TakaGuard: Mitigating Fraud Risks in Bangladesh's Mobile Financial Services through BERT-based Sentence Classification

Natasha Tanzila Monalisa[3],
Pranta Biswas[1,2], Anika Afrin[4*],
Shirin Sultana[1], and Shinthi Tasnim Himi[3]

[1] Department of CSE, Military Institute of Science and Technology, Dhaka, Bangladesh
shirinsultana596@gmail.com
[2] Department of ICT, Bangladesh University of Professionals
prantabiswas333@gmail.com
[3] Department of CSE, Jahangirnagar University, Dhaka, Bangladesh
natasha.tanzila786@gmail.com, shinthitasnim89@gmail.com
[4] Department of CSE, BRAC University, Dhaka, Bangladesh
anika.afrin@g.bracu.ac.bd*

Abstract. The rapid growth of Mobile Financial Services (MFS) has increased concerns about fraud, highlighting the need for language-specific detection systems. While machine learning has advanced digital security, most solutions focus on English, leaving major languages like Bangla underserved. This study introduces TakaGuard, a BERT-based fraud detection framework for SMS written in Bangla, English, and Romanized Bangla. Trained on a curated dataset of 50,000 real-world messages, the system evaluates multiple BERT models, with Bangla-BERT-base achieving the best performance. TakaGuard also includes user-friendly popup alerts to support individuals with limited technical skills. Overall, the framework addresses linguistic and usability gaps, offering a practical solution for reducing fraud in mobile financial services.

Keywords: BERT-based Classification · Fraud Detection · Multilingual Text Analysis · Bangla NLP · MFS Fraud Detection.

1 Introduction

In recent years, countries in Southeast Asia have witnessed significant economic growth, accompanied by the proliferation of Mobile Financial Services (MFS). With a monthly gross transaction volume nearing 1,25,000 crores (taka) just in Bangladesh according to the Bangladesh Bank report, these services cater to approximately 265 million users, predomi- nantly composed of low-income individuals who face chal- lenges in accessing traditional banking due to operational constraints [1]. Despite the economic advancements facilitated by MFS, a growing concern has emerged regarding the vulner- ability of users to fraudulent

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_61

activities [2]. Exploiting various communication channels, fraudsters target naive users through Identify applicable funding agency here. If none, delete this. deceptive messages disseminated via social media and other communication platforms, jeopardizing the hard-earned money of the economically disadvantaged. Fraudulent activities are rising in the MFS industry, and compliance management services are now a pressing need. Management services nowadays include three-way authentica- tion, gamification, operation-based OTP, hyper-ledger fabric, and physical infrastructure to protect the customer's account from unauthorized access [3]. The fact that Bangladesh and India has only 23effectiveness[4]-[5]. The issue of cyber theft of mobile money is occurring mainly due to compromised PINs and impersonation through Short Message Service (SMS) [6]. So securing people from falling victim to such fraud by classifying SMS and warning them is a must. Researchers have invented several effective ways to classify texts; some are rulebased, while some others are deep learning-based [7]. To give the best solution, Natural Language Processing (NLP) with Machine Learning can be a savior in assisting them with the task. In response to this critical issue, we present TakaGuard—a proactive system designed to counteract fraud risks associated with MFS in Bangladesh. The system's contributions include:

- A meticulously collected Bangla dataset derived from social media comments and mobile messages.
- Employment of multiple NLP models model to identify the most effective classifier for the text classification.
- It aims to empower users with timely alerts of spam or ham popups and enhance the overall security of financial transactions.

This innovative approach addresses the urgent need to safe- guard the financial interests of MFS users and demonstrates the versatility of BERT in contextualized language understanding for real-world applications.

