Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 5, October 2024, pp. 3334~3344
ISSN: 2302-9285, DOI: 10.11591/eei.v13i5.8240

Hybrid RNNs and USE for enhanced sequential sentence classification in biomedical paper abstracts

Oussama Ndama, Ismail Bensassi, El Mokhtar En-Naimi
Data Science, Artificial Intelligence, and Smart Systems (DSAI2S) Research Team, Computer Science and Smart Systems (C3S) Laboratory, Faculty of Sciences and Technologies Tangier (FSTT), Abdelmalek Essaâdi University, Tetouan, Morocco

Article Info

Article history:

Received Jan 27, 2024
Revised Feb 11, 2024
Accepted Feb 24, 2024

Keywords:

Hybrid models
Information extraction
Recurrent neural networks
Sentence classification
Universal sentence encoder

ABSTRACT

This research evaluates a number of hybrid recurrent neural network (RNN) architectures for classifying sequential sentences in biomedical abstracts. The architectures include long short-term memory (LSTM), bidirectional LSTM (BI-LSTM), gated recurrent unit (GRU), and bidirectional GRU (BI-GRU) models, all of which are combined with the universal sentence encoder (USE). The investigation assesses their efficacy in categorizing sentences into predefined classes: background, objective, method, result, and conclusion. Each RNN variant is used with the pre-trained USE as word embeddings to find complex sequential relationships in biomedical text. Results demonstrate the adaptability and effectiveness of these hybrid architectures in discerning diverse sentence functions. This research addresses the need for improved literature comprehension in biomedicine by employing automated sentence classification techniques, highlighting the significance of advanced hybrid algorithms in enhancing text classification methodologies within biomedical research.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Oussama Ndama
Data Science, Artificial Intelligence, and Smart Systems (DSAI2S) Research Team
Computer Science and Smart Systems (C3S) Laboratory, Faculty of Sciences and Technologies Tangier (FSTT), Abdelmalek Essaâdi University
Tetouan, Morocco
Email: oussama.ndama@etu.uae.ac.ma

1. INTRODUCTION

The exponential surge in biomedical literature presents a significant challenge in efficiently extracting crucial information from scientific articles. While the body of biomedical research grows, many abstracts lack a structured semantic framework, hindering effective data retrieval, and comprehension. Current solutions, including various natural language processing (NLP) techniques, have been explored to address these challenges. However, these approaches often fall short in accurately segmenting and classifying the nuanced information contained within biomedical paper abstracts, due to limitations in understanding context and managing the complexity of scientific language.

In response to these challenges, this research aims to revolutionize the comprehension and accessibility of biomedical paper abstracts through the application of cutting-edge NLP techniques. Leveraging the PubMed 200k randomized controlled trial (RCT) dataset, which comprises 2.3 million phrases from approximately 200,000 RCT abstracts, our study introduces an innovative approach. Each sentence in this dataset is mapped to specific abstract functions: background, objective, method, result, or conclusion, presenting a unique opportunity to enhance information retrieval processes.

The core of our research employs a variety of advanced recurrent neural network (RNN) architectures, including long short-term memory (LSTM), bidirectional LSTM (BI-LSTM), gated recurrent unit (GRU), and bidirectional GRU (BI-GRU), and hybrid models incorporating the universal sentence encoder (USE). These models represent the forefront of modern technology in parsing and understanding complex text structures. By fine-tuning TensorFlow-based models for text segmentation, our project seeks not only to improve the readability and accessibility of scientific content but also to introduce a novel, efficient method for literature review.

The primary goal of this study is to harness the potential of sophisticated RNN technology to categorize abstracts into their respective sections accurately. This endeavor aims to facilitate a deeper understanding and easier access to biomedical literature, thereby addressing the critical gap in current text mining methodologies. Through this research, we aspire to set a new standard for information retrieval in the biomedical domain, making significant strides beyond traditional approaches. To guide the reader, the paper is structured as follows: section 2 provides a comprehensive review of the literature, outlining the current state of NLP applications in biomedical text analysis and identifying gaps our study aims to fill. Section 3 details the research method, including the dataset, model architectures, and training procedures, ensuring reproducibility of our experiments. Section 4 presents the results of our models, followed by a discussion in section 5, where we interpret these findings within the broader context of NLP and biomedical research. Finally, section 6 concludes the paper, summarizing our contributions and suggesting directions for future research.

2. LITERATURE REVIEW

In comparison to prior works, this study introduces a unique emphasis on text segmentation within biomedical literature. While Jiang and Fan [1] redefined the sentence classification problem using a bidirectional encoder representation from transformers (BERT-based) reading comprehension model and Kesiku *et al*. [2] conducted a systematic review addressing challenges in medical text classification, our focus is on the segmentation of abstracts using a spectrum of advanced RNN architectures. Additionally, the study by Naseem *et al*. [3] delved into active learning techniques for biomedical text mining, showcasing the potential of reducing the cost of manual labeling. Our research complements this perspective by offering a novel approach centered on text segmentation, potentially addressing challenges in information retrieval more comprehensively.

As shown by Lee *et al*. [4], Zhang *et al*. [5], the adaptation of pre-trained language models like BERT for biomedical text mining offers useful insights into improving understanding of complex biomedical texts. Building on this foundation, our study explores the application of advanced RNNs, including hybrid models, in sequential sentence classification for biomedical paper abstracts, offering a nuanced approach to the challenges in the field. The benchmark study by Chen *et al*. [6] on biomedical text generation and mining using ChatGPT highlights the broader exploration of NLP technologies. Our study contributes to this exploration by focusing on the specific task of text segmentation in biomedical abstracts, providing a complementary perspective on leveraging advanced computational structures for enhancing the understanding of biomedical literature.

