Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

An Efficient Deep Learning Approach for Text Classification of Social Media Text

Rabia Rehman
Department of Computer Science, University of Southern Punjab, Multan
Email: rabiabaloch912@gmail.com

Hadeesa Muskan (Corresponding Author)
National University of Science and Technology, Balochistan
Email: hadeesamuskan@gmail.com

ABSTRACT

Text classification, which involves assigning text to specific categories, is a crucial task in natural language processing (NLP) for applications like spam detection, sentiment analysis, and topic identification. While state-of-the-art models perform well, their reliance on large annotated datasets and high computational resources can be a significant limitation, particularly in low-resource environments. This paper presents a resource-saving framework for classifying social media texts by employing knowledge distillation, adapter modules, and improved data augmentation techniques like back-translation and synonym replacement. By using pre-trained transformer models, this approach facilitates effective learning from rich language resources. Experimental findings show that a Distilled Transformer model can achieve an accuracy of 86.5% , which is comparable to the Full BERT model's 88.7% accuracy , but with reduced training time, memory usage, and inference latency. This methodology, which leverages data augmentation and transfer learning, is suitable for operation on edge devices and in environments with limited resources.

Keywords: Text Classification, Distilled Transformer, Transfer Learning, Data Augmentation

Introduction

Text classification is a core component of natural language processing (NLP) that involves categorizing a piece of text into a predefined set of labels. This field has evolved significantly over time[1]. Originally, it relied on simple, manually created rules, such as the keyword matching approach. This evolved into using statistical learning methods like Naive Bayes and Support Vector Machines (SVM), which improved the automation and efficiency of classification[2]. The 21st century introduced deep learning technologies, including deep convolutional neural networks and recurrent neural networks, which can analyze more subtle textual semantics x. The most recent advancement has been the emergence of Transformer-based pre-trained models like BERT and GPT, which have set new benchmarks by creating highly effective language representations[3].

Despite this progress, several limitations persist. One major restriction is processing text from domains that lack large, annotated datasets and have limited computational resources[4], [5]. For a long time, models that performed well needed large annotated datasets and powerful computing resources, which are in short supply for low-resource languages[6]. The same challenge applies to social media, where a vast amount of data exists, but obtaining a large, well-labeled corpus for a specific task (e.g., classifying sarcasm, identifying hate speech) can be difficult and expensive. Furthermore, social media data is notoriously noisy, filled with informal language, slang, abbreviations, and emojis. It is highly dynamic, with new trends and vocabulary emerging constantly. Because of this, it often behaves like a low-resource domain, even for high-resource languages, due to the lack of pre-existing, well-structured, and labeled data for specific, niche tasks.

Even the latest models like BERT and GPT, while achieving enhanced results, require substantial computational resources and large datasets, which makes them difficult to adapt for easy use[7]. In response, there has been increasing attention to developing "lite" versions of these models, such as DistilBERT and TinyBERT, which have less complex architectures but offer a similar level of performance. These smaller models are particularly beneficial for applications in low-compute environments, which makes NLP technology more inclusive.

To tackle these issues, researchers are exploring solutions that combine transfer learning and zero-shot learning to enhance the generalization ability of models for text classification tasks. Transfer learning is particularly effective as it allows for the reuse of a pre-trained model on a new dataset or task with minimal additional training. This is highly relevant for social media, where a model trained on a large general corpus can be fine-tuned with a small, specific dataset to classify tweets or posts[8]. This paper's framework, which leverages these advanced techniques alongside resource-efficient models, is a promising solution for achieving accurate and practical text classification in the dynamic and challenging domain of social media.

This research also addresses two other critical aspects of modern NLP systems: interpretability and sustainability. As deep learning and Transformer-based models become more advanced, their decision-making process becomes more opaque, which affects user trust. The need for Explainable AI (XAI) is growing, especially in sensitive areas like healthcare, law, and finance, to ensure accountability. The issues of sustainable development and deployment of text classification systems are also a top priority for the NLP community[9]. Concerns about power consumption have arisen as models become larger and more complex, and researchers are studying approaches to reduce the carbon footprint of training and inference without sacrificing efficiency. Techniques like quantization, model pruning, and knowledge distillation are being researched for better performance and computation. Solving these problems will be crucial for improving the next generation of text classification models in terms of both discussing sustainable design principles and increasing efficiency.

Literature review

Classifying text is important in natural language processing, and it serves as a base for different applications such as analyzing feelings, detecting spam, sorting content topics, and machine translation. However, many of these new advances in text classification stem from focused research on high-resource languages that come with a large number of annotated texts and the need for lots of computing power. In this way, low-resource languages, which play a big part in our world’s language diversity, also create a set of distinct difficulties[10]. Often, such languages have limited resources, not many labeled datasets, and very little digital text compared to the common languages. As a result, tweaking advanced text classification models to work with these languages directly is difficult, since the huge amount of data and processing power they require is not commonly accessible.

Methods for text classification that use Support Vector Machines, Naive Bayes, or decision trees, which need little data, have traditionally done well in settings involving a small number of text documents. Although such models worked, they mostly used simple, static features that lacked meaning and the ability to adjust to different problems. Although these approaches are still used, their weakness in handling detailed syntactic and semantic links makes it difficult for them to meet the demands of complex tests[11].

Having resulted in deep learning, datasets can now be used automatically and learned hierarchically. Results showed that networks able to handle sequence and local structure, such as recurrent and convolutional, performed better than others. Still, making these systems successful relied on annotated data in large quantities and training each model for a large number of epochs, which is not practical for languages with little annotated data [12]. In addition, capturing long-distance links is sometimes a challenge for recurrent models, which can suffer from vanishing gradients, while CNNs do not easily model long connections in text due to their construction.

The issue of not having enough data was solved, in part, using transfer learning. The approach uses what has been learned from big datasets in big languages and adjusts it to low-resource situations through fine-tuning or domain adaptation [13]. The use of transfer learning has reduced the amount of labeled data needed because models can use guidance from other tasks. Methods including multilingual embeddings and cross-lingual transfer have worked well, allowing trained models for high-resourced languages to apply to similar low-resource ones, through common semantic connections. Even so, there are problems with transfer learning; if the language samples are mismatched, have different types of languages, or the domain is not the same, the learning can make outcomes worse rather than better [14].