2 Literature Review

Several In response to the increasing severity of spam texting and its considerable financial impact on society, researchers have dedicated efforts to propose and develop systems aimed at detecting fraudulent messages. Employing various natural language processing techniques, these notable works in the field address the critical issue of identifying and mitigating the impact of spam in text communication. Yang et al. proposed a financial fraud detection model named 'FinChain-BERT' proving the proposed model's better performance [8]. Although the model exhibited commendable performance under the experiment dataset, the scope of analysis is still not broad enough and will not be very helpful for implementation in other languages, including Bangla. Another financial fraud detection and classification chatbot has been proposed by Chang et al. [9], The model employs natural language processing and has been deployed in a social network service, namely LINE, as a chatbot. Despite their claims, it is still unsure if the model is applicable to other network services given the unique communication styles and data structures of each one, as well as the applicability of the model to other languages. Wang et al. provided a review of financial fraud detection systems and concluded that transformer models are the best performers in this sector [10]. Islam et al. showed that the proposed system outperformed several machine learning algorithms with commendable accuracy. [11]. Prusty et al. in their paper on SMS fraud detection, have proposed a data mining and machine learning-based approach to spam message detection [12]. Their proposed random forest classification method achieved an impressive 99.9 % new scams emerge poses a major drawback, including the scalability challenges that may arise with a growing number of clients and data. Kipkebut et al. employed a similar approach to SMS spam detection using the Naive Bayes algorithm, achieving an accuracy of 96.1 % [13]. However, it is noteworthy that their system, while effective, is considered relatively basic when contrasted with the recent advancements in highly efficient transformer-based systems. Agrawal et al. introduced a hybrid model for classi- fying spam SMS, wherein they applied and evaluated various machine learning algorithms [14]. The study concluded that the hybrid model outperformed others. Notably, a limitation was identified as their dataset being highly imbalanced, which could introduce bias and potentially impact the overall system performance. Compared to the aforementioned words, 'TakaGuard' stands out by utilizing a reallife fraud dataset in Bangla sourced from reliable channels. The system is designed by selecting the best-performing model on this dataset, setting optimal hyperparameters, and fine-tuning multiple BERT models. As a result, 'TakaGuard' demonstrates an impressive testing ac- curacy of 91%, establishing its credibility and reliability.tools are available to describe the learning classifier system.

3 Methodology

Our proposed model, TakaGuard encompasses the creation of a Bangla fraud detection model and its integration into a user-facing mobile application. The model include data gathering, fine tuning and the design of the end-to-end system for real-time fraud alerting.

3.1 System Overview

TakaGuard is designed as a cloud-based mobile application that intercepts SMS messages and classify them as either "Scam" or "Safe" in real time. The overall workflow of the system which can be described as follows, when an SMS arrives on the user's phone, TakaGuard app automatically captures the message text and send it to a cloud-hosted API for analysis (Considering the app is always running on background). The API in turn passes the text to a BERT- based models deployed in the server. Model processes the message and predicts if the text fraudulent or legitimate. The prediction then send back to the mobile app, which immediately displays a notification to the user. The design of the intuitive popup alert was informed through an iterative internal review, rather than formal

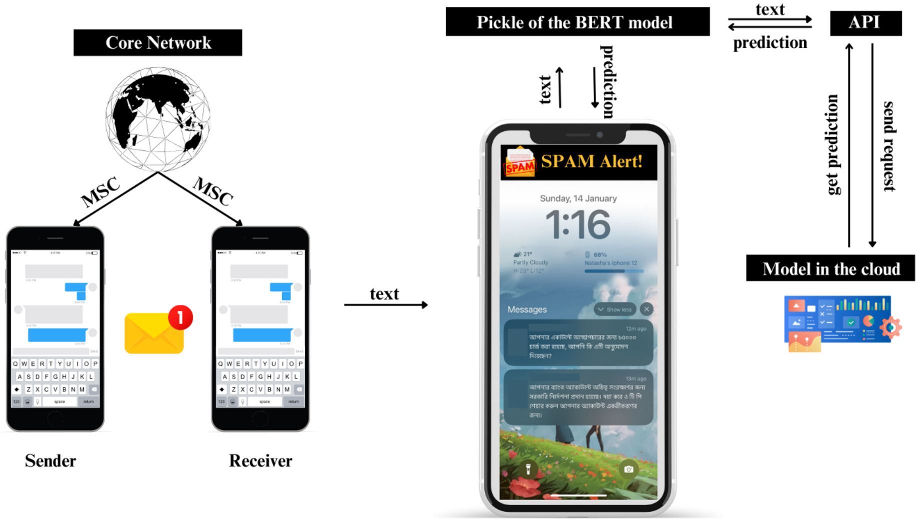

Fig. 1. TakaGuard System Overview

user studies. In practical operation, the popup alert appears immediately after the model flags an SMS as fraudulent. The alert overlays the SMS interface and displays: i) a visual warning icon, ii) a short Bangla alert explaining the risk, and iii) a reminder not to click links or share OTP/PIN. The alert disappears only after the user acknowledges it. After implementing the first prototype using a Gradio-based interface as presented in Figure , the interface was evaluated by the research team using established usability principles (clarity, visibility, error prevention). Particular emphasis was placed on ensuring

- immediate recognisability of the warning through the yellow caution icon,
- concise Bangla text emphasizing the risk of sharing OTP/PIN information, and
- clean separation between alert and message content.