The comparative analysis of text classification approaches in electronic health records (EHRs) by Mascio [7] brings attention to the unique challenges in medical language and lexicon. In contrast, our study extends this discussion by exploring text segmentation in biomedical abstracts, offering insights into how tailored, task-specific approaches can contribute to overcoming challenges in information extraction. Flores and Verschae [8] presented a generic, semi-supervised, and active learning framework tailored for biomedical text classification. The framework used semi-supervised techniques to use samples that weren't labeled, which could have made the data more representative and the classifier more accurate at telling the difference between classes. The approach combined manually annotated examples from active learning with pseudo-labels obtained from a trained classifier. The evaluation of three biomedical datasets related to obesity and smoking habits demonstrated that the framework reduced manual labeling efforts by 10% without compromising classifier performance.

Lastly, Widyantoro *et al*. [9] proposed a multiclass-based classification strategy for rhetorical sentence categorization in scientific papers. The approach employed a multi-classifier method to classify rhetorical sentences, recognizing that no single classifier excels at categorizing all rhetorical sentence types. Evaluation of sixteen rhetorical categories from the ACL-anthology reference corpus (ACL-ARC) paper collection demonstrated that this multiclass-based approach significantly enhances classification performance compared to multi-label classifiers.

3. METHOD

The method section plays a crucial role in explaining a carefully designed framework and procedural tactics to address the task of classifying sentences in a sequential manner in the abstracts of biomedical papers. In this section, we provide a thorough description of our dataset, which includes the PubMed 200k RCT dataset. We also employ rigorous preprocessing approaches to enhance the data's appropriateness for identifying biomedical abstracts.

Our method primarily focuses on carefully choosing advanced RNN architectures, namely LSTM, Bi-LSTM, GRU, and BI-GRU, enhanced with USE. This merger enables our models to effectively capture sequential linkages within biomedical language, effectively tackling the complicated problem of categorizing sentences into predetermined classifications. Furthermore, we elaborate on the complex arrangements specifically designed for dividing text into segments within models based on TensorFlow. The hybrid models incorporate LSTM, Bi-LSTM, GRU, BI-GRU, and the USE. These models are trained meticulously using the Adam optimizer and the categorical cross-entropy loss function.

This section highlights the importance of using these sophisticated models to improve the categorization of sequential information in biomedical literature, which helps to make the process of finding relevant data in biomedical research more efficient. Illustrated in Figure 1, our proposed framework architecture for classifying biomedical paper abstracts visually represents the intricate structure and components of the model. Additionally, it provides a comprehensive overview for researchers and practitioners, aiding in a better understanding of the proposed classification framework.

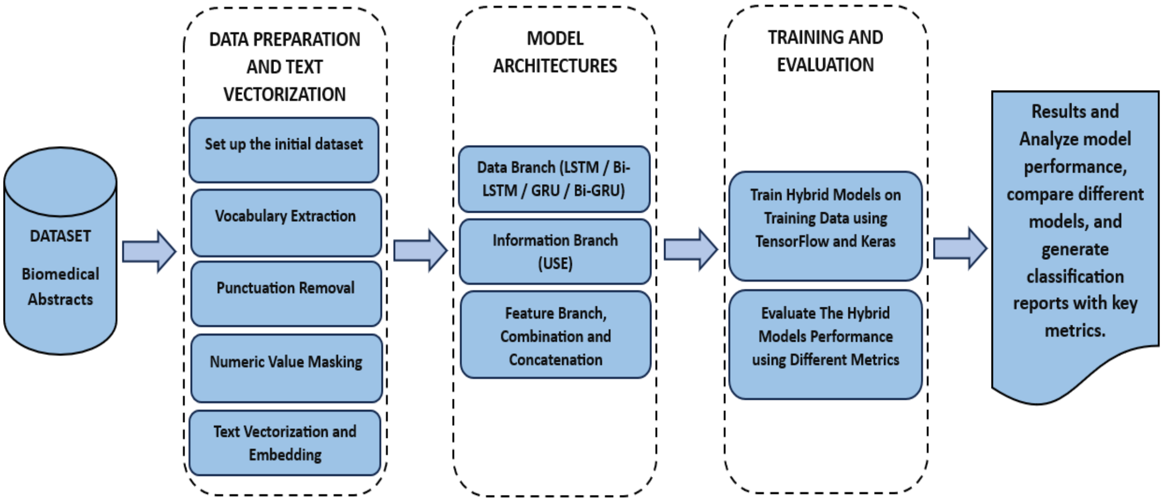

Figure 1. Proposed framework architecture for biomedical paper abstracts classification

3.1. Data preprocessing and text vectorization

We initiated the process by loading the dataset, which is the PubMed 200k RCT, a comprehensive collection comprising 2.3 million phrases extracted from around 200,000 abstracts of RCTs [10]. To facilitate seamless integration into our models, we meticulously employed preprocessing techniques [11]. Initially, we replaced tabs with spaces and removed extraneous newline characters, ensuring uniformity in the dataset's structure. Following this, we strategically handled punctuation, distinguishing between periods and other punctuation marks. This nuanced approach aimed to preserve the semantic meaning encoded in the text while preparing the data for subsequent model input. To further enhance the dataset's cleanliness, we implemented a masking strategy for numerical values. This involved replacing numeric tokens with a standardized "@" symbol, mitigating potential disruptions to the model's understanding of sequential patterns. Beyond mere token-level preprocessing, we conducted an in-depth exploration of the dataset's characteristics. This included analyzing the distribution of abstract lengths and character sequences, providing valuable insights into the data's structural nuances. These insights played a pivotal role in shaping subsequent decisions related to model configuration.

