Advanced Deep Learning Architectures for Automated Text Classification in Natural Language Processing

Abdulmohsen Algarni[*], Muhammad Anas[**], Hussein Khan[***] Muhammad Assam[****]

[*] Department of Computer Science, King Khalid University, Abha 61421, Saudi Arabia
[**] Department of Electrical and Electronic Engineering, Swinburne University of Technology Hawthorn Campus.
[***] Department of Electrical Engineering, Qatar University, Qatar
[****] Department of Software Engineering, University of Science and Technology Bannu KP Pakistan

Abstract- The primary objective of text classification in the field of Natural Language Processing (NLP) is to methodically categorize a wide range of textual input, including sentences, documents, and queries. The increasing prevalence of large digital document collections, especially in business environments seeking to enhance efficiency and profitability, has emphasized the growing importance of text classification procedures. This study aims to enhance the field of automated text classification methodologies by utilizing advanced Deep Learning (DL) frameworks such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Bi-Directional Encoder Representation from Transformers (BERT). The research carefully deals with the pretreatment of the text database by removing unnecessary special characters and non-relevant stop words, using two modern and publicly accessible databases. The subsequent process of dividing the text into tokens and providing it to the deep learning frameworks outlined before enables effective extraction of important characteristics and accurate categorization. A sequence of experimental rounds is performed to determine the most effective architectural configurations and hyperparameter values, resulting in the development of a final design that achieves an impressive validation accuracy of 98%. This architecture is ready to be used in real-time text categorization scenarios, proving its effectiveness as a powerful tool in the current field of Natural Language Processing (NLP).

Index Terms- Text Classification, LSTM, Deep Learning, BERT and Convolutional Neural Network.

I. Introduction

TExt Text categorization is a crucial task in the fields of Natural Language Processing (NLP) and Data Mining. It is becoming increasingly important because to the large amounts of textual data available online. The wide-ranging uses of this technology include several fields such as classifying news, analysing sentiment, labelling topics, and extracting information from text databases. The widespread availability of the internet and improvements in technology have led to an unprecedented increase in the creation of written content on different digital platforms, making text a valuable source of knowledge. Yet, the inherent lack of organization in languages is a difficult obstacle in extracting valuable information from written material, often requiring arduous and time-consuming manual analysis by human specialists [1].

Historically, the categorization of written information mainly depended on human involvement, which had significant drawbacks such as the likelihood of mistakes and the impracticality of managing large amounts of data within realistic time constraints. To address these issues, academics have strived to create automatic text classification approaches by harnessing the capabilities of Artificial Intelligence (AI). These automated systems try to allocate predetermined labels to input text, thus simplifying the categorization process and bypassing the constraints linked to manual intervention [2].

The significance of automated text categorization approaches has become extremely important due to the rapid increase in the quantity and variety of text data in various industries. Modern methods for automatic text classification involve both rule-based and machine learning (ML) approaches. Rule-based techniques involve the creation and implementation of predetermined sets of rules to classify text into appropriate categories. Although these strategies can be helpful in specific situations, they need a deep comprehension of the underlying subject matter and may be insufficient when dealing with complex or changing datasets [3, 4].

On the other hand, ML-based methodologies bring about a fundamental change in the way we approach problems by relying on data-driven approaches. In this approach, algorithms are taught to identify the connections between textual features and their corresponding labels. These algorithms improve their ability to classify by continuously learning and adjusting, eliminating the requirement for explicit rule creation and demonstrating better performance compared to rule-based techniques [5]. As a result, academics have been more inclined to investigate deep learning (DL) methods in recent years. These methods utilize the intrinsic ability of neural networks to extract complex patterns and representations from textual input [6]. This study focuses on automated text classification, examining the progression of methodology from rule-based to machine learning-based approaches, and investigating the emerging field of deep learning-based techniques. Our goal is to utilize AI to improve text classification methods, resulting in increased efficiency, accuracy, and scalability in analysing textual data across many fields.

Deep Learning (DL) is a subset of classic Machine Learning (ML) and Neural Network (NN) methods that stands out for its exceptional capacity to surpass conventional algorithms. The superiority of DL arises from its inherent ability to automate the process of feature extraction, which is a crucial element in text categorization jobs. DL algorithms, unlike traditional algorithms, do not require manual feature engineering. They are distinguished by their multi-layered architectures that consist of hidden neurons. These designs function similarly to the human brain, acquiring complex relationships between data patterns through multiple layers of abstraction. As a result, DL frameworks are able to autonomously perform the feature extraction process, reducing the requirement for manual intervention and specialized knowledge in a certain field [7].

