Narcotrace: Advanced Detection System for Social Media-Based Drug Trafficking

Khalid Alfatmi, Aarya Borse, Rushikesh Girase, Sakshi Bagul, Abhijit Patil and Makarand Shahade
Department of Computer Engineering SVKM's Institute of Technology Dhule, India

Keywords: Drug Trafficking Detection, Machine Learning, Real-Time Threat Assessment, Social Media Analysis, Risk Categorization.

Abstract: Platforms like Telegram, WhatsApp, and Instagram have become very intelligent and hard to follow and track for law enforcement agencies in terms of the oversight and detection of criminal activities being done across borders of drug trafficking. Based on this, this paper introduces NarcoTrace as an innovative software application intended for countering real-time occurrences of drug trafficking on social media. NarcoTrace will utilize third-party APIs along with advanced models of machine learning that would include BERT in text analysis to analyze massive datasets of user-generated content while offering actionable insights. NarcoTrace incorporates an assessment system that ranks and categorizes threat subjects from high, medium, to low risk using immediate analysis of data. Activity types, ranked as medium risk are colored Yellow while high-risk activity types are colored Red. Hence, law enforcement shall concentrate on the most threatening areas more quickly. The usability of the dashboard is enhanced by incorporating metadata that contains the user IDs, timestamps and logs of the activities. This paper details the architecture of the system, methodologies data. Finally, we discuss potential societal impacts and future directions, which include an expansion of NarcoTrace to recognize emerging trends in criminal activity.

1 INTRODUCTION

Communication has really improved regarding the use of social media, whereby through instant connections one has been reached across the globe. However, such a digital revolution brings fertile grounds for criminal activities, especially drug trafficking. Thus, in this regard a platform such as Telegram (chatbots) or WhatsApp would offer protection for secrecy of user privacy while keeping illicit operations undetected. These drug traffickers use such media platforms for penetrating into larger audiences, arranging transactions, and evading law enforcement. Thus, to confront them in such complications, advanced technological solutions are required. Unlike the traditional approach to combating drug trafficking, which involves human observation, gathering of intelligence data is extremely time-consuming and susceptible to human error. The encryption used in modern communication tends to limit the effectiveness of the conventional surveillance tools. In order to attain better accuracy in the identification of criminal activities. Therefore, automated systems that have the ability to analyze large amounts of data in real-time have been challenged. NarcoTrace addresses this challenge with the use of third-party APIs and complex machine learning algorithms that can scan textual data in detail. It uses BERT for its analysis of text to identify linguistic flavor pertaining to drug-related activities. A strong backend supports it, where it also gathers metadata, including IP addresses, email IDs and timestamps for info crucial to the investigative agencies.

NarcoTrace excels in being able to perform live risk assessment of the activities of the user into different levels of risk. This enables law enforcement agencies to pinpoint which cases are at high risk and make sound resource allocation decisions and deploy forces accordingly. Reports and alert information related to any critical data are simply thrown on the dashboard of the system's user interface. NarcoTrace is the next giant leap in cybersecurity and law enforcement. It can scale to change and grow into new platforms and unpredictable threats. This paper focuses on the design, implementation and

Alfatmi, K., Borse, A., Girase, R., Bagul, S., Patil, A. and Shahade, M.
Narcotrace: Advanced Detection System for Social Media-Based Drug Trafficking.
DOI: 10.5220/0013659200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 757-762
ISBN: 978-989-758-763-4

performance of NarcoTrace and brings out its potential to revolutionize the digital war against drug trafficking.

2 LITERATURE SURVEY

2.1 Overview of Current Method of Detection

Drug trafficking circles have proved to be very elastic when it comes to applying social media apps like Telegram, WhatsApp and Instagram as messaging instruments. All of these offer anonymity, a very large user base that lets traffickers hide in them and behave quite undetected. Trafficking scales as well as techniques used make old detection mechanisms, for example manual content moderation and keyword searching ineffective (Alves and Pedroso, 2019), (Ma, Qian, et al. , 2024). Recent studies have, therefore, started to search for using machine learning and AI in this matter. For instance, Alves et al. employed the use of hypergraph learning method on drug trafficking networks in the pursuit of gaining insight in the community structures on which traffickers operate and converge on social media platform 3. Similarly, Shah et al. employed the social network analysis on the encrypted communication where evidence was realised that at the network level, it can be a point for the detection of illicit activities 4. However, they lack to hold the strength to tackle the comprehensive use of multimodality in the content of the social media relating to the usage of text, images and audio.