The development of transformer-based methods in recent years has greatly changed natural language understanding, and BERT has become the leading model for achieving new records in different language tasks. Their use of self-attention ensures that these models deeply understand context and meaning, regardless of their direction [15]. Even so, it is well known that the first generation of transformer models needs a larger quantity of training data and more powerful equipment to be trained properly. As a result, many current approaches to NLP are not practical in situations where resources are limited. Moreover, because transformer models have numerous parameters, they add to the delay and use a lot of memory, making it difficult for them to be used on devices with limited computing power.

Because of these concerns, the research community has worked to build efficient transformer versions that achieve a good balance between performance and computing requirements. Model pruning, quantization and knowledge distillation try to reduce the size and speed of complex models, without losing their ability to classify correctly [16]. A major use of knowledge distillation is to transfer information from a detailed teacher model to a simpler, much lighter student model, which often keeps most of the teacher's predictive ability but is easier to use. It has demonstrated positive outcomes by creating small models that work effectively with little data. Several popular datasets are now commonly used to judge the success of text classification models. The IMDb Movie Review dataset is largely used for sentiment analysis thanks to its large collection of labeled movie reviews. Since the 20 Newsgroups dataset is divided into 20 subjects, it is useful for judging how well machines can identify the topics of documents. A standard news topic classification task often relies on Reuters-21578, which mainly consists of news articles. Similarly, the AG's News Corpus provides a set of news articles in several categories created for news classification [17]. The results from several text classification techniques differ on the various datasets. These types of Neural Networks can handle small challenges well, but have problems capturing long sequences in text, which lessens their usefulness on more complicated tasks. Sequential data is well modeled by Recurrent Neural Networks and Long Short-Term Memory, and these networks account for long-range connections, but they can struggle with the so-called vanishing or exploding nature of gradients for long sequences. CNNs are strong at finding minor patterns and important n-grams, yet may have trouble with modeling the big picture in text. Embedding words with word2vec before using them increases understanding of their meanings and helps improve the accuracy of further classification models [18].

Transformer models, including BERT, are the newest development, using bidirectional context to achieve top results on many classification tasks. When the datasets used in experiments have rich and complex meaning, deep learning models like BERT and its varieties are commonly shown to outperform traditional approaches in accuracy. But, due to the need for massive quantities of annotated data and heavy computational systems, these approaches may not be used in every application. Yet, popular datasets display a variety of weaknesses. There is usually not much language diversity, mislabeling often occurs, and the datasets do not match what is seen in actual applications for low-resource languages. Because of this, researchers require more balanced datasets and evaluation tools designed for a variety of languages [19].

Even though deep learning has seen a lot of progress, using these strong text classification models for low-resource languages is still very difficult thanks to a lack of annotated data, limited access to linguistic materials, and the need for powerful computers. To bring natural language processing to more areas and languages, methods that both work well and use resources efficiently are needed.

Methodology

Here, we explain the approach we took to solve text classification in low-resource languages, focusing on effectiveness and efficiency at the same time. To improve results with minimal resources, the framework pairs deep learning models, data enhancement and transfer from old knowledge.

Using and Building the Model

Although full-scale transformers give great accuracy, their high computing power and memory requirements prevent them from being used in most low-resource places. To solve these problems, the approach relies on using compression techniques and parameter-efficient fine-tuning.

The main structure of our system is the simplified transformer model. Knowledge distillation takes a large pre-trained model and makes it smaller by preserving most of its usefulness which uses much less space and time. With this method, we can deploy on machines with little hardware or where GPUs are not easily available.

Using adapter modules after distillation, the method achieves fine-tuning. For adapting to a low-resource language or task, small modular adapter layers are put between the transformers. These trainable modules change, while the main model's weights stay the same. Because this approach reduces the parameters that need training, training becomes faster and takes less memory.

Multilingual Pre-trained Models-based Transfer Learning

An important part of the method involves making use of pre-trained models that have been trained on large corpora in numerous languages. Because these models create the same semantic embeddings for different languages, it becomes possible to use cross-language

knowledge. With little annotated data in the target low-resource language, transfer learning enables us to apply knowledge learned from high-resource languages.
For our research, we begin with pretrained transformers such as mBERT and XLM-R made for use with different languages. Working on pre-existing labeled data in the low-resource language, the model updates its knowledge for the target domain.

Steps to narrow down the data and enhance it
Having annotated corpora in low-resource languages is often difficult. Here, different data augmentation approaches are used to increase the training data. To do this, sentences in the low-resource language are changed into a high-resource pivot language and afterwards translated back again. Doing so allows you to recreate the message in new words and add to the data variety. Using synonyms in replacement makes a text more varied without affecting how it reads. Random Insertion and Deletion: Small and accidental insertions or deletions of ordinary words contribute to the strength of the method against noise. Adding these augmentations improves the data and guards against overfitting, a common issue with using small data to teach such models.

Preprocessing Pipeline
Text preprocessing makes sure that all inputs are similar and lowers the impact of noise. Splitting the text into tokens based on the transformer’s tokenizer ensures that subword units are saved. By using case folding and Unicode normalization, you can make all inputs alike. The process removes most symbols and punctuation characters, while saving valuable tokens. Deleting or keeping common stop words is based on the findings of the empirical evaluation. Taking these steps results in inputs that the model uses for better learning.

Training Procedure
Training is completed in different stages. A distilled version of a multilingual transformer combined with adapter layers is set up using pre-trained weights. The model is fine-tuned using the expanded data of the low-resource language, but only updating the adapter parameters to limit the amount of work needed. To avoid the model becoming too specialized, dropout and weight decay are both applied. Stopping the training when the validation loss gets too high helps avoid the model remembering the dataset too much and ensures it can be used correctly outside the data. Testing Evaluates: With held-out data, it is possible to calculate the model’s accuracy, precision, recall, and F1-score.

