ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume 28, Number 4, 2025, 315–326
doi: 10.59277/ROMJIST.2025.4.01

112 Emergency Video Call Response Pipeline for Car Crashes Using Computer Vision and Natural Language Processing

Béatrix-May BALABAN[1,*], Ioan SACALĂ[1], and Alina-Claudia PETRESCU-NIŢĂ[2]

[1]Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, Splaiul Independenţei 313, 060042 Bucharest, Romania
[2]Faculty of Applied Sciences, University POLITEHNICA of Bucharest, Splaiul Independenţei 313, 060042 Bucharest, Romania

Email: beatrix_may.balaban@yahoo.com*, ioan.sacala@upb.ro, alina.petrescu@upb.ro

* Corresponding author

Abstract. This paper presents a comprehensive 112 emergency response pipeline that integrates multimodal technologies, such as automatic speech recognition, object detection, natural language processing, and blockchain, to enhance how critical incident information is gathered, analyzed, and acted upon. From real-world video inputs, speech content is transcribed using a speech-to-text algorithm, while computer vision models identify individuals involved. The transcribed dialogue is semantically processed to extract structured question-answer pairs, which are then evaluated to assess the medical urgency of the case. To enrich contextual awareness, the system cross-references vehicle ownership claims made in the transcript with a blockchain-based ledger of profiles linked by Vehicle Identification Number (VIN) identifiers. This verification step enables the retrieval of car owner name, age and critical medical background, such as insurance coverage, chronic conditions, and ongoing treatments. A lightweight web interface presents this information in an accessible format for first responders. The result is an intelligent, end-to-end system that prioritizes car accident emergency cases efficiently and empowers intervention teams with personalized insight, aiming to reduce response time and improve patient outcomes by classifying each emergency accordingly.

Key-words: Blockchain; computational linguistics; computer vision; emergency services; natural language processing; object detection; semantic processing.

1. Introduction

Emergency response systems are crucial for public safety, enabling quick detection and coordinated action during crises. As emergencies grow more complex, real-time collaboration becomes essential. New platforms now focus on connecting responders, dispatchers, and healthcare providers to improve communication and overall response efficiency. However, a significant gap exists in how information is collected and interpreted during emergencies. Most systems rely heavily on voice communication and sensor data, with limited integration of rich contextual inputs such as video footage.

The absence of video input in real-world emergency scenarios presents a critical limitation. While many modern vehicles and public spaces have cameras, the infrastructure to extract, analyze, and act on that visual data during live incidents is mainly undeveloped. As a result, operators often make high-stakes decisions with incomplete situational awareness, relying on verbal reports that may be ambiguous or delayed. This lack of visual context can hinder response efficiency, reduce triage accuracy, and ultimately affect victim outcomes. Emergency systems need to evolve toward more immersive, context-aware platforms that incorporate video data and intelligent interpretation tools. By embedding video analysis within collaborative decision-making frameworks, response teams can gain a more accurate and dynamic understanding of the scene, supporting faster, more informed interventions and minimizing the risk of miscommunication under pressure. The gap between modern video analytics' capabilities and their practical use in emergency collaboration platforms highlights a key limitation in current research. There is growing recognition that real-time video input, when combined with natural language processing and rule-based logic, could significantly improve how emergencies are assessed and managed. Yet, robust, scalable frameworks for fusing these modalities within active response systems are still lacking.

The remainder structure of the paper is organized as follows: Section 2 reviews existing literature to establish the foundation for this study. Section 3 outlines the proposed system, while Section 4 details the implementation. Section 5 explains the NLP pipeline developed for the system, followed by Section 6, which presents and discusses the results. Section 7 concludes the paper with a summary of key insights.

