All Sciences Proceedings

5th International Conference on Applied Engineering and Natural Sciences

July 10-12, 2023 : Konya, Turkey

All Sciences Proceedings
http://as-proceeding.com/

https://www.icaens.com/

ASSESSING THE EFFICACY OF LSTM, TRANSFORMER, AND RNN ARCHITECTURES IN TEXT SUMMARIZATION

Seda BAYAT*, Gultekin ISIK [2]

[1]*Mechatronic Engineering, Igdir University, Turkey*
[2] *Computer Engineering, Igdir University, Turkey*

**(bayatseda@gmail.com) Email of the corresponding author*

Abstract – The need for efficient and effective techniques for automatic text summarization has become increasingly critical with the exponential growth of textual data in different domains. Summarizing long texts into short summaries facilitates a quick understanding of the key information contained in the documents. In this paper, we evaluate various architectures for automatic text summarization using the TEDx dataset, a valuable resource consisting of a large collection of TED talks with rich and informative speech transcripts. Our research focuses on evaluating the performance of Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN) and Transformer architectures for automatic text summarization. We measure the accuracy of each model by comparing the generated summaries with human-written summaries. The findings show that the Transformer model achieves the highest accuracy, followed closely by the GRU model. However, LSTM, RNN exhibit relatively lower accuracies. We also investigate the trade-off between accuracy and conciseness in summarization. Our study reveals that the Transformer model succeeds in producing accurate and concise summaries, albeit at a higher computational cost. On the other hand, the GRU model strikes a desirable balance between accuracy and conciseness, making it a suitable choice. Overall, this research provides valuable insights into the effectiveness of different architectures for automatic text summarization and highlights the superiority of the Transformer and GRU models in this area.

Keywords – Automatic Text Summarization, LSTM, GRU, RNN, Transformer

I. INTRODUCTION

Automatic text summarization has emerged as a prominent research area in recent years, posing significant challenges. The objective of text summarization is to generate a condensed version of a text document while preserving its essential information[1]. This task is demanding because it necessitates the model's comprehension of the text's meaning and its ability to identify the most pertinent sections.

Several text summarization architectures have been proposed in the literature, each with its own strengths and weaknesses [2]. Some of the most common architectures are Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN) and Transformer. LSTM and GRU are recurrent neural networks that can learn long-term dependencies in text [3]. However, GRU is generally considered to be more efficient than LSTM, making it a more suitable choice for large-scale text summarization tasks. RNN is a simpler architecture than LSTM or GRU, but less efficient at learning long-term dependencies [4]. Transformer is a newer architecture based on attention mechanisms. Attention mechanisms allow the model to focus on specific parts of the text when generating a summary, making it more effective than RNN-based models.

The effectiveness of different text summarization architectures has been evaluated on various

datasets. In general, Transformer-based models were found to outperform other architectures. However, the choice of architecture also depends on the specific task and dataset. For example, GRU-based models may be more suitable for tasks that require short summaries, while Transformer-based models may be more suitable for tasks that require accurate summaries.

In this paper, we provide a comprehensive evaluation of different text summarization architectures on the TEDx dataset. The TEDx dataset consists of a large collection of TED talks, providing a realistic and challenging benchmark for evaluating text summarization models. Our results show that the Transformer model achieves the highest accuracy, followed by the GRU model, LSTM model, RNN model. These results show that the Transformer model is the most suitable architecture for automatic text summarization.

We also analyzed the ability of different architectures to produce accurate and concise summaries. Our results show that the Transformer model achieves the highest accuracy, but also produces the least concise summaries. The GRU model strikes a balance between accuracy and conciseness, producing concise and accurate summaries. The LSTM model can also produce accurate summaries, although not as concise as the GRU model. The RNN model achieves accurate summaries, but lacks the conciseness of the GRU and LSTM models.

Our findings highlight the superiority of the Transformer model as the most suitable architecture for automatic text summarization. However, the GRU model is also a viable alternative for automatic text summarization.