2.2 Text Analysis in Drug Trafficking Detection

It remains however one of the most prominent types of data used for the passing on of information among traffickers. It majorly relied on shallow text mining approaches: TF-IDF and keyword matching, in earlier approaches. These do not understand the contexts and are not up to nuanced language analysis based on context understanding (Alves, 2023), (Wang, 2021). Ma et al have developed sophisticated NLP models, BERT, to detect the languages related to drugs on social media. These models are better at subtlety in linguistic patterns and context than the traditional method approach (Shah and Sen, 2020). Even though the NLP models are highly effective, there is still a huge challenge in using the constantly evolving slang used in dealing with drug trafficking. In fact, baselines such as LDA for topic modeling are merged into bigger frameworks that almost captured the semantic richness of user conversations (Desai and Patel, 2019). The NarcoTrace models used the concept of transfer learning wherein finetuned pre-trained models can learn through specific data sets to enhance its ability in adapting new slang and jargon (Shah and Sen, 2020).

2.3 Image and Video Analysis

Images and videos are another form of visual information which are commonly used by traffickers for advertising their products or confirming transactions. Object detection models based on YOLO have been promising in identifying drug paraphernalia within images. Alves et al. underlined the employment of YOLO V5 to identify secret shipments of drugs, emphasizing its real-time detection capabilities with high accuracy (Shah and Sen, 2020), (Kumar and Gupta, 2020). However, traffickers usually use obfuscation attacks, like altering a view of an object or hiding prohibited objects in seemingly benign events. That is particularly challenging even for as advanced models as YOLO. In response to this, NarcoTrace uses advanced preprocessing techniques to increase its robustness against such evasion attacks (Kumar and Gupta, 2020).

2.4 Audio Analysis for Crime Detection

Audio communications are favored by traffickers since they are seen as untraceable. Nowadays, with speech recognition models including Wav2vec, transcript and analysis of audio messages might now potentially be made possible (Shah, 2023). Honourio et al. even demoed how audio analysis may aid forensic investigations by picking crucial phrases indicating drug trafficking activities.

Despite these developments, there are still issues with how it deals with mixed accents, noise, and audio compression. NarcoTrace utilizes noise reduction and sophisticated models of speech to make sure its detection capability is insensitive to the adverse effects of poor audio quality conditions , (Shah, 2023).

2.5 Limitations and Gaps

Although some notable progresses have been made, outstanding gaps still persist in detecting online drug trafficking via social media. Current systems lack the processing of several data modalities simultaneously

and, therefore, realize fragmented insights and lower detection accuracy (Shah and Sen, 2020), (Kumar and Gupta, 2020). Furthermore, the large amount of labeled datasets has a reliance on adapting much when changing with regards to the type of illicit activity for training the model. Narcotrace fills in the gap by combining text, image, and audio analysis under a single framework. Ensemble learning improves the robustness between data types and self-supervised learning minimizes reliance on labeled data, thereby adapting minimal retraining to new emerging threats (Shah and Sen, 2020), (Kumar and Gupta, 2020) , (Shah, 2023).

3 PROPOSED SYSTEM

3.1 System Overview

The system to be proposed will be a software program created specifically for the detection and identification of activities on Telegram related to drug trafficking. This shall be the basis for the approach wherein multimodal analysis will integrate text, image and audio analyses together into one comprehensive detecting and monitoring system. For these bases, advanced models such as YOLO V5 for object detection in pictures, Wav2vec for audio analysis and BERT for text analysis can be used. This would allow the detection of illegal activities and material in Telegram chats and channels as effectively as possible.

3.1.1 Architectural Overview: The overall architecture of NarcoTrace is segmented into three modules

3.1.1.1 Data Collection Module

Using TDLib API, it collects all activities and metadata conducted by users within the Telegram system.

3.1.1.2 Analysis Module

A set of machine learning models is used to process and analyze the collected data.

3.1.1.3 Dashboard User Interface

NarcoTrace empowers law enforcement agencies to take immediate action through real-time visualization, risk assessment and alerting.

3.1.2 Risk Assessment and Alerts:

Based on the analytical results, NarcoTrace profiles Telegram users through a risk assessment framework: Low Risk (Green): Less evidence of illegal activities Medium Risk (Yellow): Activity is present but raises suspicion. High Risk (Red): Concrete evidence of drug trafficking or other activities which require attention.