2. Literature Review

In recent years, the integration of collaborative platforms into emergency incident services has garnered significant attention within the academic community. The following section provides an overview of five pertinent academic studies published between 2023 and 2025, highlighting their contributions to this evolving field. Authors in [1] explore how Information and Communication Technology can improve the dispatch and coordination of volunteer first responders in emergencies. Based on five years of case study research, the authors propose key features like geofencing, dynamic resource management, and strong communication tools to make volunteer efforts more effective. Article [2] discusses how AI and cloud technologies can improve disaster management by enabling faster, more informed responses. The authors in [3] explore how blockchain can improve data sharing and collaboration in emergency management. Using Hyperledger Fabric, the authors develop a system that tackles issues like data silos and trust in centralized setups. Tested in epidemic prevention at customs, the study shows that blockchain can make multi-agency collaboration more secure, reliable, and efficient. Article [4] presents the design of CoSpace, a system built on service-oriented architecture to improve collaboration in emergency management. By analyzing user behavior, the authors created reusable web services that help responders communicate and coordinate more efficiently in mobile settings. The research in [5] presents a prototype GIS tool for Virtual Emergency Operation Centres, designed to help emergency teams collaborate and make decisions in real time through features like shared maps and chat. Together, these studies show how digital tools and collaboration platforms are essential for improving modern emergency response, from volunteer coordination to interagency cooperation using advanced technologies. Contemporary vehicles now often feature sophisticated safety technologies aimed at reducing the impact of traffic collisions. One notable advancement is the eCall system, a mandated telematics solution in the European Union [6] [7]. This system guarantees automatic contact with emergency services through the 112 emergency number in case of a serious accident. This system is meticulously engineered to reduce response times by offering real-time crash data to first responders, thereby enhancing survival rates and post-accident medical interventions. The eCall system activates automatically after a crash or manually through an emergency button. It uses sensors like accelerometers, gyroscopes, and airbag triggers to detect impact severity. If thresholds are exceeded, the system immediately contacts emergency services via 112 using its own cellular connection. It also sends critical data, location and occupant count, to speed up response.

3. Proposed System

There are multiple emergencies and incidents that occurs at the same time and being able to actually view the situation can drastically improve the performance of the emergency services and the response of the emergency units because they can see and evaluate witch case in extremely urgent and witch can receive a lower level of severity thus being prioritized the urgent one. The goal is to be able to provide support for citizens and to save as many lives as possible. This paper aims to propose a system to better assess the situation and provide much more information on the scene, the possibility to actually view the emergency and better understand the proportion of the case.

The proposed system is designed to automate the assessment of vehicle accidents through real-time video analysis triggered by a 112 emergency video call. In Fig. 1, an overview of the entire system flow is illustrated. When a car accident is detected, the system immediately initiates a video call and begins an intelligent analysis of the situation. The first step is to determine whether the driver is visible on camera. If the driver is visible and appears to be conscious, the emergency operator engages in a conversation with them. This dialogue is then processed using natural language processing (NLP) techniques to extract key information about the driver's condition and the nature of the accident. Based on this analysis, the system proceeds to a dedicated module that evaluates the severity of the incident and determines the appropriate prioritization level for emergency response. If the driver is visible but appears unconscious or unresponsive, the system bypasses the dialogue stage and immediately proceeds to the severity assessment module, recognizing the situation as potentially critical. In cases where no driver is visible, the system evaluates the vehicle itself. If visible damage to the car is detected, the system again proceeds to assess severity and prioritize response. If the car appears undamaged, the event may be treated as a false alarm. However, the system still logs the event and notifies the registered vehicle owner via SMS, leveraging information stored on a blockchain. The blockchain plays a key role in this system by securely storing and linking vehicle and driver data for all users of

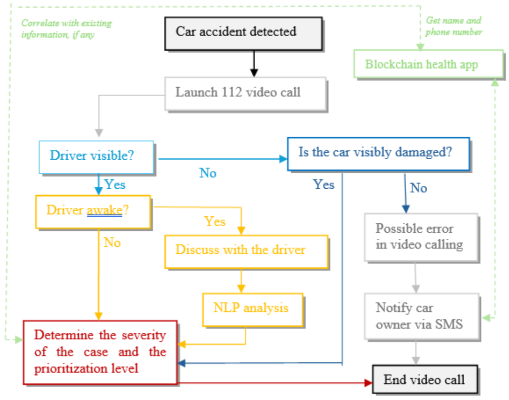


Fig. 1. Description of the proposed system for video analysis in 112 emergency services (color-coded integrated modules).