The hierarchical organization of DL systems enables the extraction of distinguishing characteristics at different levels of abstraction. The early layers of the network excel in detecting simple visual elements like edges, textures, and basic language patterns. As the layers go deeper, they gradually extract more advanced visual features that represent intricate semantic ideas and contextual connections. The hierarchical feature extraction technique of DL algorithms provides them with exceptional flexibility and adaptability, allowing them to accurately identify subtle subtleties in textual material [8].

Automated methods for text classification have become extremely important in modern commercial and enterprise environments due to the rapid increase in digital textual data. Text categorization is utilized in various domains like as sentiment analysis, question-answering systems, email filtering (spam/ham detection), and other applications. By capitalizing on technological improvements, some research have suggested automated frameworks for text classification that utilize the capabilities of artificial intelligence. An example of a text classification framework utilizing Multilayer Perceptron attained a notable accuracy of 71% on a dataset consisting of text. Nevertheless, even with the implementation of deep learning frameworks, there is still significant potential for improving the efficiency of text classification systems [9].

An important constraint seen in deep learning-based text categorization relates to the efficient encoding of extensive contextual information over extended periods, particularly in jobs using sequential input like spoken language. In order to tackle this difficulty, scholars have suggested hybrid methodologies that combine Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. These hybrid architectures aim to surpass the limitations of individual models and achieve superior performance in text classification tasks by combining the strengths of Convolutional Neural Networks (CNNs) in capturing local contextual information with the memory retention capabilities of Long Short-Term Memory (LSTM) models [10].

Furthermore, scholars have investigated the suitability of text categorization techniques in several languages, surpassing the limitations of datasets limited to the English language. An example is a research project that concentrated on categorizing Urdu editorials using Naïve Bayes and obtained encouraging outcomes, especially when rare phrases were excluded from the dataset. Nevertheless, in order to improve the effectiveness of these models in practical situations, it is crucial to broaden their ability to be used with various types of text and domains. Moreover, studies have been conducted to explore the practicality of utilising sophisticated pre-trained models like BERT (Bidirectional Encoder Representations from Transformers) and MARBERT (Multilingual ARabic BERT) on short texts such as tweets. Although these models demonstrate exceptional performance, their computational burden presents a substantial obstacle, requiring optimization methodologies for efficient implementation in real-time applications.

Researchers have thoroughly examined the incorporation of artificial intelligence (AI) in the field of news categorization, exploring various approaches such as ensemble learning techniques, which achieved a noteworthy accuracy of 87% [10], as well as the assessment of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) on news datasets [11]. In addition, attempts have been undertaken to utilize Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) structures using Czech news data, resulting in encouraging outcomes with an F1 score of 0.84 [12]. However, these efforts mainly revolve around specific forms of text categorization, with a primary emphasis on news items. Therefore, there is still a notable gap in creating a thorough and flexible framework for classifying text that utilizes AI to effectively manage various forms of textual data.

This manuscript addresses the existing gap by introducing an automated text classification framework that utilizes advanced Deep Learning (DL) algorithms. This study suggests a comprehensive strategy with multiple significant contributions by using two prominent text categorization databases for assessment. The framework presents a comprehensive text classification pipeline that encompasses five essential stages: text preprocessing, tokenization, training and validation of a deep learning framework, and a thorough comparison with existing systems. The system showcases its adaptability by utilizing various deep learning architectures, such as Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Bidirectional Encoder Representations from Transformers (BERT), on diverse text sources, including email and news datasets.

The validation trials highlight the effectiveness of the framework, as it achieves an impressive maximum accuracy of 98%. This confirms its potential for immediate implementation in many text categorization situations. The suggested strategy, which incorporates different DL frameworks, demonstrates greater performance and adaptability when compared to existing methodologies. It efficiently handles diverse forms of textual data. Overall, this research makes a substantial contribution to the development of text classification methods. It introduces a strong and flexible framework that effectively utilizes deep learning techniques to accurately and efficiently classify various types of textual data.

