Advanced Explainable AI: Self Attention Deep Neural Network of Text Classification

[1]Prasanna V, [2]Umarani S, [3]Suganthi B, [4]Ranjani V, [5]Manigandan Thangaraju and [6] Uma Maheswari P
[1]Department of Data Science and Business Systems, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.
[2]Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram, Chennai, India.
[3]Department of Electronics and Communication Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Trichy, India.
[4]Department of Computer Science and Engineering(AIML), Saranathan College of Engineering, Trichy, India.
[5]Department of Cyber Security, Sphoorthy Engineering College, Nadergul, Hyderabad , India.
[6]Department of Computer Science and Business Systems, K.Ramakrishnan College of Engineering, Trichy, India.
[1]vprasannamadhan@gmail.com, [2]ravania@gmail.com, [3]suganthib.set@dsuniversity.ac.in, [4]bhaviranjani13@gmail.com, [5]manigandant@sphoorthyengg.ac.in, [6]umamaheswarip.csbs@krce.ac.in.

Correspondence should be addressed to Prasanna V : vprasannamadhan@gmail.com

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404056
Received 12 November 2023; Revised from 10 March 2024; Accepted 30 May 2024
Available online 05 July 2024.

Abstract – The classification of texts is a crucial component of the data retrieval mechanism. By utilizing semantic details representation, and the text vector sequence is condensed, resulting in a reduction in the temporal and spatial order of the memory pattern. This process helps to clarify the context of the text, extract crucial feature information, and fuse these features to determine the classification outcome. This approach represents the preprocessed text data using character-level vectors. The self-attention mechanism is used to understand the interdependence of words in a text, allowing for the extraction of internal structure-related data. Furthermore, the semantic characteristics of text data have been extracted independently using Deep Convolutional Neural Network (DCNN) and Bi-directional Gated Recurrent Unit (BiGRU) using a Soft-Attention mechanism. These two distinct feature extraction outcomes are then merged. The SoftMax layer is employed to categorize the deep-extracted attributes, hence enhancing the accuracy of the classification model. This improvement is achieved by including a uniform distribution component into the cross-entropy loss function. Our results demonstrate that our suggested method for explainability outperforms the model that was suggested in terms of accuracy and computing efficiency. For the purpose of assessing the effectiveness of our suggested approach, we developed many baseline models and performed an evaluation their studies.

Keywords – EAI, Deep Learning, Attention, Text Classification, DCNN.

I. INTRODUCTION

Artificial intelligence (AI) has an extensive history in the field of computer science. The recent revival of AI has been largely driven by the advancements in machine learning (ML), specifically the remarkable achievements in "deep learning," which have been made during the last decade. The tremendous achievement was accompanied with further expenses and obligations: the most triumphant approaches are so intricate that it is arduous for a human to retrace, comprehend, and explain the process by which a specific outcome was attained. The advancement of deep learning (DL) techniques in natural language processing (NLP) has led to a growing interest in modelling explanation, which aims to interpret black-box models in a clear and understandable way [1].

The development of AI has seen significant advancements because to the exponential growth in data volume and improvements in computer speed. AI has successfully been implemented in several domains, including text, voice recognition, autonomous driving, and recommendation systems.

The text biomedical informatics community use machine learning, NLP, and DL-based technologies to automate the categorization of clinical records and overcome these challenges. Contemporary NLP models [2] often use attention, which is a widely used component in the construction of deep neural networks, positioned at certain locations. Nevertheless, rule-based systems are costly and unreliable due to the need of explicitly defining decision-making rules and the requirement for human updates, similar to textbooks. Furthermore, the process of coding complex connections between distinct pieces of information provided by several specialists is challenging. Additionally, the effectiveness of the structure is constrained by the extent of existing medical knowledge.

Data preparation involves the removal of unnecessary characters, the segmentation of words, and the encoding of letters or words. Feature extraction involves identifying and extracting high-frequency terms, as well as calculating the relationship of word or phrase vectors. The collected features and text categorization labels serve as the input for the classifier, which is utilized to train the optimal model using the sample data [3].