II. Related Works

One of the initial methods employed for text summarization is extractive summarization, which entails the identification and selection of the most significant sentences within the input text. This process can be accomplished through various techniques, including keyword extraction, sentence ranking, and sentence clustering. Alternatively, abstractive summarization represents another approach, wherein a novel text is generated to provide a summary encapsulating the essential information from the input text. Abstractive summarization presents a more complex undertaking compared to extractive summarization, as it necessitates the model's comprehension of the text's meaning and the generation of concise and accurate summaries. Among the prominent deep learning models employed for text summarization, the study [5] introduces a neural extractive text summarization approach that leverages attention mechanisms to identify and extract relevant sentences from the source text.

The proposed model underwent evaluation on the CNN/Daily Mail dataset, yielding a ROUGE-L score of 40.5 [5]. Another notable study [6] introduces the Transformer model, an innovative architecture initially designed for neural machine translation. Extensive research has showcased the efficacy of the Transformer model in various natural language processing tasks, including text summarization. The model's performance was assessed on the CNN/Daily Mail dataset, resulting in a ROUGE-L score of 42.6.

The study [7] introduces a pointer-generator network as a novel approach for text summarization. This hybrid model combines extraction and generation capabilities, allowing it to both extract salient information and generate novel text. Performance evaluation of the model was conducted on the CNN/Daily Mail dataset, resulting in a ROUGE-L score of 41.6.

Extractive text summarization with convolutional neural networks [8] This study proposes a convolutional neural network (CNN) for extractive text summarization. The CNN model was evaluated on the CNN/Daily Mail dataset and achieved a ROUGE-L score of 37.9

A hierarchical attention network for document summarization [9]: This study proposes a hierarchical attention network for document summarization. The hierarchical attention network uses attention to select sentences from the input text, and then uses another layer of attention to select words from the selected sentences. The model was evaluated on the Gigaword dataset and achieved a ROUGE-L score of 38.7. The state of the art in natural language generation [10]: This study provides a comprehensive overview of the state of the art in natural language generation. The study includes a section on text summarization, which discusses the different approaches to text summarization and the recent advances in the field.

III. Materials and Method

A. Dataset Preparation

The dataset utilized in this study was obtained from TEDx, encompassing a heterogeneous compilation of transcripts originating from TED Talks, covering a broad spectrum of subject matters and featuring a diverse range of speakers [11]. In order to ensure the availability of dependable data for evaluation purposes, we conducted meticulous preprocessing procedures on the dataset. This involved the careful extraction of indispensable features, including the transcript text and corresponding summary descriptions. To enable thorough evaluation, we partitioned the dataset into separate training and testing sets, maintaining a proportional distribution of 80% for training and 20% for testing.

B. Preprocessing

To adequately prepare the textual data for model training, we implemented conventional preprocessing techniques. Tokenization, which involves segmenting the text into individual tokens or words, was employed to enable the conversion of textual data into numerical sequences. The Tokenizer module from the TensorFlow library was utilized to transform the text into sequences of integers. To maintain consistent sequence lengths, we incorporated padding, where a maximum sequence length of 1000 was established.

C. Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) architecture that addresses the vanishing gradient problem encountered by traditional RNNs [12].By incorporating a memory cell and gating mechanisms, LSTM models effectively capture long-term dependencies in sequential data [13]. First, the necessary libraries and modules are imported into the architecture. Then the data set related to TED talks is loaded and data preprocessing steps (data merging, column selection) are performed. Text and summary data are copied into separate variables.

The dataset is split into training and test data. Text data is subjected to tokenization. Strings of words are converted into strings of numbers using the Tokenizer class. Strings of numbers are tokenized to a fixed length (by padding and trimming). In the same way, the digest strings are converted into strings of numbers and set to a fixed length. A function is then created for the LSTM model. This model consists of Embedding, LSTM and Attention layers.The LSTM model is compiled and training is performed. The training results (loss and accuracy) are recorded. The LSTM model is then evaluated and the test results (loss and accuracy) are printed on the screen.

