

Various Approaches for Predicting Stroke Prognosis using Magnetic Resonance Imaging Text Records

Tak-Sung Heo¹, Chulho Kim², Jeong-Myeong Choi¹, Yeong-Seok Jeong³, Yu-Seop Kim³

¹Department of Convergence Software, Hallym University, Republic of Korea

²Department of Neurology, College of Medicine, Hallym University, Republic of Korea

³College of Software, Hallym University, Republic of Korea

gjxkrtjd221@gmail.com gumdol52@naver.com jeong5905@gmail.com

dnfkdi1995@gmail.com yskim01@hallym.ac.kr

Abstract

Stroke is one of the leading causes of death and disability worldwide. Stroke is treatable, but it is prone to disability after treatment. To grasp the degree of disability caused by stroke, we use magnetic resonance imaging text records to predict stroke and measure the performance according to the document-level and sentence-level representation. As a result of the experiment, the document-level representation shows better performance.

1 Introduction

Stroke is a neurologic disorder that is characterized by an acute disruption of cerebral blood flow and corresponding symptoms lasting more than 24 hours. Stroke is one of the leading causes of death and disability worldwide (Sacco et al., 2013; Campbell et al., 2019). As the incidence of stroke is increasing and proportion of stroke survivors with a disability is also increasing recently, there are increasing demands of proper diagnosis and long-term treatment strategy to reduce the global burden of stroke (Ekker et al., 2019). Among the diagnostic tools for stroke, the most important imaging methods are brain computed tomography (CT) and magnetic resonance image (MRI). We can obtain vascular images and various functional images by brain MRI. Its usefulness can change recent paradigm of the stroke treatment (Wang et al., 2016; Atchaneeyasakul et al., 2020).

Stroke is treatable, but it is prone to disability after treatment and must be prevented in advance (Park et al., 2018). To grasp the degree of disability caused by stroke, many studies have been conducted to predict stroke prognosis in recent years (Shrestha et al., 2015; Pack et al., 2018; Monteiro et al., 2018).

Park et al. (2018) used the medical examination results of 3,605 patients as input features. For input features, a total of 76 features are extracted as results of medical tests such as 12-lead electrocardiography, chest x-ray, lipid profiles, standard blood tests, and other diagnoses. Among these features, only 19 features were used through feature selection, and these features were input to the Bayesian network to predict the prognosis of stroke.

Monteiro et al. (2018) used several medical examination records, such as CT and MRI, from admission to discharge of 425 patients, as input features. These features are input into machine learning models such as logistic regression, decision tree, support vector machine, random forest, and extreme gradient boosting (XGB) model to predict stroke prognosis.

Previous studies have predicted the prognosis of stroke using various diagnostic results. Since a vast number of tests are required for these prediction methods, a burden of a lot of time and cost may occur from the patient's point of view. In this study, the prognosis of stroke is predicted using only MRI text records created after radiologists analyze MRI images.

Stroke prognosis prediction using MRI texts corresponds to the text classification task. Recently, the text classification task is showing good performance through a convolutional neural network (CNN) which extracts local information of adjacent words, long short-term memory (LSTM), which is good in the sequential data processing, bi-directional LSTM (Bi-LSTM), which has forward-backward LSTM structure, XGB, which is level-wise tree-based learning algorithms of gradient boosting method, and light gradient boosting machine (LGBM), which is leaf-wise tree-based learning algorithms of gradient boosting method

