International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 1
ISSN: 2088-5334

Two-Stream Network for Korean Natural Language Understanding

Hwang Kim [a], Jihyeon Lee [b], Ho-Young Kwak [b,*]

[a] Department of Computer Engineering, Graduate School, Jeju National University, 102 Jejudaehakro, Jeju, 63243, Republic of Korea
[b] Department of Computer Engineering, Jeju National University, 102 Jejudaehakro, Jeju, 63243, Republic of Korea
Corresponding author: [*]kwak@jejunu.ac.kr

Abstract—**This study pioneers a dual-stream network architecture tailored for Korean Natural Language Understanding (NLU), focusing on enhancing comprehension by distinct processing of syntactic and semantic aspects. The hypothesis is that this bifurcation can lead to a more nuanced and accurate understanding of the Korean language, which often presents unique syntactic and semantic challenges not fully addressed by generalized models. The validation of this novel architecture employs the Korean Natural Language Inference (koNLI) and Korean Semantic Textual Similarity (koSTS) datasets. By evaluating the model's performance on these datasets, the study aims to determine its efficacy in accurately parsing and interpreting Korean text's syntactic structure and semantic meaning. Preliminary results from this research are promising. They indicate that the dual-stream approach significantly enhances the model's capability to understand and interpret complex Korean sentences. This improvement is crucial in NLU, especially for language-specific applications. The implications of this study are far-reaching. The methodology and findings could pave the way for more sophisticated NLU applications tailored to the Korean language, such as advanced sentiment analysis, nuanced text summarization, and more effective conversational AI systems. Moreover, this research contributes significantly to the broader field of NLU by underscoring the importance and efficacy of developing language-specific models, moving beyond the one-size-fits-all approach of general language models. Thus, this study is a testament to the potential of specialized approaches in language understanding technologies.**

Keywords—**Two-stream networks; Korean Natural Language Inference (koNLI); Korean Semantic Textual Similarity (koSTS); convolutional neural networks; long short-term memory; syntax and semantics parallel processing.**

Manuscript received 17 May. 2023; revised 29 Oct. 2023; accepted 12 Dec. 2023. Date of publication 29 Feb. 2024.

I. INTRODUCTION

Natural Language Processing (NLP) and Natural Language Understanding (NLU) are pivotal areas in the landscape of artificial intelligence research. While recent developments have mainly centered around enhancing and generalizing language models [1], yielding remarkable performances across a multitude of languages and applications [2]–[7], this broad-spectrum approach often neglects the unique linguistic nuances of specific languages [8]. This oversight is particularly evident in languages like Korean, which possess intricate syntactic and semantic structures not adequately addressed by general models. Recognizing this gap, our study introduces a specialized approach for Korean NLU. Our research introduces a novel Two-Stream network architecture designed to analyze syntactic and semantic aspects of the Korean language separately. This architectural innovation is based on the premise that a dedicated focus on syntactic and semantic elements can lead to a more profound understanding of languages with complex structures [8]–[12]. The hypothesis driving this study is that such an architecture can more effectively parse and interpret the intricate syntax and semantics inherent in the Korean language.to validate our approach, we employ the KoNLI and KoSTS datasets, chosen for their comprehensive coverage of the Korean language's linguistic complexities. These datasets include diverse linguistic expressions, ranging from simple phrases to intricate sentences, making them ideal for thoroughly evaluating our model. Through this experimentation, our goal is to provide novel insights into the domain of Korean NLU, potentially revolutionizing how this language is processed and understood in AI applications. Our study also aims to explore the broader implications of this architecture. Focusing on Korean, a language with distinct linguistic challenges, we aim to highlight the potential benefits of developing language-specific models in NLU. This approach could lead to more effective AI applications in language processing, including but not limited to improved machine translation, more accurate sentiment analysis, and advanced conversational agents that can understand and interact in Korean with greater proficiency. In addition to the practical applications, our research contributes to the theoretical understanding of NLU. By dissecting the complexities of Korean, we aim to provide a template that can be adapted and applied to other languages with unique syntactic and semantic structures. This could pave the way for more customized and effective NLU models, catering to the specific needs of different languages, thus enriching the field of NLP as a whole. In conclusion, our research stands at the intersection of linguistic specificity and AI innovation. By merging the intricate details of Korean syntax and semantics with advanced NLP techniques, we are addressing the immediate need for better Korean language processing and setting a precedent for future research in language specific NLU approaches. This study is poised to make significant contributions to the field regarding practical applications and theoretical advancements.

