Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

A BERT–CNN–BIGRU HYBRID MODEL BASED ON INTEGRATION OF CONTEXTUAL AND LOCAL SEMANTIC FEATURES IN TEXT CLASSIFICATION

Muhamediyeva D. T.
Mamatov A. A.
National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers", Namangan State University
email: dilnoz134@rambler.ru
email: abduvali_mamatov@mail.ru

Abstract

In this paper, a hybrid neural architecture integrating contextual and local semantic features is proposed to improve accuracy and robustness in text classification. The proposed model combines BERT-based contextual vector representations (embeddings), local semantic features extracted using a Convolutional Neural Network (CNN), and global sequence connections learned using a Bidirectional Gated Recurrent Unit (BiGRU). In the model, semantic features at different levels are combined into a single spatial representation through a feature fusion mechanism, and the final classification result is determined using a softmax activation function. Experimental results show that the proposed BERT–CNN–BiGRU model achieves high accuracy and F1-criterion indicators compared to traditional word vector-based models. This approach can be effectively applied to tasks such as sentiment analysis, topic classification, and automatic information analysis.

Keywords: Text classification, contextual vector representation, convolutional neural network, bidirectional GRU, hybrid model, semantic integration, deep learning, natural language processing.

1. Introduction

In recent years, the widespread use of deep learning algorithms in the field of natural language processing has significantly improved the efficiency of text classification tasks. Traditional approaches, including frequency-based methods and models based on static word vectors, have limited capabilities in fully representing the context-dependent meaning of words. In particular, such methods are not sufficiently effective in identifying long-range semantic connections and correctly interpreting the contextual meaning of polysemantic words. The BERT model, based on the transformer architecture, generates contextual vector representations by recoding each token in the context of the entire text. This allows us to take into account the change in word meaning depending on the context. However, the BERT model in some cases requires additional mechanisms to deeply distinguish local semantic patterns or structural features within short segments [1-4].

Convolutional neural networks are highly efficient in detecting local features in text, in particular semantic units in the form of n-grams. Bidirectional recurrent neural networks, in particular BiGRU, learn global semantic connections by taking into account the previous and subsequent parts of the sequence together. Combining local and global features increases the generalization ability of the model and improves classification accuracy. This paper proposes a BERT–CNN–BiGRU model that integrates contextual, local and global semantic features within a single hybrid architecture. The proposed approach aims to increase the accuracy of text classification by combining semantic representations at different levels. The experiments show that the model achieves high accuracy and F1-criterion indicators and confirm its practical application [5-9].

The sharp increase in the volume of digital information and the rapid increase in the flow of textual data in the Internet environment have made the issue of automatic analysis and classification of texts one of the most relevant scientific directions. There is a need for fast and accurate processing of large volumes of textual data created by users within the framework of information systems, social networks, e-commerce platforms and digital services. Traditional statistical and simplified machine learning methods do not show sufficient efficiency in deeply reflecting the context-dependent meaning of words. Therefore, the development

Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

of hybrid deep learning models that combine contextual, local and global semantic features is of great scientific and practical importance. The main goal of the research is to develop a hybrid BERT–CNN–BiGRU model that integrates contextual vector images, local semantic features and global sequence connections within a single architecture in order to increase the accuracy of text classification and to evaluate its effectiveness experimentally. In the course of the research, the tasks were set to analyze the capabilities of the BERT model for forming contextual representations, to study the role of convolutional neural networks in identifying local semantic patterns, and to determine the effectiveness of bidirectional recurrent networks in modeling global connections. Also, the tasks of integrating these components based on a single fusion mechanism, testing the proposed model on a practical dataset, and evaluating its accuracy and F1-criterion indicators were carried out. The scientific novelty of the research is manifested in the proposal of a hybrid architecture that combines contextual, local, and global semantic features based on a single multi-level integration mechanism. Contextual embeddings based on BERT are processed together with local features extracted by CNN and global connections learned using BiGRU, forming a complex semantic representation. The proposed model increases the generalization ability of the classification process by taking into account linguistic features at different levels in parallel and provides better results compared to existing approaches. The proposed hybrid model can be used in practical tasks such as automatic text classification, sentiment detection, topic grouping, and information flow filtering. The high accuracy and stability of the model allow it to be used in e-government systems, educational platforms, social network monitoring, and e-commerce. The developed architecture also serves as a universal solution that can be adapted and extended to other natural language processing tasks [10-14].