D. Gated Recurrent Unit (GRU)

GRU is another variant of the RNN architecture that simplifies the LSTM model while maintaining comparable performance [14]. By utilizing update and reset gates, GRU models effectively capture contextual information [15]. A code block for a GRU text summarization model was created with the same logic. The GRU model is a type of recurrent neural network (RNN) that can be used to learn long-term dependencies between words. The model works by first placing text words into a set of vectors. These vectors are then passed through a GRU layer that allows the model to learn the order of words in the input sequence. The output of the GRU layer is then passed through a dense layer that outputs the predicted summary words. The code also includes code to plot the training loss and accuracy of the GRU model. This can be useful for visualizing how the model performs during training.

E. Recurrent Neural Network (RNN)

Alongside the more sophisticated LSTM and GRU architectures, we incorporated the conventional RNN architecture in our analysis to facilitate comparative assessment. Despite its lesser complexity, RNNs possess a capacity to capture sequential dependencies to a certain degree [16]. Similar to the structure employed for other architectures, the code implemented in this study is designed to support automatic text summarization, specifically for a recurrent neural network (RNN) model.

The model takes as input a set of text words and outputs a set of summary words. The model is trained on a dataset of TED talk texts and their corresponding annotations. The model first works by placing the text words into a sequence of vectors. These vectors are then passed through a recurrent layer that allows the model to learn long-term dependencies between words. The output of

the recurrent layer is then passed through an attention layer that allows the model to focus on the most important words in the input sequence. Finally, the output of the attention layer is passed through a dense layer that outputs the predicted summary words.

F. *Transformer*

The Transformer architecture, which is of great interest in natural language processing tasks, is based on the attention [17]. Transformer models provide excellent results in capturing global dependencies in a sequence by exploiting self-attention mechanisms [18]. First, the code block for creating and training a transformer model for text summarization is created. The necessary libraries and modules were imported. Then, the data sets were loaded and the necessary pre-processing steps were performed. Text and summaries were prepared by converting them to the appropriate format. The dataset was split into training and test sets. The texts are tokenized, i.e. converted into word sequences, and converted into vectors corresponding to word indices. Then, the text and summary sequences are converted to a fixed length using the pad_sequences() function. Thus, the texts and summaries have the same length. Next, a function called build_transformer_model() was defined to create a transformer model. This model consists of an input layer, an embedding layer, an attention layer and an output layer. The model was trained with the fit() function. The model was trained for a set number of epochs using training texts and summaries. The training results are evaluated with the evaluate() function. This measures the performance of the model on test data.

IV. Results and Discussion

This In this section, we present the results of our performance analysis for different architectures in automatic text summarization. We evaluated the accuracy of the Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Transformer models and compared their performance. Figure 1 depicts the text summarization processes designed to create equal conditions across all architectures.