II. Materials and Method

A. Materials/Related Works

1) Two-Stream Network:

The Two-Stream network [13]–[21], initially developed for tasks such as recognizing motion in videos within the computer vision field, is inspired by the human visual perception system. It consists of two distinct paths: ventral (what) and dorsal (where) streams. The ventral stream recognizes "what" is being seen, contributing to shape representation, while the dorsal stream process the spatial locations of objects concerning the observer, which is crucial for motion understanding. In the context of computer vision, these two streams represent separate neural networks handling different types of information.

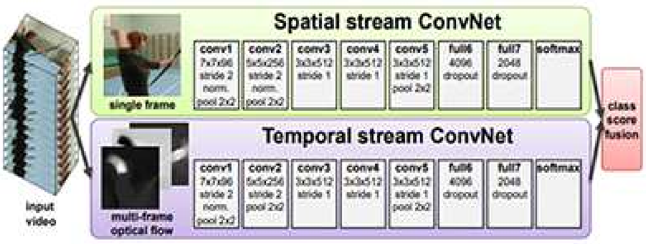

Fig. 1 Two-Stream architecture for video classification

- *Spatial Stream*: Similar to the ventral stream, processes static frames in videos to understand spatial features of the content. This lets the network learn about objects and shapes within individual video frames.
- *Temporal Stream*: Reflects the dorsal stream, dealing with the temporal aspects of videos. It processes frame sequences to learn about movement-related information, aiding in understanding how objects in the video change over time.

Combining these two streams is essential for effectively recognizing and analyzing complex patterns related to video actions, considered a significant innovation in computer vision.

2) koSTS and koNLI

Korean Sentence Textual Similarity (koSTS) is a dataset for evaluating the semantic similarity between Korean sentences. This dataset is crucial for Natural Language Understanding (NLU), specifically for Semantic Textual Similarity (STS) tasks in the Korean language. KoSTS includes various sentence pairs in Korean and their similarity scores, serving as a valuable resource for evaluating the performance of NLU systems targeting the Korean language.

Korean Natural Language Inference (koNLI) is a Korean Natural Language Inference dataset that analyzes the logical relationships between two sentences. This dataset determines whether one sentence (hypothesis) is true, false, or neutral concerning another sentence (premise). KoNLI is also a significant dataset for evaluating and improving the performance of Korean NLU systems.

3) Deep Networks and Regularization

With the rise of deep learning, various neural architectures like Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM) [22]–[24], and Convolutional Neural Networks (CNN) [25]–[27], have been employed in NLU tasks [28]–[32]. For instance, Bowman et al. [33] introduced a model using LSTM for sentence encoding. Techniques like dropout and L2 regularization have become standards for training robust neural networks [34]. This work presents a Two-Stream network for NLU tasks, explicitly applying the strengths of CNN for syntax processing and LSTM for semantic understanding to address performance gaps observed in previous models.

4) Theoretical Background

The two-stream network originated as a concept in computer vision, developed to parallelly process two different types of information - spatial and temporal - in visual information processing. This approach involves analyzing both static and dynamic aspects of images simultaneously through two separate streams, allowing for a more adequate understanding of complex visual scenes. This concept of a two-stream network can also be highly applicable to natural language understanding, such as in Korean. While the CNN-LSTM integrated model developed by Park and Kim [35] showed effectiveness by serially connecting the two models, the two-stream network in this study performs semantic and syntactic text analysis in parallel, enabling an adequate interpretation and understanding of the intricate structure of language. In this context, the semantic stream focuses on grasping the meaning of sentences or phrases, while the syntactic stream analyzes the relationships between the structure and components of sentences.