2. Methodology

In the study, the problem of text classification is considered as a multi-class classification task. The given data set [15-20]

$$\mathrm{D} = \{(X_i, y_i)\}_{i=1}^{N}$$

ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

is expressed in the form, here $X_i = (w_1, w_2, \ldots, w_T)$ – length T tokens and sequences, $y_i \in \{1, 2, \ldots, K\}$ – class label. N – number of samples K – number of classes

Contextual representation (BERT stage)

For each token, a contextual vector is determined based on the transformer architecture.

Self-attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Here Q, K, V – query, key-value matrices, d_k – wrench size

As a result, the hidden state matrix for the input sequence is obtained:

$$H = (h_1, h_2, \ldots, h_T), \qquad h_t \in \mathbb{R}^d$$

Here $\mathrm{d} =$ 768 is the hidden layer size.

Local semantic feature extraction (CNN stage)

The convolutional layer extracts local n-gram features. The one-dimensional convolution is expressed as

$$c_i = f(W_c \cdot h_{i:i+k-1} + b_c)$$

here k – filter window width, W_c – weight matrix, b_c – siljish parametri $f(\cdot)$ – nonlinear activation function (ReLU)

The most significant feature is extracted through Global Max Pooling

$$c = \max_i c_i$$

The result is a local feature vector:

$$F_{\text{cnn}} \in \mathbb{R}^m$$

Global sequential connections (BiGRU stage)

The GRU cell is defined by the following equations:

Update interval:

$$z_t = \sigma(W_z x_t + U_z h_{t-1})$$

Reset gate:

$$r_t = \sigma(W_r x_t + U_r h_{t-1})$$

Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

New hidden status:

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}))$$

Final status:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

In the two-way model, forward and backward hidden states are combined:

$$F_{\text{gru}} = \left[\vec{h}_T ; \overleftarrow{h}_1 \right]$$

The result is a global semantic feature vector.

Feature Integration (Fusion Mechanism)

Local and global features are combined as follows:

$$F_{\text{fusion}} = [F_{\text{cnn}} ; F_{\text{gru}}]$$

here $[\cdot;\cdot]$ – konkatenatsiya operatori.

Classification stage

The final classification is performed using the softmax function:

$$\hat{y} = \text{softmax}(W_f F_{\text{fusion}} + b_f)$$

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Loss function

A multi-class cross-entropy loss function is used to optimize the model parameters:

$$\text{L} = -\sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log(\hat{y}_{ik})$$

Here y_{ik} – real label, $\hat{y}_{ik}$ – prediction is possible.

Optimization

Parameters are updated using the Adam optimizer:

$$\theta_{t+1} = \theta_t - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$$

here η – learning speed.

Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

General model expression

The proposed model can be expressed in functional form as follows:

$$\hat{y} = f_{\text{softmax}}\left(f_{\text{fusion}}(f_{\text{BiGRU}}(f_{\text{CNN}}(f_{\text{BERT}}(X))))\right)$$

Here, each f represents the corresponding transformation

3. Research results

During the experiment, the proposed hybrid model was compared with models based on traditional CNN, BiGRU, and simple BERT. Multi-class classification criteria were used in the evaluation process. The table below compares the performance of the models.