the 112 video call service. It is used both for ownership verification during emergency analysis and to enable notifications in edge cases where direct human presence cannot be confirmed. By integrating real-time computer vision, NLP, and blockchain, the proposed system offers a robust, automated framework for rapid and accurate emergency response in the aftermath of a car accident. The following subsections describe the technologies used for each module and the logic behind choosing these technologies. Regarding the system proposed in this paper, the first goal is to reliably detect and count people inside vehicles from in-car video footage. Given the variability in lighting, perspective, and movement within such confined environments, a range of state-of-the-art object detection algorithms was evaluated to determine which would offer the best balance of speed, accuracy, and practical integration. For real-time person detection in tight spaces like car interiors, several algorithms were tested. YOLOv5 [8] was a good starting point due to its speed and simplicity, but it struggles in complex cases. More advanced models like EfficientDet, Faster R-CNN, and SSD MobileNet [9] bring their own strengths, yet often fail to balance speed and reliability in demanding settings. YOLOv8 [10], offering better accuracy and smoother integration, has become the preferred option for deployment. Speech-to-text systems have rapidly evolved thanks to deep learning and self-supervised techniques. Among the main tools explored are Vosk, Whisper, DeepSpeech, Wav2Vec 2.0, and SpeechRecognition. Vosk suits offline use, Whisper delivers top-tier multilingual accuracy, DeepSpeech allows customization, Wav2Vec 2.0 leads in modern learning approaches, and SpeechRecognition offers simplicity through cloud support. Choosing the right one depends on processing power, language needs, and latency constraints.

4. Implementation of the Proposed System

4.1. Video and speech dataset gathering

The dataset that was created for this paper represents a novel addition to the emergency domain as it consists of videos with voice taken inside the car, having multiple scenarios when the emergency 112 video call is launched. The dataset is called CarEM-Vid (Car Emergency Video

Dataset), and it has 49 videos, varying in length from 0.09 seconds to almost 1.49 minutes, with 1280x720 resolution. The details regarding all considered scenarios are presented as follows.

4.1.1. Driver visible and conscious

For the scenario where the driver is conscious, Fig. 2 illustrates videos from dataset and the following represent script examples followed by 112 emergency operator. English (20250403_184328_en.mp4):

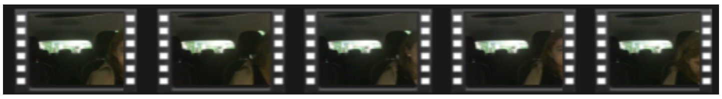


Fig. 2. Conscious driver - Example of videos from dataset.

"112 Operator: 'Hello! Can you hear me?', Driver: 'Yes.', 112 Operator: 'What is your name and age?', Driver: 'Susan, 71', 112 Operator: 'Do you know where you are?', Driver: 'Oak Street, downtown Springfield.', 112 Operator: 'Are there injuries?', Driver: 'Yes.', 112 Operator: 'What injuries?', Driver: 'My head hurts, my nose is bleeding, my hands are hurting, and I can't feel my legs.', 112 Operator: 'Where there any other persons with you in the car?', Driver: 'No.', 112 Operator: 'Do you smell gas?', Driver: 'Yes.', 112 Operator: 'Can you see fire, flames or smell smoke?', Driver: 'Yes.', 112 Operator: 'Is this your car?', Driver: 'Yes.', 112 Operator: 'Can you stay on the line until help arrives?', Driver: 'No.', 112 Operator: 'Are you on the road, on a field? What do you see?', Driver: 'A road… I don't know...'.".

