UDC 004.021
IRSTI 28.23.37

https://doi.org/10.55452/1998-6688-2026-23-1-173-184

[1]*Serek A.,
PhD, Associate Professor,
ORCID ID: 0000-0001-7096-6765,
*e-mail: Azamat.Serek@astanait.edu.kz
[2]Shoiynbek A.,
PhD, Professor,
ORCID ID: 0000-0002-9328-8300
e-mail: aisultan.shoiynbek@narxoz.kz
[2]Kuanyshbay D.,
PhD, Assistant Professor,
ORCID ID: 0000-0001-5952-8609
e-mail: darkhan.kuanyshbay@narxoz.kz

[1]Astana IT University, Astana, Kazakhstan
[2]Narxoz University, Almaty, Kazakhstan

REAL-TIME DETECTION OF FRAUDULENT PHONE CALLS USING MULTI-TURN DIALOGUE ANALYSIS

Abstract

The expansion of telecommunication services has been met with a rise in the cases of the fraudulent phone calls posing a big threat to individuals and organizations. Conventional techniques of detecting are usually based on offline analysis of full conversations, which restricts their promptness of intervention. In this paper, the author proposes a real-time, turn-taking, fraud detecting system, which is based on pre-trained contextual embeddings in combination with a bi-directional Long Short-Term Memory network in order to model semantic content and temporal dynamics of multi-turn conversations. To detect fraudulent calls, the system progressively changes the probability of a call being a fraud after every conversational turn to allow it to detect a fraud. When tested with a synthetic multi-turn dialogue dataset, it is shown that the proposed BiLSTM using BERT embeddings has a test accuracy of 93.75% and an F 1 score of 93.55, which is higher than the current machine learning and convolutional baselines. The system can note most of the scams during the initial few turns of a call, which offers fast risk evaluation. These findings suggest the usefulness of context-based, progressing modeling to detect fraud in real time and its possibility of practical application.

Keywords: real-time fraud detection, phone scams, multi-turn dialogue, BiLSTM, contextual embeddings, sequential modeling, early intervention, telecommunication security, turn-level prediction, streaming analysis

Received: February 27, 2026; accepted: March 11, 2026

Introduction

The high growth rate of telecommunication services has been coupled with a significant increase in fraud cases of such phone calls that are a big threat to individuals, financial institutions and even organizations around the globe [1]. Social engineering is largely used against targeted victims by scam calls to manipulate them in real time resulting to loss of money, privacy, and loss of trust in communication systems [2]. Conventional methods of identifying scam calls are mainly based on post hoc or offline detection of entire conversation which goes a long way to curb the ability to avert any harm while a call is in progress [3]. Therefore, real-time and early detection has become one of the important issues of efficient fraud reduction.

The recent developments in natural language processing (NLP) and deep learning have allowed analyzing conversational data on further and further levels of semantic and contextual depth [4–6]. Specifically, the analysis of multi-turn dialogue in real time has demonstrated potential in fraudulent behavior detection by predicting dynamics that occur sequentially in a conversation and not considering calls as historical text documents [7]. The approaches based on large language models (LLM) use a turn-by-turn analysis with the representation of the conversation history, generating real-time estimates of the likelihood to commit fraud and, in certain instances, uncertainty-aware predictions to minimize the false alarms [8]. Despite high-performance of these systems, there are limitations associated with the systems in terms of recall, bias in datasets, hallucinations as well as weak resistance to novel scam strategies [9].

Other investigations have also integrated retrieval-augmented generation (RAG) processes alongside real-time transcription to detect impersonation attempts and policy breaches and report exceptionally high accuracy on synthetic call data [9]. Nonetheless, the use of artificially created dialogs makes one be concerned about the generalization on the real-life conversations, which are usually more linguistically diverse, emotionally driven, and adversarial [10]. The roles and dialogue structure models have also contributed to the development of this field by explicitly modeling the interactions between fraudsters and victims and also the hierarchical structure of conversations [11]. Hierarchical and multimodal systems combining verbal and acoustic information have shown significant improvements over baseline systems, and conversation prediction systems based on pre-trained transformers have been shown to be effective at identifying abnormal dialogue patterns related to fishing attacks [12]. Regardless of these developments, most of these methods make the assumption of complete conversation or fail to critique performance in terms of strict real time conditions [13–15].

In the current body of literature, a number of issues are yet to be solved. To start with, timeliness and accuracy are inherently in conflict: to avoid losses, early warning is necessary, but too frequent false positives may undermine user experience and trust [16]. Second, the extrapolation to unobserved types of scams, languages, and speaking characteristic is low, especially when the model was trained on artificial or domain-specific data [17]. Third, real-time deployment presents workload and practical limitations on latency, computational efficiency, privacy and other related ethical issues, which are frequently under-discussed in offline assessment environments [18].