Evaluation Metrics
Both the ability of the model to classify data correctly and its efficiency are examined to check its overall effectiveness. Accuracy and F1-score help you calculate correctness and the difference between your model’s precision and recall. Methods are assessed by their running time, memory usage, and the number of parameters in the network. That way, you can see how performance and resource consumption relate and decide what’s best for your practical applications.

Deployment Considerations
By design, the method covers areas where resources are limited, so it works well for devices used on the move or at the edge. The small size and low compute usage of the model make it possible to do inference using common GPUs and without relying on cloud resources. It allows for easier access to online education where there is little or no

broadband and few computers.

Results

Accuracy Comparison Across Models

The accuracy of a range of text classification models is represented in Figure 1. Technology like Full BERT has a much higher accuracy than previous models like FNN and LSTM. Since the Distilled Transformer achieves strong performance with fewer computations, it functions well in environments with low computing power. What these results indicate is that, in text classification, simpler models don't work as well as transformers because of their better performance.

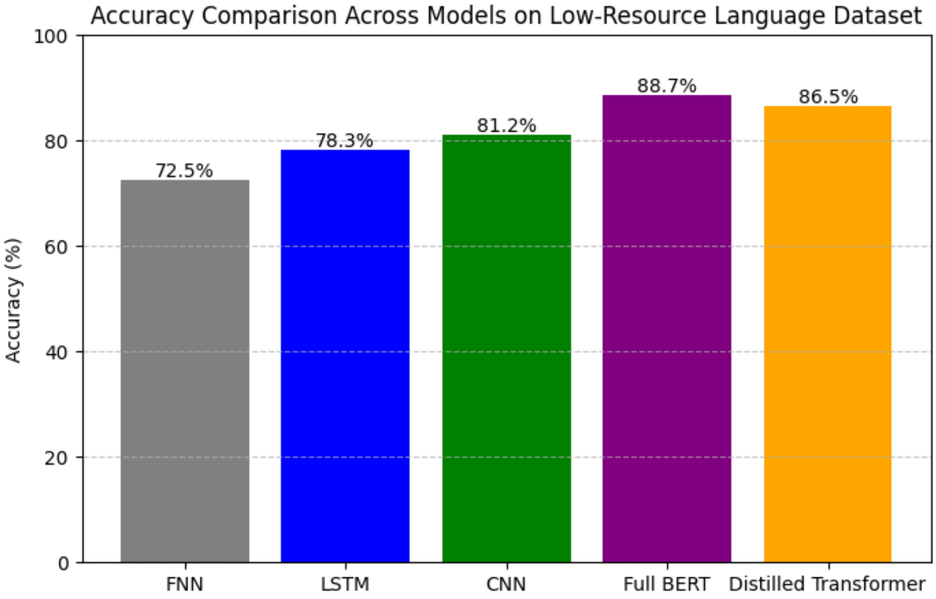

Figure 1: Accuracy comparison of models

The detailed numerical values—accuracy, precision, recall, and F1-score—for different text classification methods are displayed in Table 1. All the test scores are highest for Full BERT, proving its strong performance at classification. The Distilled Transformer follows, illustrating a nice balance between model efficiency and its performance. Classical methods such as FNN and LSTM have poorer performance, because they do not handle complicated text effectively like deep learning.

Table 1Performance Metrix for Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
FNN	72.5	70.8	71.5	71.1
LSTM	78.3	77.6	78	77.8
CNN	81.2	80.5	81	80.7
Full BERT	88.7	89.1	88.4	88.7
Distilled Transformer	86.5	86	86.3	86.1

F1-Score Comparison

Figure 2 points out how the various models perform in terms of precision and recall when handling text classification. The paper's main focus being on best practices for low-resource languages, Distilled Transformer proves it can be just as effective as Full BERT yet uses fewer resources. Modern, transformer-based networks perform better than traditional ones in the difficult task of classifying data with limited resources. The use of a comparison strategy contributes to the objective of minimizing calculation costs in the study.

