Explainable Bengali Multiclass News Classification

Md Fahim Sikder[a,1,5,∇], Md. Ferdous[b,2,5,β], Shraboni Afroz[b,2,5,γ], Uzzal Podder[5,κ], Kaniz Fatema[b,2,5,ζ], Mohammad Nahid Hossain[4,5,ϕ], Md. Tahmid Hasan[b,3,5,ω], and Mrinal Kanti Baowaly[b,2,τ]
[a]Department of Computer & Information Science (IDA)
[b]Department of Computer Science & Engineering
[1]Linköping University, Sweden
[2]Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh
[3]Northern University of Business & Technology Khulna, Bangladesh
[4]Kwangwoon University, Korea
[5]Researcher, Deep Dream Group
Email: md.fahim.sikder@liu.se[∇], md.ferdous.kuet@gmail.com[β], samapti.afroza@gmail.com[γ], uzzwalbd@gmail.com[κ], kaniz.cse11@gmail.com[ζ], nahid@kw.ac.kr[ϕ], tahmidhasan.3003@gmail.com[ω], baowaly@gmail.com[τ]

***Abstract*—The automatic classification of news articles is crucial in the era of information overflow as it assists readers in accessing relevant information in a timely manner. Even though text classification is not a new area of study, there is potential for advancement concerning the Bengali language. Unlike other languages, Bengali is a complex language, and most of the datasets available online are imbalanced in terms of class label distribution. To increase the performance of classification methods and make them robust to handle imbalanced data, in this work, we propose a model consisting of pre-trained BERT architecture. We use a publicly available dataset of Bengali news articles with nine classes and achieve 92% accuracy. Along with the classification, explaining the model and the result is necessary for the application of trustworthy Artificial Intelligence. From this motivation, we use Integrated Gradient, an explainable AI technique, to explain the outcome of our model. We show which words in a news article affect the model to choose a particular class.**

***Index Terms*—BERT, Text Classification, Bengali News, Trustworthy AI, Explainability**

I. Introduction

Text classification is a classical Natural Language Processing (NLP) task where a set of texts is classified into predefined labels, and it aids in automating real-life applications, such as recommendation systems, sentiment analysis, and spam detection. Classifying documents manually is time-consuming and requires a considerable workload, whereas automating it would help the process. Over the years, researchers have used different techniques, such as hand-crafted features and machine-learning approaches, to achieve this task. TF-IDF [1], Naïve-Bayes [2], and Neural Networks [3], [4] are examples that have been used for this task. However, most of these works have been done in the English language [5], [6]. We also see some work has been done in different languages such as Spanish [7], Arabic [8], Indonesian [9], and Hindi [10].

Bengali is a complex language with various linguistic nuances, idiomatic expressions, and regional variations [11]. Also, this is one of the most widely spoken languages in the world. Building accurate models that can handle these intricacies can be challenging as the language is complex. Furthermore, text data might incorporate longer sequences and need much preprocessing. These characteristics make it challenging to work with the Bengali language. In recent years, several works have been done on text classification using different machine learning and statistical techniques. However, there is still room for improvement.

Researchers used TF-IDF [1], [12], [13], Naïve-Bayes [2], deep-learning based method, such as Long Short Term Memory (LSTM) [14], Bi-directional Long-Short Term Memory (Bi-LSTM) [3], Convolutional Nerual Networks (CNNs) [3], Attention [15] for text classification. Recently, Transformers-based architectures [16] have become popular for NLP tasks, such as GPT [17] and BERT [4]. These models are trained using large datasets, and then researchers adapt those pre-trained models by fine-tuning them for specific tasks. On top of this, Transformers architecture enables capturing the temporal dependencies better than its predecessor (LSTM, Bi-LSTM, CNNs) [18]. Also, Transformers-based architectures are robust enough to handle imbalanced datasets [19].