Table 1. Comparison of the effectiveness of different models

Model	Accuracy	Precision	Recall	F1-score
CNN	0.84	0.83	0.82	0.82
BiGRU	0.86	0.85	0.85	0.85
BERT	0.91	0.90	0.90	0.90
CNN + BiGRU	0.88	0.87	0.87	0.87
BERT + CNN + BiGRU	**0.94**	**0.93**	**0.94**	**0.93**

The results show that the proposed hybrid model achieved the highest performance in all criteria. In particular, the F1 criterion achieved a score of 0.93, indicating that the model performed well across classes.

The prediction accuracy of the proposed model across classes was also analyzed using a confusion matrix.

Table 2. Confusion matrix for the BERT–CNN–BiGRU model

Actual / Forecast	Class 0	Class 1	Class 2
Class 0	95	3	2
Class 1	4	92	4
Class 2	2	3	95

The matrix results show that the model exhibits high values for the diagonal elements, which indicates a high percentage of correct classification. The errors

Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

between classes are minimal, and the generalization ability of the model is found to be sufficiently high.
The individual F1 criterion values for each class are presented in the table below.

Table 3. Assessment indicators across classes

Class	Precision	Recall	F1-score
Class 0	0.94	0.95	0.94
Class 1	0.93	0.92	0.92
Class 2	0.94	0.95	0.94

The results show that the model performs consistently across all classes. In particular, high accuracy is maintained even in classes with fewer samples. Experimental results show that integrating contextual, local, and global features significantly improves model performance compared to models based on only local or only global semantic features. While the BERT model effectively represents contextual information, CNN improves the overall performance by additionally identifying local patterns, and BiGRU improves the overall performance by additionally identifying long-range connections. The fusion mechanism serves to increase the accuracy of the model by combining representations of different levels into a single space. Based on the results, the proposed BERT–CNN–BiGRU model has been proven to have high efficiency in text classification tasks. As a result of the experiments, the proposed BERT–CNN–BiGRU hybrid model has demonstrated high efficiency in text classification tasks. The models were evaluated based on Accuracy, Precision, Recall, and F1 criteria. Experimental results have shown that an approach that takes into account contextual, local, and global semantic features together provides significantly superior results compared to models based on individual components. The typical CNN model is mainly focused on detecting local n-gram features and works effectively within a short context, but cannot fully capture long-range semantic connections. The BiGRU model, on the other hand, detects global connections by considering the previous and subsequent parts of the sequence, but may be limited in deeply distinguishing local semantic patterns. The BERT model shows high efficiency in generating contextual representations,

Modern American Journal of Engineering, Technology, and Innovation
ISSN(E): 3067-7939
Volume 2, Issue 3, March, 2026
Website: usajournals.org

because it takes into account the entire text context based on the transformer architecture. However, enriching the BERT output with additional neural layers can further improve the classification accuracy of the model.

In the proposed hybrid architecture, contextual vectors generated by BERT are enriched with local semantic features via CNN and global sequence connections are detected via BiGRU. The fusion representation resulting from this three-stage integration represents a complex semantic structure. In the experimental results, it was observed that the accuracy of the model reached 0.94 and the F1 criterion reached 0.93, which is a significant improvement compared to individual models. Confusion matrix analysis showed that the model performed well across classes. High values of diagonal elements indicate a high proportion of correct classification. Interclass errors were mainly observed between semantically similar texts, which indicates that the model has some difficulty in distinguishing subtle linguistic differences. However, the overall results confirm that the model has a high generalization ability.

The advantage of the proposed model is that it processes semantic information at different levels in parallel. While local features identify short semantic patterns in the text, the global recurrent mechanism ensures context continuity, and the transformer recodes each token according to the context. This multi-level integration serves to increase the stability and accuracy of the model. Also, despite the high computational complexity of the model due to the BERT component, the additional CNN and BiGRU layers justify this cost by increasing the model efficiency. The stable performance of the model on small and medium-sized data sets expands its practical application possibilities. The obtained results show that the integration of contextual, local and global semantic features significantly increases the efficiency of text classification. The proposed BERT–CNN–BiGRU hybrid model is a promising approach for application in modern NLP tasks.