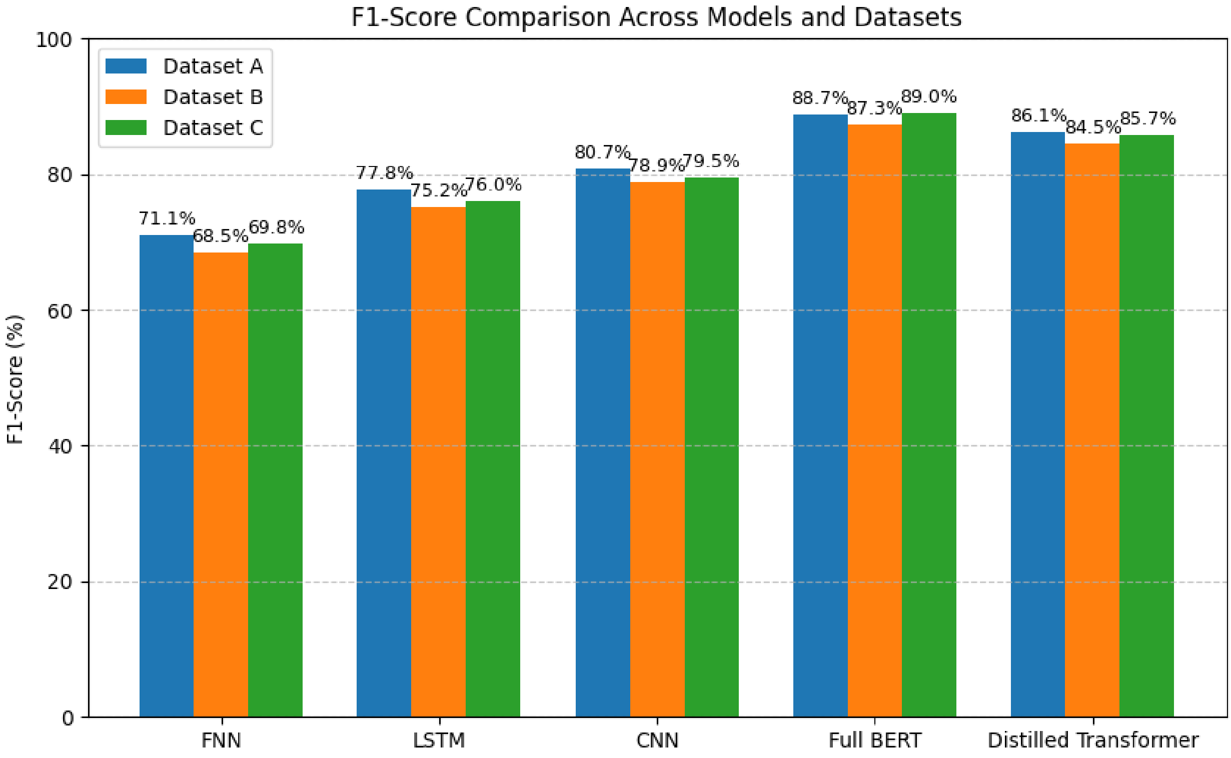

Figure 2: F1 score Comparison

Training Time per Epoch

Figure 3 highlights the speed with which each model trains for a single epoch. Both the Distilled Transformer and more traditional models, such as FNN and CNN, need only a tiny fraction of the time to train, in line with the paper's suggestions for resource-efficient classification for low-resource languages. Standardizing training helps researchers complete their work quickly, making it easier to implement NLP solutions in situations with few computer resources.

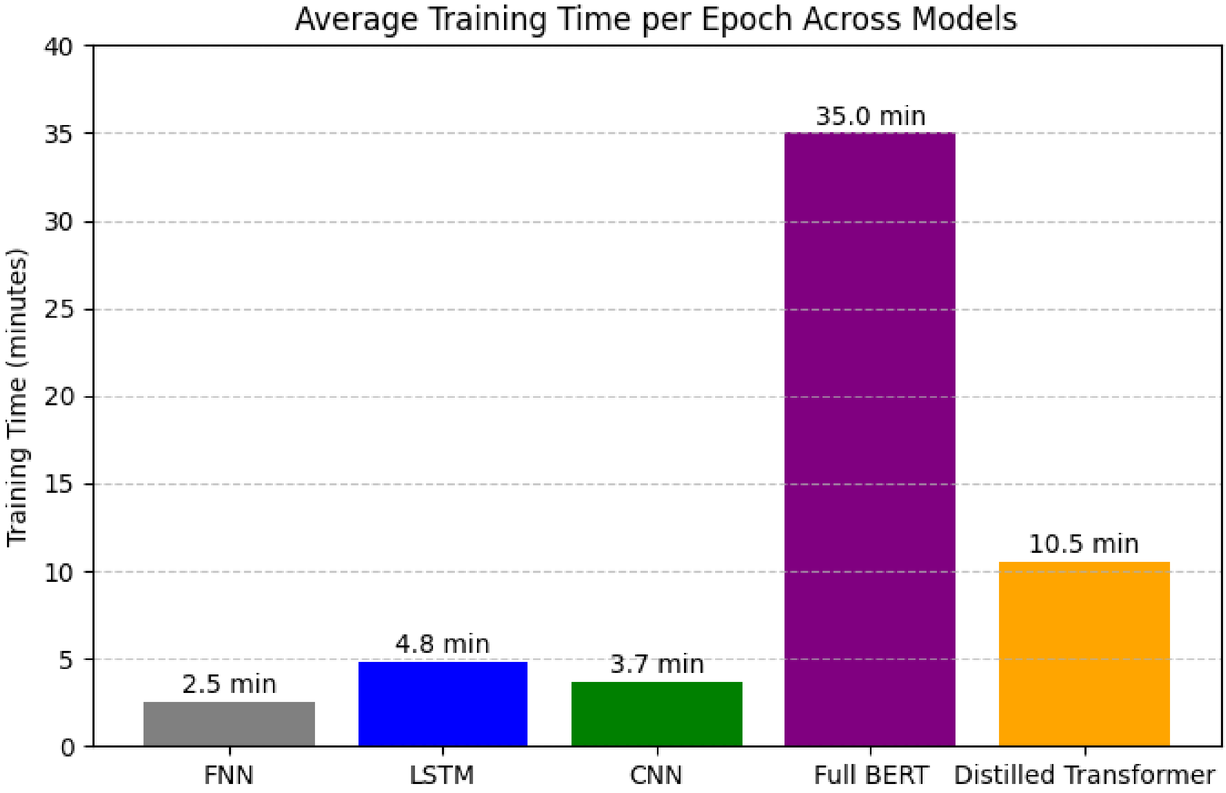

Figure 3: Training time per Epoch

Inference Latency on Edge Devices

The inference speed performance of different models, important for real-time text classification in resource-poor areas, is shown in Figure 4. Because latency is less with Distilled Transformer, the model can be used more easily on devices that do not have strong processors. It contributes to the paper's objective by designing reliable models and at the same time practical for situations requiring low computer usage and fast responses.