Besides text classification, it is essential to understand why the model makes a particular classification. This transparency ensures that the decisions are not arbitrary and can be justified. Explainable techniques allow us to identify which features are responsible for a particular classification, and this is the building block for making a trustworthy AI system. Though most models are complex, Neural Networks are known as black boxes because their internal workings are not easily interpretable. Different techniques have been proposed to interpret the model itself or the result generated by the model. Integrated Gradients [20], Shapley Additive exPlanations (SHAP) [21], Local Interpretable Model-Agnostic Explanations (LIME) [22], Grad-CAM [23] are some of the techniques/algorithms that can be used to interpret the complex models and its result.

This study proposes a model that uses pre-trained BERT architecture [24] to classify Bengali news articles. We use a publicly available Bengali news article dataset that contains over $400k$ articles [25] with nine classes. Here, we use a larger

This is a draft copy, for the published version visit: https://ieeexplore.ieee.org/document/10441218

token size of 250 (number of words per article) to see how our model handles the long-range temporal dependencies. After training, we compare our study to state-of-the-art studies and find that its performance is superior to others. Additionally, we conduct an ablation study and demonstrate how imbalanced data can give false impressions about evaluation accuracy. In addition to classification, we perform an explainability analysis using the Integrated Gradient (IG) on the news articles and highlight the words responsible for classification.

Our contribution to this study is as follows:

- We introduce a model incorporating pre-trained BERT architecture for classifying Bengali news articles that outperform state-of-the-art works.
- We perform explainability analysis using Integrated Gradient on our model to show which words are responsible for classifying a specific class.
- Additionally, we demonstrate how imbalanced datasets lead to misleading evaluation and the need for a robust model to handle imbalanced data.

The remainder of the paper is as follows: section II gives the overview of works related to this study. Section III provides the information needed to understand the paper better and how our proposed methodology works. Section IV refers to the experimental result and we finish with conclusions in section V.

The code for this paper is available on **https://github.com/fahim-sikder/explainable-bengali-news-classification**

II. Related Works

Text classification is an essential and regularly performed Natural Language Processing task that can automatically classify or predict text data, typically by employing statistical methods and supervised machine learning techniques. Much research has already been done on text classification using different methodologies for Bengali languages. Researchers used Support Vector Machine [26], Naïve-Bayes [2], TF-IDF [1], [12], [13] to classify Bengali texts over the years. As the size of the dataset grew over time, the need for more complex models became necessary. From that perspective, Neural Network-based architectures gained popularity [27]. Ahmed et al. [15] used Recurrent Neural Networks (RNNs) with Attention mechanism [16] to categorise Bengali documents where they achieved 97.72% accuracy on their dataset. In another work, Mohiuddin et al. [28] classified Bengali news headlines using a Bi-directional Gated Recurrent Unit that achieved 84% validation accuracy on their dataset.

Besides Recurrent Neural Networks-based architecture, Convolutional Neural Networks (CNNs) based architecture is gaining popularity for handling text data. Chowdhury et al. [14] classified Bengali news articles using CNNs, Long Short Term Memory (LSTM), and Glove vectorization and achieved 87% accuracy on the dataset collected from three newspapers. Another CNN-based work is from Seal et al. [3]. In their work, they used a combination of Bi-directional LSTM and CNNs to classify Bengali news articles and achieved 93.94% accuracy on the dataset they used.

III. Materials and Methods

In this section, we first provide the context and essential details to enhance the reader's comprehension of this paper, followed by a presentation of our research methodology.

A. Problem Definition

Given a set of documents $D = (D_1, D_2, ..., D_N)$ where each document contains a sequence of words, we need to classify these documents into categories $C = (c_1, c_2, ..., c_k)$ based on their contents. Here, k is the number of classes. We train a model f that takes the data $\mathbb{D} = (D, C)$ and classify each text document in a class. In this study, we use a pre-trained BERT model to classify Bengali news articles. We also show which words in an article are responsible for classifying a specific class by using explainable techniques.