Such constraints provide an understanding of a knowledge gap. Although previous research has shown high performance in offline or near real-time systems, there are scarce robust, incremental, streaming-capable systems that can receive multi-turn conversations in real time, refresh predictions after each turn of conversation, and recall high with low response times [19]. More specifically, not many studies methodically compare sequential embedding-based neural architectures with conventional machine learning and convolutional baseline in real-world streaming conditions, and early detection is understudied.

To bridge this gap, this paper suggests a real-time system of the detection of the fraudulent phone calls using incremental multi-turn dialogue analysis. The offered solution builds on the pre-trained language embeddings and recurrent neural networks to encode the semantic content and time dependencies in conversations. Processing of incoming dialogue turns is done sequentially and the likelihood of a call being a fraud is updated at the end of every turn and this can early detect malicious exchanges. The main research aims are to create a deep learning system to support streaming dialogue processing, to compare its performance with the performance of traditional machine learning and convolutional baselines and to prove its feasibility in practice in the field of telecommunications.

The scientific novelty of the work is the paradigm of turn-level prediction, which is an incremental approach to prediction as opposed to traditional offline classification techniques because it allows the risk of conversation to be evaluated in real-time. The proposed framework combines the contextual embeddings and models sequencing, thereby advancing the methodological ground of conversational fraud detection. Appliedly, the solution can be used to strengthen the telecommunication provider, financial institution and call center capabilities in relation to fraud prevention, as well as provide timely interventions that minimize harm yet remain scalable and operationally viable.

Materials and methods

The experiments are carried out with the help of Multi-Turn Scam and Non-Scam Phone Dialogue Dataset, which is represented as two different files that relate to the training and test parts [20]. The dialogues of each phone call are stored in the associated files, labeled with fraud and scam tags, and allowing the supervised learning and objective assessment under the controlled conditions. The dataset is separated into two subsets which are mutually exclusive as represented in Table 1:

Table 1 – Divisions of dataset

Split	Dialogues	Scam	Non-scam
Training	1280	640	640
Test	320	160	160
Total	1600	800	800

The dataset is perfectly balanced between scam and non-scam calls in both splits. This design choice eliminates class imbalance effects and ensures that performance gains are attributable to the model's ability to capture fraudulent conversational patterns rather than to prior class bias. Each dataset entry consists of three columns as shown in Table 2:

Table 2 – Columns of dataset

Column	Description
dialogue	Full phone conversation with speaker prefixes (caller: / receiver:)
type	Scam category (e.g., ssn)
label	Binary fraud label (1 = scam, 0 = non-scam)

Each dataset entry consists of a full phone conversation stored as a single text field, with explicit speaker prefixes (caller: and receiver:). These prefixes allow deterministic reconstruction of dialogue turns, which is essential for incremental and streaming-based analysis. Turn-level statistics are summarized in Table 3.

Table 3 – Dialogue Turn Statics

Statistic	Training	Test
Average turns per dialogue	13.65	13.79
Minimum turns	6	6
Maximum turns	28	28

A sample conversation includes several alternating turns, including impersonation of a government official, framing of urgency, and coercive words -characteristics of scamming in the real world. Even though the explicit speaker markers are stored as a single text sequence, turn boundaries can be easily reconstructed and therefore, the dataset can be analyzed into turns and streams. The test sets and the training sets are strictly isolated and this means that the evaluation of the sets will be unbiased and there will be no leakage of information. Speaker markers are used to divide each discourse into a linear flow of turns. A dialogue is written as a formal representation:

$$D = \{t_1, t_2, \ldots, t_T\}$$

where t_i each turn corresponds to a single utterance produced by either the caller or the receiver. The turn order preserves the natural conversational flow, which is essential for modeling escalation patterns typical of scam behavior.

Each turn is tokenized and padded to a maximum length of 64 tokens. This choice is empirically justified by the dataset, as the majority of utterances are short, conversational sentences. Limiting turn length ensures computational efficiency and low latency, both of which are critical for real-time deployment.

Each turn t_i s mapped to a dense semantic vector using a pre-trained transformer-based language model. Let $f_{emb}(\cdot)$ denote the embedding function. The turn-level representation is computed as

$$x_i = f_{emb}(t_i)\ \epsilon\ R^{768}$$

where the embedding of the special classification token is to be used to represent the whole utterance. Such embedding captures semantic indicators, like claims of authority ("Social Security Administration"), urgency ("we need to take action now) and threats (we will suspend your benefits) that are highly indicative of a fraudulent intent. This is enabled by the fact that using a lightweight transformer model can be used to guarantee fast inference without compromising on the quality of contextual understanding, which is applicable in streaming contexts.