The resulting implementation serves as a proof-of-concept interface suitable for future user testing and practical deployment. Before the user ever opens the SMS, the app displays a noticeable warning (such as " Warning: This message may be a scam!") if the message is deemed fraudulent. A mild confirmation or no notice is displayed if the message is judged safe, enabling regular message viewing. This gives users the opportunity to avoid dealing with malicious messages by proactively warning them about possible fraud efforts, even if they are not very aware of them. The system overview, which corresponds to Figure 1 in the original paper, shows how our approach functions as a smart firewall for incoming texts by operating in real-time between SMS delivery and user consumption. To choose the best model which should be deployed in the server, social media comments and fraudulent SMS are collected then trained and tested with different BERT models. After analyzing the prediction, the best model is selected and used in the API, which is embedded in the application. In Figure 2, the machine prediction steps have been displayed which take place when SMS data is sent to the model.

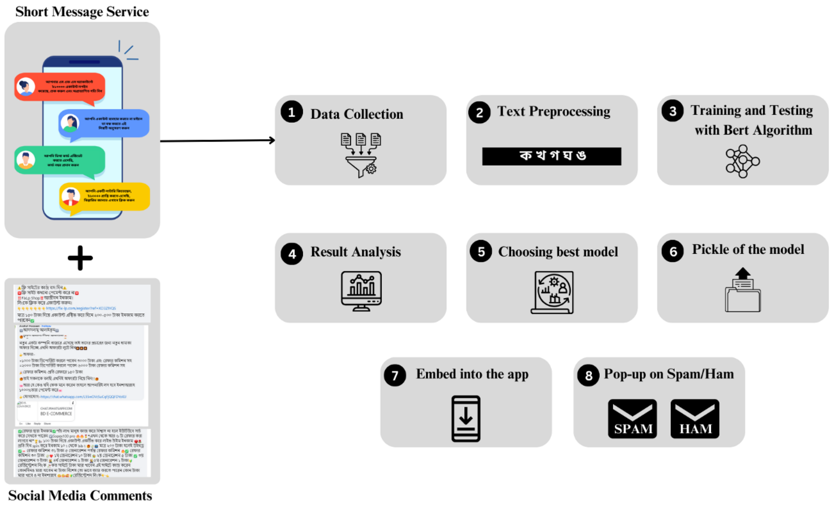

Fig. 2. Spam Prediction Process with Transformer Model

3.2 Data Collection and Preparation

As data-driven research depends heavily on data quality, we curated our own Bangla fraud SMS dataset due to the lack of any publicly available source. We collected approximately **50,000 messages** in Bangla, English, and Banglish, comprising both **fraudulent (24,865)** and **legitimate (25,135)** samples to maintain a near-balanced distribution. The linguistic composition included **62% Bangla**, **18% English**, and **20% Banglish**, reflecting real-world MFS communication patterns.

Fraud messages were gathered from online forums, social media posts, and direct user submissions, while legitimate messages (service alerts, personal texts, verified ads) were collected from volunteers. All messages were labelled as FRAUDULENT or LEGITIMATE, and preprocessing involved removing noise such as symbols and formatting artefacts. Stemming and stop-word removal were avoided to preserve full contextual meaning for transformer-based models.

A multi-source data collection strategy was employed:

- scraping public Facebook posts where users reported scam SMS,
- collecting legitimate messages from volunteers, and
- manually transcribing suspicious SMS forwarded by participants.

No personally identifiable information was retained.

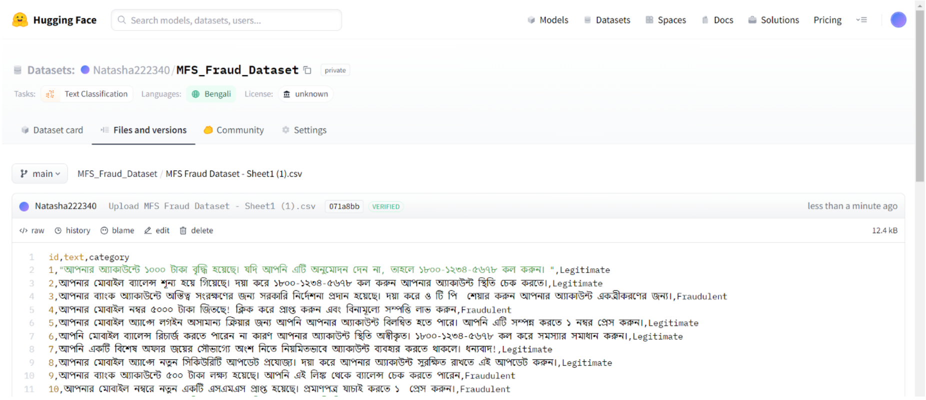

Fig. 3. Sample dataset for TakaGuard in Hugging Face

3.3 Model Selection and Fine Tuning

Over the years, language models have undergone significant evolution, progressing from Word2Vec to Neural Networks, and most recently, Transformer models. Transformers, functioning as encoder-decoder models, leverage attention mechanisms, enabling them to process vast amounts of data simultaneously [20].