> (*Semantic analysis involves understanding a sentence or utterance's direct or abstract meaning. This includes comprehending how combinations of words go beyond simple concatenation to form specific meanings. For example, the sentence "It's raining" provides direct information about the weather.*
>
> *Syntactic analysis involves analyzing the structural relationships between words or phrases in a sentence. It focuses on understanding the roles of each word or phrase and the overall structure of the sentence. For instance, in the sentence "It's raining," "It" serves as the subject, and "raining" functions as the predicate.*)

The Two-Stream network integrates these two analyses in a parallel and unified manner, allowing for a more profound understanding of Korean text.

B. Methods

1) Model Architecture:

The model proposed in this research represents a significant leap in computational linguistics, merging the robust capabilities of Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). This integration caters to a holistic approach in text analysis by dissecting and interpreting language's semantic and syntactic dimensions.

The LSTM network is meticulously engineered to unravel the semantic intricacies embedded within textual data. Semantics, in the realm of language processing, involves a complex array of tasks, such as discerning context, unraveling implied meanings, and appreciating the subtleties and nuances inherent in language. The quintessential design of LSTMs equips them with the capacity to maintain a sustained memory, relating information across lengthy text sequences. This characteristic is indispensable for grasping the convoluted sentence structures and rich meanings that often elude more transient models.

On the contrary, the CNN is harnessed for its proficiency in syntactic dissection. Syntax, the scaffold of language structure, is concerned with the orderly arrangement of words and phrases to construct coherent sentences. CNNs excel in pattern recognition, a trait that translates into identifying grammatical structures and syntactic patterns when transposed to the domain of Natural Language Processing (NLP). The convolutional layers of CNNs are adept at processing textual data that uncovers these underlying structural patterns with remarkable efficiency. Within the architecture of the proposed model, these networks function as two independent but complementary streams. This dual-stream approach is a systematic choice, ensuring that each aspect of language is analyzed with an unwavering focus. The LSTM network plunges into the depths of meaning, interpreting and connecting themes across the textual fabric. Concurrently, the CNN traverses the surface structure of the text, delineating its grammatical contours and parsing its syntactic architecture.

The culmination of this approach is the fusion of the two streams. Once generated independently, the outputs are synthesized through a carefully calibrated weighting process. This process is not merely additive but integrative, with weights learned adaptively based on each stream's contributions to the text's overall understanding. The goal is to achieve a synergistic effect where the combination is greater than the sum of its parts, yielding a richer and more nuanced interpretation of the text. This comprehensive output captures the essence of the text in a multidimensional space—encompassing its semantic content and syntactic form. The integration promises an analysis that is thorough and exceptionally accurate. The combined output is anticipated to surmount the limitations of individual analyses, providing insights into the text that are both deep in meaning and precise in structure.

The implications of such a model are profound. By achieving a more detailed understanding of the text, we pave the way for advancements in various applications of NLP. These range from improving the naturalness of conversational AI and enhancing the accuracy of machine translation to refining sentiment analysis and expanding the capabilities of information extraction systems. Moreover, this model sets a foundation for future research where the convergence of semantic and syntactic analysis could lead to breakthroughs in machine understanding of human languages.

2) Data Processing:

For this study, the datasets selected are tailored explicitly for Sentence Textual Similarity (STS) and Natural Language Inference (NLI) tasks [36]. These tasks are central to evaluating the model's proficiency in understanding and interpreting language. STS involves assessing the degree of similarity between pairs of sentences, which tests the model's ability to comprehend and compare the meaning of different texts. NLI, conversely, involves determining the relationship between a pair of sentences, such as whether one sentence logically follows from the other, contradicts it, or is unrelated.