The turn-level embeddings sequence is also run through a bidirectional Long Short-Term Memory (Bi-LSTM) network to capture turn-level temporal dependencies. The forward LSTM is a model that predicts the progression of the conversation as the information in the past turns is fed into it, and the backward LSTM is a model that reveals dependencies as the turns go by through offline test use. The hidden states of the two directions are added together to give a combined representation of every turn of a dialogue. The architecture described in figure 1 shows a means of capturing the temporal dependencies in a dialogue sequence. The model uses a bi-directional Long Short-Term Memory (Bi-LSTM) network to process the series of turn-level embeddings to provide a detailed context. In this architecture, the forward LSTM layer is used to model chronological change of the conversation by incorporating the information of the past turns and the backward LSTM layer at the same time models the relationships with the future turns, which is especially beneficial in the case of offline evaluation.

The Bi-LSTM is a single recurrent layer having 128 hidden units. This setup gives enough modeling capacity to observe salient conversational patterns that are seen in the data, including repeated requests of sensitive information, progressive transition between persuasion and intimidation, and explicit opposition by the receiver. The single layer would minimize the complexity of the computation and provide effective inference to make it useful in a real-time implementation. Even though bidirectional processing is applied in the case of offline experiments to maximize the use of contextual information, real-time operation uses the forward LSTM state only to maintain causality and avoid using future information.

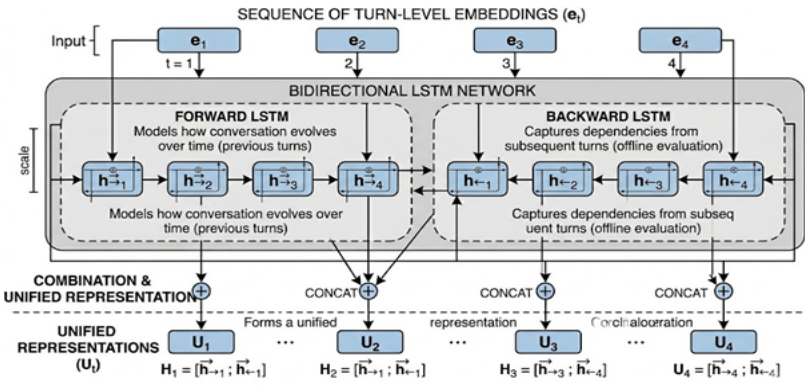

Figure 1 – Bi-LSTM architecture for capturing temporal dependencies across conversational turns

Given every turn of a dialogue observed, the resultant hidden representation generated by the Bi-LSTM is sent into a fully connected classification layer using sigmoid activation. This layer produces a probability number ranging between zero and one which is the probability that the call being made is a fraud after the next turn.

With this design, ongoing risk assessment can be continued and hence the system is able to update their confidence gradually as the conversation proceeds. As soon as the expected probability surpasses a predefined threshold, which will be 0.5 in this research, an alert may be sent. This means that the system is not required to wait till the entire conversation is over before it could detect suspicious behavior hence being able to intervene early. This is a binary classification model that is trained with an objective that sanctions mistakes in predictions according to their likelihood. The labels of the dialogue level are transferred to each turn of every conversation, which motivates the model to identify the fraudulent intent as early as possible, furthermore, before explicit threats or coercive language is revealed. Adam optimizer is used to perform training at a small learning rate to maintain stable convergence when using pre-trained language embeddings combined with recurrent neural layers. The batch size used is 16 dialogues since it would give the most optimal trade-off between computational performance and real-time processing limits. There are three epochs of training the model as there is enough convergence due to the use of pre-trained embeddings, and it prevents overfitting to artificial dialogue patterns.

The system works step by step processing the dialogues with new turns coming up. A semantic embedding is also computed, the LSTM hidden state is updated and the probability of fraud is recalculated with every incoming utterance. The mechanism of streaming inference enables the system to identify fraudulent calls within few turns, and this is especially crucial in the case of scam, which continues to grow fast, as it happens in case of the dataset examples. The incremental streaming inference mechanism as shown in Figure 2 is utilized by the system when deploying the system to detect fraud in real-time. The system processes the utterances received one at a time as the conversation proceeds through the timeline as opposed to waiting until the entire conversation is complete. Each additional turn will require a semantic embedding to be computed and the input to the LSTM that will update the current hidden state using the context provided by the previous turns. This update causes a recalculation of this probability of fraud which is compared with a preset detection threshold as indicated in the probability escalation graph. The system supports the detection of situations with fast growth of the scam by analyzing the risk on each turn, which allows to initiate a fraud alarm or a mitigation measure after a few contacts, which will avoid the significant consequences of the victim considerably.