The present investigation examines label-attention techniques that integrate implicit or explicit supplementary information into various text-encoder designs, including convolutional neural networks (CNN) [21]and recurrent neural networks (RNN) [22]. In addition, we evaluate the efficacy of our label-attention mechanisms in comparison to target-attention and other conventional approaches for producing document embeddings. The achievements that we have made are as follows:

- Our work is the initial effort to thoroughly compare the label-attention process in clinical text categorization. Our main objective is to enhance the automated encoding of clinical texts with medical symbols.
- We investigate the impact of various attention techniques on document embeddings unique to DCNN and BiGRU models. and additionally demonstrates that setting up the reference material of an attention mechanism with explicit label-specific auxiliary information enhances the effectiveness of categorization.

II. RELATED WORK

S. Liu et al., [4] research presents two innovative explanation strategies, AGrad and RePAGrad, that generate directional applicability scores using attention weights. Both of these methods aim to overcome the constraint of attention weights in giving directional data that provides significance. The research paper proposes three criteria for assessing exposition methods: fidelity, resilience, and consistency. It also presents specific tests to measure each criterion, highlighting the significance of these factors in the context of explainability. The work emphasizes the significance of the structure of models in achieving explainability and proposes further research on identifying multiple interconnected words or sentences as explanations. It takes into account the constraints of single-word explanations, particularly in the interpretation of adjectives and adverbs without taking into account the items that they outline. The research mainly assesses the effectiveness of explanation methods on Transformer models and pre-trained BERT models, perhaps neglecting the effectiveness of these strategies on other model topologies.

H. Chefer et al.,[5] research presents a novel approach to elucidate predictions made by Transformer-based architectures, that include bi-modal Transformers and Transformers featuring co-attentions. This approach significantly surpasses previous techniques derived from single modality explainability.

E. Hashmi et al.,[6] presents a resilient strategy for identifying fake news by including several openly accessible datasets: WELFake, Fake Newsnet, and Fake News Prediction. This methodology combines Fast Text word embeddings using a range of Machine Learning and Deep Learning techniques. The proposed approach presents a hybrid strategy that merges Convolutional Neural Networks with Long Short-Term Memory, enhanced through Fast Text embeddings. The proposed approach surpassed previous methods in terms of classification efficacy on all datasets, obtaining exceptional accuracy and F1-scores. The investigation advances upon advanced transformer-based theories such as BERT, XLNet, and RoBERTa by making hyperparameter modifications. This surpasses typical RNN-based structures in effectively handling syntactic intricacies, leading to enhanced semantic comprehension. The research utilizes explainable AI modelling methodologies such as Local Interpretable Model-Agnostic Explanations and Latent Dirichlet Allocation to acquire a more profound understanding of how the system makes decisions. This approach improves the transparency and interpretability of the fake news detection algorithms.

H. Sebbaq and N. El Faddouli, [7] The research presents MTBERT-Attention, an innovative model that combines multi-task [24] Learning (MTL), BERT [23], and the attention mechanism to perform cognitive text categorization. This framework gains the cognitive categorization of texts in addition to appropriate subtasks, which improves its ability to apply knowledge to new situations and enables the expansion of the available data. The recommended approach surpasses the baseline models in terms of loss, F1-score, and accuracy, obtaining an overall classification accuracy of 97.71% with the test set. It accurately categorizes learning goals by using ambiguous action verbs from Bloom's taxonomy. The research introduces a paradigm for explainability that relies on the attention mechanism. The effectiveness of this framework is assessed via both qualitative and quantitative investigations. The explainability technique outperforms the LIME explainer in terms of accuracy and processing resources, therefore showcasing the efficacy of the suggested methodology.

L. H. Baniata and S. Kang, [8] presents a unique text categorization framework for Arabic dialects called Switching Self-Attention. The approach utilizes Reverse Positional Encoding (RPE) to break down the work into shorter sub-tasks, hence enhancing accuracy and efficacy in sentiment analysis. The present research tackles the difficulties of recognizing texts in Arabic by using a switching self-attention shared encoder combined with Mixture of Experts (MoE) and RPE approaches. This approach enhances flexibility and accuracy in representing sentences, which is especially advantageous for Arabic text classification problems. Implemented a Transformer model including Multi-Head Attention (MHA) and Feedforward Neural Networks (FFN) to perform text classification tasks. This approach allows for the assessment of word importance by considering semantic connections and improving presentations via non-linear changes. Arabic dialects are characterized by substantial linguistic variety and complexity, which presents difficulties for deep-learning networks. These issues arise from differences in syntax, word order, and limited resources, which impede the effectiveness of text categorization tasks. The scarcity of information poses a challenge in obtaining significant training data for deep learning algorithms across Arabic dialects. This challenge arises from the unstructured nature and limited availability of resources in Arabic dialect datasets, which in turn affects information retrieval and the effectiveness of models. **Table 1** present some previous work discussion.