II. The Acquisition and Preparation of Datasets

For two distinct analyses, we utilized textual data gathered from publicly accessible online sources. The initial study utilized a dataset of emails categorized as "ham" or "spam." Within this context, the term "ham" refers to genuine emails, while "spam" pertains to counterfeit or suspicious emails. The primary objective

was to accurately categorize the text data obtained from these communications [13]. For the second phase of our research, we utilized the comprehensive BBC news dataset, encompassing stories from five distinct domains: sports, entertainment, politics, business, and technology. The objective was to accurately classify the textual material into specific news groups [14].
Common stop words and special symbols were found in the text databases utilized in both studies. These elements typically have no effect on text classification systems, but if they are not addressed, they could impact the overall performance of the proposed framework. As part of our professional research process, we took the necessary steps to remove these components before proceeding with data processing. In order to achieve the objective of obtaining a more refined dataset, unnecessary words and special characters were eliminated. In addition, we standardized the case of the entire English text to eliminate any potential biases caused by differences in letter casing. The text databases were prepared for the next stages of research through essential preprocessing processes. These processes played a crucial role in achieving more accurate and robust classification results [15].

III. Enhanced Preprocessing through Tokenization

Tokenization is an essential initial step in natural language processing (NLP) that helps convert unprocessed text material into a well-organized format that can be easily understood by machines. Tokenization allows machines to comprehend the specific elements and the contextual connections within the text by breaking down sentences into their constituent parts. Having a detailed knowledge is crucial because it enables algorithms to not only understand the meanings of individual words but also their functions and importance within the larger textual context. The capacity to examine the frequencies of words and their positional data is crucial for later phases of NLP research.

In this study, we used a word-level tokenizer during the tokenization process. This allowed us to divide the text data into separate tokens that represent specific words. The tokenizer, which is incorporated through the Kera's library, is highly skilled at analysing textual inputs and producing a series of tokens that represent each word. Afterwards, these tokens are given distinct numerical identities and transformed into sequential representations prior to being fed into deep learning (DL) algorithms for the goal of categorization. Through the utilization of tokenization, we efficiently preprocess the textual input, converting it into a structured format that is suitable for examination by deep learning algorithms. This methodological approach improves the understandability and computing efficiency of following NLP tasks. It also emphasizes the significance of careful preprocessing in attaining precise and reliable classification results.

IV. Developed Deep Learning Model Framework

A. Long Short-Term Memory LSTM

Long Short-Term Memory (LSTM) is an adaptable artificial intelligence framework that is known for its feedback links, which allow it to effectively handle many types of input, such as sequences seen in text, films, and audios. LSTM, known for its proficiency in managing both long and short-term memory, has been well recognized in the field of sequence-based data classification. In contrast to conventional feedforward networks, LSTM networks experience modifications to their weights and biases during each training session, resembling the synaptic alterations involved in the storage of long-term memories. Additionally, the activation patterns of LSTM networks change over time, similar to the variations observed in short-term memory in the brain. The fundamental components of LSTM architecture consist of three gates: the input gate, the output gate, and the forget gate. These gates control the flow of information into, out of, and inside the cell, respectively. Through the selective retention or discarding of information, these gates augment the memory retention capabilities of LSTM in comparison to traditional RNN designs. The schematic depiction of the LSTM architecture, as shown in Figure 1, depicts its constituent elements and operational process. The suggested LSTM architecture includes embedding layers, which enable data entry through tokenization. Afterwards, the LSTM layer, consisting of 16 cells, analyses the embedded data and extracts important properties. The Fully Connected layer retains the retrieved features, while the dropout layer contributes to network regularization by randomly deactivating neurons. Ultimately, the SoftMax layer carries out the task of classification, marking the end of the text categorization process.

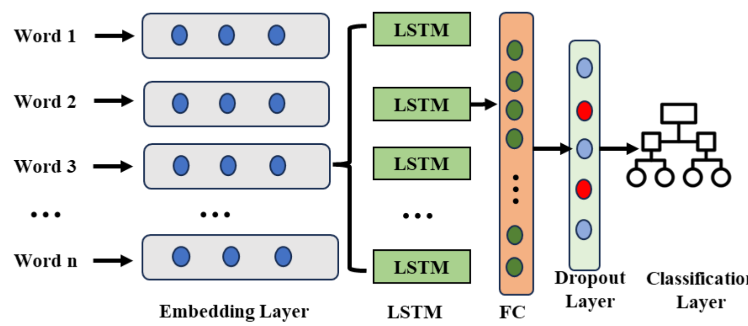

Figure 1. LSTM Proposed Architecture

B. Convolutional Neural Network

Deep Learning (DL) frameworks have led to the use of Convolutional Neural Networks (CNNs) in text classification research applications. CNNs are designed to imitate the arrangement of sensory neurons in the human brain and visual cortex, which allows for effective processing of information [17]. The architecture of Convolutional Neural Networks (CNNs) consists of convolutional, fully connected, and pooling layers, which enable the extraction of profound information from textual material. Within a Convolutional Neural Network (CNN), individual layers have distinct roles, where early layers extract basic features and subsequent layers generate more intricate features based on the input data. Pooling layers are crucial in minimizing computing complexity through dimensionality reduction. The collected features are combined using completely linked layers prior to the final classification [18].
The CNN design, as shown in Figure 2, demonstrates the sequence of steps where word vectors are input into the embedding layer. Convolutional layers perform the task of extracting features, while max pooling layers are used to reduce the dimensionality. The

dropout layer contributes to regularization by stochastically deactivating nodes, whereas the SoftMax layer carries out the final classification.