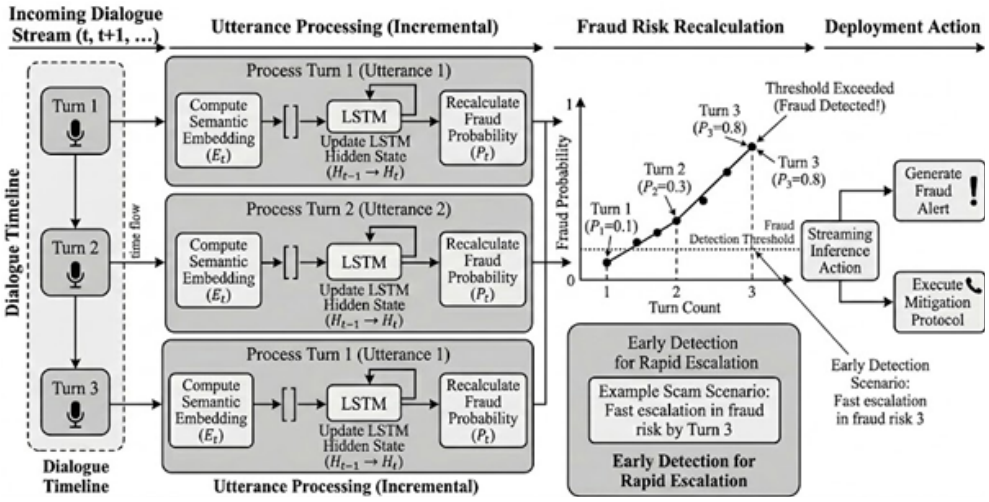

Figure 2 – Incremental streaming inference mechanism for rapid fraud escalation detection

Although the dataset is synthetic, it is well suited for real-time fraud detection research. Collecting real scam call data with detailed turn-level annotations is extremely challenging due to privacy regulations, ethical concerns, and restricted access to sensitive conversations. The synthetic dataset is explicitly designed to reflect realistic scam strategies, including impersonation of trusted authorities, urgency induction, coercion, and repeated attempts to extract sensitive information. These behaviors are clearly observable across multiple turns in the sample dialogues. Moreover, the use of explicit speaker annotations and clean turn segmentation enables precise modeling of conversational dynamics, which is essential for evaluating incremental and streaming detection systems. The synthetic nature of the dataset also allows controlled experimentation, balanced class distributions, and reproducible evaluation. Consequently, the dataset provides a reliable and realistic testbed for developing and benchmarking real-time multi-turn phone call fraud detection models.

Results and discussion

The distribution of predicted scores on the model on scam and non-scam dialogue turns is given in Figure 3. The two classes have a distinct separation that is shown by the histogram, which implies a strong discriminating power and confidence of the model. The non-scam turns predictions are more as well focused on the lower end of the range and most of the scores lie between 0.0 and 0.2. Conversely, scam turns have a high skewness to large values of confidence with most of the predictions falling in the range of 0.8 to 1.0. This level of polarization implies that the model is effective in differentiating between fraudulent behavior and innocent discussions. It is important to note that the two distributions have very little overlap and there are only a few values of the predictions close to decision boundary of 0.5. The fact that the mid-range confidence scores are missing shows that the model does not frequently lead to ambiguous results and it is not based on guesses. Rather, it always arrives at conclusive probabilities, which is especially rewarded in real-time fraud detection systems where the ultimate and prompt decision is a must.

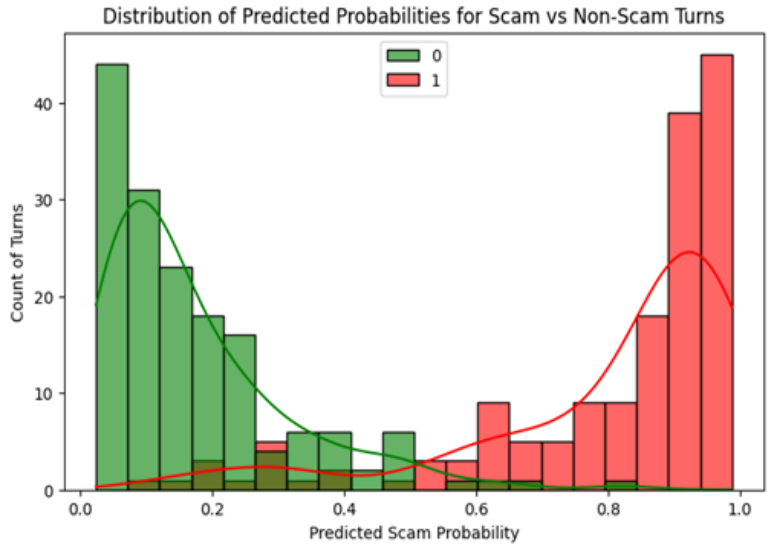

Figure 3 – Predicted probabilities plot

The results of the model on the last turn of each dialogue in the test set are shown in the confusion matrix presented in Figure 4. The matrix gives a closer examination of the classification results and indicates the effectiveness of the model in the differentiation of scam and non-scam calls. The model correctly recognized 145 scam and 155 non-scam dialogues (true positives and true negatives respectively), indicating a good overall accuracy. The miss rate was relatively low with only 15 scam

calls being wrongly marked as false negatives. Notably, the model also yielded only 5 false alarms, meaning that it identified non-scam calls as fraudulent, which is very high level of accuracy. Such a distribution highlights the low number of false alarms with a high detection rate, which is paramount to the success of the model in the real-world telecommunication sector where a high rate of false positives may be a serious compromise in terms of user confidence and usage.