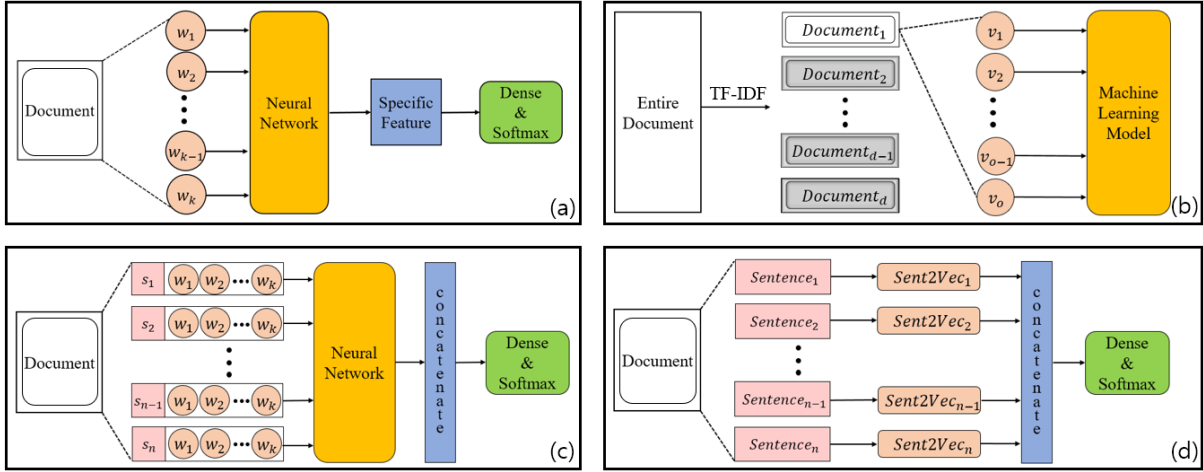


Figure 1: (a). Representation of document-level, (b). Representation of TF-IDF, (c). Representation of sentence-level using word embedding, (d). Representation of sentence-level using sentence embedding

(Kim, 2014; Zhou et al., 2015; Alzamzami et al., 2020).

Few widely known studies have used MRI text recordings to predict stroke prognosis. In this study, rather than improving the performance, the performance difference according to the document representation method is identified and analyzed. In this paper, MRI text is represented at the document level or sentence level.

Document-level representation is an approach that represents the entire document as a vector and considers it as an input and analyzes it. And sentence-level representation is an approach that analyzes the entire document by representing each sentence constituting the document as individual vectors and inputting them.

This paper is composed as follows. Section 2 is about a dataset and describes the MRI text records used in this study. Section 3 describes the architecture of the model corresponding to each representation regarding Figure 1. Section 4 shows experimental results, including the hyperparameters used in the experiment, and compares and analyzes the results of various representation levels. Section 5 describes a conclusion, which outlines the summary and plans for this study.

2 Dataset

MRI text records used in this study are collected from people hospitalized with acute ischemic stroke at Hallym University Chuncheon Sacred

ID	Text of MRI	Label
A	1. focal old petechial hemorrhage in left parietal subcortical WM 2. No diffusion restriction 3. magnetic resonance angiography: No gross abnormal finding.	0
B	1. multiple diffusion restriction in left temporo-parietal subcortical WMs. ; multiple acute infarction. 2. Old small infarct in left parietal cortex.	1

Table 1: Examples of MRI text records

Heart Hospital¹ from February 2010 to October 2019. Patients performed MRI scans several times during hospitalization. However, we only used here for text reports of brain MRI that were examined immediately after hospitalization. Text of MRI in Table 1 means text records of MRI images analyzed by radiologists, and the order of sentences shown here is meaningless. A label is determined according to the score of modified rankin scale² (mRS) which is a clinical outcome measure of the degree of disability. The mRS used in this study is the prognostic score of stroke patients after 3 months, defined as 0-6 points. Patients with scores of 0-2 mRS are grouped as ‘good outcome’, and patients with a score of 3-6 are grouped as ‘poor outcome’ (Powers et al., 2015; Rangaraju et al., 2017). In this study, label 0 is defined as ‘good outcome’ and label 1 as ‘poor outcome’.

The total number of MRI text records collected is 2,071. In MRI text records, the number of MRI

¹ <https://chuncheon.hallym.or.kr>

² https://en.wikipedia.org/wiki/Modified_Rankin_Scale

text records with label 0 is 1,317, and the number of MRI text records with label 1 is 754. In the MRI text record, the average number of sentences is 10, the minimum number of sentences is 1, and the maximum number of sentences is 40.

3 Document Representation

In this study, two document representation methods, document-level and sentence-level, are used to predict the prognosis of stroke.

3.1 Document-Level Representation

In the document-level representation, the document's entire contents are represented like a single sentence and entered into the model. In this representation, word embedding and TF-IDF are used. The model structure using word embedding is shown in (a) of Figure 1, and the model structure using TF-IDF is shown in (b) of Figure 1.