Table 1. Previous Studies

Author(s)	**Method**	**Contributions**	**Limitations**
S. Liu et al., [4]	AGrad and RePAGrad: explanation strategies generating directional applicability scores using attention weights.	Proposed three criteria (fidelity, resilience, and consistency) for assessing explanation methods. Developed specific tests to measure these criteria. Emphasized the importance of model structure in explainability. Highlighted the limitation of single-word explanations and the potential of using multiple interconnected words or sentences.	Focused mainly on Transformer and BERT models, potentially neglecting other model topologies.
H. Chefer et al., [5]	Novel approach to elucidate predictions made by Transformer-based architectures, including bi-modal Transformers and those with co-attentions.	Surpassed previous single modality explainability techniques. Improved interpretability of Transformer-based models.	Specific details of contributions and limitations not detailed in provided text.
E. Hashmi et al., [6]	Resilient strategy combining FastText word embeddings with Machine Learning and Deep Learning techniques (CNN and LSTM).	Utilized datasets: WELFake, FakeNewsNet, and FakeNewsPrediction. Enhanced fake news detection accuracy and F1-scores. Advanced transformer-based theories (BERT, XLNet, RoBERTa) with hyperparameter modifications. Used Local Interpretable Model-Agnostic Explanations (LIME) and Latent Dirichlet Allocation (LDA) for better decision transparency.	Challenges in handling syntactic intricacies and limited datasets for extensive validation.
H. Sebbaq and N. El Faddouli, [7]	MTBERT-Attention: Combines Multi-Task Learning (MTL), BERT, and attention mechanism for cognitive text categorization.	Achieved high classification accuracy (97.71%). Effectively categorized learning goals using ambiguous action verbs. Introduced a new explainability paradigm relying on the attention mechanism. Outperformed LIME in terms of accuracy and processing resources.	Specific limitations not detailed in provided text.
L. H. Baniata and S. Kang, [8]	Switching Self-Attention framework for Arabic dialect text categorization using Reverse Positional Encoding (RPE).	Enhanced accuracy and efficacy in sentiment analysis for Arabic dialects. Utilized a shared encoder with Mixture of Experts (MoE) and RPE. Implemented Transformer model with Multi-Head Attention (MHA) and Feedforward Neural Networks (FFN). Addressed challenges of linguistic variety and complexity in Arabic dialects.	Scarcity of training data and limited resources in Arabic dialect datasets. Challenges in information retrieval due to unstructured data and linguistic variety.

III. METHODOLOGY

The self-attention procedure, known for its ability to extract relevant information from the text structure, is combined with distinct methods for obtaining features of the DCNN network and the Soft-Attention-based BiGRU system in **Fig 1**. The text information undergoes tokenization, where it is divided into individual characters. A marker is subsequently included and a vectorized depiction is created, which includes the semantic data of the text [9].

