911 Call Analyzer: A Vital Tool for Detecting Critical Emergencies

Paresh Patil[1], Sushant Gaikwad[2], Akash Hatkangane[3]

[1]*Department of Computer Science, Dr Vishwanath Karad MIT World Peace University, Paud Road, Kothrud, Pune, Maharashtra 411038*

[2] *Department of Production sandwich, AISSMS College of Engineering, Shivajinagar, Pune, Maharashtra 411001*

[3]*Department of electrical engineering, All India shri Shivaji memorial society college of engineering, Shivajinagar, Pune, Maharashtra 411001*

Abstract - Emergency response systems must be able to promptly and accurately evaluate emergency calls. We provide a machine learning-based method in this study, called the "911 Call Analyzer," to automate the process of identifying serious crises from 911 call audio recordings. Mel-frequency cepstral coefficients (MFCCs) are used by the system to extract features, and machine learning and deep learning architectures are used for classification. To forecast the urgency and severity of each emergency call, the collected features are fed into a model that has been trained on a dataset of labelled calls. We assess the 911 Call Analyzer's performance using a test dataset, and we obtain a 91% accuracy rate with RF and XG Boost model followed by SVM with 90% accuracy, CNN with 69% accuracy and lastly LSTM with 64% accuracy. These findings show how well the suggested method works to reliably identify important crises, which helps emergency dispatchers prioritize calls and allocate resources more wisely. The 911 Call Analyzer is a tool that holds great potential for improving emergency response systems' efficacy and efficiency, which will eventually benefit those who are in need.

Key Words: 911 calls, MFCCs, LSTM, CNN, SVM, RF, XG Boost.

1 Introduction

The backbone of the infrastructure for public safety are emergency response services, which offer quick aid to people in need and in dire circumstances. The 911 emergency call centre, which acts as the main point of contact for anyone in need of immediate assistance, is at the centre of these systems. However, the capacity to appropriately determine the seriousness and urgency of incoming 911 calls is crucial to the efficacy of emergency response—a task that is frequently complicated and challenging. The process of call evaluation and prioritization in manyemergency call centres is primarily manual, depending on dispatchers' experience and discretion to assess incoming calls and distribute resources appropriately. Despite receiving extensive training to manage a wide range of emergency situations, dispatchers may take longer to dispatch vital resources to lifethreatening situations due to the subjective nature of call assessment, which can bring variability and inconsistent decisionmaking. Advanced technical solutions that can automate call analysis and prioritize important situations in real-time are desperately needed to address these issues and improve the capabilities of emergency response systems. These systems, which make use of artificial intelligence and machine learning, have the ability to completely change the way emergency calls are handled by promptly and accurately evaluating incoming 911 calls.

The goal of this project is to create a powerful machine learning-based system known as the "911 Call Analyzer," which can accurately identify serious crises from audio recordings of 911 calls. The 911 Call Analyzer is designed to help emergency dispatchers prioritize replies, triage calls, and allocate resources as efficiently as possible in emergency scenarios by utilizing machine learning algorithms and advanced signal processing techniques. The main goal of this study is to create and apply a scalable and reliable machine learning model that can reliably prioritize 911 calls according to their urgency and seriousness. By accomplishing this goal, we hope to increase emergency response system's efficiency, which will eventually benefit people who are in need. We provide a thorough examination of the 911 Call Analyzer's creation, use, and assessment in this report. We go over the approach used to gather and preprocess the dataset, identify pertinent features in audio files, train and assess the learning model. In addition, we explore the implications of our findings for emergency response procedures and offer the experiment data, including performance metrics and comparative analysis.

2 Related Work

Several works have described the creation of tools to help operators prioritize calls. A system for supporting emergency centres is designed [1], combining multiple modules. An Automatic Speech Recognition (ASR) system is used to first transcribe the Spanish calls. A Named Entity Recognition (NER) module then processes the transcriptions to extract relevant entities. An extra classifier module uses methods like Support Vector Machines (SVMs) and TF-IDF (Term Frequency-Inverse
Document Frequency) to determine the service type and priority of a particular call's transcription. In a different study [2], the emergency calls classifier is covered in full.

The texts undergo a number of pre-processing procedures, including lemmatization, stop-word removal, and format conversion to lowercase. Moreover, "word pruning" is applied to the texts to lower the dimensionality of the characteristics. Recall, accuracy, and $F1$-score are all 86%, 75%, and 80% for the best model, respectively. This approach draws upon another technique developed for the benefit of French emergency call operators [3]. Emergency calls are handled using audio processing blocks such as speech activity detection first in the French language. Another block automatically converts the calls to text. The latter is then utilized to train a BERT model [4] on 904 emergency calls in order to forecast the victim's degree of harm at the time of the call. In terms of severity classification, the latter is accurate to 71.2%. Authors such as Blomberg et al. [5] suggested training a machine learning framework for calls of type "cardiac arrest" for high priority calls. Evaluation metrics including sensitivity, specificity, and positive predictive value were used to assess the effectiveness of their methods in identifying extra-hospital cardiac arrest among 108,607 emergency calls. Their approach exhibited a much higher sensitivity and a lower specificity when compared to medical dispatchers.