These tasks present a comprehensive challenge for the model, testing its ability to interpret language in various contexts. The datasets consist of pairs of sentences, each accompanied by corresponding labels. These labels are either similarity scores, in the case of STS tasks, or inference relationships for NLI tasks. Including these labels is crucial as they provide the ground truth against which the model's predictions are compared. The preprocessing stage of the data is critical in preparing the text for analysis by the model. The first step involves handling missing values in the text data, ensuring the dataset is complete and reliable for training and testing the model. Following this, the text is tokenized using a tokenizer. Tokenization converts the text into smaller units, such as words or phrases.

This step is essential for transforming the raw text into a format the model can process. After tokenization, the text undergoes padding. Padding involves transforming the tokenized text into sequences of a fixed length, ensuring uniformity in the input data. This uniformity is necessary for efficient processing by the neural networks. The processed data, now in a structured and consistent format, serves as the input for the model. The data's journey through the model, from input to the combined analysis by the LSTM and CNN streams, is depicted in Table 1.

TABLE I
DATASETS

	STS	NLI
Training data	sts-train.tsv	multinli.train.ko.tsv, snli_1.0_train.ko.tsv
Verifying data	sts-dev.tsv	xnli.dev.ko.tsv
Testing data	sts-test.tsv	xnli.test.ko.tsv

This table visually represents the data processing and analysis workflow, illustrating how the model transforms and interprets the text. In the following sections, we will delve into the specifics of the LSTM and CNN architectures, discuss the criteria for dataset selection, and outline the detailed steps involved in the data preprocessing phase. We will also explore the implementation of the dual-stream approach and its impact on the model's performance in STS and NLI tasks.

3) Model Training and Evaluation:

The model's training is a crucial phase where the theoretical architecture is implemented and tested. We use the Adam optimization algorithm [37], a popular choice in machine learning for its efficiency in handling large datasets and adaptability in adjusting learning rates. The choice of the loss function, either Mean Squared Error or Sparse Categorical Cross-entropy, is determined based on the specific task at hand. Mean Squared Error is typically used for regression tasks, where the goal is to predict continuous values. At the same time, Sparse Categorical Cross-entropy is suited for classification tasks, where the objective is to categorize data into distinct classes.

Training is conducted separately for each dataset to ensure that the model is finely tuned to the nuances of each task. This isolation of training allows for a more focused and tailored learning process, providing that the model optimally learns the specific features of each dataset. The model's performance is constantly monitored and evaluated after each epoch using a validation dataset during training. This evaluation provides insights into the model's learning progress, helping to identify and rectify any issues early in the training process.

After the training phase, the model's general performance is rigorously assessed using a test dataset. This dataset is separate from the training and validation datasets and objectively evaluates the model's effectiveness. Using a test dataset is crucial in machine learning, as it offers a realistic measure of how the model will perform in real-world scenarios.

Several key metrics are monitored throughout the training and evaluation process to assess the model's performance. These metrics include accuracy, precision, recall, and F1 score, each providing different insights into the model's capabilities. Accuracy measures the overall correctness of the model's predictions, while precision and recall assess the model's ability to identify positive cases and avoid false positives, respectively, correctly. The F1 score balances precision and recall, offering a comprehensive view of the model's performance.

The model's training and evaluation process also includes fine-tuning of hyperparameters, such as learning rate, batch size, and the number of epochs. Hyperparameter tuning is essential for optimizing the model's performance, ensuring it learns effectively and efficiently.

```
VOCAB_SIZE = 10000
MAX_LEN = 100
EMBEDDING_DIM = 100
HIDDEN_DIM = 50
LEARNING_RATE = 0.0001
```

Fig. 2 Hyperparameters

In conclusion, the model training and evaluation phase is critical in developing our dual-stream network architecture. We aim to create a robust and effective model for Korean NLU through rigorous training, constant monitoring, and comprehensive evaluation. This phase tests the model's capabilities and provides valuable insights that guide further refinements and improvements, ultimately contributing to the advancement of NLU technologies.