B. Transformers and BERT

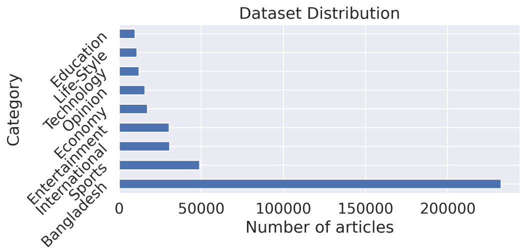
Transformers is one of the most popular architectures for Natural Language Processing and Computer Vision domains. It has been used in various state-of-the-art models. It is an Encoder-Decoder-based architecture where self-attention is used to capture long-term dependencies in the data whereas Recurrent and Convolutional Neural Network-based architecture failed to capture it due to limitations in their architectures. Transformers is capable of handling parallel input sequences which makes it superior to other previous models.

Bi-directional Encoder Representations from Transformers (BERT) [4] is a popular language model based on Transformers architecture. It uses a Transformer's Encoder and stacks them on top of each other. BERT focuses on pre-training a transformer model on a large corpus of text data using a masked language model (MLM) objective. In MLM, some words in the input are masked, and the model is trained to predict those masked words based on the context provided by the surrounding words. What sets BERT apart is its bidirectional training approach, where it considers both the left and right context of each word during training. This contrasts with earlier models that were unidirectional and could only consider one direction of context. BERT's bidirectional training allows it to capture more comprehensive contextual information, making its embeddings more contextually rich and useful for a wide range of downstream tasks when fine-tuned.

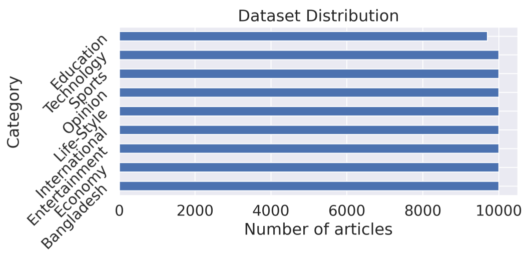
C. Integrated Gradients

Explainable Artificial Intelligence (XAI) refers to the process of explaining the model and why it makes any particular decision. This transparency is needed for trusting the model and its outcome and for the application of trustworthy AI. Several explainable algorithms have been proposed in recent years for different domains. In this study, we use the Integrated Gradient (IG) [20] technique to explain our model's output. IG attempts to explain the relationship between a model's predictions and its characteristics.

Integrated Gradients (IG) aims to address this by attributing the prediction to various input features while considering their values along a path from a baseline (e.g., all-zero input) to



(a) Distribution of imbalanced data



(b) Distribution of balanced data

Fig. 1: Data Distribution of the dataset

the actual input. Integrated Gradients work by calculating the integral of gradients along a straight path in the input space from a baseline to the actual input. The resulting integrated gradients provide a measure of feature importance. Positive values indicate that increasing the feature value contributed positively to the prediction, while negative values suggest the opposite. Larger absolute values indicate higher importance. We compute the IG for a given input x and its baseline x' by using the equation 1.

$$IG_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} \, d\alpha \quad (1)$$

Here, $\frac{\partial F(x)}{\partial x_i}$ is the gradient of $F(x)$.

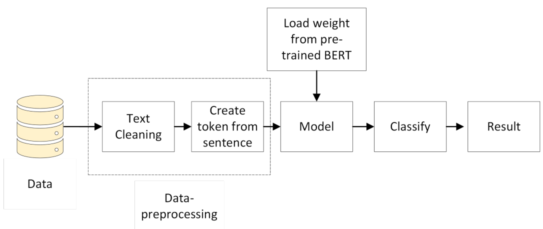


Fig. 2: Our proposed model

D. Dataset

In this study, we use a publicly available Bengali newspaper dataset. This dataset contains more than $400k$ news article from the Daily Prothom-Alo which ranges in nine categories. Due to this dataset's imbalanced distribution of classes, we only take equal number of news articles from every class and discard the rest. We run two experiements with both the balanced and imbalanced datasets. Figure 1 show the class distribution of the dataset.

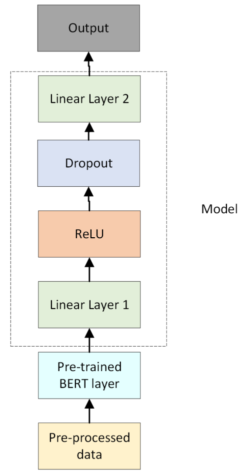


Fig. 3: Model architecture

E. Methodology

In this study, we use a BERT-based model to classify Bengali news articles. Besides this, we also use Integrated Gradients (IG) to explain which word in the articles is responsible for classifying a particular class. In this section, we describe our model's workflow and illustrate it in Figure 2.