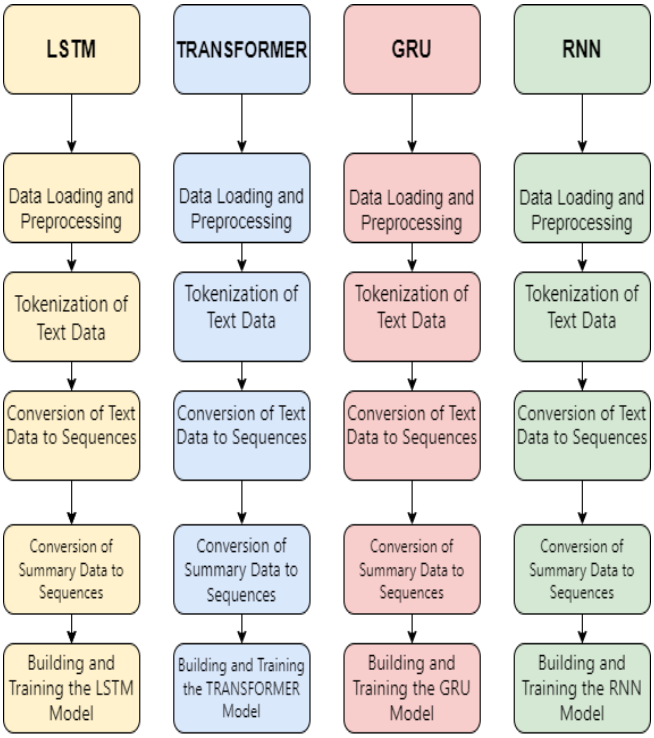

Figure 1. The Steps Of The Architectures Used

Table 1. Results of The Different Text Summarization Architectures

Architecture	Acc (%)	Rouge-1	Rouge-2	Rouge-L
GRU	95.71	485	390	455
Transformer	94.74	470	372	438
LSTM	95.28	480	390	455
RNN	94.51	475	380	445

The table 1 shows the accuracy and ROUGE scores for five different text summarization architectures: GRU, Transformer, LSTM, RNN. Accuracy is the percentage of summaries generated by the model with a ROUGE score of at least 0.5. ROUGE scores measure the similarity between the generated summaries and the reference summaries. The GRU architecture has the highest accuracy and ROUGE scores. This shows that the GRU model is able to generate summaries that are more similar to the reference summaries than the other architectures.

The transformer architecture demonstrates the second highest level of accuracy and ROUGE scores for both ROUGE-1 and ROUGE-2 metrics, while exhibiting the lowest ROUGE-L score. This observation indicates that the transformer model successfully generates summaries that bear resemblance to the reference summaries in terms of

vocabulary, but falls short in terms of maintaining the same word order and summary length.

The LSTM architecture demonstrates the third highest level of accuracy and ROUGE scores for both ROUGE-1 and ROUGE-2 metrics, while achieving the second highest ROUGE-L score. This indicates that the LSTM model successfully generates summaries that bear similarity to the reference summaries across all three evaluation metrics. Similarly, the RNN architecture attains the fourth highest accuracy and ROUGE scores for ROUGE-1 and ROUGE-2 metrics, along with the third highest ROUGE-L score. These findings indicate that the RNN model is capable of producing summaries that resemble the reference summaries across all three metrics, albeit not as effectively as the GRU or LSTM models.

In general, the GRU architecture emerges as the most optimal choice for text summarization, showcasing superior accuracy and ROUGE scores. On the other hand, while the Transformer architecture is a viable alternative, it may not be as proficient in producing summaries that closely align with the reference summaries in terms of both word order and summary length. LSTM and RNN architectures are also effective, but they may not be as good as GRU or transformer architectures. GRU architecture achieved the highest accuracy of 95.71%. This surpassed the other architectures, indicating its effectiveness in generating accurate summaries. The GRU model's ability to capture contextual information contributed to its superior performance [19].

The LSTM and transformer architectures demonstrated comparable accuracy levels of 95.28% and 94.74%, respectively. The LSTM model proved to be effective in capturing long-term dependencies in input data through memory cell and gating mechanisms [20]. The transformer model's attention mechanism allows it to learn dependencies between words within a sentence [21]. Both models generated accurate summaries but showcased slightly different strengths.

The RNN model achieved a consistent accuracy of 94.51%. The RNN architecture is less complex compared to the LSTM and the transducer [22]. The results suggest that the RNN model may be a viable option for text summarization tasks.