The results of this in-depth exploration were utilized to tailor the preprocessing and modeling strategies to the specific nuances of the biomedical abstracts. Understanding the distribution of abstract lengths and character sequences informed the configuration of our text vectorization and embedding processes, ensuring that our models were optimized for the actual content structure encountered within the dataset. This preparation was crucial for the subsequent application of text vectorization and embedding processes. Using TensorFlow's `TextVectorization` layer, we transformed the refined text into numerical sequences at both character and word levels, adjusting sequence lengths based on the insights gained from the dataset analysis. Character-level vectorization identified unique character identifiers, while word-level vectorization processed textual content. `Embedding` layers then converted these sequences into dense, 256-dimensional vectors, capturing semantic details at multiple linguistic levels. This comprehensive approach, informed by our initial dataset analysis, enabled our hybrid models to classify sentences with high accuracy into categories like background, objective, method, result, and conclusion. The strategic use of dataset characteristics, as presented in Table 1, laid a robust foundation for the effective implementation and performance evaluation of our advanced RNN architectures and hybrid models, ensuring the dataset's alignment with the requirements of sequential sentence classification in biomedical research [12], [13].

Table 1. Example of the preprocessed dataset

Text	Label	Id	Order
This paper describes the design and evaluation…	Background	24491034	2
The program is based on self-efficacy theory…	Methods	24491034	5
The educational program was rated highly…	Results	16042514	10
In postmenopausal women with low bone density...	Conclusions	1922205	11
Bone loss at the median forearm site was...	Results	1922205	8

3.2. Sequential sentence classification

The sequential sentence classification emerges as a pivotal task within the realm of NLP, particularly in domains like biomedical research. This method involves the systematic categorization of individual phrases within a text based on their contextual roles or functions [14]. The significance lies in its ability to organize data within seemingly chaotic textual contexts, such as summaries or essays, by assigning distinct labels to sentences and elucidating their positions within the hierarchical structure of the text [15]. This approach empowers a more nuanced understanding of written content, facilitating targeted examination and the extraction of crucial information. In domains like medicine and healthcare, sequential sentence classification proves indispensable for distilling intricate insights from scientific publications. It streamlines the comprehension process and expedites knowledge access. The categorization technique plays a pivotal role in information extraction by discerning the sequential roles of phrases, contributing to advancements in various disciplines that heavily rely on organized textual analysis [16], [17]. Its impact reverberates across fields, enhancing the efficiency of extracting meaningful insights from vast volumes of text.

3.3. Recurrent neural networks

A RNN stands out as a specialized variant of artificial neural networks uniquely designed to handle sequential data, demonstrating exceptional proficiency in tasks involving language translation, NLP, and time series analysis [18]. Unlike conventional neural networks, RNNs possess an inherent memory mechanism that enables them to consider preceding inputs, proving invaluable in tasks requiring contextual comprehension. The operational principle of RNNs involves iteratively using outputs from previous instances as inputs, coupled with the exchange of parameters within a sequence to amplify efficiency [19]. While RNNs excel at capturing short-term dependencies, challenges arise when dealing with long-range dependencies due to issues like vanishing or exploding gradients [20]. Despite these challenges, RNNs remain foundational in the domain of deep learning, playing a pivotal role in representing diverse sequential data. Their utility extends to comprehending temporal patterns and structures, establishing them as indispensable tools for tasks demanding a nuanced understanding of sequential information. In applications such as biomedical abstract classification, where contextual understanding is paramount, leveraging the capabilities of RNNs becomes integral to achieving accurate and meaningful results.

a. LSTM and BI-LSTM

LSTM is a specialized variant of RNNs designed to overcome the challenges associated with capturing long-range dependencies in sequential data [21]. LSTMs introduce memory cells and gating mechanisms, allowing them to selectively retain and forget information over extended sequences. This architecture proves highly effective in handling biomedical abstracts, where understanding the sequential relationships among sentences is crucial [22]. In our study, LSTM models were employed to grasp the nuanced contextual dependencies within medical text, demonstrating their proficiency in classifying sentences into predefined categories.

BI-LSTM further enhances the capabilities of LSTM by processing the input data in both forward and backward directions [23]. This bidirectional approach enables the model to capture contextual information from both preceding and succeeding sentences, providing a more comprehensive understanding of the sequential structure [24]. In our investigation, BI-LSTM models played a pivotal role in improving the sequential sentence classification task, allowing for a more thorough analysis of the hierarchical relationships within biomedical abstracts. It consists of memory cells and gating mechanisms, which are expressed mathematically as (1) to (6):

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

$$\widetilde{c_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

where f_t, i_t, o_t are the forget, input, and output gates, respectively, $\widetilde{c_t}$ is the candidate cell state, c_t is the cell state, and h_t is the hidden state.