Romanian (20250403_185635_ro.mp4): *"112 Operator: 'Alo? Mă auziţi? Sunt operatorul 112 de urgenţă.', Driver: 'Da, vă aud.', 112 Operator: 'Care este numele şi vârsta dumneavoastră?', Driver: 'Numele meu este Beatrix şi am 29 de ani.', 112 Operator: 'Ştiţi unde vă aflaţi?', Driver: 'Da, bulevardul Pantelimon numărul 33, Bucureşti.', 112 Operator: 'Sunteţi rănită?', Driver: 'Nu.', 112 Operator: 'Mai sunt şi alte persoane care vă însoţesc în maşină?', Driver: 'Nu.', 112 Operator: 'Simţiţi miros de benzină?', Driver: 'Nu.', 112 Operator: 'Vedeţi foc, flăcări sau simţiţi miros de fum?', Driver: 'Nu.', 112 Operator: 'Este maşina dumeavoastră?', Driver: 'Da.', 112 Operator: 'Puteţi rămâne cu mine pe linie până când ajung echipajele?', Driver: 'Da, pot rămâne.', 112 Operator: 'Sunteţi pe stradă sau pe spaţiul viran? Ce anume vedeţi?', Driver: 'Sunt pe stradă.' "*.

4.1.2. Driver visible in video but unconscious

Fig. 3 illustrates videos from dataset for this scenario, with the following examples of scripted text for the 112 emergency operator:

English: *"Hello! Ms. ? This is 112 emergency number operator. Can you hear me? Ms.? You were in a car accident. Can you hear me? Help is on the way. Don't worry! "*.

Romanian: *"Alo? Sunt operatorul serviciului de urgență 112. Mă auziţi? Doamnă! Mă auziţi? Aţi fost implicată într-un accident rutier. Echipajele sunt pe drum! Nu vă faceți griji! Mă auziţi?"*

4.1.3. No driver visible inside the car

In this case there are two types of scenarios, shown in Fig. 4, and 112 operator-voiced scripts:

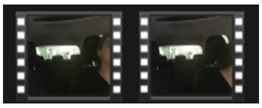


Fig. 3. Unconscious driver - Example of videos from dataset.

- **Car looks untouched:** The voice of the 112-emergency operator can be heard using English and Romanian language as both languages are part of this proposal. Examples of 112 operator-voiced scripts, related to this scenario:

  English: *"Hello! This is 112 emergency number operator. Can anyone hear me? Hello! Can anyone hear me? It seems no one is in the car. I will send a police unit to check just in case. The owner of the car will be notified of this call using agreed contact method."*

  Romanian: *"Alo? Sunt operatorul serviciului de urgență 112. Mă aude cineva? Alo? Mă aude cineva? Se pare că nu este nimeni în mașină. Voi trimite un echipaj de poliție să verifice pentru orice eventualitate. Proprietarul mașinii va fi anunțat despre acest apel folosind metoda de contact convenită."*

- **Car looks damaged, appearing just like after a crash:**

  English: *"Hello! This is 112 emergency number operator. Can anyone hear me? Hello! Can anyone hear me? If somebody hears me, don't worry, help is on the way."*

  Romanian: *"Alo? Sunt operatorul serviciului de urgență 112. Mă aude cineva? Alo? Mă aude cineva? Dacă sunt persoane care mă aud, nu vă faceţi griji, echipajele sunt pe drum."*

Fig. 4. Unconscious driver - Example of videos from dataset.

4.2. Training algorithms used in detecting the driver's body

This module processes in-car video footage to identify and count unique individuals with accuracy. It began using YOLOv5 for object detection, which performed well frame-by-frame but couldn't reliably distinguish the same person across multiple frames, often resulting in duplicate counts. To improve performance, the system transitioned to YOLOv8, which offers better speed, accuracy, and integration support, though it remains limited by its frame-based detection approach. To solve this, Deep SORT was added to provide real-time tracking, allowing the system to assign consistent identities to individuals over time. This ensures each person is counted only once, even if they appear in multiple frames. To avoid counting the same person multiple times, the system applies additional filtering: a person must appear in enough frames to be considered valid, and overlapping tracks are compared and merged using IoU to keep only the most reliable detection. The final results include one high-quality image per person.

The main goal was to detect, track, and accurately count people without duplication, and then to extract one image per unique individual in video. Regarding this goal, the accuracy reached with YOLOv5 was under 50%. On the other hand, with Yolov8 the accuracy reached is 83,67%, a much better result.