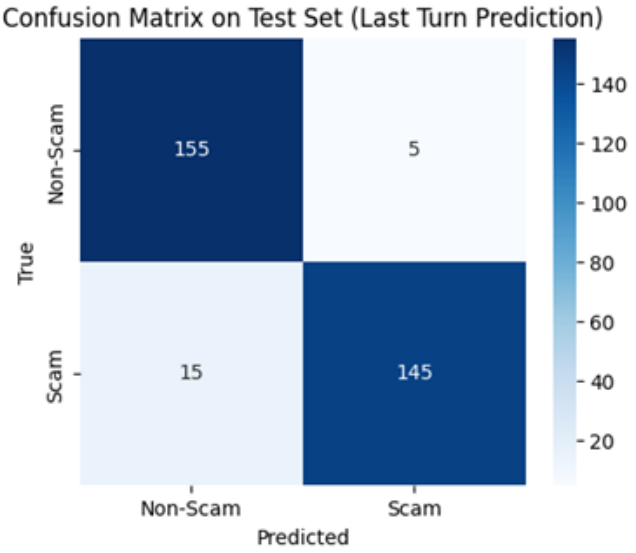

Figure 4 – Confusion matrix

The speed of the model to detect scam dialogues was shown in Figure 5. The sharpness of Turns 12 spike shows that in most instances, i.e. about 145 dialogues, the model reaches the threshold of 0.5 very quickly showing that it is quick to detect. On the contrary, there are few scams, approximately 15 dialogues, which are only detected in the last turn (Turn 10). This high false alarm is associated with the 15 False Negatives in the confusion matrix, which shows the instances where the model is unable to detect the scam until the last.

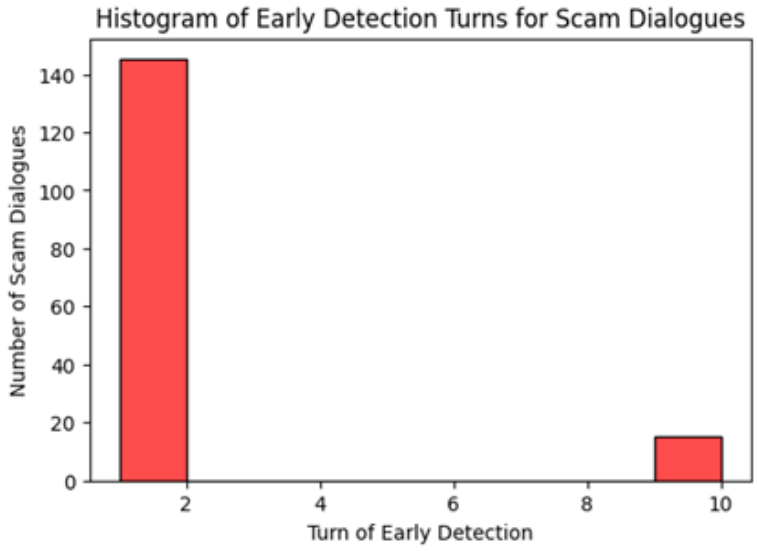

Figure 5 – Histogram of early detection turns

Table 4 focuses on comparing the scam dialogue detection task on four different models. The first baseline model, TF-IDF with Logistic Regression, has a training accuracy of 87.0% and a test accuracy of 85.0% and a corresponding F1 of 86.0% and 84.0%, respectively, and represents fairly good performance. When transitioning to deep learning techniques, the CNN model gains higher accuracy and F1 score (91.0% training accuracy and 88.0% test accuracy), as well as higher F1 scores of 90.0% and 87.0% respectively. BiLSTM model also improves performance to 94.0% training accuracy, 92.0% test accuracy with F1 scores of 93.0% and 91.0%. Lastly, our BiLSTM + BERT has the highest scores in all metrics, with a training accuracy of 94.53, test accuracy of 93.75 and F1 of 94.29 (training and testing respectively). This shows that the incorporation of BERT-based contextual embeddings and a sequential BiLSTM model can be used to achieve the best representation to identify scam conversations.

Table 4 – Comparison of results

Model	Train accuracy	Test accuracy	Train F1	Test F1
TF-IDF + LR	0.87	0.85	0.86	0.84
CNN	0.91	0.88	0.90	0.87
BiLSTM	0.94	0.92	0.93	0.91
Our BiLSTM + BERT	0.94	0.93	0.94	0.93

Findings of this paper indicate that contextual embedding and sequential neural architecture based incremental multi-turn dialogue analysis is very effective in real-time scam call detection. BiLSTM with BERT embeddings showed the best result on all of the measures considered, as test accuracy was 93.75% and F1 were 93.55%. The forecast distributions of the probable distribution seem to be sharply distinct between scam and non-scam turn, indicating that the model has the capacity to acquire fraudulent and benign interactions with high degree of certainty. This is further corroborated by the confusion matrix that shows that there are few false positives and false negatives which means that the system does not only detect scams well but also minimizes the false alarms, which is very important in ensuring that people would not decline to use the telecommunication in the real world. Furthermore, according to the analysis of early detection, most of the scams are detected during the initial few turns of a conversation, which shows the possibility of intervention and prevention of financial or privacy damage in a timely manner [1, 2].