1) Dataset Pre-processing: As mentioned, we use a publicly available Bengali news dataset from Kaggle [25], which contains more than $400k$ news articles with nine classes. We need to preprocess this dataset before using it in our model. To begin with, we clean the data by removing punctuation and unnecessary characters. After that, we separate each word from the articles (tokenisation). For this study, we use a token size of 250, which means we take 250 words from each article. In cases where news articles are less than 250 words, we use padding.

2) Classification Methods: After pre-processing the data, we feed it to the pre-trained BERT model. We choose BanglaBert-Large [24] as the pre-trained model. BanglaBert-Large has been trained with a large Bengali dataset, which is collected from 110 popular Bengali sites [24]. In order to fine-tune the BanglaBert-Large with our dataset, we use the rest of the model. Figure 3 shows the entire architecture of our model. After the data passes through the pre-trained BERT, it goes through a *Linear layer* followed by *ReLU* activation function. To avoid overfitting, we use a *Dropout layer* and then through a final *Linear layer*, which maps the model dimension with our class dimension. Hyperparameters to reproduce our work can be found in Table I.

3) Explainability: To explain which token/word in an article affects the model to choose a particular class, we use the Integrated Gradient (IG) algorithm. The first step is to convert an article into tokens. Then, we create a placeholder

TABLE I: Hyperparameters to reproduce the results

Linear Layer 1	128 Units
Linear Layer 2	9 Units
Dropout	0.2
Learning rate	1e-5
Optimizer	AdamW
Epochs	50

called *baseline*, which is the same shape as the input token. The baseline does not contain any information. Afterwards, we pass the tokens into our classification models and record the gradients. Then, using the baseline and the gradients, IG calculates a score called the *attribution score* for each token using equation 1. Ultimately, every token receives an attribution score, and the token with the higher score is responsible for classifying a specific class. Details of this method can be found in the Sundararajan et al. article [20].

4) Evaluation Metrics: We use macro average F1 score and accuracy to determine the classification performance of our model. The F1-Score is calculated using equation 2, and by calculating the mean for all classes F1-score, we get the macro average F1 score.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} = \frac{TP}{TP + 0.5 \times (FP + FN)} \tag{2}$$

Here, TP, FP, FN stands for number of true positives, false positives and false negatives respectively.

We also use the accuracy metrics to see how are model performing the classification tasks.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Here, TN stands for the number of true negatives in our classification.

To measure explainability, we calculate the attribution score described in section III-E3 and highlight the token responsible for the classification.

IV. Experimental Results and Discussion

After training our model with the balanced dataset, we compare the results with three relevant and recent works [3], [15], [28]. We use the same dataset when training these baselines. The result shows that our model outperforms these architectures by a significant margin. As we use a token size of 250, these architectures fail to capture the long-temporal dependencies. On the contrary, our model is comprised of Transformers architecture, which can capture these dependencies. That is why we achieve better performance than others. Table II shows the macro average F1-score and accuracy of our model and others. Figure 4 shows the confusion matrix of our model using the held-out dataset. We observe that our model successfully classifies all the classes.

TABLE II: Comparison of our work with the others using the same dataset (we use balanced dataset for this comparison)

Model Name	Macro avg F1-Score	Accuracy (%)
RNN + Attention [15]	.09	28
Bi-GRU [28]	.0001	0.001
BI-LSTM + CNN [3]	.08	27
Ours	**.92**	**92**