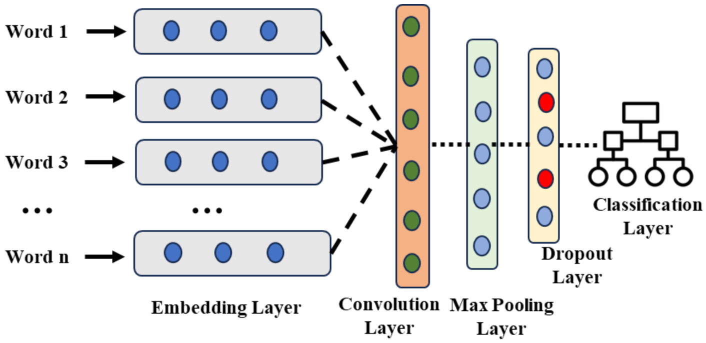

Figure 2. Architecture for proposed CNN

C. Transformative Capabilities of BERT in NLP

BERT, which stands for Bidirectional Encoder Representations from Transformers, is an innovative AI model designed primarily for Natural Language Processing (NLP) tasks. BERT, based on the transformer architecture and attention processes, signifies a significant change in the capabilities of natural language processing (NLP) [19, 20].

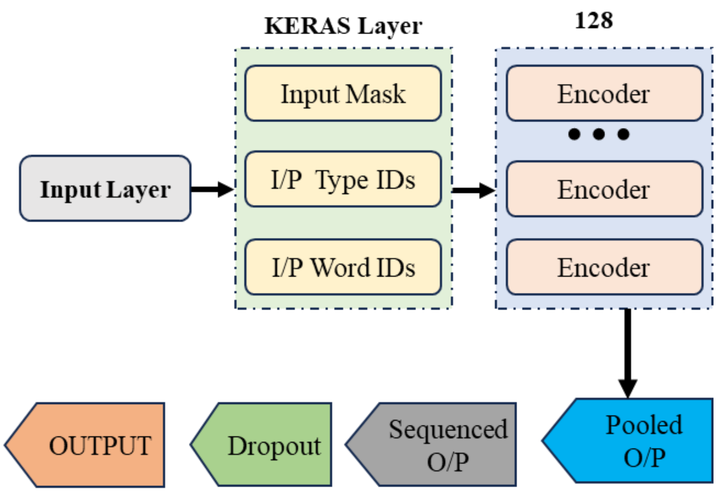

Figure 3. BERT Developed Architecture.

BERT, unlike standard language models that can only process text in a sequential manner, incorporates bidirectionality, allowing for simultaneous comprehension of material in both directions. BERT's capacity to handle information in both directions allows it to utilize advanced approaches like the Masked Language Model (MLM). When a word is "masked," BERT uses contextual information from the tokens that come before and after it to anticipate the hidden word accurately [19, 20]. This is different from previous models that just used one-way context to predict words. The BERT architecture, shown in Figure 3, is an extension of the BERT-Base framework. It consists of 12 transformer layers, each having 768 hidden neurons. The design possesses significant computational capability, with around 110 million trainable parameters. In order to improve the resilience of the model and mitigate the risk of overfitting, a dropout layer is implemented to stochastically eliminate 50% of the neurons. Ultimately, a classification layer is implemented to allocate labels to the given text data, so concluding the BERT-based classification procedure.

V. Results

To evaluate the efficiency of the proposed method, we employed essential characteristics obtained from the confusion matrix to measure performance. These measurements, namely accuracy, precision, and recall, offer unique perspectives on the effectiveness of the categorization system. Accuracy, a commonly employed measure, provides a comprehensive evaluation of the performance of the categorization system. The computation is performed using Equation (1), which yields a measure of the ratio of accurately classified cases to the total number of occurrences. Precision measures the system's capacity to correctly classify instances that belong to a particular class. On the other hand, recall quantifies the rate at which instances of a specific class are accurately predicted. Equations (2) and (3) can be used to compute both precision and recall, respectively.

$$Precision = {}^{TP}/_{FP + TP}$$

$$Recall = {}^{TP}/_{FN + TP}$$

$$Accuracy = {}^{TP + TN}/_{FP + TP + FN + TN}$$

These metrics are essential tools for assessing the appropriateness of the suggested approach for real-time text classification scenarios. Through the quantification of the system's accuracy, precision, and recall, we offer a thorough study of its performance across several categorization tasks.