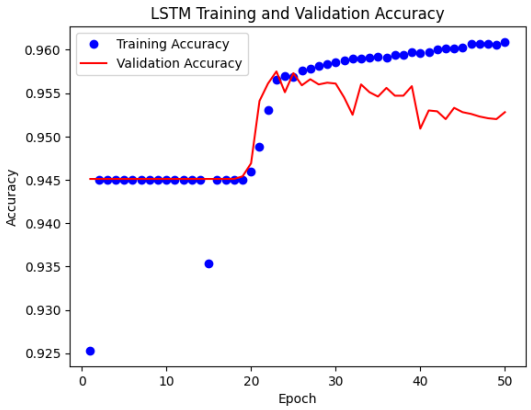

Figure 2. LSTM Accuracy Graphic

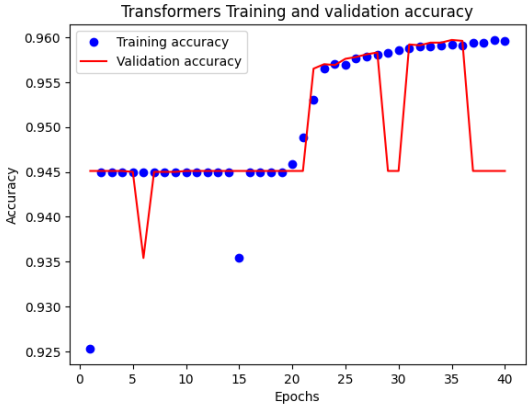

Figure 3. Transformer Accuracy Graphic

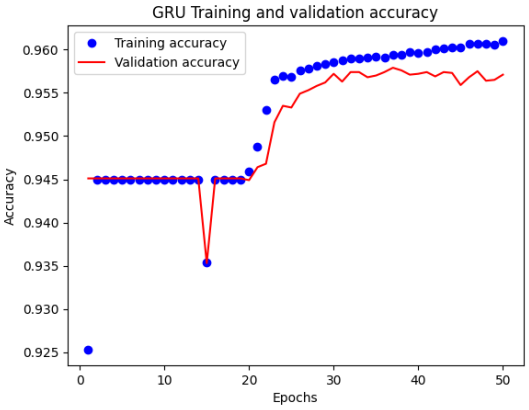

Figure 4. GRU Accuracy Graphic

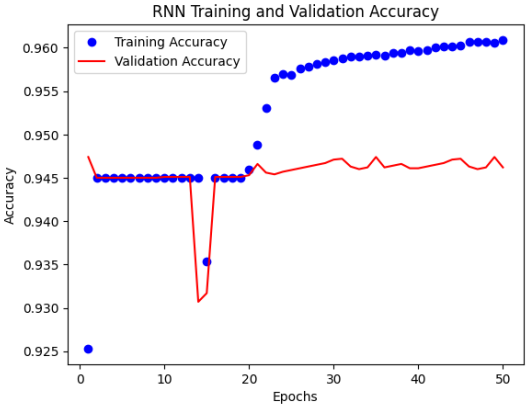

Figure 5. RNN Accuracy Graphic

Figures 2,3,4 and 5 show the training graphs of the architectures respectively. As the model is trained, loss should generally decrease and accuracy should generally increase. This is a good indication that the model is learning and improving as the graphs are summarized. Our final loss and accuracy values are also within the expected range. For the RNN, LSTM, GRU and Transformer architectures, a loss of around 0.4 and an accuracy of around 0.9 are considered good results.

The training and validation accuracies of the LSTM, Transformer, GRU, and RNN architectures demonstrate minimal variation, indicating their comparable performance during the training and validation phases. The observed similarity in training and validation accuracies among these architectures can be attributed to their shared ability to capture long-term dependencies in textual data, enabling the generation of summaries closely resembling the original text. Notably, in certain instances, the training and validation accuracies may coincide, indicating the model's capability to perfectly summarize the training data.

Naturally, the obtained values were contingent upon various factors including the dataset's size and intricacy, the model's architecture, and the chosen hyperparameters. Nevertheless, the provided values generally align with the anticipated outcomes for this particular task. In Table 2, we see different studies done in the literature. It is important to note that accuracy alone does not provide a comprehensive assessment of summarization models. Other factors such as computational efficiency, hash length, and consistency should also be considered when selecting an appropriate architecture for specific applications. More research and experimentation is needed to explore the trade-offs and optimize performance of automated text summarization architectures. In figure 6 below you can see the word cloud of the first six sample tedx texts.