The BI-LSTM processes input data in both forward and backward directions. The forward and backward hidden states are concatenated, providing a more comprehensive understanding of sequential structures [25]. The mathematical representation of BI-LSTM involves two sets of LSTM computations, capturing information from past ($h_t(\rightarrow)$) and future ($h_t(\leftarrow)$) contexts.

$$h_t = [h_t(\rightarrow), h_t(\leftarrow)] \tag{7}$$

b. GRU and BI-GRU

The GRU represents another variant of RNNs designed for sequential data processing. Similar to LSTMs, GRUs incorporate gating mechanisms, facilitating the retention and updating of information over sequential inputs [26], [27]. In our study, GRU models were applied to biomedical abstracts, demonstrating their efficiency in capturing short and long-term dependencies within the text.

Introducing the BI-GRU further enhances the capabilities of GRU by processing data in both directions. This bidirectional approach, similar to BI-LSTM, allows the model to capture contextual information from both past and future sentences [28]. In the context of our research, BI-GRU models proved valuable in enhancing the comprehension of sequential relationships within biomedical literature, contributing to the accurate classification of sentences into predefined categories. Mathematically, GRU operations include:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{8}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{9}$$

$$\widetilde{h_t} = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \tag{10}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t} \tag{11}$$

where z_t and r_t are update and reset gates, respectively, and $\widetilde{h_t}$ is the candidate hidden state.

BI-GRU extends GRU by processing data in both directions, enhancing the model's ability to capture contextual information [29]. The mathematical formulation is analogous to BI-LSTM, incorporating forward and backward GRU computations.

3.4. Universal sentence encoder

USE stands as a significant advancement in the field of NLP, offering a powerful and adaptable mechanism for transforming sentences into vector representations of predetermined dimensions [30]. Developed by Google, USE has shown remarkable effectiveness across various NLP tasks, including text classification, semantic similarity assessment, and sentiment analysis. At its core, the encoder employs a transformer-based architecture, renowned for its ability to discern intricate connections and nuances within sentences [31], [32]. This capability enables the generation of contextually rich embeddings, providing a deep understanding of sentence meanings.

One of the key strengths of USE is its pre-training on a diverse array of data sources, which has significantly bolstered its comprehension across multiple languages and domains. Such extensive pre-training allows USE to capture the essence of texts, from their broad themes down to their subtlest details [33]. This feature is particularly beneficial in tasks demanding a nuanced grasp of context, as USE can adeptly handle sentences of varying lengths without compromising on the semantic depth [34]. In this study, we integrate the USE as a pre-trained word embedding model within various RNN architectures, including LSTM, BI-LSTM, GRU, and BI-GRU, to refine the classification of sequential sentence structures in biomedical paper abstracts. This integration employs a novel approach, as demonstrated in the provided code, where the USE acts as a pivotal layer within an expansive hybrid network configuration.

Our method establishes three distinct branches within the hybrid model: a data branch that processes words through text vectorization and embedding followed by RNN layers; an information branch utilizing the USE to extract high-level semantic embeddings from entire sentences; and a feature branch that adds an analytical dimension based on sentence sequencing. These branches converge, combining the deep semantic embeddings from the USE with the sequential processing capabilities of different RNNs, resulting in a powerful model capable of categorizing sentences into predefined classifications such as background, objective, method, result, and conclusion. By amalgamating the outputs from the token model, order model, and sentence model—each offering unique insights into textual structure and meaning—the approach achieves a synergistic improvement in classification accuracy. This innovative amalgamation highlights the study's application of advanced hybrid algorithms, blending the semantic analysis depth provided by USE with the dynamic sequence modeling of various RNNs to address the complex challenge of biomedical literature comprehension.

3.5. Hybrid model architecture

Our innovative hybrid model, strategically combines advanced technologies to address the complexities of sequential sentence classification within biomedical paper abstracts. The architecture comprises three key branches: a data branch responsible for processing text data at the sentence level using text vectorization, embedding, and different RNNs, an information branch for capturing contextual information using USE, and a feature branch handling numerical data.

a. Data branch
   - Input: the model takes sentence data as input, processed through text vectorization and embedding layers to capture semantic information.
   - Processing: utilizing LSTM/Bi-LSTM/GRU/BI-GRU, the model captures sequential dependencies and nuanced patterns within the embedded text data.
   - Output: the output from this branch serves as a foundation for understanding the sequential structure of biomedical abstracts.

b. Information branch
   - Input: this branch receives additional text data and utilizes the pre-trained USE to extract contextual information efficiently.
   - Processing: dense layers refine features, and dropout layers prevent overfitting, contributing to the model's comprehension of the abstracts' contextual nuances.
   - Output: the output from this branch enhances the model's ability to grasp the broader context of biomedical information.

c. Feature branch
   - Input: the numerical branch processes additional numerical data, contributing to a holistic understanding of abstracts.
   - Processing: dense layers handle numerical information, enhancing the model's capability to consider diverse data types.
   - Output: the output from this branch enriches the overall feature set, aiding in comprehensive sequential sentence classification.

As shown in Figure 2, our proposed hybrid model architecture incorporates three key branches that collectively contribute to the model's nuanced understanding of sequential information within biomedical literature. Further dense layers refine features, and dropout layers prevent overfitting, ensuring an effective fusion of information. The output layer, employing a softmax activation function, classifies sentences into predefined categories such as background, objective, method, result, and conclusion [35]. The hybrid model is trained using the Adam optimizer and categorical cross-entropy loss function, providing a sophisticated solution for efficient text segmentation and classification [36], [37].