3 Methodology

Data Collection: The late Gary Allen, the editor and publisher of Dispatch Monthly magazine and the 911 Dispatch website, gathered these 911 logging tapes from open sources. This dataset was downloaded from Kaggle. The dataset consists of 708 audio files.
The recordings, all in wav format, and brief text summaries of every event were taken from a now-defunct 911 Dispatch website that was captured by Internet Archive. Based on the descriptions, recordings, and additional

real news sources, additional metadata (date, state, citizen started, deaths, probable death, and false alert) were entered by hand and coded. The majority of recordings feature a single, complete, uncut 911 call. Some, however, start in the middle, contain content that has been removed, or have several calls.

Data Preprocessing: Initially, the dataset was examined to gain insights about its structure. To understand the distribution of the variables and find any missing values, descriptive statistics and visualizations were used.

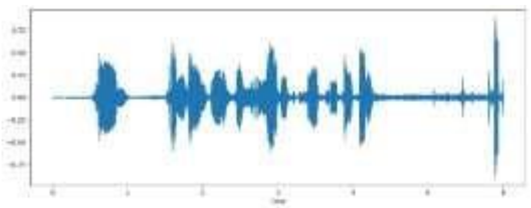


Fig 1. Waveform plot of an audio file

Feature Extraction: The Python librosa library was used to load the audio files associated with the 911 emergency calls. The WAV format was used to store the audio files. Mel-frequency cepstral coefficients (MFCCs) are used to extract the characteristics from the audio files. MFCCs offer a simplified representation of the spectral envelope of an audio signal and are frequently utilized in speech and audio signal processing applications. Iteratively, the features are extracted for every audio file. The dataset is divided into independent variables (X) containing the MFCC features and dependent variables (y) containing the labels after all features and labels have been extracted. SMOTE (Synthetic Minority Over-sampling Technique) is used to balance the training data because there may be a class imbalance problem in the dataset. By following these procedures, you can be sure that the data is ready for training the neural network model and effectively classify audio files according to their attributes.

Model Development: Once MFCC features are extracted from audio files and the data are split into train-test data, an SVM classifier with linear kernel is trained using the training data. Similarly, the training data is utilized to train Random Forest, XG Boost, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) models. Predictions are then generated for both the training and testing data sets. Subsequently, a classification report is produced to assess the models' performance using various evaluation metrics. The Keras API provided by TensorFlow is used to initialize a sequential model. The input layer is the first layer that is introduced to the model in a sequential manner, followed by hidden layers and the output layer. Activation functions are added after each layer to introduce nonlinearity, and each layer is made up of densely linked neurons. In order to minimize overfitting, dropout layers are added after each activation function. During training, a random portion of the neuron outputs are dropped. By defining the optimizer and loss function, the model is compiled. Because binary crossentropy loss works well for binary classification problems, it is selected as the loss function. The model parameters are optimized using the Adam optimizer, which has a set learning rate of 0.01. The compiled model is then trained on the training data. The fit() method was used to train the model. The model's performance is assessed using a number of measures following training, including confusion matrix, accuracy, precision, recall, and F1-score. Classification metrics offer valuable information on how well the model classifies cases into the appropriate classes and help detect biases or incorrect classifications. Confusion Matrix was created to show how well the model performed in terms of true positive, false positive, true negative, and false negative predictions.