Document-Level using Word Embedding Word vectors extracted through word embedding can be inputs of CNN, multi-filter CNNs, LSTM, or Bi-LSTM, and the structures of these models can be described as follow:

- When CNN is used, the features of adjacent words are entered into global max-pooling to extract the most prominent features. Finally, these features predict the prognosis of stroke through the softmax layer.
- When multi-filter CNNs are used, the features of adjacent words for each filter size are entered into global max-pooling to extract the most prominent features, and then these features are concatenated into one vector. Finally, concatenated one vector predicts the prognosis of stroke using the softmax.
- When LSTM is used, the last hidden state that contains the entire information of the document predicts the prognosis of stroke using the softmax layer.
- When Bi-LSTM is used instead of LSTM, the last hidden states of forward LSTM and backward LSTM are concatenated. The concatenated hidden states predict

the prognosis of stroke using the softmax layer.

Document-Level using TF-IDF The weights of the words extracted from each document through TF-IDF are finally entered into the level-wise tree-based learning algorithm XGB³ or the leaf-wise tree-based learning algorithm LGBM⁴ to predict the prognosis of the stroke.

3.2 Sentence-Level Representation

In the sentence-level representation, a document is divided into sentences and represented with several sentences. Since these divided sentences need to use information about the entire document to predict stroke prognosis, all the sentences about the document are entered into the model at once. In this representation, word embedding and sentence embedding are used. The model structure using word embedding is shown in (c) of Figure 1, and the model structure using sentence embedding is shown in (d) of Figure 1.

Sentence-Level using Word Embedding When word embedding is used, to make information about divided sentences into the document, LSTM is applied to each sentence, then hidden states about each sentence are concatenated into one vector. Finally, the concatenated one vector predicts the prognosis of stroke using the softmax layer. LSTM used in this representation uses the siamese network that shares weights (Mueller and Thyagarajan, 2016). In the siamese network, weights are learned to approximate the prediction values by sharing the weights to several unordered sentences. When using Bi-LSTM instead of LSTM, the last hidden states of forward LSTM and backward LSTM are concatenated, and the method after that is the same as LSTM.

Sentence-Level using Sentence Embedding

When sentence embedding is used, to make information about divided sentences into entire information about the document, sentence vectors about each sentence are concatenated into one vector. Finally, concatenated one vector predicts the prognosis of stroke using the softmax layer.

4 Experiments

In this study, MRI text records remove all non-

³ <https://github.com/dmlc/xgboost>

⁴ <https://github.com/microsoft/LightGBM>

No.	Model	Precision	Recall	F1-Score	Accuracy
a.1	CNN + Dense	0.796	0.7577	0.7689	79.81%
a.2	Multi-Filter CNNs + Dense	0.8045	0.761	0.7733	80.29%
a.3	LSTM + Dense	0.7525	0.7292	0.7329	76.53%
a.4	Bi-LSTM + Dense	0.7643	0.7373	0.7458	77.5%
b.1	XGB	0.7846	0.7468	0.7578	78.85%
b.2	LGBM	0.7682	0.742	0.7506	77.88%
c.1	LSTM + Dense	0.7732	0.7291	0.7404	77.6%
c.2	Bi-LSTM + Dense	0.7196	0.7052	0.7096	73.85%
d	Dense	0.7333	0.7012	0.7089	74.62%

Table 2: Performance comparison according to various representations

English alphabet and stopwords, and all the texts are changed to lower case letters and then experimented. For the experiment, the entire data is divided in the ratio of training set: test set = 9: 1, and the training set is again divided into 5-fold. Among the training set divided into folds, four folds learn a model, and the other fold validates a trained model. The performance of the model is measured by applying the test set to the trained model.

4.1 Hyperparameters

Word embedding uses BioWordVec, a pre-trained FastText model that represents the meaning of words in 200 dimensions (Zhang et al., 2019), and sentence embedding uses BioSentVec, a pre-trained Sent2Vec model that represents the meaning of sentences in 700 dimensions (Chen et al., 2019). These models are pre-trained models for large clinical corpus.

In CNN, the number of filters is 256, the filter size is 3, and the stride is 1. In multi-filter CNNs, all hyperparameters are the same as CNNs, except that the filter sizes are 3, 4, and 5. LSTM is used as the number of units of 100. Bi-LSTM is used as the number of units of 200. And XGB uses the basic hyperparameters provided by the Distributed Machine Learning Community⁵, and LGBM uses the basic hyperparameters provided by Microsoft⁶. Moreover, the loss function uses binary-cross entropy, and the optimizer uses Adam.