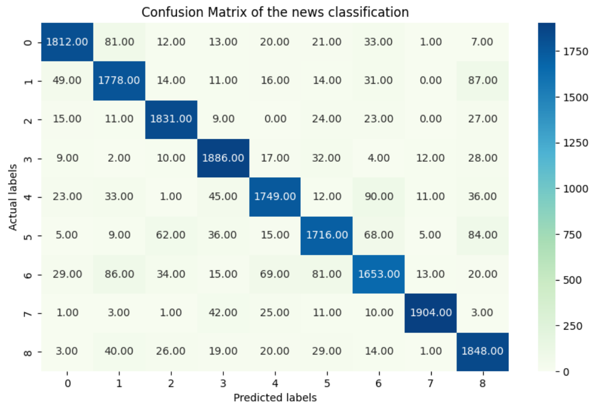


Fig. 4: Confusion Matrix of our model using held-out dataset

A. Ablation Study

We conduct an ablation study to see the performance of our model and the baselines on the imbalanced data. We train our model and two of the baselines using the original dataset where the number of articles in each class is imbalanced, and we record the macro average F1-score and the accuracy. Table III shows the ablation study with the imbalanced data. We observe that while the baselines struggle for the balanced dataset, their score is better for the imbalanced dataset. As the macro average F1-score is calculated by taking the mean of each class score, it can easily be fooled by this non-generalised baseline architecture's F1-score. In this case, the class *Bangladesh* in the imbalanced dataset has more than $250k$ articles, whereas other classes have less than $50k$. So, the F1-score of *Bangladesh* class is higher than the other classes. That is why when we calculate the mean of the F1-score for all the classes, we see a higher overall F1-score due to an imbalance in certain classes, thus giving the false impression of a reasonable classification score. Table IV shows the per class F1 score from one of the baseline [3].

On the contrary, our model performs better in both balanced and imbalanced datasets, proving our model's robustness.

B. Explainability Analysis

We take the news articles and pass them to the Integrated Gradient (IG) algorithm to get individual tokens' attribution score that plays a critical role in the classification. Figure 5 shows two examples of explainability analysis. In each article,