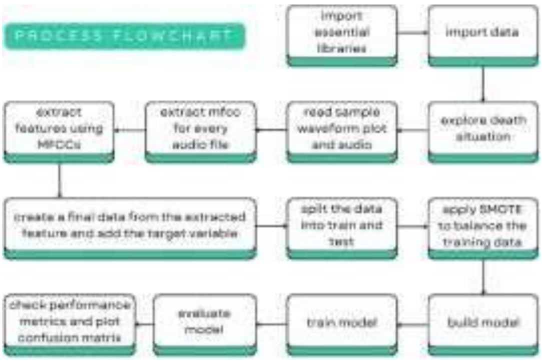


Fig 2. Flowchart of Methodology

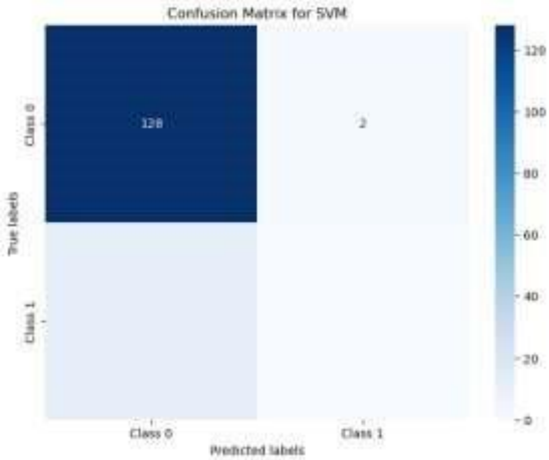


Fig 5. Confusion Matrix for SVM

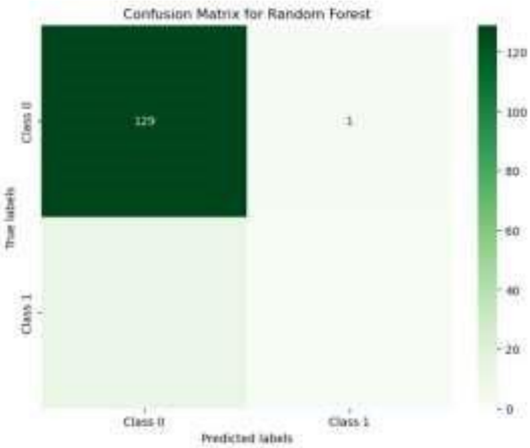


Fig 3. Confusion Matrix for RF

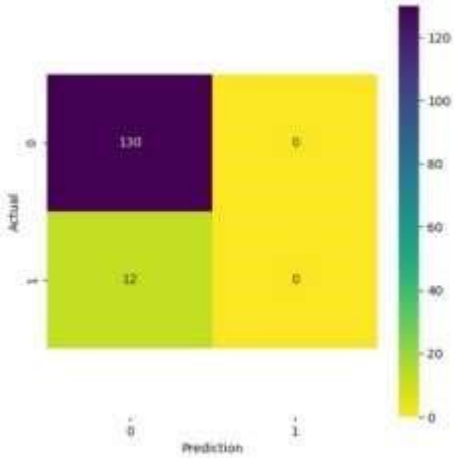


Fig 6. Confusion Matrix for CNN

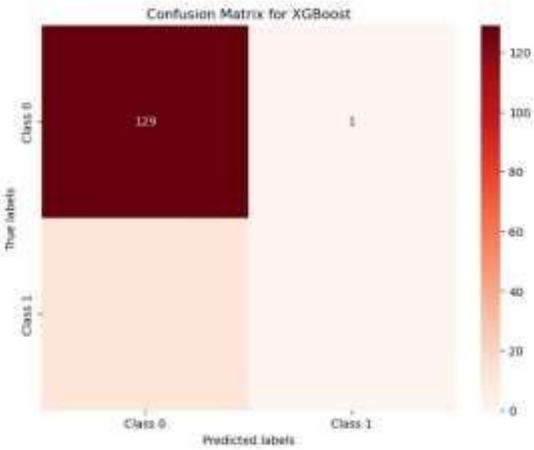


Fig 4. Confusion Matrix for XG Boost

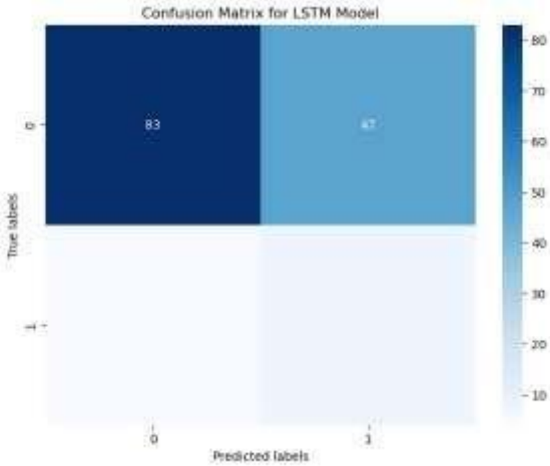
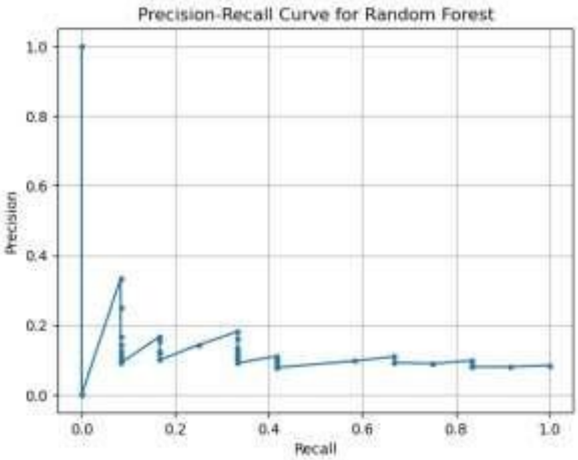