4) Weighting and Integration:

Integrating outputs from Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) streams is a pivotal aspect of our study's methodology. LSTM and CNN are deep learning models but fundamentally differ in processing and interpreting data. The LSTM, a type of recurrent neural network, is well-suited for sequence prediction problems because it can store past information. This is particularly useful for understanding textual data's context and semantic nuances. On the other hand, CNNs, commonly used in image processing, are adept at extracting hierarchical features. When applied to text, CNNs can capture syntactic structures and patterns effectively.

In our approach, the outputs of the LSTM and CNN streams are generated independently to capitalize on their respective strengths. The LSTM stream processes the textual data to comprehend the underlying semantic content - the meaning conveyed by word sequences and the overall context of the text. Meanwhile, the CNN stream focuses on syntactic analysis, identifying patterns and structures in sentence construction. This dual approach ensures a comprehensive text analysis, covering both its semantic meaning and syntactic form.

The subsequent stage involves the combination of these independent outputs. This is achieved through a sophisticated weighting process. In this process, weights are learned for each stream's output, which allows for a dynamic adjustment of the influence of each model based on the specific characteristics of the text being analyzed. Learning these weights is a critical part of our methodology, as it determines how the outputs from the LSTM and CNN models are merged to form a cohesive analysis. Integrating the LSTM and CNN outputs through this weighting process is not just a mere combination of results. It is a deliberate strategy to balance the semantic depth provided by the LSTM with the syntactic precision offered by the CNN. By fine-tuning the weights, we can control the emphasis placed on the text's semantic or syntactic aspects, depending on what is more relevant or significant in a given context. This adaptability is crucial for handling various text types and structures, ensuring our analysis remains robust across different textual formats and styles. Furthermore, this weighting and integration technique offers several advantages. Firstly, it enhances the accuracy of text analysis by ensuring that both semantic and syntactic aspects are considered. This is particularly important in complex text analysis tasks where neglecting either element could lead to incomplete or skewed interpretations. Secondly, this method allows for greater flexibility in handling diverse text types. Whether dealing with straightforward, syntactically-driven texts or complex, semantically-rich narratives, our approach can adaptively adjust to capture the essential characteristics of the text effectively.

III. Results and Discussion

In our comprehensive experiments, we scrutinized the performance of the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models alongside a Combined model integrating the Two-Stream network approach on tasks of Korean Natural Language Understanding (NLU). These models underwent rigorous training and evaluation processes utilizing the Korean Natural Language Inference (KoNLI) and Korean Semantic Textual Similarity (KoSTS) datasets to ascertain their efficacy in processing and understanding the Korean text. The results, meticulously detailed in Tables 2 and 3, delineate the comparative performance across diverse evaluation metrics, providing crucial insights into the strengths and weaknesses of each model type.

For the KoSTS dataset, which is designed to gauge textual similarity, the CNN model exhibited a notable level of proficiency in capturing the syntactic features of the Korean language. The intricate structure of Korean sentences, with their agglutinative morphology, presents a unique challenge that CNN models tackle by discerning patterns in sentence construction. The LSTM model, known for capturing long-term dependencies, proved more adept at grasping the semantic context embedded within Korean text. Its architecture, designed to mitigate the vanishing gradient problem, allows for effective learning of context over longer stretches of text, thus excelling in semantic understanding. However, the Combined model, which leverages syntactic and semantic streams, demonstrated an enhanced ability to process textual similarity.

By integrating the outputs of the LSTM and CNN models through a sophisticated weighting mechanism, the Combined model outperformed the individual models in most metrics. This integration allows the model to balance and optimize the unique strengths of each approach, resulting in a more comprehensive understanding of the text.

Similarly, the KoNLI dataset, aimed at evaluating language inference, served as a robust benchmark to test the models' capabilities in deducing logical relationships between sentence pairs. In this context, the LSTM model's superior performance in understanding context was again evident, outshining CNN in most metrics. Yet, the Combined model excelled, reinforcing the strength of integrating syntactic and semantic analyses. The robustness of the Combined model in KoNLI tasks suggests that it is better equipped to handle the complexities involved in discerning the implied meanings and relationships within Korean text, an essential component of language inference.