The advantages of the suggested approach are mostly in its capability to process the conversation on an incremental basis and refresh the risk assessment on a turn-by-turn basis. The system can learn the semantic meaning and temporal relationships of a conversation by using pre-trained contextual embeddings and sequencing the model using BiLSTM networks, an element that is crucial in identifying the escalation patterns that characterize the typical social engineering scam. Also, it can be extended to include turn-level predictions in order to operate the system in streaming conditions, which is appropriate to practical implementation to telecommunication networks or call centers. High basis of discriminative operation and early detection of the model highlight its applicability in operational conditions where timely and accurate decisions are important [3, 4].

Along with these strengths, there are a number of limitations that can be identified. The fact that a synthetic dataset is required to allow controlled experimentation and reproducibility is not expected to represent the diversity of language, emotional subtleties, and counter-argumentative strategies present in real scam calls. As a result, performance of the models may be worse when used with natural dialogues with more elaborate patterns of speech, code- mixing, or domain-specific differences. Furthermore, whereas the BiLSTM using BERT has good performance, the computational complexity of transformer embeddings and sequential processing can be a challenge to implement in a resource-constrained setting or other contexts, where very low latency is required.

The existing mechanism is also based on preset probability threshold of alerts that might require dynamic adjustment in the heterogeneous operational environment to strike a balance between precision and recall [5, 6].

Extrapolating the given approach to the current work, it can be observed that it is distinctly better than traditional machine learning tools like TF-IDF with logistic regression and convolutional models. Although CNNs performed better, depicting local patterns, sequential and contextual representation of BiLSTM with BERT embeddings can better recognize multi-turn frauds. In contrast to some other systems previously, which expect to access the entire conversation or do it offline, the incremental, streaming-based architecture allows it to be detected early without having to wait until the end of the conversation, overcoming a major weakness many have noted in literature. Furthermore, the fact that temporal dependencies are explicitly modeled helps distinguish the given approach, provided compared to purely embedding-based or retrieval-augmented ones that can ignore the sequential development of dialogue cues essential in the detection of fishing attacks [7, 8].

Further studies need to be done on the model to validate the model using real-world data to evaluate the generalization and ability to work across diverse linguistic, cultural, and emotional contexts. Additional features like prosodic or voice emphasis or background noise, represented as multimodal data, can also increase detection accuracy and robustness to adversarial attacks. Also, studies on adaptive thresholding, predictions using uncertainty, and lightweight transformer models might be enhanced to enhance applicability in real time and decrease computation. Studying transfer learning in other languages, other areas of calls, and typologies of scams could broaden the scope of the model. Lastly, longitudinal research focusing on the process of implementing such detection systems within the live telecommunication infrastructure would be of great value on the levels of user trust, alerts handling and the actual effectiveness of early intervention mechanisms [9, 10].

In general, the research offers a strong piece of evidence that multi-turn dialogues can be incrementally, contextually, and sequentially modeled to make important contributions to the current state of research in the area of real-time scam detection which can both contribute to improvements in the methodology and be of practical use in telecommunication and financial security contexts [1–10].

Conclusion

This paper introduces a turn-level, real-time fraud detection system of telephone calls, which is based on pre-trained contextual embeddings and a BiLSTM network to ensure that the semantic content and temporal dynamics of multiple turn conversations can be reflected. The suggested framework has good performance, a test accuracy of 93.75% and F1 of 93.55, which is better than the baseline of traditional machine learning and convolutional. Notably, the system is capable of detecting most scams during the initial several turns of the conversation, which will allow intervening and minimizing possible losses (financial and privacy). Although the experimental dataset was synthetic, and thus the controlled experiment and reproducible evaluation were possible, more validation on actual call data should be performed before declaring the robustness in a variety of linguistic, cultural, and emotional contexts. Future studies must examine multimodal features, adaptive thresholding, and lightweight transformer models in order to improve scalability and generalization. Comprehensively, the current research indicates that context-sensitive and incremental dialogue modeling is a viable and powerful framework at detecting scam in real-time and has huge potential of application in telecommunication and financial security systems.

Information on funding. This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP27510301 “Development of technology for recognizing fraudulent actions during a telephone conversation and/or text message exchange in messengers based on artificial intelligence algorithms”).