4.2 Results

Table 2 shows the performance comparison of various representation methods and shows the experiments' results with various models of each repre-

sentation method. In the first column, **a** is a document-level representation, **b** is the TF-IDF representation, **c** and **d** are sentence-level representations through word embedding and sentence embedding, respectively. Moreover, precision, recall, F1-score, and accuracy are Marco-Averaged scores and mean average values of 5 experiment results.

As a result of the experiment, **a.2**, which considers the relationships of adjacent word representations in the document-level, shows the best performance. And, **b**, which is a method that uses the frequency of words appearing in the documents, was expected to show lower performance than **a**, but it is visible that the performance is higher than **a.3** and **a.4**. With these results, **a.3** and **a.4** consider word orders for each sentence, but due to the characteristic of MRI text records, the order of sentences was irregular, so we can infer that the performance was relatively low. **c** and **d**, which are sentence-level representations that sufficiently consider sentence's information, expected to show better performance than **a** and **b**, but they show low performance. Through this, we can infer that in MRI text records, which do not have the order of sentences, it does not mean much to reconstruct documents into sentences. Based on these analyses, we confirm that the document-level representation method showed better performance in predicting the prognosis of stroke when using MRI text records and that the relationship of adjacent words is more important than the sequential information.

5 Conclusion

In this study, the performance is compared and analyzed according to the document-level and the sentence-level representation methods using MRI

⁵ <https://github.com/dmlc/xgboost/blob/master/doc/parameter.rst>

⁶ <https://github.com/Microsoft/LightGBM/blob/master/docs/Parameters.rst>

text records to predict stroke prognosis. In the document-level representation, the entire document is used as an input value of the model using word embedding and TF-IDF, and in the sentence-level representation, the document is divided into sentences and used as the input value of the model using word embedding and sentence embedding. As a result of the experiment, it is better to consider the information of adjacent words using the document-level representation compared to the sentence-level representation.

Since the number of datasets used in this study is unbalanced, it may adversely affect performance improvements. Therefore, in future studies, we will include data used by other hospitals in the data used by this study to make more exquisite models. Also, based on the results of this study, we will investigate studies that detect a variety of diseases using neural networks based on a document-level representation. Accordingly, we will apply the latest language model and various neural networks to improve performance for stroke prognosis prediction.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C2 006010) and the National Research Fund of Korea (NRF-2019R1G1A1097707).

References

Fatimah Alzamzami, Mohamad Hoda, and Abdulmotaleb El Saddik. 2020. [Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation](https://doi.org/10.1109/access.2020.2997330). *IEEE Access*. <https://doi.org/10.1109/access.2020.2997330>.

Kunakorn Atchaneeyasakul, Ty Shang, Diogo Haussen, Gustavo Ortiz, and Dileep Yavagal. 2019. [Impact of MRI selection on triage of endovascular therapy in acute ischemic stroke: the MRI in Acute Management of Ischemic Stroke \(MIAMIS\) Registry](https://doi.org/10.1159/000490580). *Interventional Neurology*, 8(2-6): 135-143. <https://doi.org/10.1159/000490580>.

Bruce C. V. Campbell, Deidre A. De Silva, Malcolm R. Macleod, Shelagh B. Coutts, Lee H. Schwamm, Stephen M. Davis, and Geoffrey A. Donnan. 2019. [Ischemic stroke](https://doi.org/10.1038/s41572-019-0118-8). *Nature Reviews Disease Primers*, 5(1):70. <https://doi.org/10.1038/s41572-019-0118-8>.

Qingyu Chen, Yifan Peng, and Zhiyong Lu. 2019. [BioSentVec: creating sentence embeddings for bio-](https://arxiv.org/abs/1908.08114)

[medical texts](https://arxiv.org/abs/1908.08114). In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1-5. <https://doi.org/10.1109/ichi.2019.8904728>.