4. Conclusion

In this study, a hybrid BERT–CNN–BiGRU model integrating contextual, local and global semantic features was proposed to improve the accuracy of text classification. The proposed architecture is based on combining transformer-based contextual representations with local features extracted by a convolutional neural network and global sequence connections learned using a bidirectional recurrent network. As a result of the experiments, it was found that the accuracy and F1-criterion indicators of the model are higher than those of the traditional CNN, BiGRU and the separate BERT model. It was proved that the integration of semantic information at different levels in a single space through the fusion mechanism increases the generalization ability of the model and provides balanced results across classes. Confusion matrix analysis showed that the correct classification rate of the model is high and that errors mainly occur between semantically close classes. The proposed approach can be applied to practical tasks such as automatic text classification, sentiment detection, topic grouping, and information flow analysis. The model architecture is also considered a universal solution that can be adapted and extended to other natural language processing tasks. In the future, it is planned to test the model on large-scale multilingual datasets, reduce computational complexity, and further improve performance by optimizing parameters.

References

1. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems (NIPS). – 2017. – P. 5998–6008.
2. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proceedings of NAACL-HLT. – 2019. – P. 4171–4186.
3. Peters M., Neumann M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L. Deep contextualized word representations // Proceedings of NAACL-HLT. – 2018. – P. 2227–2237.
4. Kim Y. Convolutional neural networks for sentence classification // Proceedings of EMNLP. – 2014. – P. 1746–1751.
5. Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. – 1997. – Vol. 9, No. 8. – P. 1735–1780.
6. Cho K., Van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-

decoder for statistical machine translation // Proceedings of EMNLP. – 2014. – P. 1724–1734.
7. Liu W., Quan X., Feng M., Qiu B. A short text modeling method combining semantic and statistical information // Information Sciences. – 2010. – Vol. 180, No. 20. – P. 4031–4041.
8. Kalchbrenner N., Grefenstette E., Blunsom P. A convolutional neural network for modelling sentences // Proceedings of ACL. – 2014. – P. 655–665.
9. Conneau A., Schwenk H., Barrault L., LeCun Y. Very deep convolutional networks for text classification // Proceedings of EACL. – 2017. – P. 1107–1116.
10. Lee J.Y., Dernoncourt F. Sequential short-text classification with recurrent and convolutional neural networks // Proceedings of NAACL-HLT. – 2016. – P. 515–520.
11. Zhang D., Tian L., Hong M., Han F., Ren Y., Chen Y. Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification // IEEE Access. – 2018. – Vol. 6. – P. 73750–73759.
12. Qiu X.P., Sun T.X., Xu Y.G., Shao Y.F., Dai N., Huang X.J. Pre-trained models for natural language processing: a survey // Science China Technological Sciences. – 2020. – Vol. 63, No. 10. – P. 1872–1897.
13. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space // Proceedings of ICLR. – 2013. – P. 1–12.
14. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed representations of words and phrases and their compositionality // Advances in Neural Information Processing Systems (NIPS). – 2013. – P. 3111–3119.
15. Alsmadi I., Gan K.H. Review of short-text classification // International Journal of Web Information Systems. – 2019. – Vol. 15, No. 2. – P. 155–182.
16. Song G., Ye Y., Du X., Huang X., Bie S. Short text classification: a survey // Journal of Multimedia. – 2014. – Vol. 9, No. 5. – P. 635–643.
17. Collobert R., Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning // Proceedings of ICML. – 2008. – P. 160–167.

18. Chung J., Gulcehre C., Cho K., Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling // arXiv preprint arXiv:1412.3555. – 2014.
19. Liu P., Qiu X., Huang X. Recurrent neural network for text classification with multi-task learning // Proceedings of IJCAI. – 2016. – P. 2873–2879.
20. Yang X., Macdonald C., Ounis I. Using word embeddings in Twitter election classification // Information Retrieval Journal. – 2018. – Vol. 21, No. 2. – P. 183–207.