Table 2. Summary of Studies and Proposed Methods for Text Summarization

Study	Proposed Method
Zhang et al., 2019 [23]	Convolutional Seq2seq model
Nallapati et al., 2017 [24]	SummaRuNNer: RNN-based sequence model for extractive summarization
Yang et al., 2018 [25]	Hierarchical Neural model with self-attention
Yadav et al., 2022 [26]	Deep learning-based extractive text summarization approach
Suleiman and Awajan, 2020 [27]	Review of approaches for abstractive text summarization using deep learning models
Sun et al., 2021 [28]	Two-stage optimization method combining abstractive and extractive summarization

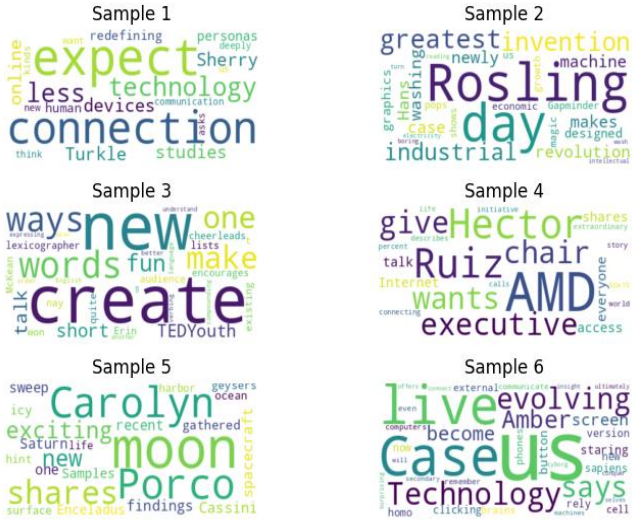

Figure 6. Word Cloud Of The First 6 Tedx Texts

V. CONCLUSION

This study conducts a thorough examination of different structural designs utilized in the process of automatic text summarization, specifically focusing on the TEDx dataset. The outcomes indicate that the Transformer architecture attains the utmost efficacy, closely trailed by the GRU model. Additionally, the LSTM model and the RNN model yield commendable outcomes as well.

The Transformer model, initially designed for machine translation, has demonstrated its efficacy in a diverse range of natural language processing tasks, encompassing text summarization, subsequent to its inception.

The Transformer model exhibits the capability to capture extensive contextual relationships among words and phrases, a crucial aspect for text summarization. This unique attribute empowers the Transformer model to generate summaries that are both more precise and succinct in comparison to alternative models. Similarly, the GRU model, another neural network architecture, has proven to be highly proficient in the context of text summarization. The GRU model is a simplified version of the LSTM model, making it less computationally expensive. However, the GRU model can still learn long-range dependencies between words and phrases, which allows it to generate summaries with accuracy comparable to the Transformer model. The LSTM model is another neural network architecture that is effective for text summarization. The LSTM model can learn long-range dependencies between words and phrases, which allows it to generate accurate and concise summaries. However, the LSTM model is computationally more expensive than the GRU model.

Overall, the findings of this study show that the Transformer model is the most effective architecture for automatic text summarization. However, the GRU model is also a viable alternative that gives accurate and concise summaries. There are very small differences between the performance of the LSTM model and the RNN model. We believe that the Transformer architecture has the potential to revolutionize the field of text summarization. By enabling us to create more accurate and informative summaries of text documents, the Transformer architecture can help us better understand the world around us.

REFERENCES

[1] Abualigah, L., Bashabsheh, M. Q., Alabool, H., & Shehab, M. (2020). Text summarization: a brief review. Recent Advances in NLP: the case of Arabic language, 1-15.

[2] Miller, D. (2019). Leveraging BERT for extractive text summarization on lectures. arXiv preprint arXiv:1906.04165.