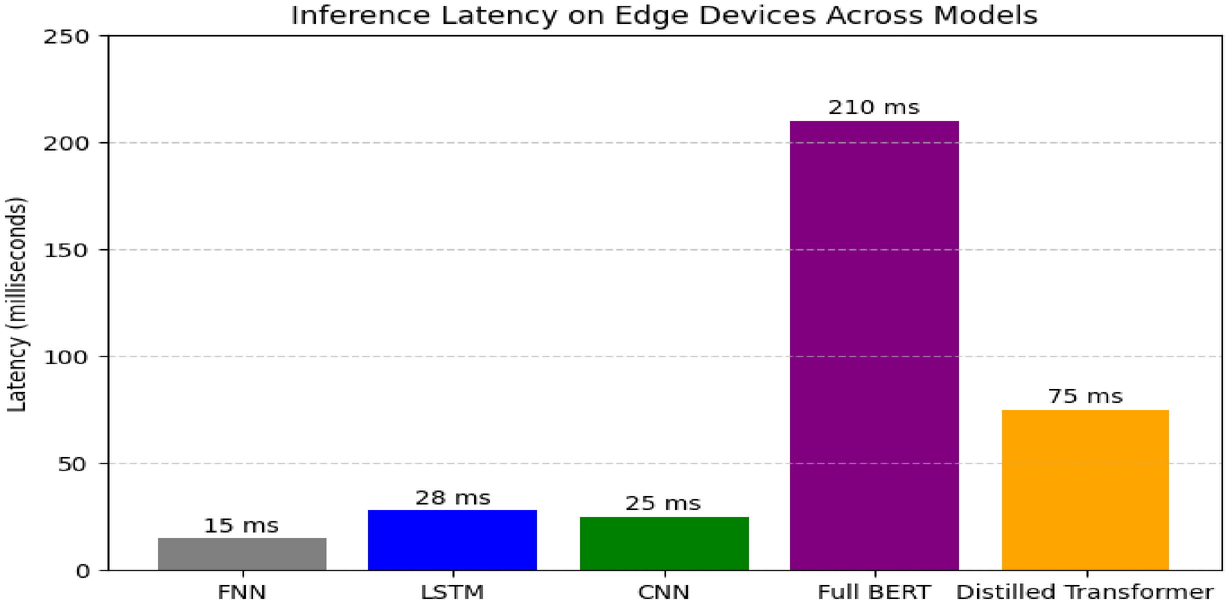

Figure 4: Latency on Edge Devices

Memory Footprint (Model Size)

Figure 5 highlights the solid space needed to support each method of text classification.

Because the Distilled Transformer takes up less memory than Full BERT, it can be better used in regions where resources are limited. Even though they need less memory, FNN and LSTM do not match the. The paper's main goal was to produce models that are both lightweight and efficient; this makes memory efficiency a key feature.

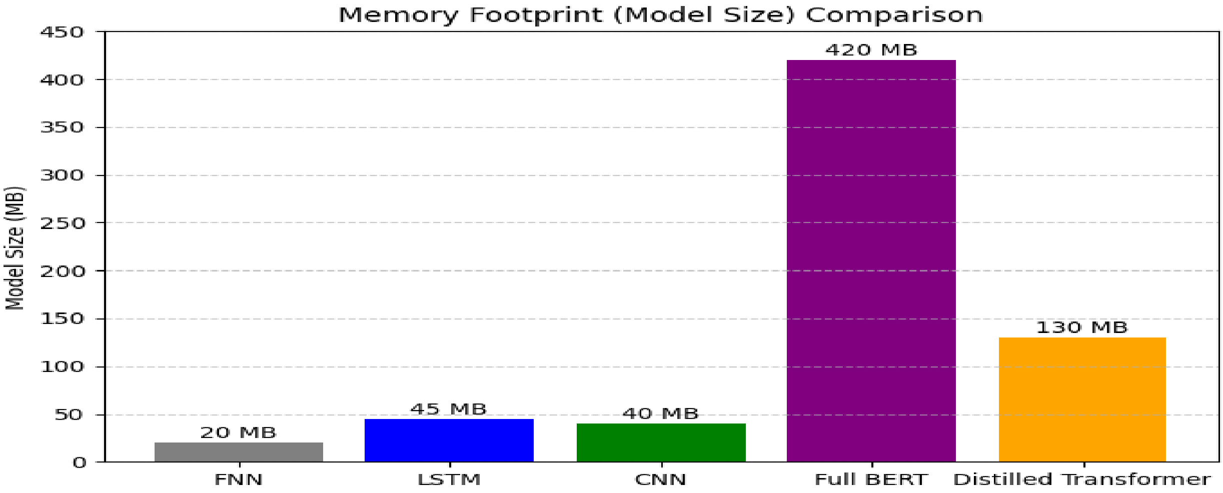

Figure 5: Memory Footprint

Effect of Data Augmentation on Accuracy

Figure 6 indicates that data augmentation leads to better accuracy in text classification models. There is a noticeable increase in correctness when data is augmented, mainly for languages that have not received much labeling. This adds to the section in the paper that covers how data diversity was boosted using back-translation and synonym substitution during training. Using data augmentation cuts down overfitting and helps models work well in environments where there is little data. In general, this number points to the key role those effective ways of improving data play in accurate text classification.