4.3. Speech-to-text module

The most effective way to test algorithms is to implement them individually and then compare the results. From the analysis conducted about the algorithms Wav2Vec2.0, Whisper, Vosk, SpeechRecognition and DeepSpeech [11], [12], SpeechRecognition was eliminated because, in terms of accuracy, it ranked last. DeepSpeech was also eliminated because it does not provide multilingual support. The three algorithms left are: Wav2Vec 2.0, Whisper and Vosk. Therefore, for each mentioned algorithm, a brief script was written using Python and provided with videos from the created dataset. A comparative analysis is needed to analyze the results obtained upon implementing the mentioned speech-to-text algorithms, for each of the above-mentioned scenarios. As the results obtained with the three implemented algorithms had the same behavior in all four scenarios, to better exemplify the obtained results, Table 1 illustrates an brief example of results obtained for Scenario 1, with red color marking the errors in transcribed text.

The resulting transcriptions from Vosk algorithm consist solely of words that do not construct a coherent sentence in any of the four scenarios, regardless of the language employed. An example of such Vosk transcription can be seen in the supplementary material [13]. Vosk was excluded for producing incoherent output, and Wav2Vec2.0 struggled with Romanian and punctuation. Whisper, on the other hand, handled both languages well and delivered accurate, punctuated transcriptions, making it the best fit for this system.

Separately, in the supplementary material [13], Table 2 contains full examples of speech-to-text result for each scenario using the three mentioned algorithms.

5. NLP Pipeline

- Step 1: Parse the .txt Transcript File

  In this stage, emergency call transcripts from a plain-text file are structured by linking each one to its corresponding video using the filename as an identifier. The result is a dictionary format that keeps each transcript tied to its video, allowing for clear, independent processing in later analysis.

- Step 2: Segment Questions and Answers

  After parsing the transcripts, semantic similarity is used to identify question-and-answer pairs, avoiding reliance on exact wording. By comparing segments of the transcript to known emergency questions using a SentenceTransformer model, questions can be matched even if they are phrased differently or contain transcription errors. This makes the system more reliable in capturing key information from natural, imperfect speech.

5.1. Determine the severity of the emergency

- Step 3: Classify Each Answer's Severity

  Once having extracted question-answer pairs from the transcript, the severity of each answer is assessed using rule-based logic. Each known question has a set of possible answer patterns mapped to severity levels such as **"Low", "Medium"** or **"High"**. These mappings allow the system to interpret the urgency of specific responses—for example, distinguishing between "I'm fine" and "I can't feel my legs." Any unrecognized or missing answer defaults to a high-severity rating for caution reasons.

- Step 4: Calculate Overall Urgency Score

  Once all relevant answers are classified, the system calculates a single, overall urgency score for the case based on rule-based logic: **"Critical", "Urgent"** or **"Non-Critical"**. The implemented system also checks for monologue-only cases, where no response is given. If the transcript contains familiar reassurance phrases and no interaction, the urgency can be overridden to "Non-Critical". Fig. 5 illustrates how the overall urgency score is computed for each case. The logic of the overall urgency score is designed to prioritize safety and quick response when the situation appears severe. If there are three or more "High" severity answers, the case is marked "Critical." A single "High" is enough for an "Urgent" severity, meaning there is a clear risk but perhaps not an immediate life-threatening emergency. The absence of any high-severity indicators results in a "Non-Critical" classification, implying that while a response may still be needed, the situation appears stable.

Table 1. Example of speech-to-text results - Scenario 1

Language	Wav2Vec2.0	Whisper
English	Hello can you hear me? Yes.what is your name and nage?i am jessica i am fifty years old. do you know where you are? It's parking.A parking in Sunset Boulevard.are there kinjuries?yes whatkinjuries?my hand hurts a little bitand Iam dizzy [...]	Hello, can you hear me? Yes. What is your name and age? I am Jessica. I am 50 years old. Do you know where you are? It's a parking. A parking in... Sunset Boulevard. Are there injuries? Yes. What injuries? My hand hurts a little bit. And I am dizzy [...]
Romanian	Allo mosit sontotorato u unoda deugento da vad carristenaminithipostade mnavastro berno chaptedesho stiton te vaflat da raa bustovali bocorist sontetranit da cheron cevodwari nupotzarespir ma doremonaston ofartatari son gere my soon chak the pastoni cariponsotiscon [...]	Alo, ma auziţi?sunt operatorul 1-1-2 de urgență. Da, va aud. Care este numele și vârsta dumneavostra? Berna. 78. Știți unde va aflați? Da. Stratda postăvarului. București. Sunteti rănit? Da. Ce răni? Ce vădoare? Nu pot să respir. Mă doare mâna stângă foarte tare. [...]