A. Email Dataset Classification

Here, we showcase the outcomes derived from the suggested approach implemented in Study I, with a specific emphasis on categorizing text inside a dataset of emails. The dataset consists of two distinct categories: "spam" which denotes fraudulent and deceptive emails, and "ham" which denotes authentic and genuine emails. The suggested methodology seeks to categorize emails by examining textual information through three separate frameworks: LSTM, CNN, and BERT. Figure 4 presents confusion matrices that provide a detailed representation of the categorization findings. The training graph, illustrated in Figure 5, displays the accuracy scores attained by our framework: 97.9% for CNN, 97.9% for LSTM, and 86.2% for BERT. Similarly, the validation findings (Figure 5) demonstrate that LSTM, CNN, and BERT achieved high accuracy scores of 95.9%, 95.2%, and 85.5% correspondingly.

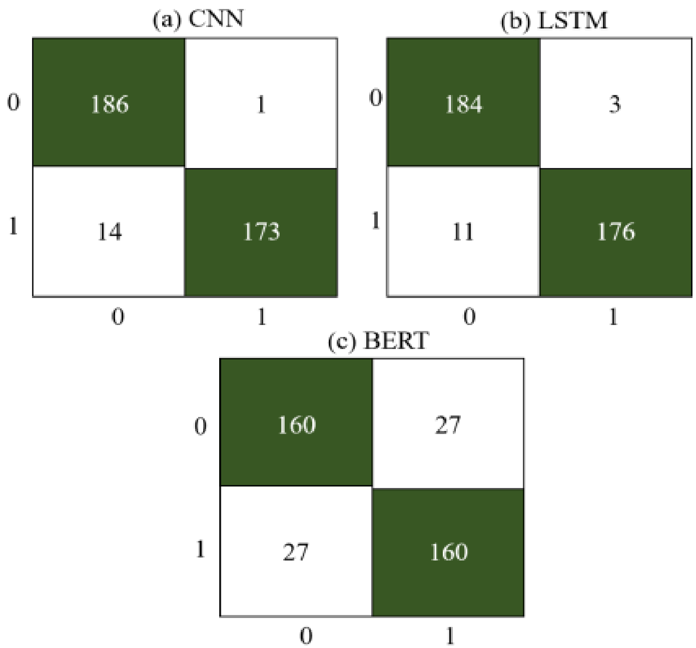

Figure 4. Email Dataset Confusion Matrix.

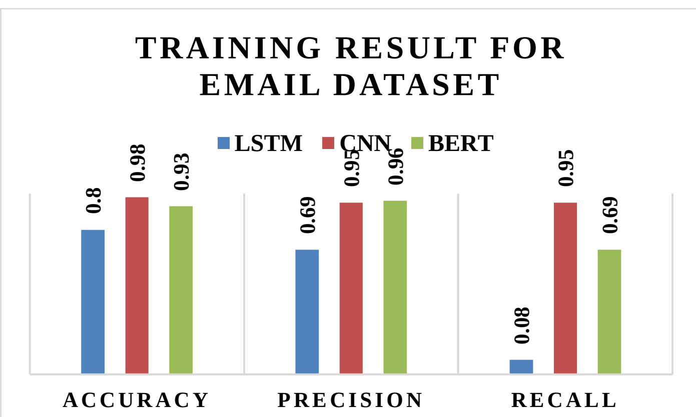

Figure 5. Training result for email dataset.

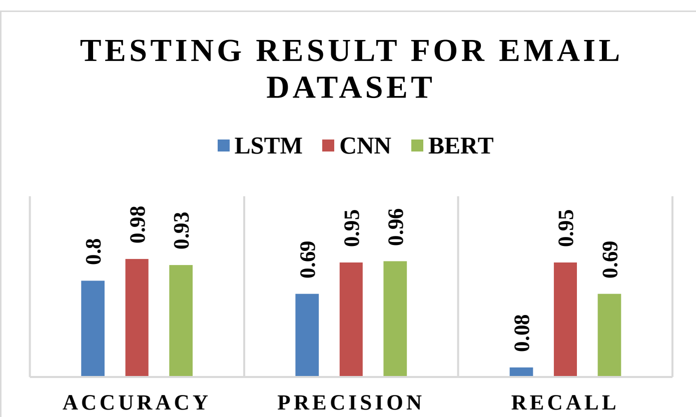

Figure 6. Testing result for email dataset.