The experiment's findings highlight the significance of a dual-analytical approach in NLU tasks for the Korean language. Applying separate streams to evaluate distinct linguistic features, followed by a nuanced fusion of these analytical paths, offers a more robust understanding of the text. The success of this approach in our experiments lays a strong foundation for future explorations into language-specific processing models and sets a precedent for more nuanced NLU systems.

In-depth analysis of the experimental results also provides invaluable insights into the practical applications of such models in real-world scenarios. These models can potentially revolutionize automated translation services by improving the accuracy of translations between Korean and other languages. They can enhance advanced dialogue systems, allowing for more natural and contextually aware interactions in Korean. Moreover, they can be instrumental in developing intelligent text analysis tools that provide insights into consumer behavior, public opinion, and social trends based on Korean text data.

The implications of this research extend beyond academic interest, indicating a future where NLU systems can be finely tuned to accommodate the intricate nature of Korean semantics and syntax. We can aim for greater accuracy and efficiency in language processing tasks by further refining these models and their integration mechanisms. The nuanced understanding of the Korean language that these models facilitate could lead to the development of more sophisticated NLU systems capable of interpreting not only the text's literal meaning but also its subtler, culturally nuanced implications. Our study also underscores the importance of specialized datasets like KoNLI and KoSTS.

The development of such resources is crucial for advancing NLU research, particularly for languages that have been underrepresented in the field. By building and utilizing datasets tailored to specific languages, researchers can better train and test NLU models, ensuring that these systems are as effective and inclusive as possible. In conclusion, the research presented in this paper represents a significant step forward in Korean NLU.

The superior performance of the Combined model across our experiments suggests that integrating syntactic and semantic analyses is beneficial and perhaps essential for developing effective NLU systems for complex languages like Korean. As we continue to refine these models and explore their applications, we pave the way for a future in which NLU systems can process language with the same depth and nuance as human communicators.

TABLE II
KOSTS RESULTS

koSTS	Test loss	Test MAE
CNN Model	2.38803554	1.280205607
LSTM Model	2.28155065	1.257431030
Combined Model	2.20477033	1.243777394

TABLE III
KONLI RESULTS

koNLI	Test loss	Test accuracy
CNN Model	1.16071367	0.524124146
LSTM Model	0.97025299	0.543543518
Combined Model	0.92268127	0.556356370

IV. CONCLUSION

In the culmination of this paper, we have conducted an exhaustive exploration into applying a Two-Stream network architecture for Korean Natural Language Understanding, seamlessly integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Our research embarked on a comprehensive journey to unravel the intricacies inherent in the Korean language, explicitly focusing on elevating the performance of NLU systems by addressing the syntactic and semantic nuances unique to Korean.

The outcomes of our experiments, meticulously conducted through rigorous training and testing on the KoNLI and KoSTS datasets, underscore the effectiveness of the Two-Stream network architecture. Notably, the CNN models excelled in decoding syntactic structures, while the LSTM models exhibited superior performance in grasping semantic comprehension. Remarkably, the Combined model, synergizing the strengths of both CNN and LSTM networks, emerged as the most adept in comprehending and interpreting the complexities of the Korean language, as substantiated by the detailed metrics in Tables 2 and 3.

This research makes a noteworthy contribution to the NLU field, reinforcing the assertion that language-specific models can outshine general language models in specific contexts. Furthermore, it paves the way for future investigations into specialized models for languages with distinctive linguistic characteristics. Subsequent research endeavors could expand upon our work by exploring the adaptability of the Two-Stream network to other NLU tasks and incorporating more intricate linguistic features into the analysis.