In addition to this question-based logic, the system includes a safety mechanism for handling monologue situations, transcripts where only the operator is speaking, and no one responds. If the transcript contains signs that no one is present (for example, "it seems no one is in the car," or "the owner of the car will be notified"), and those signs match a

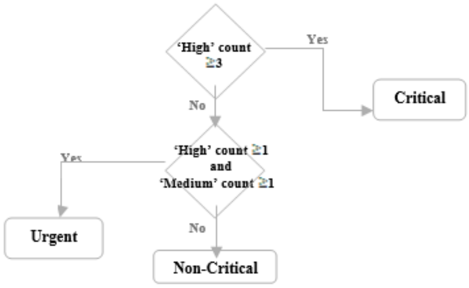


Fig. 5. Computing Overall Score of Urgency of a case.

predefined list of non-critical clues, the system overrides the question-based logic. It classifies the case as Non-Critical, even if the Q&A matching returned minimal or misleading results. This makes the algorithm more robust to real-world scenarios where not every call includes a two-way conversation.

5.2. Blockchain implementation

The blockchain links each vehicle's VIN to the owner's medical records, adding crucial context in emergencies when victims can't speak. Unlike transcript analysis, which shows urgency, the blockchain provides medical history, insurance info, and treatments tied to the vehicle, as shown in Fig. 6. If the transcript includes "yes" to "Is this your car?" and the name matches the VIN record, the system pulls the medical data and adds it to the output. This happens in real time, with results shown through a Flask web app, helping responders consider pre-existing risks for better, faster decisions.

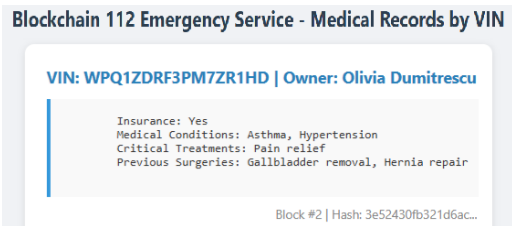


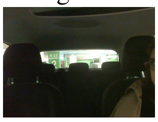
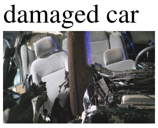
Fig. 6. Sample of data found in the implemented blockchain.

6. Results

The results obtained provide a comprehensive view of how the system performs across different accident scenarios. For each case, the input video was analyzed to determine whether a person was visible in the car, and if so, how many unique individuals were present. The detection module ensured that each person was counted only once, as tracking logic that maintained identity consistency across frames was integrated. Next, the English or Romanian audio, whether a dialogue or a monologue, was transcribed using a speech-to-text algorithm. The resulting text was processed through the NLP pipeline, which recognized key emergency-related questions and assessed the urgency level of each answer.

Results in Table 3 display for each scenario what the input video was, from where the detection module identified if a person was visible or not, identified unique people in each video, counting them only once. Then the English and Romanian dialogue or monologue from the video was transcribed using a speech-to-text algorithm, and then the transcription was fed to the NLP pipeline to analyze each scenario and each question and give each question an urgency level. At the end, the entire scenario is analyzed and an overall urgency level is given for each video. Thus, a classification of emergency car crashes was achieved in the aid of first responders. This layered

Table 3. Implemented proposed system with NLP Pipeline results.