REFERENCES

1 Sergeevna, P.E., Timurovich, G.S., and Gennadievich, B.P. The factor of complex interaction in responding to telephone fraud. Voprosy Bezopasnosti, 1, 1–9 (2023). https://doi.org/10.25136/2409-7543.2023.1.39274

2 Syafitri, W., Shukur, Z., Mokhtar, U.A., Sulaiman, R., and Ibrahim, M.A. Social engineering attacks prevention: A systematic literature review. IEEE Access, 10, 39325–39343 (2022). https://doi.org/10.1109/ACCESS.2022.3162594

3 Anuar, H.A., et al. Phone scam unveiled: Insights from a systematic literature review. Journal of Financial Crime (2025). https://doi.org/10.1108/JFC-03-2025-0078

4 Soudani, H., et al. A survey on recent advances in conversational data generation. ACM Computing Surveys (2024). https://doi.org/10.1145/3795686

5 Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., and Abraham, A. AI-based conversational agents: A scoping review from technologies to future directions. IEEE Access, 10, 92337–92356 (2022). https://doi.org/10.1109/ACCESS.2022.3201144

6 Singh, S., and Beniwal, H. A survey on near-human conversational agents. Journal of King Saud University – Computer and Information Sciences, 34 (10), 8852–8866 (2022). https://doi.org/10.1016/j.jksuci.2021.10.013

7 Rim, D., et al. To chat or task: A multi-turn dialogue generation framework for task-oriented dialogue systems. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Industry Track), 576–592 (2025). https://doi.org/10.18653/v1/2025.acl-industry.41

8 Pattnayak, P., et al. Hybrid AI for responsive multi-turn online conversations with novel dynamic routing and feedback adaptation. In: Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing, 215–229 (2025). https://doi.org/10.18653/v1/2025.knowledgenlp-1.20

9 Li, X., et al. Proactive guidance of multi-turn conversation in industrial search. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Industry Track), 706–717 (2025). https://doi.org/10.18653/v1/2025.acl-industry.50

10 Pandey, A. Retrieval augmented fraud detection. In: Proceedings of the 5th ACM International Conference on AI in Finance, 328–335 (2024). https://doi.org/10.1145/3677052.3698692

11 Perera, L., et al. AE-RAGX: Combining autoencoders with retrieval-augmented generation for explainable anomaly detection using LLMs. In: 2025 IEEE Latin-American Conference on Communications (LATINCOM), 1–6 (2025). https://doi.org/10.1109/LATINCOM67778.2025.11345384

12 Xu, J. Enhancing financial risk management with retrieval-augmented large language models. In: Proceedings of the 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID), 138–141 (2025). https://doi.org/10.1109/ICAID65275.2025.11034536

13 Li, Z. Knowledge-grounded detection of cryptocurrency scams with retrieval-augmented LMs. In: Proceedings of the 3rd Workshop on Towards Knowledgeable Foundation Models (KnowFM), 40–48 (2025). https://doi.org/10.18653/v1/2025.knowllm-1.4

14 Chang, Y.C. Scam detection with large language models: Multimodal risk analysis of URLs and chat messages (2025). https://doi.org/10.71781/239

15 Akbar, K.A., et al. Retrieval augmented generation-based large language models for bridging transportation cybersecurity legal knowledge gaps. Transportation Research Record (2025). https://doi.org/10.1177/03611981251372471

16 Malhotra, S., Arora, G., and Bathla, R. Detection and analysis of fraud phone calls using artificial intelligence. In: Proceedings of the International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON), 592–595 (2023). https://doi.org/10.1109/REEDCON57544.2023.10150631

17 Cazzolato, M., et al. CallMine: Fraud detection and visualization of million-scale call graphs. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 4509–4515 (2023). https://doi.org/10.1145/3583780.3614662

18 Wahid, A., et al. NFA: A neural factorization autoencoder based online telephony fraud detection. Digital Communications and Networks, 10 (1), 158–167 (2024). https://doi.org/10.1016/j.dcan.2023.03.002

19 Ren, L., et al. Dynamic graph neural network-based fraud detectors against collaborative fraudsters. Knowledge-Based Systems, 278, 110888 (2023). https://doi.org/10.1016/j.knosys.2023.110888

20 Wu, Y., et al. Fraud-agents detection in online microfinance: A large-scale empirical study. IEEE Transactions on Dependable and Secure Computing, 20 (2), 1169–1185 (2022). https://doi.org/10.1109/TDSC.2022.3151132

[1]*Серек А.,
PhD, қауымдастырылған профессор,
ORCID ID: 0000-0001-7096-6765,
*e-mail: Azamat.Serek@astanait.edu.kz
[2]Шойынбек А.,
PhD, профессор,
ORCID ID: 0000-0002-9328-8300,
e-mail: aisultan.shoiynbek@narxoz.kz
[2]Қуанышбай Д.,
PhD, ассистент-профессор,
ORCID ID: 0000-0001-5952-8609,
e-mail: darkhan.kuanyshbay@narxoz.kz

[1]Astana IT университеті, Астана қ., Қазақстан
[2]Narxoz университеті, Алматы қ., Қазақстан