Beyond the academic realm, the practical implications of this study are significant, with potential applications spanning various domains requiring Korean language processing. The insights derived from this study can enhance systems involving machine translation, content analysis, and conversational AI, fostering more natural and efficient human-computer interactions. In essence, our research is a testament to the promise of language-specific NLU models. It underscores the importance of tailoring NLP technologies to the nuanced intricacies of individual languages. As we persist in refining these models, we progress closer to realizing AI systems endowed with a deeper and more intuitive understanding of human language.

ACKNOWLEDGMENT

This study was supported by a research grant from Jeju National University in 2023.

REFERENCES

[1] SKTBrain, KoBERT. 2021. [Online] Available: https://github.com/SKTBrain/KoBERT.

[2] J. Y. Choi, H. S. Rim, "E-commerce databased Sentiment Analysis Model Implementation using Natural Language Processing Model," *Journal of Korea Convergence Society*, Vol. 11, No. 13, pp.33-39, 2020, doi:10.15207/JKCS.2020.11.11.033.

[3] Muhammad Abdul-Mageed, AbdelRahim Elmadany, El Moatez Billah Nagoudi, "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic,", *arXiv* 2020: 2101.01785 [cs.CL], doi:10.48550/arXiv.2101.01785.

[4] Zhiheng Huang, Peng Xu, Davis Liang, Ajay Mishra, Bing Xiang, "TRANS-BLSTM: Transformer with Bidirectional LSTM for Language Understanding," Mar, *arXiv,* 2020:2003.07000 [cs.CL], doi:10.48550/arXiv.2003.07000.

[5] A. Aggarwal, A. Chauhan, D. Kumar, M. Mittal, and S. Verma, "Classification of Fake News by Fine-tuning Deep Bidirectional Transformers based Language Model," *ICST Transactions on Scalable Information Systems*, p. 163973, Jul. 2018, doi: 10.4108/eai.13-7-2018.163973.

[6] Deepa, Ms D. "Bidirectional Encoder Representations from Transformers (BERT) Language Model for Sentiment Analysis task: Review," *Turkish Journal of Computer and Mathematics Education*, Vol.12, No. 7, pp. 1708-1721, Apr. 2021.

[7] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D Manning, Percy S. Liang, J. Leskovec, "Deep Bidirectional Language-Knowledge Graph Pretraining," *Advances in Neural Information Processing Systems,* 2022.

[8] K. Seok, D. Hwang, and J. H. Kim, "Syntactic and se mantic analysis for the discrimination of similar documents," *Journal of Korean Content Association*, Vol. 14, No. 3, pp. 40–51, 2014

[9] Soukaina Fatimi, Chama EL Saili, Larbi Alaoui, "A Framework for Semantic Text Clustering," *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 6, 2020, pp.451-459.

[10] K. Jacksi, R. Kh. Ibrahim, S. R. M. Zeebaree, R. R. Zebari, and M. A. M. Sadeeq, "Clustering Documents based on Semantic Similarity using HAC and K-Mean Algorithms," *2020 International Conference on Advanced Science and Engineering (ICOASE)*, Dec. 2020, doi: 10.1109/icoase51841.2020.9436570.

[11] C. Little, D. Mclean, K. Crockett, and B. Edmonds, "A Semantic and Syntactic Similarity Measure for Political Tweets," *IEEE Access*, vol. 8, pp. 154095–154113, 2020, doi: 10.1109/access.2020.3017797.

[12] A. M. Rinaldi, C. Russo, and C. Tommasino, "A semantic approach for document classification using deep neural networks and multimedia knowledge graph," *Expert Systems with Applications*, vol. 169, p. 114320, May 2021, doi: 10.1016/j.eswa.2020.114320.

[13] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in neural information processing systems*, pp. 568– 576, 2014.

[14] S. H. Kim, J. H. Kim, "Two-Stream Fall Detection using Temporal and Spatial Contexts," *Conference on KCC*, pp. 421 – 423, Dec. 2020

[15] D. Zhu, B. Du, and L. Zhang, "Two-Stream Convolutional Networks for Hyperspectral Target Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, pp. 6907–6921, Aug. 2021, doi: 10.1109/tgrs.2020.3031902.