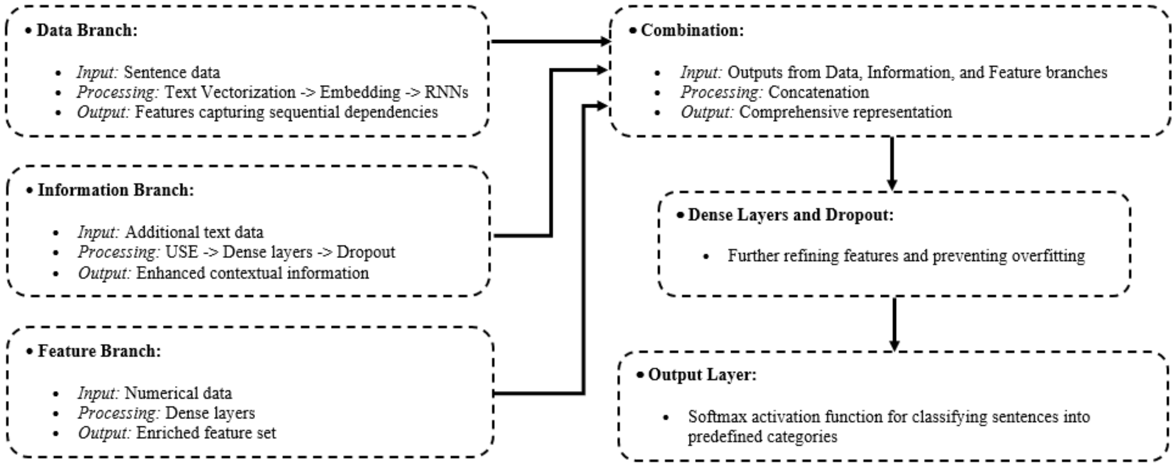

Figure 2. The hybrid model architecture

4. TRAINING AND EVALUATION

In our study, aimed at evaluating the performance of hybrid models for sequential sentence classification in biomedical paper abstracts, we meticulously divided our dataset into training, validation, and test sets, adopting a balanced distribution that facilitates both comprehensive learning and unbiased evaluation. Initially, the dataset underwent an 80-20 split, designating 80% to training and 20% to testing. To further enhance the training process and allow for precise hyperparameter tuning, we subdivided the training set with an additional 75-25 split, resulting in a final distribution of 60% for training, 20% for validation, and 20% for testing.

To evaluate our models, we employ key metrics, including precision, recall, F1 score, and accuracy. These metrics play a pivotal role in assessing the models' effectiveness across designated categories such as background, objective, method, result, and conclusion. By analyzing precision, we gauge the accuracy of positive predictions, while recall assesses the completeness of correctly identified instances. Additionally, F1 score serves as a harmonized metric that balances precision and recall, offering a comprehensive assessment of classification accuracy within the specified categories [38], [39].

Precision, indicating the accuracy of positive predictions, is calculated as the ratio of true positives to the sum of true positives and false positives:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{12}$$

The ratio of true positives to the sum of true positives and false negatives serves as a measure of recall, which evaluates the completeness of correctly identified instances:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{13}$$

The F1 score, a harmonized metric balancing precision and recall, offers a comprehensive assessment of classification accuracy:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

Accuracy, representing the overall correctness of predictions, is calculated as the ratio of correct predictions to the total number of predictions:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \tag{15}$$

5. RESULTS AND DISCUSSION

The results and discussion section presents a comprehensive exploration of the performance of various hybrid models developed for sequential sentence classification in biomedical abstracts. The evaluation focuses on key metrics, with a central emphasis on accuracy at first, to provide insights into the efficacy of each model. Table 2 showcases the accuracy scores of distinct hybrid models, each integrating LSTM, BI-LSTM, GRU, or BI-GRU with USE. These models play a pivotal role in enhancing text segmentation and classification within the intricate domain of biomedical literature.

Table 2. Accuracy of different hybrid models

Hybrid model	Accuracy
LSTM-USE	**0.8841**
BILSTM-USE	0.8769
GRU-USE	0.8758
BIGRU-USE	0.8707

The results obtained indicate noticeable differences in accuracy metrics among the various hybrid models, highlighting their unique abilities in the complex task of sequential sentence classification. The LSTM-USE model is the most accurate among the models considered, achieving an accuracy score of 0.8841. This result highlights the model's skill in accurately capturing the sequential relationships and complex patterns that are naturally present in biomedical text data. The BI-LSTM-USE model achieves an accuracy of 0.8769 while closely following, highlighting the significant impact of bidirectional processing in enhancing contextual understanding. In addition, the GRU-USE and BI-GRU-USE models obtain impressive accuracies of 0.8758 and 0.8707, respectively, indicating the strong and reliable performance of GRU architectures in this field.

Table 3 provides an in-depth examination of the macro and weighted average outcomes for four different hybrid models, revealing their performance across many evaluation metrics. The LSTM-USE model demonstrates strong performance in recognizing sequential dependencies within biomedical text data, as evidenced by its robust macro average precision, recall, and F1-score values of 0.8429, 0.8325, and 0.8363. The weighted average measures, with precision at 0.8838, recall at 0.8841, and F1-score at 0.8833, highlight the model's ability to accurately classify data into specified categories. The BI-LSTM-USE model exhibits powerful macro average precision of 0.8371, recall of 0.8212, and F1-score of 0.8248, which greatly enhances its overall accuracy. The weighted average metrics of 0.8765, 0.8758, and 0.8744 for precision, recall, and F1-score, respectively, highlight the model's effectiveness. Notably, it is seen that bidirectional processing does not significantly enhance accuracy in these models. This highlights the subtle complexity of improving models and underscores the importance of making thoughtful decisions regarding architectural choices. The GRU-USE model demonstrates its proficiency with macro average precision, recall, and F1-score values of 0.8354, 0.8247, and 0.8277. The model's overall performance is further confirmed by the related weighted average metrics of 0.8774, 0.8770, and 0.8761. The BI-GRU-USE model achieves macro average metrics with a precision of 0.8280, recall of 0.8193, and F1-score of 0.8205. The weighted average metrics for precision, recall, and F1-score are 0.8718, 0.8707, and 0.8699, respectively.