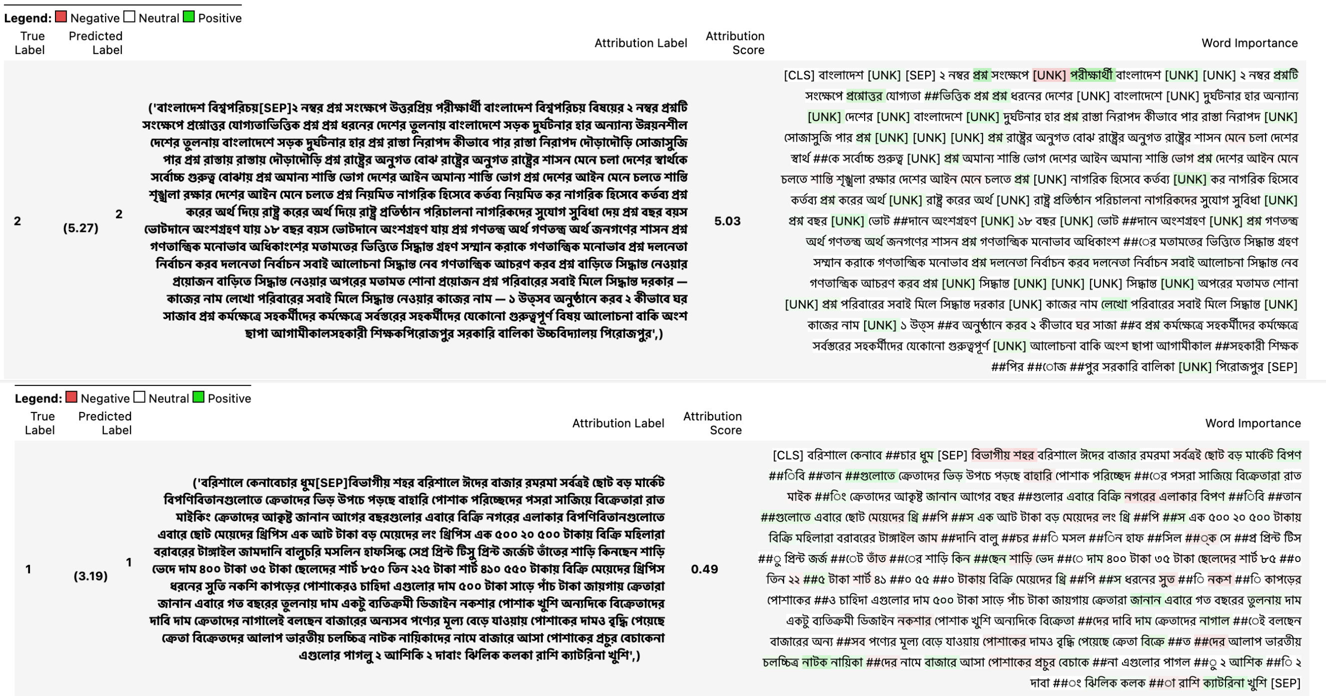


Fig. 5: Explainability Analysis: The first example is from the *Education* class and the second example is from the *Economy* class

TABLE III: Ablation Study using the imbalanced data

Model Name	Macro avg F1-Score	Accuracy (%)
Ours	**0.93**	**91**
RNN + Attention [15]	0.74	87
BI-LSTM + CNN [3]	0.73	86

TABLE IV: Ablation Study: Class per F1-score of the baseline [3] using imbalanced held-out dataset

Label	F1-Score	Label	F1-Score
0	0.920	5	0.667
1	0.6722	6	0.651
2	0.285	7	0.967
3	0.874	8	0.758
4	0.804		

we highlight the tokens that affect the classification decision, with green highlights representing positive impact and red highlights showing negative impact. In the first example in Figure 5, the words like প্রশ্ন (Question), পরীক্ষার্থী (Examinee), প্রশ্ন-উত্তর (Answer-Question) are highlighted green. These words represent the *Education* class.

Moreover, for the second example, মার্কেট (Market), ঈদের মার্কেট (Eid Market) is highlighted where the class is *Economy*. Besides these, we also see some wrong highlights in the example, such as the word **[UNK]** is highlighted more often

TABLE V: Corresponding English meaning of the Bengali words

Words	English Meaning	Words	English Meaning
প্রশ্ন	Question	পরীক্ষার্থী	Examinee
প্রশ্নটি	The Question	লেখো	Write
বিভাগীয় শহর	Divisional City	ঈদের বাজার	Eid Market
বাহারি	Different	বড় মার্কেট	Big Market
বিক্রেতারা	Seller	কিনছেন	Buying
নগরের	Suburb	এলাকার	Area

because the vocabulary size is not enough to accommodate all the words in the article. As a result, the IG algorithm sometimes misinterprets the word's attribution score. This is one of the critiques that IG has received over time. Table V shows some of the corresponding English meaning of the Bengali words that appear in Figure 5.

V. Conclusions

Due to the vast amount of text data being generated daily, the necessity of automating it based on their labels has increased. Though much research has been done in this field to advance this process, there is still scope for the Bengali language. In this study, we propose a BERT-based architecture to classify Bengali news articles. We train the model using a publicly available news dataset and show that our model can capture the long-term dependency and classify articles correctly where state-of-the-art studies fail to do that. We also show using an ablation study that an imbalanced dataset can give a false impression of good accuracy if the model is not robust. Besides classification, we perform explainability analysis to see which words in an article are responsible for classifying specific classes using the Integrated Gradient technique. In our future work, we will try to improve the accuracy of our model and will compare different explainable methods.

Acknowledgment

This work was funded by the Knut and Alice Wallenberg Foundation, the ELLIIT Excellence Center at Linköping-Lund for Information Technology, and TAILOR - an EU project with the aim to provide the scientific foundations for Trustworthy AI in Europe. The computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre. This research was also partially supported by the research fund of Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh. The authors would like to thank Sehrish Qummar from LiU for her support during the implementation phase. Deep Dream Group [1] is a non-profit research group that focuses on constructing state-of-the-art machine learning models for Natural Language Processing, Computer Vision, Time-Series analysis and Trustworthy AI. Md Fahim Sikder is the corresponding author.

References

[1] Z. Yun-tao, G. Ling, and W. Yong-cheng, "An improved tf-idf approach for text classification," *Journal of Zhejiang University-Science A*, vol. 6, pp. 49–55, 2005.

[2] A. N. Chy, M. H. Seddiqui, and S. Das, "Bangla news classification using naive bayes classifier," in *16th Int'l Conf. Computer and Information Technology*, pp. 366–371, IEEE, 2014.