[16] Z.-M. Wang, M.-H. Li, and G.-S. Xia, "Conditional Generative ConvNets for Exemplar-Based Texture Synthesis," *IEEE Transactions on Image Processing*, vol. 30, pp. 2461–2475, 2021, doi: 10.1109/tip.2021.3052075.

[17] Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie LIU, Brian Mak, "Two-Stream Network for Sign Language Recognition and Translation," *Advances in Neural Information Processing Systems* 2022.

[18] S. Arora et al., "ESPnet-SLU: Advancing Spoken Language Understanding Through ESPnet," ICASSP 2022 - *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, doi: 10.1109/icassp43922.2022.9747674.

[19] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, doi: 10.18653/v1/2021.acl-long.494.

[20] P. Zhang, R. Zhao, B. Yang, Y. Li, and Z. Yang, "Integrated Syntactic and Semantic Tree for Targeted Sentiment Classification Using Dual-Channel Graph Convolutional Network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1109–1124, 2024, doi: 10.1109/taslp.2024.3350877.

[21] T. Liu, Y. Hu, B. Wang, Y. Sun, J. Gao, and B. Yin, "Hierarchical Graph Convolutional Networks for Structured Long Document Classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 8071–8085, Oct. 2023, doi: 10.1109/tnnls.2022.3185295.

[22] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Procedia CIRP*, vol. 99, pp. 650–655, 2021, doi: 10.1016/j.procir.2021.03.088.

[23] Z. Su and J. Jiang, "Hierarchical Gated Recurrent Unit with Semantic Attention for Event Prediction," *Future Internet*, vol. 12, no. 2, p. 39, Feb. 2020, doi: 10.3390/fi12020039.

[24] H. Fan, M. Jiang, L. Xu, H. Zhu, J. Cheng, and J. Jiang, "Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation," *Water*, vol. 12, no. 1, p. 175, Jan. 2020, doi: 10.3390/w12010175.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[26] S.-H. Kim and Y.-H. Park, "Adaptive Convolutional Neural Network for Text-Independent Speaker Recognition," Interspeech 2021, Aug. 2021, doi: 10.21437/interspeech.2021-65.

[27] Park, S.U, "Analysis of the Status of Natural Language Processing Technology Based on Deep Learning," *The Korea Journal of Big Data*, Vol. 6 No. 1, pp.63-81. doi:10.36498/KBIGDT.2021.6.1.63.

[28] Y. Belinkov and J. Glass, "Analysis Methods in Neural Language Processing: A Survey," *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 49–72, Apr. 2019, doi: 10.1162/tacl_a_00254.

[29] D. W. Otter, J. R. Medina, and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/tnnls.2020.2979670.

[30] Amirsina Torfi, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, Edward A. Fox, "Natural Language Processing Advancements By Deep Learning: A Survey," *arXiv*:2003.01200. 2020 doi:10.48550/arXiv. 2003.01200.

[31] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, Jan. 2022, doi: 10.1016/j.neucom.2021.05.103.

[32] A. S. Pillai and R. Tedesco, “Introduction to Machine Learning, Deep Learning, and Natural Language Processing,” *Machine Learning and Deep Learning in Natural Language Processing*, pp. 3–14, Aug. 2023, doi: 10.1201/9781003296126-2.

[33] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," *arXiv* preprint arXiv:1508.05326, 2015.

[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929– 1958, 2014. [Online] Available: http://jmlr.org/papers/v15/srivastava14a.html.

[35] H. Y. Park & K. J. Kim. "Sentiment Analysis of Movie Review Using Integrated CNN-LSTM Model," *Journal of Intelligence and Information Systems,* Vol. 25(4), pp 141-154. 2019.

[36] J. Ham, Y. J. Choe, K. Park, I. Choi, and H. Soh, "Kornli and korsts: New benchmark datasets for korean natural language understanding," *arXiv* preprint arXiv:2004.03289, 2020.

[37] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv* preprint arXiv:1412.6980, 2014.