КӨПҚАДАМДЫ ДИАЛОГТЫ ТАЛДАУ НЕГІЗІНДЕ ТЕЛЕФОН АРҚЫЛЫ ЖАСАЛАТЫН АЛАЯҚТЫҚ ҚОҢЫРАУЛАРДЫ НАҚТЫ УАҚЫТ РЕЖИМІНДЕ АНЫҚТАУ

Аңдатпа

Телекоммуникациялық қызметтердің кеңеюі жеке тұлғалар мен ұйымдарға елеулі қауіп төндіретін алаяқтық телефон қоңыраулары санының артуымен қатар жүреді. Анықтаудың дәстүрлі әдістері әдетте толық әңгімелерді офлайн талдауға негізделеді, бұл олардың жедел әрекет ету мүмкіндігін шектейді. Бұл мақалада автор көп айналымды диалогтардың семантикалық мазмұны мен уақыттық динамикасын модельдеу үшін алдын ала үйретілген контекстік ендірмелермен біріктірілген екі бағытты ұзақ қысқа мерзімді жад желісіне (BiLSTM) негізделген нақты уақыттағы, кезең-кезеңімен жүзеге асатын алаяқтықты анықтау жүйесін ұсынады. Алаяқтық қоңырауларды анықтау үшін жүйе әрбір диалог кезеңінен кейін қоңыраудың алаяқтық болу ықтималдығын біртіндеп жаңартып отырады, бұл алаяқтықты ерте анықтауға мүмкіндік береді. Синтетикалық көп айналымды диалогтық деректер жиынтығында жүргізілген сынақ нәтижелері BERT ендірмелерін пайдаланатын ұсынылған BiLSTM моделінің дәлдігі 93,75% және F1 көрсеткіші 93,55% екенін көрсетті, бұл қазіргі машиналық оқыту мен конволюциялық базалық модельдердің нәтижелерінен жоғары. Жүйе алаяқтық қоңыраулардың көпшілігін әңгіменің алғашқы бірнеше кезеңінде анықтай алады, бұл тәуекелді жедел бағалауға мүмкіндік береді. Алынған нәтижелер нақты уақыт режимінде алаяқтықты анықтау үшін контекстке негізделген прогрессивті модельдеудің тиімділігін және оның практикалық тұрғыдан қолдануға жарамдылығын көрсетеді.

Тірек сөздер: нақты уақыт режиміндегі алаяқтықты анықтау, телефон арқылы алаяқтық жасау, көп айналымды диалог, BiLSTM, контекстік енгізулер, тізбекті модельдеу, ерте араласу, телекоммуникациялық қауіпсіздік, кезекті болжау, ағындық талдау

[1]*Серек А.,
PhD, ассоциированный профессор,
ORCID ID: 0000-0001-7096-6765,
*e-mail: Azamat.Serek@astanait.edu.kz
[2]Шойынбек А.,
PhD, профессор,
ORCID ID: 0000-0002-9328-8300,
e-mail: aisultan.shoiynbek@narxoz.kz
[2]Куанышбай Д.,
PhD, ассистент-профессор,
ORCID ID: 0000-0001-5952-8609,
e-mail: darkhan.kuanyshbay@narxoz.kz

[1]Astana IT университет, г. Астана, Казахстан
[2]Narxoz университет, г. Алматы, Казахстан

ОБНАРУЖЕНИЕ МОШЕННИЧЕСКИХ ТЕЛЕФОННЫХ ЗВОНКОВ В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ МНОГОХОДОВОГО АНАЛИЗА ДИАЛОГА

Аннотация

Расширение телекоммуникационных услуг сопровождалось ростом числа мошеннических телефонных звонков, представляющих серьезную угрозу для отдельных лиц и организаций. Традиционные методы обнаружения обычно основаны на офлайн-анализе полных разговоров, что ограничивает оперативность реагирования. В данной работе автор предлагает систему обнаружения мошенничества в реальном времени, основанную на предварительно обученных контекстных встраиваниях в сочетании с двунаправленной сетью долговременной кратковременной памяти (LSTM) для моделирования семантического содержания и временной динамики многоходовых разговоров. Для обнаружения мошеннических звонков система постепенно изменяет вероятность мошенничества после каждого хода разговора, что позволяет обнаружить мошенничество. При тестировании на синтетическом наборе данных многоходовых диалогов показано, что предложенная BiLSTM с использованием встраиваний BERT имеет точность 93,75% и показатель F1 93,55, что выше, чем у существующих базовых моделей машинного обучения и сверточных нейронных сетей. Система способна выявлять большинство мошеннических схем на начальных этапах звонка, что обеспечивает быструю оценку риска. Эти результаты свидетельствуют о полезности контекстно ориентированного, последовательного моделирования для обнаружения мошенничества в режиме реального времени и о возможности его практического применения.

Ключевые слова: обнаружение мошенничества в реальном времени, телефонные аферы, многоходовый диалог, BiLSTM, контекстные вложения, последовательное моделирование, раннее вмешательство, телекоммуникационная безопасность, прогнозирование уровня хода диалога, потоковый анализ.