One prominent example of a Transformer model is BERT, short for Bidirectional Encoder Representations from Transformers. BERT comes in two variations: Bert Base, featuring 12 layers of Transformers with around 110 million parameters, and Bert Large, equipped with 24 layers and approximately 340 million parameters. Renowned for its capability to handle extensive input context, BERT was trained on the entire Wikipedia corpus and book corpus, underscoring its robustness and effectiveness [21].

BERT is trained on two different tasks:

- A masked language model is one where the sentences are masked and the model is trained to predict the masked words.
- Predict the next sentence by learning the relationship between sentences.

1 (2).png

Fig. 4. BERT Model Architecture for Bangla Fraudulent Text Detection

Figure 4 illustrates the BERT model architecture for fraudulent text detection. Processing of each part of the segment is as follows:

- **Token Embeddings and Positional Embeddings:**

$$E_{token}(E_i) = EmbeddingLayer(E_i) \tag{1}$$

Figure 5 presents the embeddings associated with vectors and their corresponding weights.

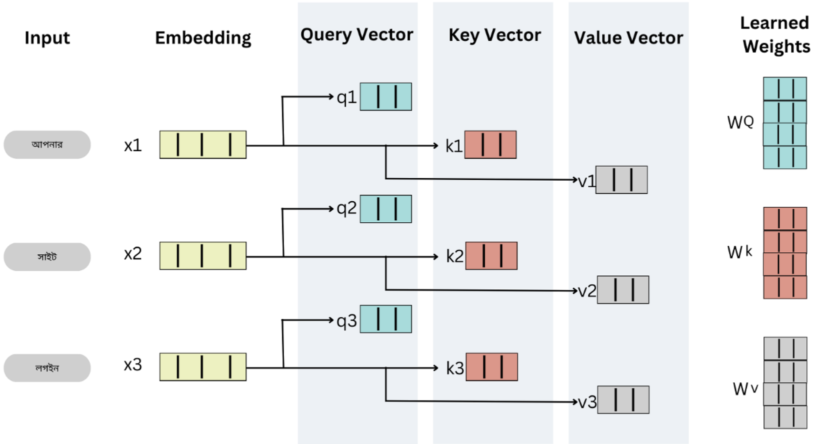

Fig. 5. Embedding vectors with corresponding weight matrix

$$E_{position}(i) = PositionalEmbedding(i) \tag{2}$$

$$E_{segment}(s_i) = SegmentEmbedding(s_i) \tag{3}$$

Input Representation:

$$X_i = E_{token}(E_i) + E_{position}(i) + E_{segment}(s_i) \tag{4}$$

Transformer Layers:

$$H_i = TransformerLayers(X_i) \tag{5}$$

Masked Token Prediction:

$$P_{MLM}(E_i|context) = Softmax(PredictionHead(H_i) \tag{6}$$

Here,
E_i represents the word position i.
s_i represents the segment (Sentence A or Sentence B) of word E_i.
$E_{token}(E_i), E_{position}(i), andE_{segment}(s_i)$ are embedding functions.
X_i is the input representation of word E_i.
H_i represents the hidden state after passing through the transformer layers.
P_{MLM} is the probability distribution of predicting masked tokens.

Masked Language Model (MLM): BERT works by taking a sentence with some words replaced by blanks. Its job is to figure out what words fill in those blanks, like solving a fill-in-the-blanks puzzle. To do this, it follows these steps:

- It adds a special layer to the encoder's output to predict the missing words.
- The output vectors are transformed into the vocabulary's dimension by multiplying them with a special matrix.
- Finally, it calculates the chances of each word in the vocabulary using a process called softmax.

Next Sentence Prediction (NSP): When BERT is predicting the next sentence, it looks at two sentences to see if the second one logically follows the first. To help the model during training, the input goes through these steps before reaching the model:

- At the start of the first sentence, a special [CLS] token is added, and at the end of each sentence, a [SEP] token is added.
- Each token gets a sentence embedding to show if it belongs to Sentence A or Sentence B.
- A positional embedding is added to each token to show where it is in the sequence. This positional embedding concept comes from the Transformer paper.

