6i4.4005

The Table-4 and Figure-5 above shows that there is a difference between the results with BiLSTM without WE and BiLSTM using WE. Evaluation value using the BiLSTM using WE outperformed the evaluation value on the BiLSTM without WE. Context BiLSTM and WE 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. This model identifies disease concepts more correctly out of the predicted disease name mentions.

Discussion

The aim of this paper is to address the lack of standard word embedding in NLP (compared to vision) by introducing a series of simple operations that may serve as the basis for future investigations. With the rate at which NLP research has progressed in recent years, we suspect that researchers will soon discover a higher-performance word embedding technique that will also be easy to use.

In particular, much of the recent work at NLP has focused on making neural models larger or more complex. Simple operation has been introduced in this study. With this study, it is hoped that it can inspire new approaches to universal or task-specific word embedding.

4. Conclusion

In this paper a BiLSTM using WE model is introduced to a clinical text case. The experiment compares the BiLSTM without WE and BiLSTM using WE, which finds that evaluation can be improved after using BiLSTM using WE. In addition, our model was evaluated by applying it to the classification of clinical trial texts. The models are applied to clinical trial texts and then the corresponding effects are compared with each other. It turns out that our model is more suitable for clinical text. Apart from that, through our evaluation it has also been shown that our BLSTM using WE model achieves excellent results and also outperforms various basic models.

References

[1] W. J. Gradishar *et al.*, "Clinical practice guidelines in oncology," *JNCCN J. Natl. Compr. Cancer Netw.*, vol. 16, no. 3, pp. 310–320, 2018.

[2] S. Jin, R. Pazdur, and R. Sridhara, "Re-evaluating eligibility criteria for oncology clinical trials: Analysis of investigational new drug applications in 2015," *J. Clin. Oncol.*, vol. 35, no. 33, pp. 3745–3752, 2017.

[3] J. Jasmir, S. Nurmaini, and B. Tutuko, "Fine-grained algorithm for improving knn computational performance on clinical trials text classification," *Big Data Cogn. Comput.*, vol. 5, no. 4, 2021.

[4] J. Zhang, F. Liu, W. Xu, and H. Yu, "Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism," *Futur. Internet*, vol. 11, no. 11, 2019.

[5] R. Pavitra and P. C. D. Kalaivaani, "Weakly supervised sentiment analysis using joint sentiment topic detection with bigrams," *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015*, pp. 889–893, 2015.

[6] 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.

[7] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review Spam Detection via Temporal Pattern Discovery," pp. 823–831, 2012.

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

[9] 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.

[10] Jasmir *et al.*, "Breast Cancer Classification Using Deep Learning," *Proc. 2018 Int. Conf. Electr. Eng. Comput. Sci. ICECOS 2018*, vol. 17, pp. 237–242, 2019.

[11] 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.

[12] Y. Li, X. Wang, and P. Xu, "Chinese text classification model based on deep learning," *Futur. Internet*, vol. 10, no. 11, 2018.

[13] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1724–1734, 2014.

[14] W. C. Annisa Darmawahyuni, Siti Nurmaini, Sukemi and M. N. R. and F. Vicko Bhayyu, "Deep Learning with a Recurrent Network Structure in the Sequence Modeling of Imbalanced Data for," pp. 1–12.

[15] 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.

[16] A. Gulli and S. Pal, "Long Short Term Memory - LSTM," *Deep Learn. with Keras*, pp. 187–195, 2017.

[17] A. Darmawahyuni, S. Nurmaini, and Sukemi, "Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification," *Proc. 2019 6th Int. Conf. Instrumentation, Control. Autom. ICA 2019*, no. August 2019, pp. 19–23, 2019.

[18] C. Zuheros, S. Tabik, A. Valdivia, E. Martínez-cámara, and F. Herrera, "Deep recurrent neural network for geographical entities disambiguation on social media data," *Knowledge-Based Syst.*, 2019.

[19] V. Menger, F. Scheepers, and M. Spruit, "Comparing deep learning and classical machine learning approaches for predicting inpatient violence incidents from clinical text," *Appl. Sci.*, vol. 8, no. 6, 2018.

[20] Z. Liu, B. Tang, X. Wang, and Q. Chen, "De-identification of clinical notes via recurrent neural network and conditional random field," *J. Biomed. Inform.*, vol. 75, pp. S34–S42, 2017.

[21] 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.

[22] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, "Efficient processing of GRU based on word embedding for text classification," *Int. J. Informatics Vis.*, vol. 3, no. 4, pp. 377–383, 2019.

[23] P. Wang, B. Xu, J. Xu, G. Tian, C. L. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, pp. 806–814, 2016.

[24] Z. H. Kilimci and S. Akyokus, "Deep learning- and word embedding-based heterogeneous classifier ensembles for text classification," *Complexity*, vol. 2018, 2018.

DOI: https://doi.org/10.29207/resti.v6i4.4005

[25] 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.

[26] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning Sentiment-Specific Word Embedding," pp. 1555–1565, 2014.

[27] J. Jasmir *et al.*, "Improving Eligibility Classification on Clinical Trials Document using Bidirectional Long Short Term Memory Recurrent Neural Network," vol. 12, no. 14, pp. 2784–2792, 2021.

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

[29] T. Demeester, T. Rocktäschel, and S. Riedel, "Distributed Representations ofWords and Phrases and their Compositionality," *EMNLP 2016 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 1389–1399, 2016.

[30] 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.

[31] K. S. Tai, R. Socher, and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks."

[32] 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.

DOI: https://doi.org/10.29207/resti.v6i4.4005