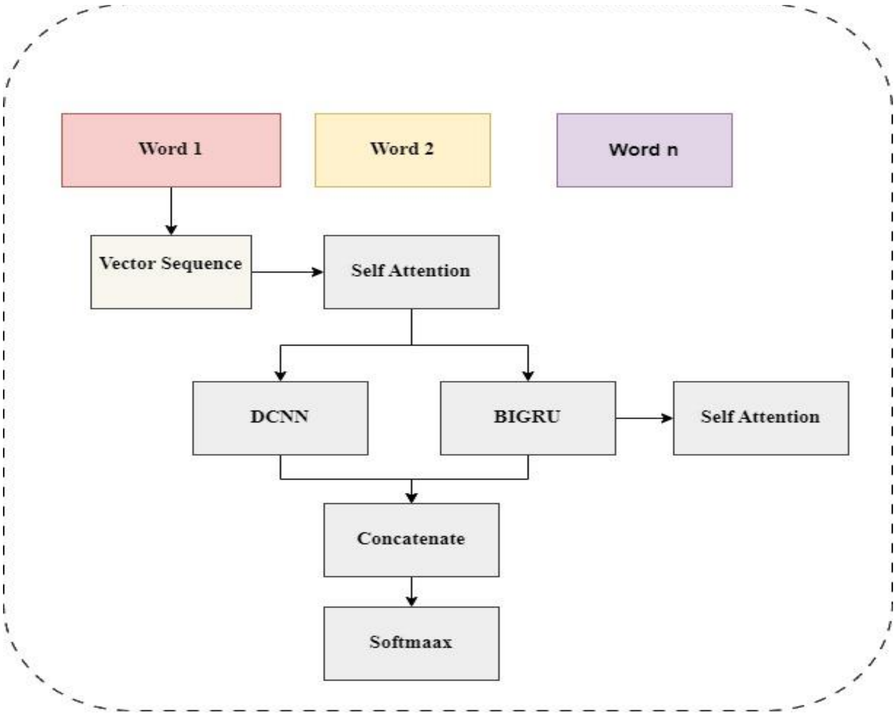

Fig 1. Proposed Model.

The self-attention framework is composed of 3 matrices, namely Q (Query), K (Key), and V (Value), they are used to convert text vectors. The precise formulations are listed below*:*

$$Q^i = W^{q*}W^*Y_i$$
$$K^i = W^k * W^*Y_i \tag{1}$$
$$V^i = W^v * W^*Y_i$$

Y_i is a vector
W, W^q, W^k and W^v ia a model matrices weight
Q^i, K^i and V^i – Single text character vector

$$Q = [Q^1 \quad \cdots \quad Q^m]^r$$
$$K = [K^1 \quad \cdots \quad K^m]^r \tag{2}$$
$$V = [V^1 \quad \cdots \quad V^m]^r$$

Equation 2 give dimension of matrix Z

The primary purpose of the self-attention model **Fig 3** is to compute similarities between two vectors using the dot product similarity attention formula. To mitigate the impact of data dimensionality, it is essential to use a conventional normal distribution flattening procedure, denoted as follows:

$$B = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

B gives feature outcome of self-attention layer.

DCNN

Recently, sentiment orientation categorization has used deep learning frameworks to automatically choose characteristics using algorithms, making it very applicable in the industry. I performed text categorization using the CNN convolutional network approach and obtained specific local key information by configuring several convolution kernels.

The conventional CNN approach [10] has a significantly improved performance in text classification challenges. However, it lacks the ability to effectively capture semantics when dealing with lengthy text extraction characteristics. Hence, it is essential to use lots of convolutional layers and modify the size of the convolution window in order to collect and combine local semantic data from various places, hence enhancing the comprehension of text semantics.

Thus, it is suggested to use the DCNN model for extracting text attention characteristics. The multi-layer convolution feature extraction layer is established to combine the multi-layer convolution and initial convolution layers. This combination enhances important local features [11] and works in conjunction with the global maximum pooling layer (GlobalMaxPooling) to extract the highest value feature from the text vector. This process achieves deep superimposed convolution feature extraction. **Fig 2** displays the structural diagram.

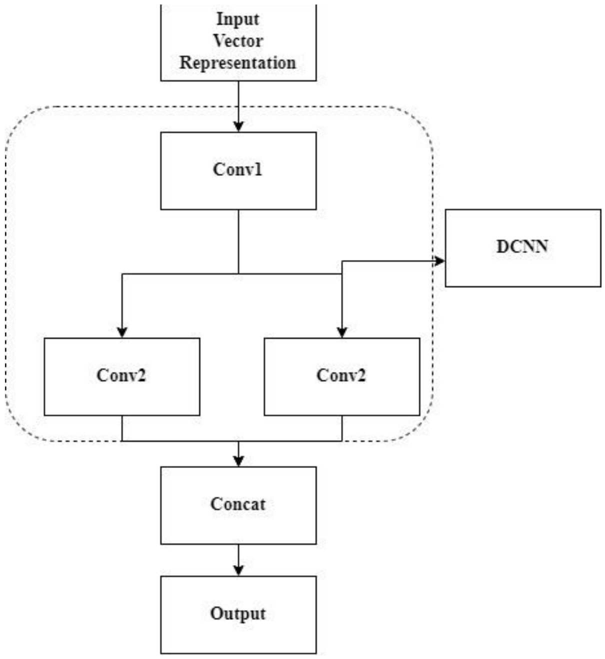

Fig 2. DCNN.