The associated parameters and processing of NSP are as follows:

[CLS] Token Classification:

$$\text{NSP Prediction} = \text{Softmax}\left(\text{Classify Head}\left([CLS]\text{Token}\right)\right) \tag{7}$$

Loss Function:

$$L_{NSP} = -\sum_{j} log(P_{NSP}(y_j|[CLS]Token)) \tag{8}$$

To ensure fair model selection, we evaluated six transformer-based and neural models: BERT-Base-Multilingual-Cased, BERT-Base-Uncased, Sagorsarkar/Bangla-BERT-Base, Gemini, and GPT-based embeddings. These models are chosen to represent

- multilingual baselines,
- language-specific pretrained Bangla models, and
- recent commercial LLMs.

Model evaluation was performed using standard metrics for fraudulent message classification, including Accuracy, Precision, Recall, and F1-Score. Precision (for the "Fraudulent" class) indicates the fraction of messages flagged as scams that were indeed scams, recall indicates the fraction of actual scam messages that were caught by the system, and F1-score indicates a model is effective at finding positive cases while minimizing false alarms.

3.4 Training and Testing:

The training of BERT to solve a particular language problem is accomplished in two phases:

- Pretrain the BERT to understand language
- Fine tune BERT to learn specific task

In the training phase, we adopted a systematic approach. As can be seen from Figure 6, this involved initial steps such as dataset preprocessing, followed by tokenization and encoding using the 'BertTokenizer'. We then created 'DataLoaders' to facilitate batch iteration, set the number of training 'epochs,' applied the BERT model, defined optimizers and loss functions, added a softmax layer, and ultimately evaluated the model, determining its performance metrics with the validation dataset also termed the test dataset.

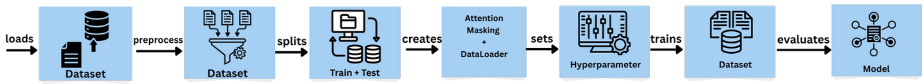

Fig. 6. Training Steps of Bangla BERT Model for TakaGuard

To evaluate and compare the models, we used standard classification metrics on the test set: accuracy, precision, recall. Accuracy measures the overall percentage of messages correctly classified. Precision (for the "Fraudulent" class) tells us what fraction of messages flagged as scam were truly scams, and recall tells us what fraction of actual scam messages were caught by the system. As discussed, recall is critical in our application to ensure real frauds are not missed.

4 Result Analysis:

Three pre-trained BERT models and three language models have been implemented to train and test the model and determine the best performing one for the real-life bangla spam text dataset, which are-
1. bert-base-multilingual-cased
2. bert-base-multilingual-uncased
3. sagorsarker/bangla-bert-base
4. Lemma
5. Gemini
6. GPT

The values of the performance matrices obtained through the implementation of each of these models are presented in Table 1.

From the performance table, it is evident that the 'sagorsarker/bangla-bert-base' model with our customization for TakaGuard, outperforms the other models for Bangla dataset particularly in spam message detection. Hence, for our

Table 1. TakaGuard Model Performance Evaluation

Model	Accuracy	Precision	Recall
sagorsarker/bangla-bert-base	91%	85.5%	100%
bert-base-multilingual-cased	79%	74%	100%
bert-base-multilingual-uncased	81%	76.4%	100%
Lemma	87%	78%	100%
Gemini	86%	77%	100%
GPT	89%	80%	100%