Scenario	Image from video and Whisper transcription output	Overall Urgency score
Driver unconscious 	*Hello, Ms. This is 112 emergency number operator. Can you hear me? Ms. You were in a car accident. Can you hear me? Help is on the way. Don't worry.*	"Urgent"
Non-Critical dialogue 	*Hello, can you hear me? Yes. What is your name and age? My name is Hope and I'm 44. Do you know where you are? Yes, I'm on sunset street. Las Vegas. Are there any injuries? No, I'm fine. Were any other persons in the car? No. Do you smell gas? No. Can you see fire flames or smell smoke? No. Is this your car? Yes. Can you stay on the line with me until help arrives? Yes, I can. Are you on the road? On a field? What do you see? I'm on the road.*	"Non-Critical"
No driver, damaged car 	*Hello, this is 112 emergency number operator. Can anyone hear me? Hello, can anyone hear me? If somebody hears me, don't worry. Help is on the way.*	"Critical"
No driver, car untouched 	*Hello, this is 112 emergency number operator. Can anyone hear me? Hello, can anyone hear me? It seems no one is in the car. I will send a police unit to check just in case. The owner of the car will be notified of this call using agreed contact method.*	"Non-Critical"

analysis allowed the system to assign a final urgency score to each case, helping to classify the severity of the car accident. By doing so, the system supports emergency responders in quickly identifying which situations require immediate attention and which may be less critical.

Beyond accuracy, the results show the system's ability to handle diverse, multilingual input and extract real-time actionable insight. It goes beyond basic detection by interpreting behavior, speech, and visual cues to support life-saving decision-making. This means a faster, more informed response for first responders, a reduction in false alarms, and a deeper contextual understanding of each situation, even when the driver is unconscious or unable to communicate. By merging video analysis, speech transcription, and semantic interpretation, the system classifies emergencies with high accuracy and brings automation to a critical space where time and clarity are essential.

One of the most meaningful outcomes was the system's overall urgency classification accuracy of 83.67%. Achieving such accuracy in a real-world context, where data comes from noisy, spontaneous speech under stressful conditions, demonstrates the robustness of the combined vision and NLP pipeline.

7. Conclusions

The pipeline proposed in this paper represents a significant step forward in how emergency response systems can intelligently leverage technology. From the initial extraction of information using Whisper to transcribe speech from video environments, to object detection for identifying individuals involved, each component contributes to building a situationally aware and data-enriched emergency pipeline. Examples of system obtained results can be observed in Table 3. The proposed pipeline's emergency final classification has an accuracy of 83,6%.

The introduction of NLP for interpreting transcriptions enables structured understanding from unstructured dialogue, allowing us to assess the severity of each case automatically. This speeds up triage and reduces the cognitive load on emergency dispatchers who often work under intense pressure. The urgency classification system ensures that critical cases are never delayed due to uncertainty. Aiming to create a collaborative space, integrating a blockchain-based medical information system adds a layer of trust, security, and context often missing in real-world emergency calls. Verifying vehicle ownership and retrieving relevant medical records, such as existing conditions or critical treatments, can mean the difference between a generalized response and a life-saving, personalized intervention. One direction in future work is to achieve a better speech-to-text result for the Romanian language, and there are several papers that are identified for future research, namely [14], [15], and [16].

The results presented in this paper are important because they demonstrate how technologies often used independently, such as machine learning, computer vision, natural language processing, and blockchain, can be orchestrated into a unified, meaningful tool. In emergency services, where every second counts, this kind of intelligent automation has the potential to transform outcomes, protect responders, and ultimately save lives.

References

[1] S. PILEMALM, A. FOLLIN and E. PRYTZ, *Digitalized co-production of emergency response: ICT-enabled dispatch and coordination of volunteers at the emergency site*, Journal of Humanitarian Logistics and Supply Chain Management **15**(1), 2025, pp. 34–47.

[2] S. GUPTA, A. MODGIL, A. KUMAR, U. SIVARAJAH and Z. IRANI, *Artificial intelligence and cloud-based collaborative platforms for managing disaster, extreme weather and emergency operations*, International Journal of Production Economics **254** (1), 2022, pp 108642.