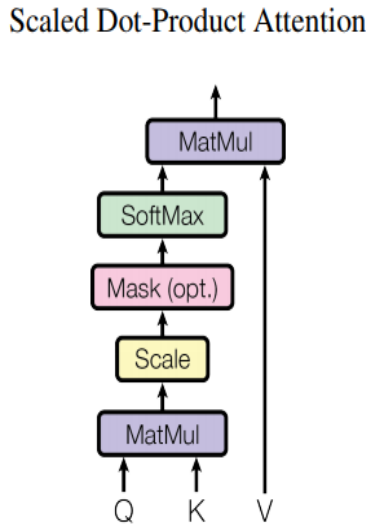

Fig 3. Self-Attention.

BiGRU

The BiGRU framework enhances the ability to represent bidirectional semantic dependence by including a reverse operation into the GRU paradigm. The total architecture may be partitioned into a forward GRU and a backward GRU. By concatenating the bidirectional GRU features at the same position in the hidden state, the resulting text depiction may be comprehended within its context.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{4}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{5}$$

z_t – update gate
r_t – reset gate
h_t – hidden layer
h_{t-1} – previous hidden layer
x_t - input data

$$\tilde{h}_t = tanh\,(W \cdot [r_t * h_{t-1}, x_t]) \tag{6}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{7}$$

$\tilde{h}_t$ - current candidate set

The text vector is processed by the DCNN layer to get the feature vector X_c. It then goes through the BiGRU-Attention layer to produce the feature vector X_s. X_c *and* X_s are combined to form the semantic characteristic X_c. The precise calculating procedure is as outlined below:

$$X_{cs} = X_c \oplus X_s \tag{8}$$

$$\mathrm{y} = \mathrm{softmax}\ (X_{cS}) \tag{9}$$

$$\bar{\mathrm{y}} = \mathrm{argmax}\ (\mathrm{y}) \tag{10}$$

y is a feature vector output
$\bar{\mathrm{y}}$ is argmax function index

$$S(q \mid p) = -\sum q_i log\ p_i \tag{11}$$

p_i is the predicted value
q_i is the actual value
Final loss function calculated by

$$J(\theta) = -(1-\eta)(-\sum_i q_i \log p_i) - \eta\left(-\sum_i \frac{1}{n} \log p_i\right) \tag{12}$$

n is the number of classification

IV. RESULTS AND DISCUSSION

This test checked how well the text sorting model worked. It took true Q & A info from a site about health and drugs. There were 32636 text questions. They got split into 0 to 6 groups. These matched with sickness spread, kids' health, baby birth and women's health, body health, skin issues, cutting and fixing body parts, and face parts. The structure of the question's wording is difficult to comprehend. The discourse inside it is excessively mundane, making it arduous to delve into the essence and significance of the vocabulary. In order to have an accurate understanding of the question's wording, it is necessary to engage in thorough contemplation. Evaluating the model's performance and assessing the applicability of the procedures utilized on this kind of information may provide an accurate measure of the model's effectiveness.