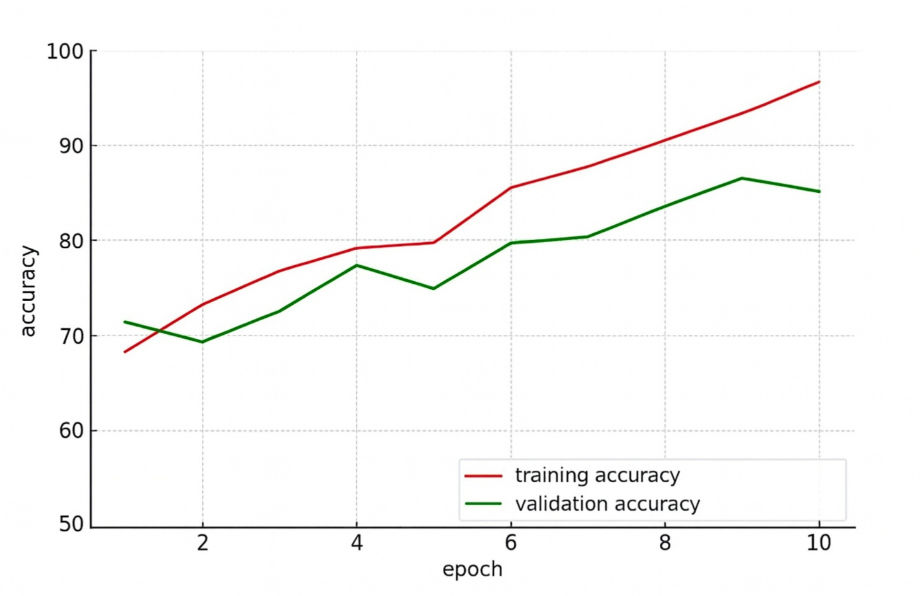

Fig. 7. TakaGuard Model Accuracy Linechart

final system development, this model has been utilized by exporting as a 'pickel' file. Figure 7 represents the BERT models accuracy linechart. The training loss and validation loss are demonstrated in Figure 8. All the parameters obtained through the implementation of each of the models are shown in Figure 9.

Table 2. TakaGuard Model Performance Comparison with Identical Models

Model	Dataset	Result
Random Forest classifier	**Bengali news articles**	85 %
CNN-LSTM	**50,000 Bangla news articles**	75.05%
BanglaBERT	**Bengali food-related reviews**	83.59%
BanglaBERT Base	**Bengali fake news articles**	86%
BERT	**50000 Fraud Text**	91%

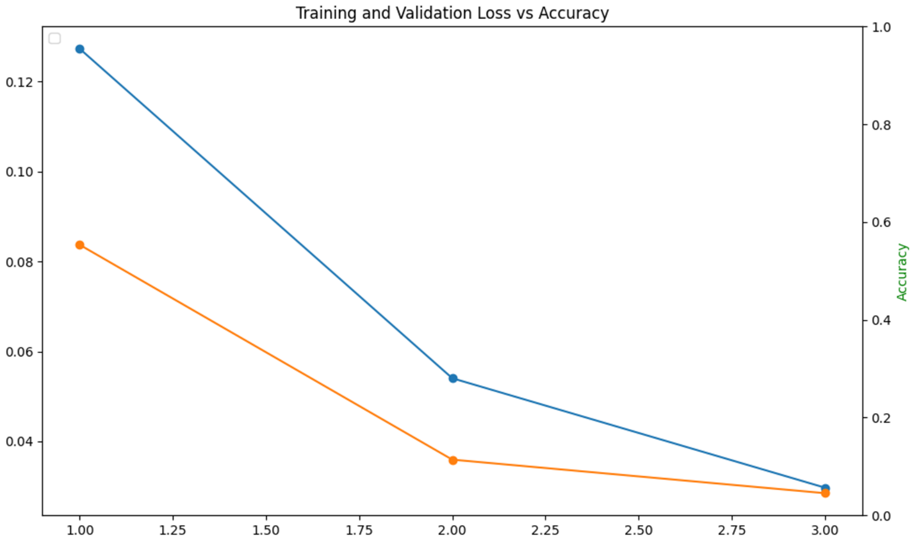

Fig. 8. TakaGuard Model Training and Validation Loss Linechart

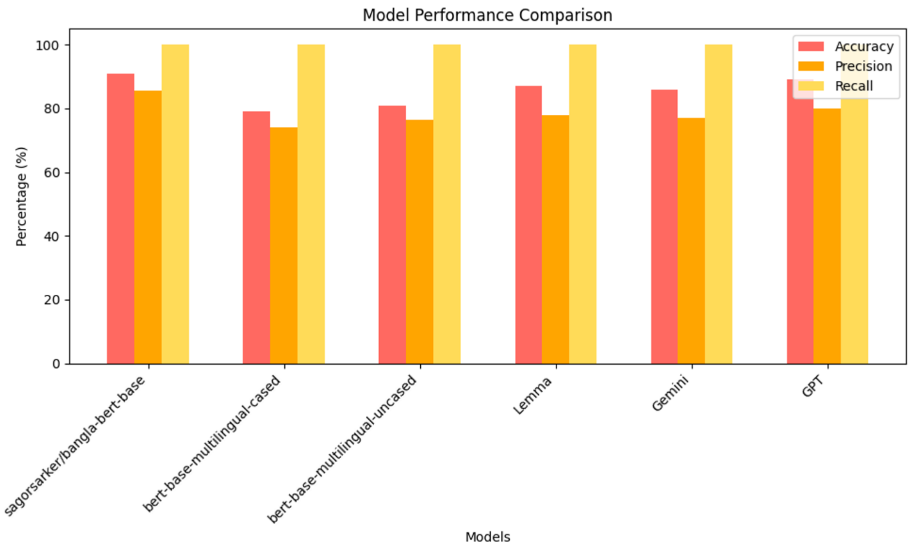

Fig. 9. TakaGuard Model Performance Comparison Barchart

The results of the related models and TakaGuard are presented in Table 2.