Table 3. Macro and weighted average results of different hybrid models

Metric	Average	LSTM-USE	BILSTM-USE	GRU-USE	BIGRU-USE
Precision	Macro-average	0.8429	0.8371	0.8354	0.8280
	Weighted-average	0.8838	0.8765	0.8774	0.8718
Recall	Macro-average	0.8325	0.8212	0.8247	0.8193
	Weighted-average	0.8841	0.8758	0.8770	0.8707
F1 score	Macro-average	0.8363	0.8248	0.8277	0.8205
	Weighted-average	0.8833	0.8744	0.8761	0.8699

Table 4 offers a detailed insight into the class-wise performance metrics of the best-performing hybrid model, LSTM-USE. Precision, recall, and F1-score metrics are presented for each predefined category, providing a granular evaluation of the model's capabilities. These class-wise metrics contribute significantly to understanding the model's performance variations across different categories. It is evident that the LSTM-USE model excels in certain categories, such as methods and results, where precision, recall, and F1-score are notably high. The model's ability to accurately classify information in the conclusions category is also commendable. However, some challenges are observed in the objective and background categories, where recall and F1-score are comparatively lower.

Table 4. Class-wise performance of the best performing hybrid model (LSTM-USE)

Class	Precision	Recall	F1 Score
Background	0.7181	0.7863	0.7507
Conclusions	0.8951	0.8796	0.8873
Methods	0.9193	0.9445	0.9317
Objective	0.7672	0.6420	0.6990
Results	0.9151	0.9101	0.9126

Our study has demonstrated the significant potential of advanced NLP techniques, with a particular emphasis on hybrid models like LSTM-USE, in enhancing the accessibility and readability of biomedical literature. The LSTM-USE model, in particular, showcases robust performance across various categories, indicating its potential for effective sequential sentence classification in biomedical texts. This performance is highlighted by variations in class-wise metrics, providing valuable insights for further refining the model to enhance its overall effectiveness and address specific challenges within each category. Such a nuanced evaluation underscores the importance of considering both global performance and class-specific metrics for a comprehensive understanding of the model's capabilities. The LSTM-USE model, with its notable accuracy and detailed class-wise performance, not only stands out but also paves the way for continued innovation in utilizing NLP within biomedical research.

6. CONCLUSION

In conclusion, our exploration into advanced NLP techniques, particularly through the deployment of hybrid models like LSTM-USE, underscores a significant advancement in enhancing the comprehension and accessibility of biomedical literature. The robust performance of the LSTM-USE model across various classification categories highlights its potential for effective sequential sentence classification in biomedical abstracts, promising a substantial improvement in how biomedical literature can be accessed and understood. This achievement, derived from a nuanced evaluation that considered both overall and category-specific metrics, provides valuable insights into optimizing model performance and addressing the unique challenges of the biomedical domain. The integration of LSTM, BI-LSTM, GRU, and BI-GRU architectures with the USE has not only demonstrated varying degrees of efficacy across different models but also illustrated the potential applications and extensions of our methodologies beyond the initial scope of the PubMed 200k RCT dataset. This suggests the possibility of broadening the impact of our findings by applying these techniques to a wider array of biomedical texts, thereby enhancing the robustness and applicability of our research and contributing to the ongoing discourse on NLP's pivotal role in biomedical research.

Looking forward, the potential for further exploration is vast. Future research could extend the application of our models to diverse datasets, explore alternative neural network architectures, and integrate different pre-trained embeddings to uncover additional improvements in model performance. Such endeavors promise not only to refine the capabilities of NLP technologies but also to expand their application in navigating the complex landscape of biomedical literature. By continuing to innovate and explore, we aim to enrich the tapestry of NLP applications, facilitating more efficient and effective access to biomedical knowledge and ultimately advancing the field of biomedical research. Our study lays the groundwork for these advancements, emphasizing the importance of detailed model evaluations and the pursuit of sophisticated, context-aware models that can meet the evolving challenges of the biomedical domain.

REFERENCES

[1] C.-Y. Jiang and Y.-C. Fan, "Biomedical Abstract Sentence Classification by BERT-Based Reading Comprehension," *SN Computer Science*, vol. 4, no. 4, p. 395, May 2023, doi: 10.1007/s42979-023-01830-0.

[2] C. Y. Kesiku, A. Chaves-Villota, and B. Garcia-Zapirain, "Natural Language Processing Techniques for Text Classification of Biomedical Documents: A Systematic Review," *Information*, vol. 13, no. 10, p. 499, Oct. 2022, doi: 10.3390/info13100499.

[3] U. Naseem, M. Khushi, S. K. Khan, K. Shaukat, and M. A. Moni, "A Comparative Analysis of Active Learning for Biomedical Text Mining," *Applied System Innovation*, vol. 4, no. 1, p. 23, Mar. 2021, doi: 10.3390/asi4010023.