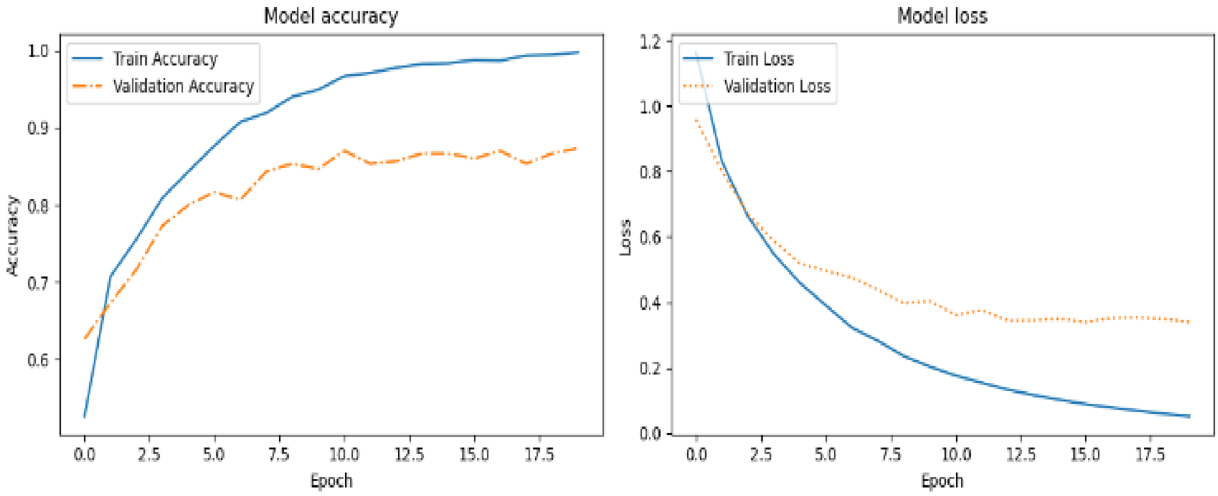

Fig 4. Evaluation Accuarcy and Loss.

Comparison Results Analysis

In order to validate the effectiveness of our suggested methodology, we do experiments using approaches for the purpose of text categorization [18]. The comparing computational models consist of the following: CNN + BILSTM, DCNN, CNN + TABAS series networking with Attention, and BLTCN-BLSTM [19] parallel network with Attention. Additionally, there are deep learning systems based on Self-Attention DCNN and BiGRU series structure with attention [20]. The empirical assessment findings are shown in **Table 2.**

Table 2. Comparative results of our proposed approach

Model	*Accuracy*	*Precision*	*Recall*	*F1*
CNN + BILSTM [14]	74.05	73.50	74.10	74
BERT + Self Attention [13]	80.01	79	79.85	81
DCNN [12]	86.80	86.00	86.50	86.70
CNN + TABAS [15]	88.26	87.90	87.80	88.00
BLTCN-BLSTM [16]	89.10	89.40	89.20	89.30
OUR Proposed Model (**DCNN + BIGRU**)	94.50	94.70	94.10	94.25

In order to address the issue of over-fitting in the framework, the text introduces a uniform distribution component to the cross-entropy loss equation. This addition serves to mitigate the problem of over-fitting. **Fig 4** demonstrate that the loss value of the hypothesis using the uniformly distributed cross-entropy loss function has risen, but it remains within the acceptable range. The loss product that is uniformly distributed might come together at a quicker rate and marginally enhance the accuracy of the strategy classification.

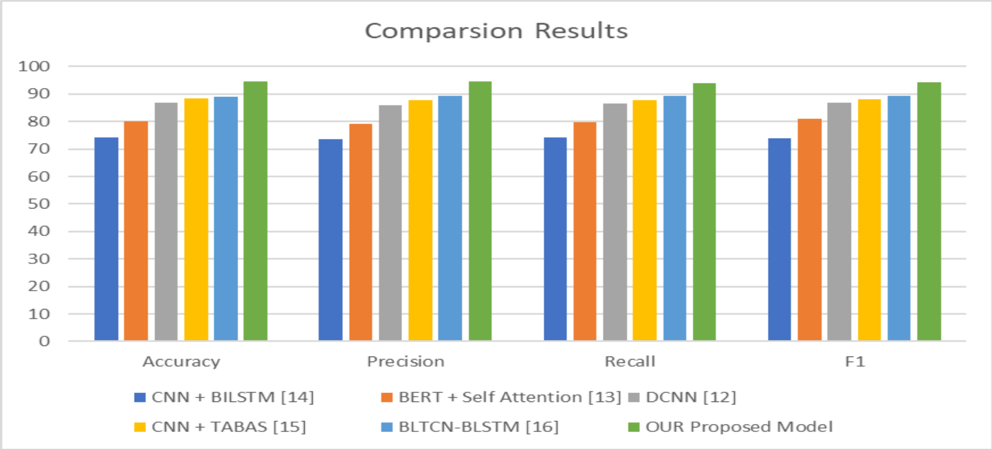

Fig 5 . Comparison Results Analysis.

The table illustrates the results of a comparative experiment between CNN + TABAS [15] [18] and BLTCN-BLSTM [16], which indicates the text classification approach with self-attention mechanism [17] has achieved significant progress. This research presents an approach that employs a cross-entropy loss function that is uniformly distributed on the Self-Attention Deep Convolutional Neural Network and BiGRU architecture. The F1 value has achieved its maximum. As a consequence of the high value of 94.25%, it is clear that the classification results produced by this framework are more accurate, and the accuracy of classification is greater. **Fig 5** shows the comparison results analysis.