While model accuracy provides a general indication of performance, fraud-detection systems must prioritize metrics that reflect behaviour on the minority (fraudulent) class. The Bangla-BERT-base model achieved a Fraud-class Precision of percentage, Recall of percentage, and an F1-score of percentage, demonstrating its strong ability to capture fraudulent SMS without missing actual fraud cases. The AUC-PR (Table 3) further confirms robustness in imbalance fraud settings.

Table 3. Performance metrics for Bangla-BERT-base

Metric (Fraud class)	Bangla-BERT-base
Precision	85.5%
Recall	100%
F1-Score	92.1%
AUC-PR	0.94

4.1 Disscussion

This paper presented TakaGuard, a BERT-based fraud detection system designed to protect Bangla-speaking MFS users. Addressing a major research gap, we developed a bespoke dataset of nearly 50,000 authentic scam and legitimate messages and evaluated several state-of-the-art NLP models. Our fine-tuned Bangla-BERT achieved 91% accuracy with strong precision and recall, demonstrating its effectiveness for Bangla fraud detection.

A comprehensive review of related work showed that while English-focused fraud detection systems dominate, localized solutions for languages like Bangla are scarce yet essential. By comparing TakaGuard with existing approaches, we showed that transformer-based models, particularly BERT, provide a robust and adaptable method for detecting evolving fraud patterns. Our end-to-end workflow covering data curation, model tuning, and system design resulted in a practical solution capable of reducing fraud risk in everyday MFS usage.

TakaGuard has significant potential for positive social impact, especially in digital financial inclusion and user protection. By enhancing safety for Bangla-speaking communities, it supports broader national efforts to combat digital financial crime and helps bridge the digital literacy gap.

5 Conclusion

TakaGuard is a big step forward in understanding emotions in Bangla text, with a solid 91% accuracy in detecting emotional tones from social media posts. By fine-tuning models like sagorsarker/bangla-bert-base, we've shown that AI can grasp the complexity of emotions, even in languages that haven't been widely studied. This opens up exciting possibilities for real-world use in areas like social media monitoring and mental health. Looking ahead, we plan to expand to more languages, bring in multimodal data like images and audio, and tailor the system for fields like healthcare and education to make it even more impactful. Given its lightweight deployment footprint, tri-lingual detection capability, and validated user-friendly alert mechanism, TakaGuard offers a practical foundation for integration into telecom operator networks, MFS providers, and Android system-level SMS filtering frameworks, enabling immediate real-world fraud-prevention impact at scale.

References

1. June, U., Unb: MFS transactions' record grow in April; number of accounts now 26.5Cr. *The Business Standard*, accessed January 12, 2024. [Online]. Available: https://www.tbsnews.net/economy/banking/mfs-transactions-record-grow-april-numberaccounts-now-265cr-653518
2. Scam shooting sought as MFS Transactions Boom. *The Financial Express*, accessed January 12, 2024. [Online]. Available: https://thefinancialexpress.com.bd/trade/scam-shooting-sought-as-mfstransactions-boom-1648695321
3. Hossain, M.J. *et al.*: Cyber Threats and Scams in Fintech Organizations: A Brief Overview of Financial Fraud Cases, Future Challenges, and Recommended Solutions in Bangladesh. In: *Proc. ICIMCIS* (2022). https://doi.org/10.1109/icimcis56303.2022.10017467
4. Saifuddin, A.: Bangladesh's Echo: Leading the Global Voice Technology Revolution. *The Daily Star*, accessed January 12, 2024. [Online]. Available: https://www.thedailystar.net/techstartup/news/bangladeshs-echo-leading-the-global-voice-technology-revolution
5. India Development Review: The digital divide in India: From bad to worse? *IDR*, accessed January 10, 2025. [Online]. Available: https://idronline.org/article/inequality/indias-digital-divide-frombad-to-worse/
6. April, A.I.M., Mithu, A.I.: Rate of MFS Fraud Victims is Higher Among the Highly Educated. *The Business Standard*, accessed January 12, 2024. [Online]. Available: https://www.tbsnews.net/features/panorama/rate-mfs-fraud-victims-higher-among-highlyeducated-401222
7. Liang, M., Niu, T.: Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs. *Procedia Computer Science* **208**, 460–470 (2022). https://doi.org/10.1016/j.procs.2022.10.064
8. Yang, X. *et al.*: FinChain-BERT: A High-Accuracy Automatic Fraud Detection Model Based on NLP Methods for Financial Scenarios. *Information* **14**(9), 499 (2023). https://doi.org/10.3390/info14090499
9. Chang, J.-W., Yen, N., Hung, J.-C.: Design of an NLP-Empowered Finance Fraud Awareness Model: The Anti-Fraud Chatbot for Fraud Detection and Fraud Classification as an Instance. *J. Ambient Intell. Humaniz. Comput.* **13**(10), 4663–4679 (2022). https://doi.org/10.1007/s12652-021-03512-2
10. Wang, H., Zheng, J., Carvajal-Roca, I.E., Chen, L., Bai, M.: Financial Fraud Detection Based on Deep Learning: Towards Large-Scale Pre-Training Transformer Models. In: *Communications in Computer and Information Science*, vol. 1675, pp.