[4] J. Lee *et al*., "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Sep. 2019, doi: 10.1093/bioinformatics/btz682.

[5] X. Zhang, X. Song, A. Feng, and Z. Gao, "Multi-Self-Attention for Aspect Category Detection and Biomedical Multilabel Text Classification with BERT," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–6, Nov. 2021, doi: 10.1155/2021/6658520.

[6] Q. Chen *et al*., "An extensive benchmark study on biomedical text generation and mining with ChatGPT," *Bioinformatics*, vol. 39, no. 9, Sep. 2023, doi: 10.1093/bioinformatics/btad557.

[7] A. Mascio, "Comparative Analysis of Text Classification Approaches in Electronic Health Records," *arXiv*, May 2020, doi: 10.48550/arXiv.2005.06624.

[8] C. A. Flores and R. Verschae, "A Generic Semi-Supervised and Active Learning Framework for Biomedical Text Classification," *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, Scotland,

United Kingdom, 2022, pp. 4445-4448, doi: 10.1109/EMBC48229.2022.9871846.
[9] D. H. Widyantoro, M. L. Khodra, B. Riyanto, and E. A. Aziz, "A Multiclass-based Classification Strategy for Rhetorical Sentence Categorization from Scientific Papers," *Journal of ICT Research and Applications*, vol. 7, no. 3, pp. 235–249, Dec. 2013, doi: 10.5614/itbj.ict.res.appl.2013.7.3.5.
[10] F. Dernoncourt, "PubMed 200k RCT: a Dataset for Sequential Sentence Classification in Medical Abstracts," *arXiv*, Oct. 2017, doi: 10.48550/arXiv.1710.06071.
[11] C. P. Chai, "Comparison of text preprocessing methods," *Natural Language Engineering*, vol. 29, no. 3, pp. 509–553, Jun. 2022, doi: 10.1017/s1351324922000213.
[12] S. Hemtanon, K. Phetkrachang, and W. Yangyuen, "Classification and keyword extraction of online harassment text in Thai social network," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3837–3842, Dec. 2023, doi: 10.11591/eei.v12i6.5939.
[13] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organizational Research Methods*, vol. 25, no. 1, pp. 114–146, Nov. 2020, doi: 10.1177/1094428120971683.
[14] A. Hassan and A. Mahmood, "Convolutional Recurrent Deep Learning Model for Sentence Classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018, doi: 10.1109/access.2018.2814818.
[15] I. Alsmadi and K. H. Gan, "Review of short-text classification," *International Journal of Web Information Systems*, vol. 15, no. 2, pp. 155–182, Jun. 2019, doi: 10.1108/ijwis-12-2017-0083.
[16] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–40, Apr. 2021, doi: 10.1145/3439726.
[17] J. Liu, H. Ma, X. Xie, and J. Cheng, "Short Text Classification for Faults Information of Secondary Equipment Based on Convolutional Neural Networks," *Energies*, vol. 15, no. 7, p. 2400, Mar. 2022, doi: 10.3390/en15072400.
[18] Z. Z., P. A. E. A., and H. M. H., "Predicting machine failure using recurrent neural network-gated recurrent unit (RNN-GRU) through time series data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 870–878, Apr. 2021, doi: 10.11591/eei.v10i2.2036.
[19] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306.
[20] H. Salehinejad, "Recent Advances in Recurrent Neural Networks," *arXiv*, Dec. 29, 2017, doi: 10.48550/arXiv.1801.01078.
[21] I. Triyadi, B. Prasetiyo, and T. L. Nikmah, "News text classification using Long-Term Short Memory (LSTM) algorithm," *Journal of Soft Computing Exploration*, vol. 4, no. 2, May 2023, doi: 10.52465/joscex.v4i2.136.
[22] G. Dou, K. Zhao, M. Guo, and J. Mou, "Memristor-based LSTM network for text classification," *Fractals*, vol. 31, no. 06, Jan. 2023, doi: 10.1142/s0218348x23400406.
[23] M. P. Kantipudi, S. Kumar, and A. K. Jha, "Scene Text Recognition Based on Bidirectional LSTM and Deep Neural Network," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–11, Nov. 2021, doi: 10.1155/2021/2676780.
[24] R. Yunida *et al*., "LSTM and Bi-LSTM Models For Identifying Natural Disasters Reports From Social Media," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 4, pp. 241–249, Sep. 2023, doi: 10.35882/jeeemi.v5i4.319.
[25] J. Jiang *et al*., "Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization," *IEEE Access*, vol. 9, pp. 123660–123671, 2021, doi: 10.1109/access.2021.3110143.
[26] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, "Efficient processing of GRU based on word embedding for text classification," *JOIV: International Journal on Informatics Visualization*, vol. 3, no. 4, pp. 377–383, Nov. 2019, doi: 10.30630/joiv.3.4.289.
[27] P. Eswaraiah and H. Syed, "A Hybrid Deep Learning GRU based Approach for Text Classification using Word Embedding," *EAI Endorsed Transactions on Internet of Things*, vol. 10, Dec. 2023, doi: 10.4108/eetiot.4590.
[28] Q. Tang, J. li, J. Chen, H. Lu, Y. Du and K. Yang, "Full Attention-Based Bi-GRU Neural Network for News Text Classification," *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2019, pp. 1970-1974, doi: 10.1109/ICCC47050.2019.9064061.
[29] E. Ahmadzadeh, H. Kim, O. Jeong, N. Kim, and I. Moon, "A Deep Bidirectional LSTM-GRU Network Model for Automated Ciphertext Classification," *IEEE Access*, vol. 10, pp. 3228–3237, 2022, doi: 10.1109/access.2022.3140342.
[30] D. Cer, "Universal Sentence Encoder," *arXiv.org*, Mar. 2018, doi: 10.48550/arXiv.1803.11175.
[31] I. E. Fattoh, F. K. Alsheref, W. M. Ead, and A. M. Youssef, "Semantic Sentiment Classification for COVID-19 Tweets Using Universal Sentence Encoder," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/6354543.
[32] O. V. Orlovskiy, K. Sohrab, S. E. Ostapov, K. P. Hazdyuk, and L. M. Shumylyak, "Multilingual Text Classifier Using Pre-Trained Universal Sentence Encoder Model," *Radio Electronics, Computer Science, Control*, no. 3, p. 102, Oct. 2022, doi: 10.15588/1607-3274-2022-3-10.
[33] D. Cer *et al*., "Universal Sentence Encoder for English," *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2018, pp. 169-174, doi: 10.18653/v1/d18-2029.
[34] Y. Yang, "Multilingual Universal Sentence Encoder for Semantic Retrieval," *arXiv*, Jul. 2019, doi: 10.48550/arXiv.1907.04307.
[35] A. N. Holm, D. Wright, and I. Augenstein, "Revisiting Softmax for Uncertainty Approximation in Text Classification," *Information*, vol. 14, no. 7, p. 420, Jul. 2023, doi: 10.3390/info14070420.
[36] K. K. Chandriah and R. V. Naraganahalli, "RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting," *Multimedia Tools and Applications*, Apr. 2021, doi: 10.1007/s11042-021-10913-0.
[37] C. -H. Chen, P. -H. Lin, J. -G. Hsieh, S. -L. Cheng and J. -H. Jeng, "Robust Multi-Class Classification Using Linearly Scored Categorical Cross-Entropy," *2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, Kaohsiung, Taiwan, 2020, pp. 200-203, doi: 10.1109/ICKII50300.2020.9318835.
[38] M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.
[39] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/ijacsa.2021.012067.