Both plots offer insights into the precision and recall measures achieved by different deep learning systems on email data. The results unambiguously establish the efficacy of the suggested technique in precisely categorizing emails, even on previously unknown samples. Therefore, this method shows potential for effectively identifying and removing spam emails, highlighting its suitability for implementation in email categorization systems.

B. News Classification Dataset

This section outlines the suggested structure for news classification using deep learning algorithms. The methodology is formulated by utilizing a wide range of news data from several areas, such as technology, business, sports, entertainment, and politics. The data is sourced from the BBC news database. Figure 7 displays the confusion matrices obtained from the CNN, LSTM, and BERT models. Figure 6 illustrates the training outcomes acquired using the news classification technique. The proposed system attained the highest levels of accuracy, reaching 80.9%, 99.5%, and 93.4% using LSTM, CNN, and BERT architectures, respectively. Figure 7 demonstrates that the LSTM, CNN, and BERT models had the highest validation accuracies on the test set, which were 80.9%, 98%, and 93.3% respectively.

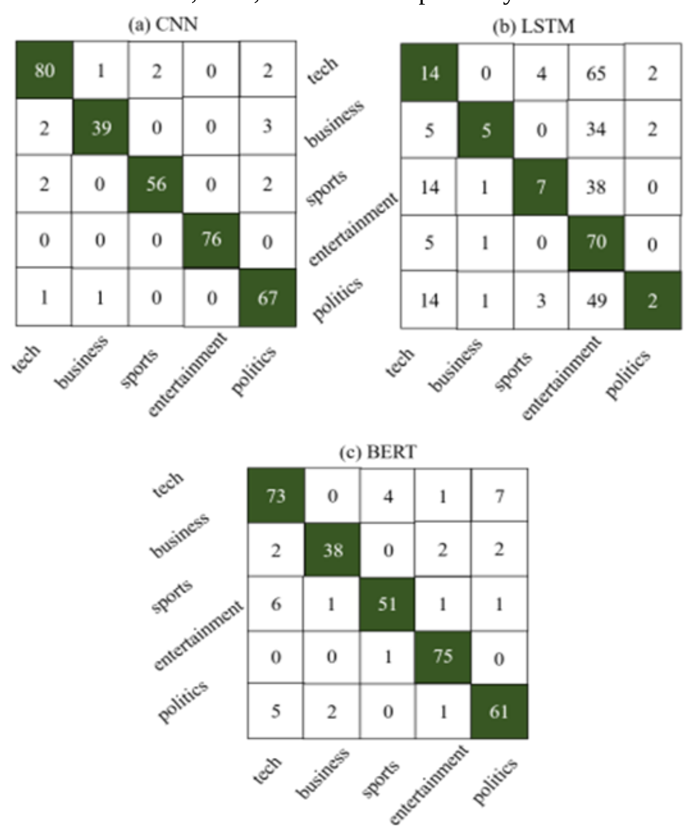

Figure 7. News Dataset Confusion Matrix.

The validation findings highlight the efficacy of the suggested strategy in precisely categorizing various news genres. Therefore, this method has the capability to be used in real-time for classifying news based on textual data. The ultimate result is derived by the execution of numerous experiments and meticulous fine-tuning of parameters. Table 1 summarizes the many optimizations performed in terms of epochs, learning rate, number of layers, activation functions, drop rates, and so on. These optimizations ultimately led to the discovery of the most effective CNN, LSTM, and BERT designs.

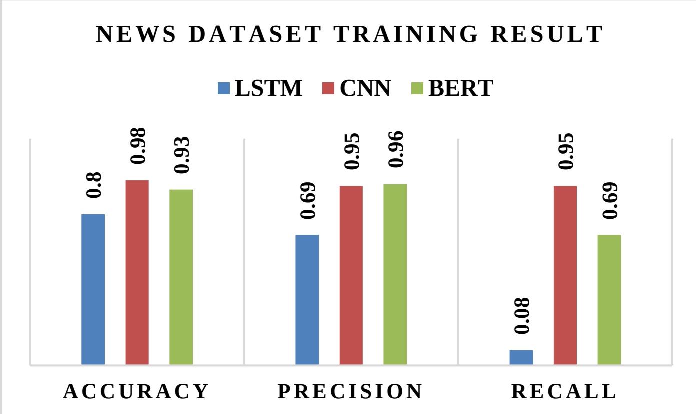

Figure 8. Training result for news dataset.