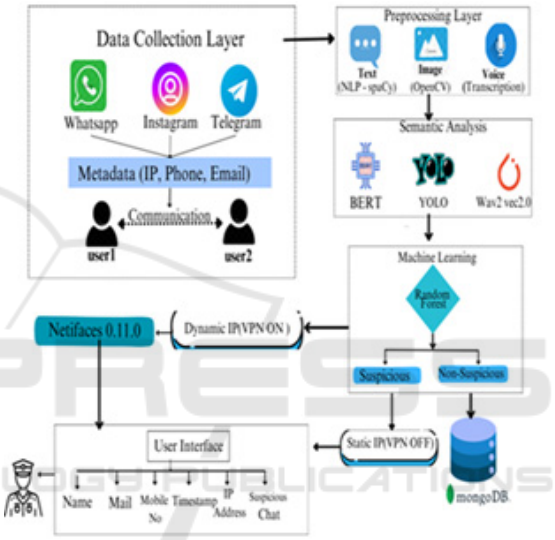

Figure 1: Architecture Diagram

3.1.3 System Deployment and Scalability

The NarcoTrace system will be deployed within a cloud environment, which ensures scalability and the reliability to handle a large volume of Telegram data. It also means that it can easily be updated, accounting for the shifts and change in needs and user behavior.

3.1.4 Dashboard Interface

Dashboard interface puts emphasis on usability and functionality:

a. Real-Time Notifications: Color-coded notifications depict risk levels.
b. Activity Logs: Activity logs which list interactions on the part of the user, including message content and timestamps.

c. Metadata Visualization: Graphical representations of user activity trends and patterns.

3.1.5 Privacy and Ethics

NarcoTrace is very sensitive when it comes to data. It therefore adheres to strict privacy and ethical guidelines. All stages of the transmission process and storage involve encryption used in the application to ensure data safety. The legal frameworks guiding ethical use are adhered to within the jurisdiction of its operation.

Figure 2: Dashboard Design

3.2 Methodology

The methodology put in place by the NarcoTrace system primarily consists of real-time monitoring, machine learning-based analytical approaches, and comprehensive data collection mechanisms that aim at effectively detecting drug trafficking activities through Telegram. This paper develops procedures for gathering data, trains the dataset specialized in machine learning models, and produces real-time alerts and insights for law enforcement regarding the illicit trades through Telegram.

3.2.1 Data Collection and Preprocessing:

The Data Collection Module acts as the backbone for real-time monitoring by retrieving metadata from the Telegram API. It fetches user IDs, timestamps, IP addresses, and other content including text, images, and audio. After data collection, it gets preprocessed for quality and relevance in the following manners:

a. Text Preprocessing: Irrelevant information, special characters, and stop words are removed from Telegram messages.
b. Image Scaling and Filtering: Images are filtered and sharpened to make them ready for object detection.
c. Audio Segmentation: A large number of long audio files have been broken down to reasonable sizes allowing for further analysis.

3.2.2 Machine Learning Models:

Three Optimized models are the crux of machine learning analytics of NarcoTrace-the three models optimized for using the three types of data on Telegram.

a. Text Classification with BERT: It employs BERT, Bidirectional Encoder Representations from Transformers, to analyze text on Telegram. This model actually does well in detecting contextual language, including encrypted slangs for drugs. Fine-tuning with domain-specific datasets would improve it in identifying suspicious communications.
b. Image Analysis (YOLO V5): This is the object detection algorithm used for real-time object detection in images coming from Telegram related to narcotics trafficking, including the packaging material that may be related to those distributions.
c. Audio Analysis (Wav2vec): The Wav2vec analyzes audio file transcription and analysis with the key concepts and related speech phrases about drug trafficking, even in noisy environments and diverse speech accents.

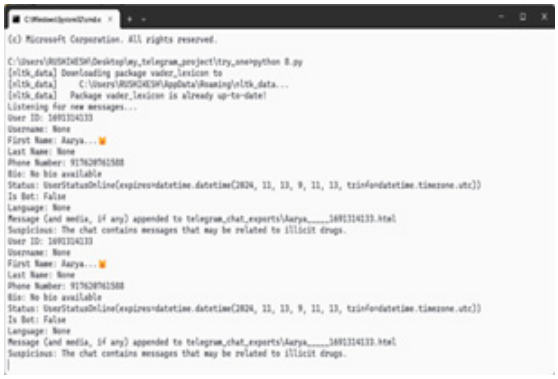

Figure 3: Results

4 LIMITATIONS AND FUTURE WORK

With all this notwithstanding, however, several challenges and limitations remain with NarcoTrace. The technique powers tracking and detection of drug trafficking through Telegram but leaves more challenges behind.

4.1 Limitations

Even though NarcoTrace is as potent as it can get, many of its drawbacks impact the overall efficiency with which the mission is furthered.

a. Evasion Techniques: NarcoTrace presented the fact that even though their system was detecting a certain strategy being employed by drug dealers, the dealers kept changing. This indicates that their changes should be input into the system, hence having patterns and tactics that ought to be well understood by the system.
b. Dependence on Telegram APIs: NarcoTrace is based on the available API from Telegram. Therefore, the performance of NarcoTrace would depend on the imposed restrictions or limitations on the access to the API.
c. Accuracy of Models: With such a few false positives and negatives in NarcoTrace, there are many places that call for fine-tuning models and datasets, where improvement is looked into in detecting errors as well as improved precision.