BIOGRAPHIES OF AUTHORS

Oussama Ndama is a Ph.D. student in Data Science, Artificial Intelligence and Smart Systems Research Team (DSAI2S), C3S Laboratory, Faculty of Sciences and Technologies (FST), Tangier, Morocco. He had his master in Computer Science and Big Data, Laureate of FST of Tangier. He is also a Business Intelligence Engineer with more than 5 years of experience in different multinational companies. The research topics of interest are smart systems, machine learning, deep learning, NLP, ANN, sentiment analysis, and smart cities. He can be contacted at email: oussama.ndama@etu.uae.ac.ma.

Ismail Bensassi is a Ph.D. student in Data Science, Artificial Intelligence and Smart Systems Research Team (DSAI2S), C3S Laboratory, Faculty of Sciences and Technologies (FST), Tangier, Morocco. He is an Engineer in Computer Science, Laureate of FST of Tangier. The research topics of interest are smart connection of user profiles in a big data context, multi-agent systems (MAS), case-based reasoning (CBR), ontology, machine learning, smart cities, and eLearning/MOOC/SPOC. He can be contacted at email: bensassi.ismail@gmail.com.

El Mokhtar En-Naimi is a Full Professor in the University of Abdelmalek Essaâdi (UAE), Faculty of Sciences and Technologies of Tangier (FSTT), Department of Computer Sciences. He was Temporary Professor: from 1999 to 2003 and Permanent Professor: since 2003/2004 until now. Actually, he is a Full Professor in UAE, FST of Tangier. He was a Head of Department of Computer Sciences, since October 2016 until the end of December 2020. He was responsible for a Licence of Science and Technology, LST Computer Engineering (“Licence LST-GI”), from January 2012 to October 2016. He is a Chief of Data Science, Artificial Intelligence, and Smart Systems (DSAI2S) Research Team since the academic year 2022/2023. He is also a founding member of the Both Laboratories: LIST (Laboratoire d'Informatique, Systèmes et Télécommunications) Laboratory (From 2008 To 2022) and C3S (Computer Science and Smart Systems) Laboratory since the academic year 2022/2023 until now, the University of Abdelmalek Essaâdi, FST of Tangier, Morocco. He is also an expert evaluator with the ANEAQ, since the academic year 2016/2017 until now, that an Expert of the Private Establishments belonging to the territory of the UAE and also an Expert of the Initial or Fundamental Formations and Formations Continuous at the Ministry of Higher Education, Scientific Research and Executive Training and also at the UAE University and the FST Tangier since 2012/2013 until Now. He is an author/co-authors of several articles, published in The International Journals in Computer Sciences, in particular, in multi-agent systems (MAS), cases based reasoning (CBR), artificial intelligent (AI), machine learning (ML), deep learning (DL), eLearning, MOOC/SPOC, big data, data-mining, wireless sensor network, VANet, MANet, and smart city. He was/is also Director of several Doctoral Theses in Computer Sciences. He has too served as a general chair, technical program chair, technical program committee member, organizing committee member, session chair, and reviewer for many international conferences and workshops. In addition, he is an associate member of the ISCN-Institute of Complex Systems in Normandy, the University of the Havre, France, since 2009 until now. He can be contacted at email: en-naimi@uae.ac.ma.