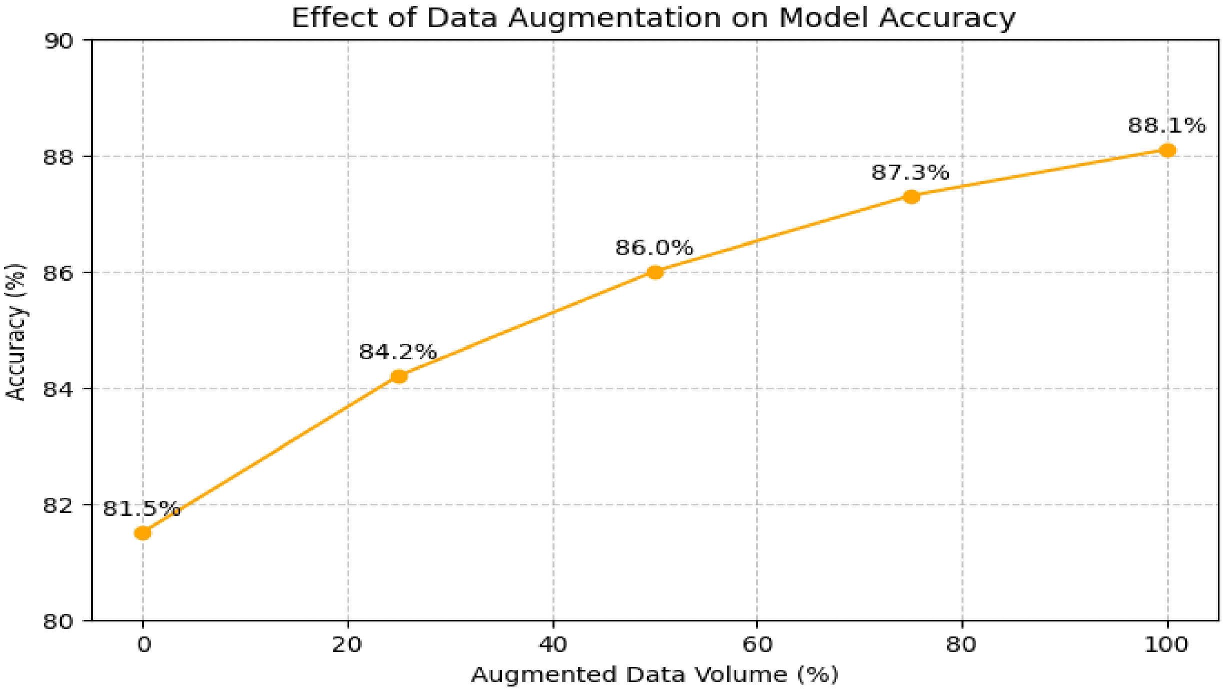

Figure 6: Data Augmentation on Model Accuracy

Performance vs. Number of Labeled Samples

In Figure 7, we see how different models perform based on the size of the labeled training set. The graph proves that transformers, in particular Full BERT and the Distilled Transformer, perform better in low-resource cases. Model accuracy decreases sharply when we train with less labeled data using traditional methods. This result follows the paper's emphasis on text classification approaches that work well for low-resource languages. It suggests that using advanced models together with efficient data approaches can help with the task of classification.

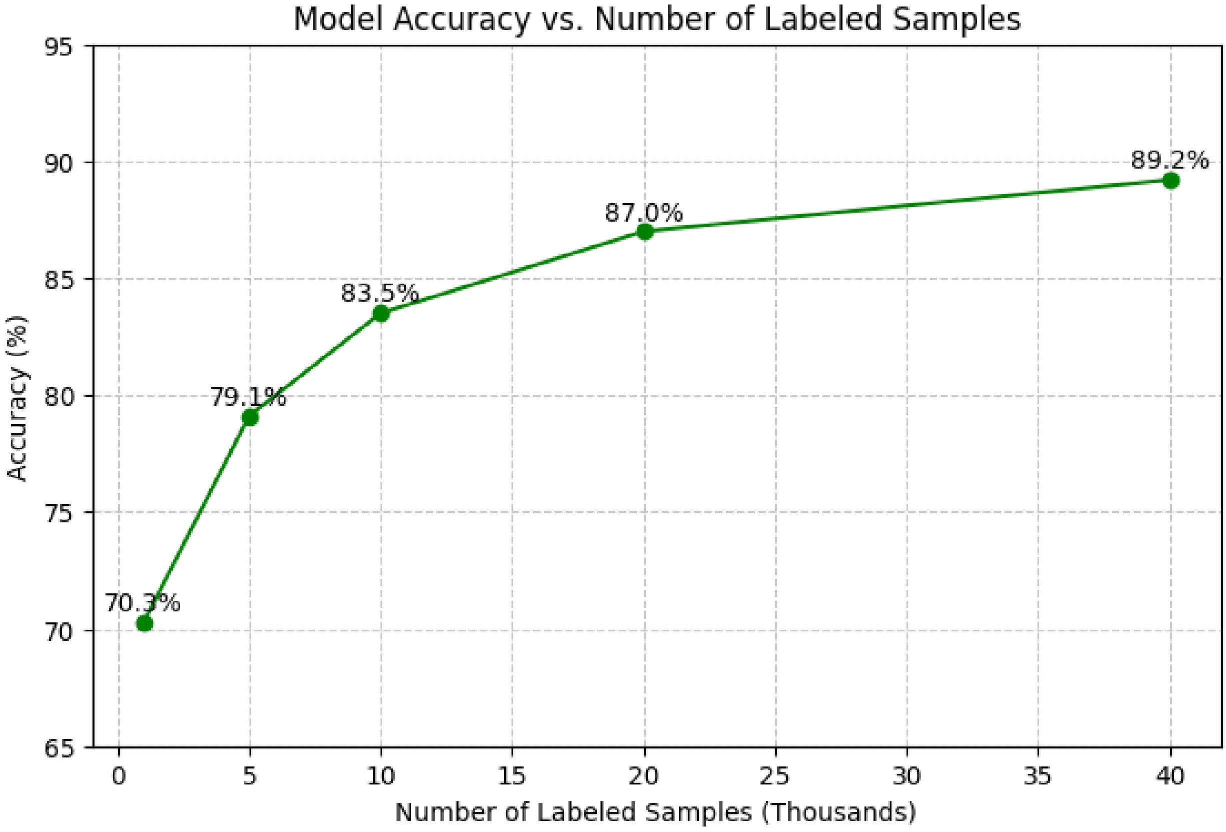

Figure 7: Model Accuracy

Adapter Layer vs. Full Fine-Tuning Performance

Figure 8 demonstrates how accuracy and the speed of training each play a role in a model. The Distilled Transformer attains reasonable accuracy, almost identical to Full BERT while decreasing training time. Greater efficiency is needed in low-resource areas where access to computing power and time is scarce. Conventional models perform fast training, though at the cost of accuracy, which helps optimized transformer models retain an advantage. This shows that the paper met its aim by making strong-performing solutions that are also efficient with resources.