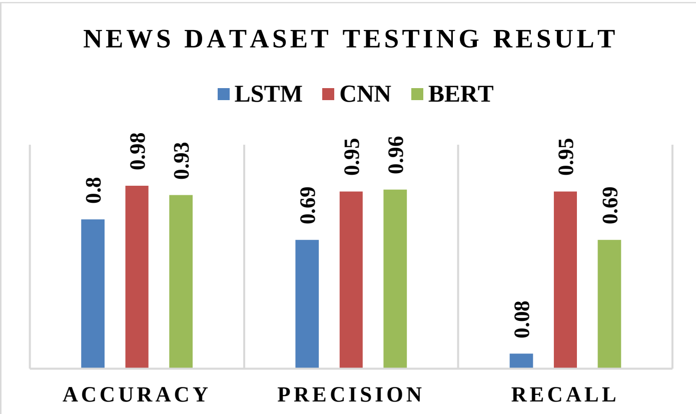

Figure 9. News Dataset Testing result

Table 1. Parameters for Model Tuning

Parameters	**Values**
Optimizer	**Adam, RMS Prop**
Drop rate	0.25, **0.5**
No. of CNN /LSTM layer	**1**, 2, 3
Learning Rates	**0.001**, 0.0001, 0.01
Batch Size	16, 64, **128**
Epochs	5, **10**, 15

C. Comparison with Existing Literature

This section entails a thorough examination of the suggested text categorization algorithms' performance in comparison to previous research, as outlined in Table 2. The method we proposed attained a validation accuracy of 95.9% and 98% on email and news databases, respectively, which were the highest recorded. In contrast, previous systems frequently depended on only one deep learning architecture or one source of text data. The restricted range of applications limits the resilience of these systems for immediate deployment of text classification. In contrast, our method utilizes various deep learning architectures and integrates a wide range of textual data sources, leading to a substantial enhancement in classification accuracy and resilience.

Table 2. Comparison of developed methodology with existing Literature

Reference	**Methodology**	**Accuracy**
J. Briskilala et.al; [3]	BERT	85%
R. Pappagari et.al;[4]	CNN	86%
B. Jang et. al; [5]	MLP	71%
Developed Methodology	**LSTM**	**95.2%**
	CNN	**98%**
	BERT	**93.3%**

In contrast, the proposed system takes a comprehensive approach by utilizing various deep learning approaches on different text databases. As a result, it shows efficacy in categorizing unfamiliar text samples. In terms of performance comparison, the suggested methodology consistently produces higher accuracy scores. Specifically, it reaches 95.2% accuracy using LSTM, 98% accuracy using CNN, and 93.3% accuracy using BERT. The results highlight the effectiveness of the suggested method for classifying text, especially when dealing with unexpected samples, achieving amazing accuracy.

VI. CONCLUSION

Text classification algorithms have become the subject of much research due to their crucial importance in business and enterprise environments. These algorithms provide useful insights into categorizing textual data. This analytical ability has a wide range of applications, including email filtering, sentiment analysis, and news segmentation. Although various automated approaches have been developed in the subject, many of them have been hampered by their dependence on a single deep learning framework or a small amount of textual data. To address these limitations, we introduce an innovative, completely automated framework for text classification that utilizes the advantages of deep learning techniques. Our technique starts by thoroughly cleaning the data, which involves the rigorous elimination of stop words and unnecessary letters. Afterwards, the clean datasets are tokenized, allowing for improved contextual understanding within deep learning algorithms. Significantly, our approach is completely self-governing, without any human involvement. Through our practical efforts, we achieved an impressive 98% accuracy by utilizing CNN architecture. This highlights the strength and effectiveness of our proposed framework compared to previous research. By conducting thorough comparisons and analysis, we provide solid evidence of the superiority of our models in text classification tasks. This demonstrates their potential for immediate implementation in real-life situations. In the future, our research seeks to go beyond the limitations of language by improving the effectiveness of our model in languages other than

English. This undertaking guarantees to strengthen the adaptability and dependability of our framework, establishing it as an essential instrument in the ever-changing realm of business and enterprise applications.

References

[1] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, "Deep Learning-Based Sentiment Classification of Evaluative Text Based on Multi-Feature Fusion", Information Processing & Management, vol. 56, no. 4, pp. 1245-1259, 2019.