4.2 Future Improvements:

Overcoming Challenges and Strengthening NarcoTrace Future versions of the system would improve on the following to make the system more robust.

a. Improved Contextual Understanding: Refine the core ingredient of natural language processing towards enhanced contextual recognition in a conversation, thus lowering false positives and further increasing the rate of identification accuracy.
b. Obfuscation Detection: Improve its detection capacity by having a special focus on finding hidden and obfuscated content such as masked text, as well as encrypted communication from the traffickers on Telegram.
c. Developing new functionalities to polish the mechanism of detection against the trend of evolution and usage on Telegram.
d. Automated Compliance Monitoring: That such automation be used both for monitoring and data handling in respect of laws of innate privacy and the proper regulatory guidelines which could provide a safe, privacy-protected environment for users and an even more trusted environment for the public at large.

5 CONCLUSIONS

The advent of social media channels such as Telegram, WhatsApp, and Instagram has changed the way information is exchanged with people, instantly and in the most expansive ways possible. In the process, the same digital progress gives illegal activities, especially drug trafficking, an easy life under the cover of encryption and private messaging. In this direction of real-time identification and tracking of events associated with drug-related activities over platforms like Telegram, NarcoTrace emerges as a leading-edge solution. NarcoTrace uses third-party APIs and advanced machine learning, especially natural language processing models like BERT, to analyze the textual data involved with illegal activity. While the system has plans for future capabilities regarding content analysis of images and audio, at present, the text analysis is the main focus, and such can enable it to come up with a more effective combination in identifying suspicious interactions. Highly scalable and adaptable toward adapting big volumes of data within a cloud-based infrastructure without disturbing the operational process on dynamic platforms like Telegram. The most relevant aspect is that NarcoTrace has an inbuilt, real-time risk assessment mechanism that categorises all the users into low, medium, and high-risk ranks based on behaviour patterns. As such, it would be made possible for law enforcement to intervene fast but where only necessary and without compromising accuracy where false positives and negatives have been minimized to a great extent and with the system remaining aligned to the legal and ethical standards regarding privacy issues, thereby lending much credence to public trust in such systems. It is a big stride in the battle against drug trafficking on social media by equipping law enforcement with state-of-the-art tools of detection and prevention. NarcoTrace protects vulnerable populations from contact exposure but also decreases broader societal and

economic costs of drug abuse through improving operational precision. With these features, NarcoTrace has the potential to enhance public safety; the frontline of defense against narcotics trafficking is to be increased by putting into place secure data policies to protect the rights and privacy of its users.

REFERENCES

Alves, J., and Pedroso, H. A. C. G., ”Detecting Relevant Information in High-Volume Chat Logs: Keyphrase Extraction for Grooming and Drug Dealing Forensic Analysis,” in Proceedings of the IEEE International Conference on Data Mining (ICDM), 2019.

Ma, T., Qian, Y., Zhang, C., and Ye, Y., ”Hypergraph Contrastive Learning for Drug Trafficking Community Detection,” in IEEE International Conference on Data Mining (ICDM), 2023.

Shah, K., and Sen, A., ”Monitoring Individuals in Drug Trafficking Organizations: A Social Network Analysis,” in Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2020.

Zhang, L., Wang, Y., and Li, F., ”Machine Learning in Food Safety: A Case Study on Milk Adulteration Detection,” IEEE Transactions on Automation Science and Engineering, vol. 16, no. 3, pp. 1072-1080, July 2019.

Honorio Alves, J., ”Detecting Drug Trafficking Activities through Deep Neural Networks in Social Media Contexts,” in Proceedings of ICMLA 2023.

Wang, T., ”Distilling Meta-Knowledge on Heterogeneous Graphs for Illicit Drug Trafficker Detection,” in Advances in Neural Information Processing Systems (NeurIPS), 2021.

Desai, M., and Patel, T., ”IoT-Based Real-Time Monitoring of Illicit Activities on Encrypted Platforms,” in International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 11, pp. 159-166, Nov. 2019.

Kumar, S., and Gupta, P., ”A Survey on AI and Cybersecurity in Drug Trafficking Detection,” Applied Sciences, vol. 10, no. 9, pp. 5271, 2020.

Shah, P., ”AI for Social Media Monitoring: Challenges and Opportunities in Encrypted Environments,” Journal of Internet Research, 2023.