[3] T. Seal, D. D. Dawn, A. Khan, S. K. Setua, and R. K. Pal, "Ben-cnn-bilstm: A model of consequential document set identification of bengali text," in *Applied Computing for Software and Smart Systems: Proceedings of ACSS 2022*, pp. 175–190, Springer, 2023.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[5] J. Wen and P. Liu, "A classification method for english texts based on hybrid recurrent neural network and graph construction in social recommendation systems," *IEEE Systems Journal*, 2023.

[6] H. Wu, Y. Liu, and J. Wang, "Review of text classification methods on deep learning.," *Computers, Materials & Continua*, vol. 63, no. 3, 2020.

[7] H. A. Ceccatto, R. A. Calvo, J. J. García Adeva, and U. Cerviño Beresi, "Automatic classification of new articles in spanish," in *X Congreso Argentino de Ciencias de la Computación*, 2004.

[8] O. El-Barbary, "Arabic news classification using field association words," *Advances in research*, vol. 6, no. 1, pp. 1–9, 2016.

[9] M. A. Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 2, no. 3, pp. 401–408, 2016.

[10] K. Dutta, S. Kaushik, and N. Prakash, "Machine learning approach for the classification of demonstrative pronouns for indirect anaphora in hindi news items," *The Prague Bulletin of Mathematical Linguistics*, vol. 95, p. 33, 2011.

[11] B. Pal and P. Bhattacharjee, "What is translated; what is not translated: studying the translation process of select bengali dalit short stories," *The Translator*, vol. 28, no. 1, pp. 41–53, 2022.

[12] M. T. Alam and M. M. Islam, "Bard: Bangla article classification using a new comprehensive dataset," in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pp. 1–5, IEEE, 2018.

[13] A. Dhar, N. S. Dash, and K. Roy, "Classification of bangla text documents based on inverse class frequency," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1–6, IEEE, 2018.

[14] P. Chowdhury, E. M. Eumi, O. Sarkar, and M. F. Ahamed, "Bangla news classification using glove vectorization, lstm, and cnn," in *Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021*, pp. 723–731, Springer, 2022.

[15] M. Ahmed, P. Chakraborty, and T. Choudhury, "Bangla document categorization using deep rnn model with attention mechanism," in *Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021*, pp. 137–147, Springer, 2022.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] OpenAI, "GPT-4 Technical Report," Mar. 2023. arXiv:2303.08774 [cs].

[18] M. F. Sikder, R. Ramachandranpillai, and F. Heintz, "Transfusion: Generating long, high fidelity time series using diffusion models with transformers," *arXiv preprint arXiv:2307.12667*, 2023.

[19] S. Paul and P.-Y. Chen, "Vision transformers are robust learners," in *Proceedings of the AAAI conference on Artificial Intelligence*, vol. 36, pp. 2071–2081, 2022.

[20] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*, pp. 3319–3328, PMLR, 2017.

[21] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.

[22] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[24] A. Bhattacharjee, T. Hasan, W. Ahmad, K. S. Mubasshir, M. S. Islam, A. Iqbal, M. S. Rahman, and R. Shahriyar, "BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla," in *Findings of the Association for Computational Linguistics: NAACL 2022*, (Seattle, United States), pp. 1318–1327, Association for Computational Linguistics, July 2022.

[25] Z. A. Nazi, "Bangla newspaper dataset," 2020.

[26] S. A. Taher, K. A. Akhter, and K. A. Hasan, "N-gram based sentiment mining for bangla text using support vector machine," in *2018 international conference on Bangla speech and language processing (ICBSLP)*, pp. 1–5, IEEE, 2018.

[27] M. F. Sikder, "Bangla handwritten digit recognition and generation," in *Proceedings of International Joint Conference on Computational Intelligence*, pp. 547–556, Springer, 2020.

[28] E. Mohiuddin and A. Matin, "Multilevel categorization of bengali news headlines using bidirectional gated recurrent unit," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, pp. 1–6, IEEE, 2021.

[1] https://deep-dream-group.github.io