V. CONCLUSION

This article examines the categorization model for complicated text sentences. To ensure that the model is oriented towards the text sentences that need comprehension, it incorporates a self-attention mechanism to examine the textual structure between words. Additionally, it utilizes convolutional networks to extract semantic information from the text and employs bidirectional gated loops. The neural network comprehends contextual semantics, integrates many characteristics to form the ultimate text semantic features, and categorizes sentences based on the classifier. In comparison investigations have demonstrated that the self-attention mechanism significantly enhances classification accuracy. DCNN enables the extraction of more semantic features. Additionally, the uniform distribution of the cross-entropy loss function properly mitigates over-fitting. The subsequent stage will include examining the use of a multi-view text classification model applied to a single text multilingual dataset in order to enhance the model's adaptability capability.

VI. LIMITATION & FUTURE WORK

Although the model excels at classifying complex textual phrases, its ability to generalize to other kinds of textual data or languages may be restricted. The efficacy of the machine learning algorithm may be affected by the distinct qualities of the training data, which might result in possible bias in the classification outcomes. The computationally complex nature of the algorithm may see a substantial rise when dealing with bigger datasets or more intricate text structures, which might possibly hinder its capacity to scale effectively. Although the model achieves a high level of accuracy in classifying data, its interpretability is restricted, making it difficult to comprehend the underlying logic within the categories.

Improve the flexibility of the framework to multilingual datasets by exploring strategies such as including language-agnostic features or using cross-lingual embeddings. For the intent of improving the model's performance on specialized tasks or domains, domain adaptation involves doing research into various strategies that may be used to adapt the model to certain domains or kinds of text input. Improvements in the area to Interpretability: In order to give insights into the decision-making process of the model, it is necessary to include methods that may improve the interpretability of the judgments made by the model. Example approaches include attention visualization and feature significance analysis.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. X. Luo and X. Wang, "Research on multi-feature fusion text classification model based on self-attention mechanism," Journal of Physics: Conference Series, vol. 1693, no. 1, p. 012071, Dec. 2020, doi: 10.1088/1742-6596/1693/1/012071.

[2]. Talukdar and S. Kr. Sarma, "Hybrid Model for Efficient Assamese Text Classification using CNN-LSTM," International Journal of Computing and Digital Systems, vol. 14, no. 1, pp. 10183–10192, Oct. 2023, doi: 10.12785/ijcds/140191.

[3]. J. Zhang, F. Liu, W. Xu, and H. Yu, "Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism," Future Internet, vol. 11, no. 11, p. 237, Nov. 2019, doi: 10.3390/fi11110237.

[4]. S. Liu, F. Le, S. Chakraborty, and T. Abdelzaher, "On Exploring Attention-based Explanation for Transformer Models in Text Classification," 2021 IEEE International Conference on Big Data (Big Data), Dec. 2021, doi: 10.1109/bigdata52589.2021.9671639.

[5]. H. Chefer, S. Gur, and L. Wolf, "Generic Attention-model Explainability for Interpreting Bi-Modal and Encoder-Decoder Transformers," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2021, doi: 10.1109/iccv48922.2021.00045

[6]. Hashmi, S. Y. Yayilgan, M. M. Yamin, S. Ali, and M. Abomhara, "Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI," IEEE Access, vol. 12, pp. 44462–44480, 2024, doi: 10.1109/access.2024.3381038.

[7]. H. Sebbaq and N. El Faddouli, "MTBERT-Attention: An Explainable BERT Model based on Multi-Task Learning for Cognitive Text Classification," Scientific African, vol. 21, p. e01799, Sep. 2023, doi: 10.1016/j.sciaf. 2023.e01799.

[8]. L. H. Baniata and S. Kang, "Switching Self-Attention Text Classification Model with Innovative Reverse Positional Encoding for Right-To-Left Languages: A Focus on Arabic Dialects," Feb. 2024, doi: 10.20944/preprints202402. 0332.v1.