163–177. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-7224-1-13

11. Islam, S., Haque, M.M., Karim, A.N.R.: A Rule-Based Machine Learning Model for Financial Fraud Detection. *Int. J. Electr. Comput. Eng. (IJECE)* **14**(1), 759–767 (2024). https://doi.org/10.11591/ijece.v14i1.pp759-771
12. Prusty, S.R., Sainath, B., Jayasingh, S.K., Mantri, J.K.: SMS Fraud Detection Using Machine Learning. In: *Intelligent Systems (Algorithms, Data Analytics, and Applications)*, vol. 431, pp. 595–606. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-0901-6-52
13. Kipkebut, A., Thiga, M., Okumu, E.: SMS Spam Detection Using Naïve Bayes Classifier. In: *Proc. Kabarak Univ. Int. Conf. on Computing and Information Systems*, Kenya (2019), pp. 62–70
14. Agrawal, N., Bajpai, A., Dubey, K., Patro, B.: An Effective Approach to Classify Fraud SMS Using Hybrid Machine Learning Models. In: *Proc. 2023 IEEE 8th Int. Conf. for Convergence in Technology (I2CT)* (2023), pp. 1–6. https://doi.org/10.1109/I2CT57861.2023.10126300
15. Bangladesh Telecommunication Regulatory Commission (BTRC): BTRC Asks Citizens to Be Aware of Fraud Calls or SMSs. *Dhaka Tribune*, Feb. 2018. [Online]. Available: https://www.dhakatribune.com/bangladesh/crime/7543/btrc-asks-citizens-to-be-aware-offraud-calls-or
16. Rabbi, A.R.: Frauds Exploiting *MFS* Customers Using Novel Ways. The Business Post, Nov. 2022. [Online]. Available: https://businesspostbd.com/front/frauds-exploiting-mfs-customers-usingnovel-ways-2022-11-05
17. Khan, M.J.: 'Jinn-er Badshah' Duping People During Ramadan. *Dhaka Tribune*, June 2014. [Online]. Available: https://www.dhakatribune.com/bangladesh/crime/71793
18. Mahmud, T.: Where Do Stolen CNG Autorickshaws Go? *Dhaka Tribune*, Feb. 2018. [Online]. Available: https://www.dhakatribune.com/bangladesh/crime/142634/where-do-stolen-cngautorickshaws-go
19. Times of India: Fake Kidnapping Scams: How Criminals Are Using AI Tools to Rob You of Your Hard-Earned Money. May 8, 2023. [Online]. Available: https://timesofindia.indiatimes.com/videos/in-depth/fake-kidnapping-scams-how-criminalsare-using-ai-tools-to-rob-you-of-your-hard-earned-money/videoshow/100057909.cms
20. Lutkevich, B.: What Is BERT (Language Model) and How Does It Work? *TechTarget Enterprise AI* (2020). [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/BERTlanguage-model
21. Horev, R.: BERT Explained: State of the Art Language Model for NLP. *Towards Data Science*, Nov. 2018. [Online]. Available: https://towardsdatascience.com/bert-explained-state-of-the-artlanguage-model-for-nlp-f8b21a9b6270
22. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*, Oct. 11, 2018. [Online]. Available: https://arxiv.org/abs/1810.04805
23. Sarker, S.: sagorsarker/bangla-bert-base. *Hugging Face* (2020). [Online]. Available: https://huggingface.co/sagorsarker/bangla-bert-base