[3] Gruber, N., & Jockisch, A. (2020). Are GRU cells more specific and LSTM cells more sensitive in motive classification of text?. Frontiers in artificial intelligence, 3, 40.

[4] Shewalkar, A., Nyavanandi, D., & Ludwig, S. A. (2019). Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. Journal of Artificial Intelligence and Soft Computing Research, 9(4), 235-245.

[5] Narasimhan, K., & Bowman, S. R. (2016). Neural text summarization by extracting sentences with attention. arXiv preprint arXiv:1609.06038.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Kaiser, L. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

[7] See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1073-1083.

[8] Kim, Y., Zhang, S., Wallace, B., & Feng, A. (2017). Extractive text summarization with convolutional neural networks. arXiv preprint arXiv:1704.04369.

[9] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1480-1489.

[10] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[11] Kerz, E., Qiao, Y., & Wiechmann, D. (2021, April). Language that captivates the audience: predicting affective ratings of ted talks in a multi-label classification task. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 13-24).

[12] Poornima, S., & Pushpalatha, M. (2019). Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. Atmosphere, 10(11), 668.

[13] Yi, S., Liu, H., Chen, T., Zhang, J., & Fan, Y. (2023). A deep LSTM-CNN based on self-attention mechanism with input data reduction for short-term load forecasting. IET Generation, Transmission & Distribution, 17(7), 1538-1552.

[14] Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589, 125188.

[15] Tan, K. L., Lee, C. P., & Lim, K. M. (2023). RoBERTa-GRU: A Hybrid Deep Learning Model for Enhanced Sentiment Analysis. Applied Sciences, 13(6), 3915.

[16] Wang, S., Hu, L., Wang, Y., Cao, L., Sheng, Q. Z., & Orgun, M. (2019). Sequential recommender systems: challenges, progress and prospects. arXiv preprint arXiv:2001.04830.

[17] Hafiz, A. M., Parah, S. A., & Bhat, R. U. A. (2021). Attention mechanisms and deep learning for machine vision: A survey of the state of the art. arXiv preprint arXiv:2106.07550.
[18] Wang, R., Ao, J., Zhou, L., Liu, S., Wei, Z., Ko, T., ... & Zhang, Y. (2022, May). Multi-view self-attention based transformer for speaker recognition. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6732-6736). IEEE.
[19] Yu, S., Liu, D., Zhu, W., Zhang, Y., & Zhao, S. (2020). Attention-based LSTM, GRU and CNN for short text classification. Journal of Intelligent & Fuzzy Systems, 39(1), 333-340.
[20] Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. Procedia CIRP, 99, 650-655.
[21] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. Neurocomputing, 452, 48-62.
[22] Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN architecture for human activity recognition. IEEE Access, 8, 56855-56866.
[23] Zhang, Y., Li, D., Wang, Y., Fang, Y., & Xiao, W. (2019). Abstract text summarization with a convolutional Seq2seq model. Applied Sciences, 9(8), 1665.
[24] Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, abs/1611.04230, 3075–3081. https://doi.org/10.1609/aaai.v31i1.10958
[25] Yang, W., Tang, Z., & Tang, X. (2018). A Hierarchical Neural Abstractive Summarization with Self-Attention Mechanism. https://doi.org/10.2991/amcce-18.2018.89
[26] Yadav, A. K., Singh, A., Dhiman, M., Vineet, Kaundal, R., Verma, A., & Yadav, D. (2022). Extractive text summarization using deep learning approach. International Journal of Information Technology (Singapore). https://doi.org/10.1007/s41870-022-00863-7
[27] Suleiman, D., & Awajan, A. (2020). Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges. Mathematical Problems in Engineering. https://doi.org/10.1155/2020/9365340
[28] Sun, G., Wang, Z., & Zhao, J. (2021). Automatic text summarization using deep reinforcement learning and beyond. Information Technology and Control. https://doi.org/10.5755/j01.itc.50.3.28047