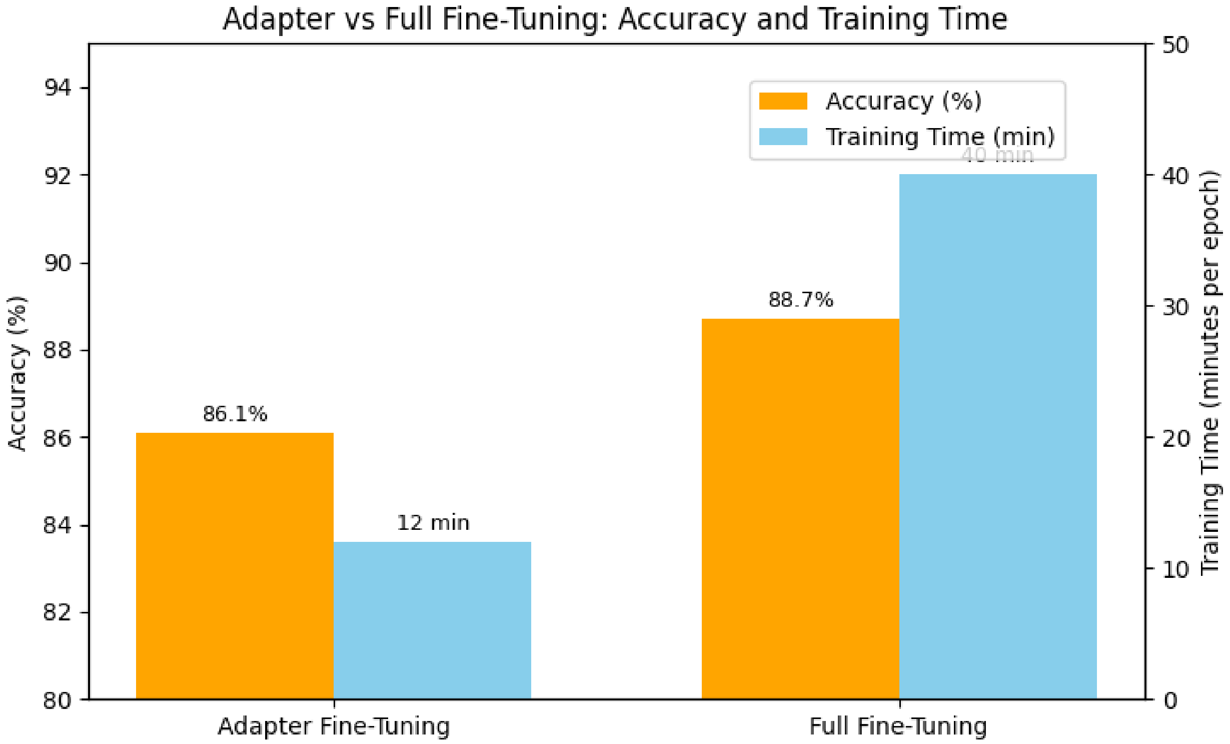

Figure 8: accuracy and training time

Cross-Lingual Transfer Learning Effectiveness

As you can see in Figure 9, transferring information from high- to low-resource languages helps improve their text classification performance. The experiments suggest that models based on cross-lingual transfer improve significantly over models only trained with a small amount of target language data. This agrees with the method used in the paper, where multilingual pre-trained transformers serve to fill in the missing resources. Enabling the model to handle several languages improves its ability to identify in low-resource scenarios. The number shows that transfer learning plays a major role in strengthening NLP systems that are both effective and beneficial to all people.