[2] S. Minaee, E. Azimi and A. A. Abdolrashidi, "Deep-Sentiment: Sentiment Analysis using Ensemble of CNN and BiLSTM Models", arXiv preprint arXiv:1904.04206, pp. 1-6, 2019.

[3] J. Briskilal C. N. Subalalitha, "An Ensemble Model for Classifying Idioms and Literal Texts Using BERT and RoBERTa", vol. 59, no. 1, 102756, 2022.

[4] R. Pappagari, P. Zelasko, J. Villalba, Y. Carmiel and N. Dehak, "Hierarchical Transformers for Long Document Classification", IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 14-18 December, 2019, Singapore, pp. 838-844.

[5] B. Jang, M. Kim, G. Harerimana, S-u. Kang and J. W. Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism", Applied Sciences, vol. 10, no. 17, 5841, 2020.

[6] X. She and D. Zhang, "Text Classification Based on Hybrid CNN-LSTM Hybrid Model", 11th International Symposium on Computational Intelligence and Design (ISCID), 08-09 December, 2018, China, pp. 185-189.

[7] S. A. Sattar, S. Hina, N. Khursheed, A. Hamid, "Urdu Documents Classification using Naïve Bayes", International Journal of Science and Technology, vol. 10, no. 29, pp. 1-4, 2017.

[8] W. Antoun, F. Baly and H. Hajj, "Arabert: Transformer-Based Model for Arabic Language Understanding", arXiv preprint arXiv:2003.00104, pp. 1-7, 2020.

[9] M. A-Mageed, A. Elmadany and E. M. Nagoudi, "ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic", arXiv preprint arXiv:2101.01785, pp. 1-18, 2020.

[10] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi and L. E. Barnes, "RMDL: Random Multimodel Deep Learning for Classification", Proceedings of the 2nd International Conference on Information System and Data Mining, 09-11 April, USA, 2018, pp. 19-28.

[11] J. Cai, J. Li, W. Li and J. Wang, "Deeplearning Model used in Text Classification", 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 14-16 December, 2018, China, pp. 123-126.

[12] L. Lenc and P. Král, "Deep Neural Networks for Czech Multi-Label Document Classification", International Conference on Intelligent Text Processing and Computational Linguistics, 03-09 April, 2016, Turkey, pp. 460-471.

[13] H. Mehnatkesh, S. M. J. Jalali, A. Khosravi and S. Nahavandi, "An Intelligent Driven Deep Residual Learning Framework for Brain Tumor Classification using MRI Images", Expert Systems with Applications, vol. 213, no. C1, 119087, 2023.

[14] K. S. Kumar, A. Bansal, and N. P. Singh, "Brain Tumor Classification Using Deep Learning Techniques," Proceedings of Machine Learning, Image Processing, Network Security and Data Sciences: 4th International Conference (MIND), 19-20 January, 2023, India, pp. 68-81.

[15] W. Hillier, "What Is Data Cleaning and Why Does It Matter?", Available at: https://careerfoundry.com/en/blog/data-analytics/what-is-data-cleaning/

[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[17] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks", arXiv preprint arXiv:1511.08458, pp. 1-11, 2015.

[18] L. V. Haar, T. Elvira and O. Ochoa, "An Analysis of Explainability Methods for Convolutional Neural Networks", Engineering Applications of Artificial Intelligence, vol. 117, 105606, 2023.

[19] J. Devlin and M.-W. Chang. "Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing", Available at: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

[20] J. Mutinda, W. Mwangi, and G. Okeyo, "Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network", Applied Sciences, vol. 13, no. 3, 1445, 2023.

Authors

Abdulmohsen Algarni– Abdulmohsen Algarni received the PhD degree from Queensland University of Technology, Australia, in 2012. He was a research associate in the School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia, in 2012. He is currently an associate professor at the College of Computer Science, King Khalid University. His research interests include Artificial intelligence, data mining, text mining, Machine Learning, information retrieval, and information filtering.

Muhammad Anas – Department of Electrical and Electronic Engineering, Swinburne University of Technology Hawthorn Campus.

Hussein khan – Department of Electrical Engineering, Qatar University, Qatar.

Arbab Masood – Muhammad Assam working as Lecturer in Department of Software Engineering, University of Science and Technology Bannu KP Pakistan. His Research Interests include Brain Computer Interface, Computer Vision, Artificial Intelligence, Natural Language Processing and Medical Image Processing.

Correspondence Author – Abdulmohsin Algarni.