[3] Q. WANG and Y. LIU, *Enhancing collaborative emergency management: Leveraging blockchain distributed ledger for inter-agency data sharing*, Proceedings of 8th ACM SIGSPATIAL International Workshop on Security Response using GIS (EM-GIS '23), SIGSPATIAL '23: The 31st ACM International Conference on Advances in Geographic Information Systems, Hamburg, Germany, 2023, pp. 13–17.

[4] F. YOU, R.-P. ZHANG, P. LI and J.-M. WANG, *Service-oriented emergency management collaboration system design*, in Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management. DHM 2014, V. G. Duffy, Ed., Springer, Cham, Lecture Notes in Computer Science **8529**, pp. 649–660, 2014.

[5] Z. CHANG and S. LI, *Collaboration enabled GIS tools for emergency operation centre*, in Geomatics Solutions for Disaster Management, J. Li, S. Zlatanova and A. G. Fabbri, Eds., Springer-Verlag, Berlin, Heidelberg, New York, Lecture Notes in Geoinformation and Cartography, pp. 305–320, 2007.

[6] A. BONYAR, A. GECZY, O. KRAMMER, H. SANTHA, B. ILLES, J. KAMAN, Z. SZALAY, H. PETER and G. HARSANYE, *A review on current eCall systems for autonomous car accident detection*, Proceedings of 40^{th} International Spring Seminar on Electronics Technology, Sofia, Bulgaria, 2017, pp. 1–8.

[7] G. LELOW and F. MARINCIONI, *The European emergency number 112: Exploring the potential of crowd-sourced information for emergency management*, International Journal of Disaster Risk Reduction **120**, 2025, paper 105376.

[8] R. KHANAM and M. HUSSAIN, *What Is YOLOv5: A Deep Look into the Internal Features of the Popular Object Detector*. Accessed: April 2025. arXiv [Preprint]. Available: https://doi.org/10.48550/arXiv.2407.20892.

[9] H. FAYSAL, R. RAIHAN-AL-MASUD and M. RUBAIYAT HOSSAIN MONDAL, *Effectiveness of deep learning in real time object detection*, International Journal of Computer Applications **176**(41), 2020, pp. 55–60.

[10] M. SOHAN, T. SAI RAM and C.-V. RAMI REDDY, *A review on YOLOv8 and its advancements*, in Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems, I. J. Jacob, S. Piramuthu and P. Falkowski-Gilski, Eds., Springer, Cham, pp. 529–545, 2024.

[11] R. S. A. PRATAMA and A. AMRULLAH, *Analysis of Whisper automatic speech performance on low resource language*, Jurnal Pilar Nusa Mandiri **20**(1), 2024, pp. 1–8.

[12] A. MEHRISH, R. BHARDWAJ, S. PORIA and N. MAJUMDER, *A review of deep learning techniques for speech processing*. Accessed: April 2025. ResearchGate [Preprint]. Available: https://doi.org/10.48550/arXiv.2305.00359.

[13] B.-M. BALABAN, I. SACALĂ and A.-C. PETRESCU-NIŢĂ, Supplementary material of the paper Beatrix-May BALABAN, Ioan SACALĂ and Alina-Claudia PETRESCU-NIŢĂ, *112 Emergency Video Call Response Pipeline for Car Crashes Using Computer Vision and Natural Language Processing*, Romanian Journal of Information Science and Technology, 2025. Accesed: April 2025 [Online]. Available at: https://1drv.ms/b/s!Ajn0ZsYKsVKMkvFwQDqhmf-lpwuqBA?e=F0Zb8y.

[14] V. PAIS, R. ION, A.-M. AVRAM, M. MITROFAN and D. TUFIS, *In-depth evaluation of Romanian natural language processing pipelines*, Romanian Journal of Information Science and Technology **24**(4), 2021, pp. 384–401.

[15] M. NITU and M. DASCALU, *Natural language processing tools for Romanian – going beyond a low-resource language*, Interaction Design and Architecture(s) Journal- IxD&A **60**, 2024, pp. 7–26.

[16] G.-A. DIMA, A.-M. AVRAM, C.-G. CRĂCIUN and D.-C. CERCEL, *RoQLlama: A lightweight Romanian adapted language model*, Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, FL, USA, 2024, pp. 4531–4541.