[9]. N. R, N. K, S. R, S. Banu S, S. P, and B. P, "Graph-Based Rumor Detection on Social Media Using Posts and Reactions," International Journal of Computing and Digital Systems, vol. 15, no. 1, pp. 173–182, Jul. 2024, doi: 10.12785/ijcds/160114.

[10]. R. Nareshkumar and K. Nimala, "Interactive Deep Neural Network for Aspect-Level Sentiment Analysis," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Jan. 2023, doi: 10.1109/iceconf57129.2023.10083812.

[11]. L. Wang, R. Chen, and L. Li, "Knowledge-Guided Prompt Learning for Few-Shot Text Classification," Electronics, vol. 12, no. 6, p. 1486, Mar. 2023, doi: 10.3390/electronics12061486.

[12]. Z. Wang, "Deep Learning Based Text Classification Methods," Highlights in Science, Engineering and Technology, vol. 34, pp. 238–243, Feb. 2023, doi: 10.54097/hset. v34i.5478.

[13]. M. Y. Hailemariam, S. Lynden, A. Matono, and T. Amagasa, "Self-Attention-based Data Augmentation Method for Text Classification," Proceedings of the 2023 15th International Conference on Machine Learning and Computing, Feb. 2023, doi: 10.1145/3587716.3587779.

[14]. S. Lee, L. Liu, and W. Choi, "Iterative Translation-Based Data Augmentation Method for Text Classification Tasks," IEEE Access, vol. 9, pp. 160437–160445, 2021, doi: 10.1109/access.2021.3131446.

[15]. Y. J. Yu, S. J. Yoon, S. Y. Jun, and J. W. Kim, "TABAS: Text augmentation based on attention score for text classification model," ICT Express, vol. 8, no. 4, pp. 549–554, Dec. 2022, doi: 10.1016/j.icte.2021.11.002.

[16]. T. Cai and X. Zhang, "Imbalanced Text Sentiment Classification Based on Multi-Channel BLTCN-BLSTM Self-Attention," Sensors, vol. 23, no. 4, p. 2257, Feb. 2023, doi: 10.3390/s23042257.

[17]. X. Li and H. Ning, "Deep Pyramid Convolutional Neural Network Integrated with Self-attention Mechanism and Highway Network for Text Classification," Journal of Physics: Conference Series, vol. 1642, no. 1, p. 012008, Sep. 2020, doi: 10.1088/1742-6596/1642/1/012008.

[18]. K. Zahoor, N. Zakaria Bawany, and T. Qamar, "Evaluating text classification with explainable artificial intelligence," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 13, no. 1, p. 278, Mar. 2024, doi: 10.11591/ijai. v13.i1. pp278-286.

[19]. R. Jing, "A Self-attention Based LSTM Network for Text Classification," Journal of Physics: Conference Series, vol. 1207, p. 012008, Apr. 2019, doi: 10.1088/1742-6596/1207/1/012008.

[20]. K. J, A. H, P. V, Y. S, and G. N, "A Study Of Dynamic Thresholds Power Detection Spectrum Sensing Techniques In CRN," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar. 2023, doi: 10.1109/icaccs57279.2023.10112923.

[21]. R. Nareshkumar and K. Nimala, "An Enhanced BERT Model for Depression Detection on Social Media Posts," Lecture Notes in Networks and Systems, pp. 53–64, 2024, doi: 10.1007/978-981-99-8479-4_5.

[22]. R. Nareshkumar and K. Nimala, "An Exploration of Intelligent Deep Learning Models for Fine Grained Aspect-Based Opinion Mining," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Jul. 2022, doi: 10.1109/icses55317.2022.9914094.

[23]. S. Durairaj and R. Sridhar, "Coherent virtual machine provisioning based on balanced optimization using entropy-based conjectured scheduling in cloud environment," Engineering Applications of Artificial Intelligence, vol. 132, p. 108423, Jun. 2024, doi: 10.1016/j.engappai.2024.108423.

[24]. M. Amanullah, S. Thanga Ramya, M. Sudha, V. P. Gladis Pushparathi, A. Haldorai, and B. Pant, "Data sampling approach using heuristic Learning Vector Quantization (LVQ) classifier for software defect prediction," Journal of Intelligent Fuzzy Systems, vol. 44, no. 3, pp. 3867–3876, Mar. 2023, doi: 10.3233/jifs-220480.