- Merel S. Ekker, Jamie I. Verhoeven, Ilonca Vaartjes, Koen M. van Nieuwenhuizen, Catharina J.M. Klijn, and Frank-Erik de Leeuw. 2019. [Stroke incidence in young adults according to age, subtype, sex, and time trends](https://doi.org/10.1212/wnl.0000000000007533). *Neurology*, 92(21): e2444-e2454. <https://doi.org/10.1212/wnl.0000000000007533>.
- Yoon Kim. 2014. [Convolutional Neural Networks for Sentence Classification](https://www.aclweb.org/anthology/D14-1181/). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746-1751. <https://www.aclweb.org/anthology/D14-1181/>.
- Miguel Monteiro, Ana Catarina Fonseca, Ana Teresa Freitas, Teresa Pinho e Melo, Alexandre P. Francisco, José M. Ferro, Arlindo L. Oliveira. 2018. [Using machine learning to improve the prediction of functional outcome in ischemic stroke patients](https://doi.org/10.1109/tcbb.2018.2811471). *IEEE/ACM transactions on computational biology and bioinformatics*, 15(6): 1953-1959. <https://doi.org/10.1109/tcbb.2018.2811471>.
- Jonas Mueller and Aditya Thyagarajan. 2016. [Siamese Recurrent Architectures for Learning Sentence Similarity](https://dl.acm.org/doi/10.5555/3016100.3016291). In *thirtieth AAAI conference on artificial intelligence*. <https://dl.acm.org/doi/10.5555/3016100.3016291>.
- Eunjeong Park, Hyuk-jae Chang and Hyo Suk Nam. 2018. [A bayesian network model for predicting post-stroke outcomes with available risk factors](https://doi.org/10.3389/fneur.2018.00699). *Frontiers in neurology*, 9. <https://doi.org/10.3389/fneur.2018.00699>.
- William J. Powers, Colin P. Derdeyn, José Biller, Christopher S. Coffey, Brian L. Hoh, Edward C. Jauch, Karen C. Johnston, S. Claiborne Johnston, Alexander A. Khalessi, Chelsea S. Kidwell, James F. Meschia, Bruce Ovbiagele, and Dileep R. Yavagal. 2015. [2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association](https://doi.org/10.1161/str.0000000000000074). *Stroke*, 46(10):3020-3035. <https://doi.org/10.1161/str.0000000000000074>.
- Srikant Rangaraju, Diogo Haussen, Raul G. Nogueira, Fadi Nahab, and Michael Frankel. 2017. [Comparison of 3-Month Stroke Disability and Quality of Life across Modified Rankin Scale Categories](https://doi.org/10.1159/000452634). *Interventional neurology*, 6(1-2):36-41. <https://doi.org/10.1159/000452634>.
- Ralph L. Sacco, Scott E. Kasner, Joseph P. Broderick, Louis R. Caplan, J.J. (Buddy) Connors, Antonio Culebras, Mitchell S.V. Elkind, Mary G. George, Allen D. Hamdan, Randall T. Higashida, Brian L. Hoh, L.

- Scott Janis, Carlos S. Kase, Dawn O. Kleindorfer, Jin-Moo Lee, Michael E. Moseley, Eric D. Peterson, Tanya N. Turan, Amy L. Valderrama, and Harry V. Vinters. 2013. **An Updated Definition of Stroke for the 21st Century.** *Stroke*, 44(7): 2064-2089. <https://doi.org/10.1161/str.0b013e318296aeca>.
- Shakti Shrestha, Ramesh Sharma Poudel, Dipendra Khatiwada, and Lekhjung Thapa. 2015. **Stroke subtype, age, and baseline NIHSS score predict ischemic stroke outcomes at 3 months: a preliminary study from Central Nepal.** *Journal of multidisciplinary healthcare*. 8: pages 443-448. <https://doi.org/10.2147/jmdh.s90554>.
- Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. 2016. **Accelerating magnetic resonance imaging via deep learning.** In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514-517. <https://doi.org/10.1109/isbi.2016.7493320>.
- Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. 2019. **BioWordVec, improving biomedical word embeddings with subword information and MeSH.** *Scientific data*, 6(1): 1-9. <https://doi.org/10.1038/s41597-019-0055-0>.
- Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis C.M. Lau. 2015. **A C-LSTM neural network for text classification.** *arXiv preprint arXiv:1511.08630*. <https://arxiv.org/abs/1511.08630>.