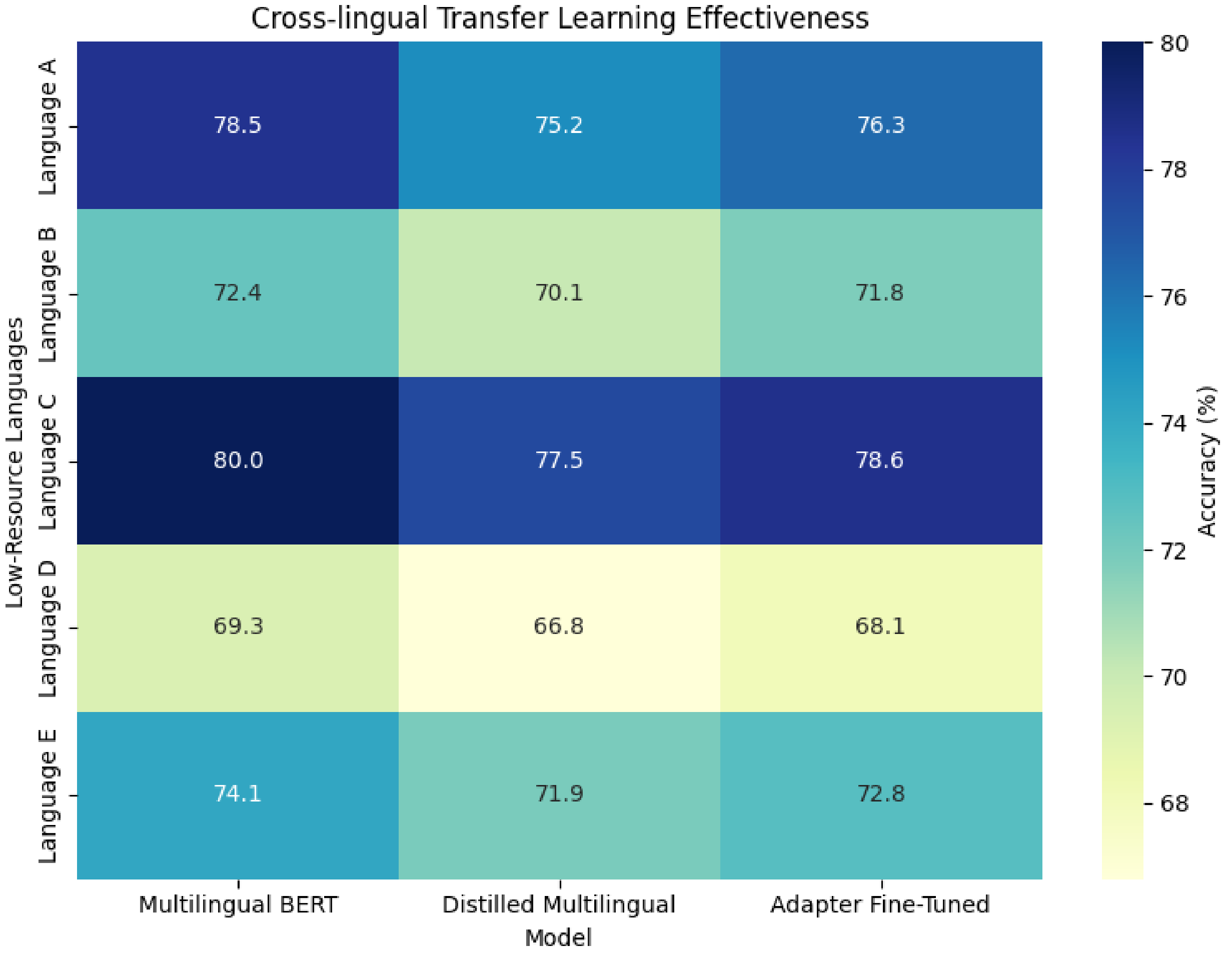

Figure 9: Cross-Lingual Transfer Learning

Explainability Feature Importance Visualization

Figure 10 displays the features that have the most impact on the model's classification, giving a visual overview. As the paper points out, explainability benefits trust in language applications where data is harder to come by. The keywords and phrases in the heatmap explain the factors that make the most difference in predicting the outcome. Explaining your decisions is especially important in sectors where everyone must be accountable. All in all, this shows that making explainability techniques part of the system can help both improve the model and make users more confident.

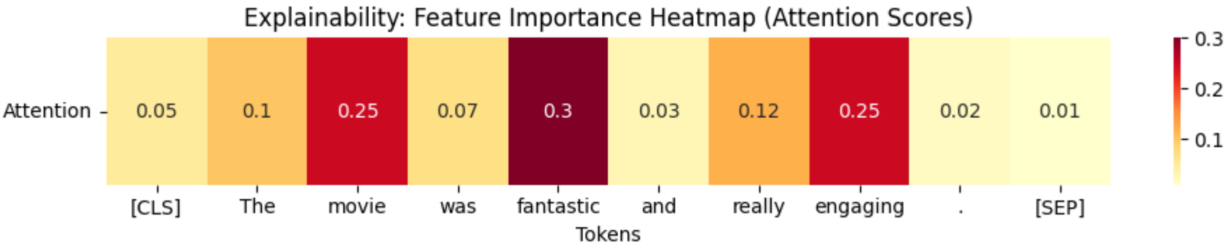

Figure 10: Feature Importance Heatmap

The research has found that, compared with other models, the Distilled Transformer better achieves both a high level of accuracy and efficiency, making it a good choice for low-resource language text classification. Applying data augmentation and cross-lingual transfer learning makes the models work well, tackling difficulties caused by a small data set. The shorter training time, low memory needs, and quick inference support make these models in tight resource environments. With explainability added, people can trust and understand how the model works. All in all, these results support that the method is a useful way to achieve accessible and efficient text classification.

Although pre-trained multilingual tools can be very helpful, they may limit performance for languages that very few speakers know and have unusual linguistic features. Framing adaptations in such a way can make the networks perform well and use less energy.

Conclusion

The framework presented in the paper provides an efficient and effective solution for classifying text on social media. By using distilled transformer models, data augmentation, and cross-language transfer learning, the approach successfully addresses the challenges of limited computational resources and data scarcity in this domain. While the Full BERT model achieved the highest accuracy, the Distilled Transformer model's comparable performance at a fraction of the computational cost makes it a more practical choice. The methodology's ability to significantly reduce training time and inference latency makes it well-suited for deployment on a variety of devices, including those with limited computing power. The smaller memory footprint also makes it easier to use in places with limited computing resources. The research demonstrates that combining methods for model compression, efficient fine-tuning, and data strategies can effectively tackle the problem of social media text classification. This research contributes to making important natural language processing technologies accessible to more areas, which supports greater inclusion in the field.

References

K. Saifullah, M. I. Khan, S. Jamal, and I. H. Sarker, “Cyberbullying Text Identification: A Deep Learning and Transformer-based Language Modeling Approach,” EAI Endorsed Trans. Ind. Networks Intell. Syst., vol. 11, no. 1, pp. 1–12, 2024, doi: 10.4108/EETINIS.V11I1.4703.

A. Palanivinayagam, S. S. Gopal, S. Bhattacharya, N. Anumbe, E. Ibeke, and C. Biamba, “An Optimized Machine Learning and Big Data Approach to Crime Detection,” Wirel. Commun. Mob. Comput., vol. 2021, 2021, doi: 10.1155/2021/5291528.

M. E. Maron, “Automatic indexing: an experimental inquiry,” J. ACM, vol. 8, no. 3, pp. 404–417, 1961.

T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in Machine Learning: ECML-98, 10th European Conference on Machine Learning, in Lecture Notes in Computer Science, vol. 1398. Chemnitz, Germany, 1998, pp. 137–142.

I. J. Unanue, G. Haffari, and M. Piccardi, “T3l: Translate-and-test transfer learning for cross-lingual text classification,” arXiv Prepr. arXiv2306.04996, 2023.

[D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv Prepr. arXiv1412.6980, 2015.

A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers, Valencia, Spain, 2017, pp. 427–431.

S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” IEEE Trans. Neural Networks, vol. 20, no. 1, pp. 61–80, 2008.

Y. Wu et al., “Google’s neural machine translation system: Bridging the gap between human and machine translation," arXiv Prepr. arXiv1609.08144, 2016.

Q. Li et al., "A Survey on Text Classification: From Traditional to Deep Learning," ACM Trans. Intell. Syst. Technol., vol. 13, no. 2, 2022, doi: 10.1145/3495162.

F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv., vol. 34, no. 1, pp. 1–47, 2002.

F. Zhuang et al., "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, 2020.

C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?," arXiv Prepr. arXiv1905.05583, 2019.

Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V Le, "XLNet: Generalized autoregressive pretraining for language understanding," in Advances in neural information processing systems, 2019, pp. 5753–5763.

J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 2018, pp. 328–339.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Improving Language Understanding by Generative Pre-Training," arXiv Prepr. arXiv1801.06146, 2018.

[18] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Adv. Neural Inf. Process. Syst., vol. 26, 2013.

[19] Z. Hu, D. S. McNamara, A. Graesser, and Z. Cai, "Open-domain question answering with pre-trained language models," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.