Fig 7. Confusion Matrix for LSTM



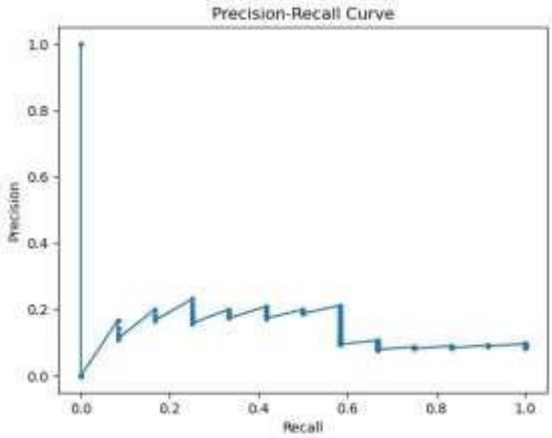


Fig 9. Precision-Recall Curve (CNN)

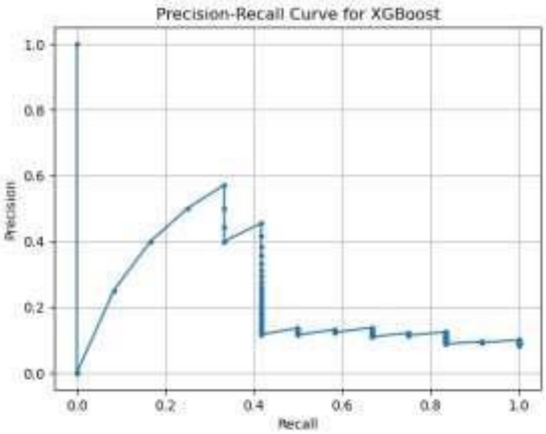


Fig 8. Precision-Recall curve for XG Boost

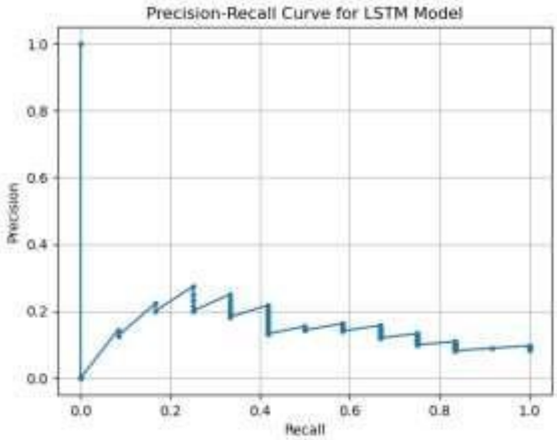


Fig 11. Precision-Recall curve for CNN

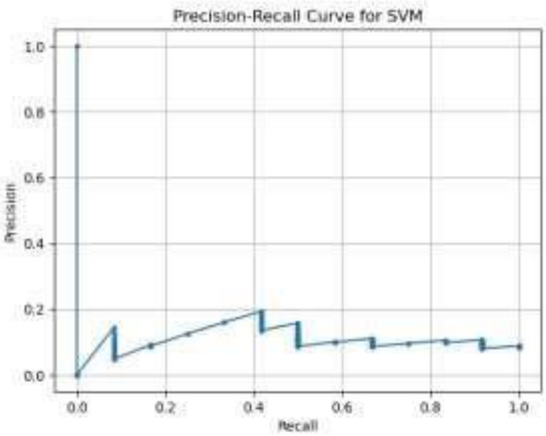


Fig 10. Precision-Recall curve for SVM

4 Result

The analysis results demonstrate the effectiveness of the 911 Call Analyzer in accurately distinguishing emergency calls based on their urgency and severity. Among the models evaluated, Random Forest (RF) along with XG Boost exhibit the highest accuracy rate of 91%. The Support Vector Machine (SVM) model trails slightly with an accuracy of 90%, while the Convolutional Neural Network (CNN) model achieves 69% and Long Short-Term Memory (LSTM) model achieves 64% accuracy.

These findings underscore the diverse capabilities of the analyzed models, showcasing their potential in aiding emergency response systems and optimizing resource allocation for critical situations.

Model	Accuracy
RF	91%
XG Boost	91%
SVM	90%
CNN	69%
LSTM	64%

Fig 12. Comparison table for various models

	precision	recall	f1-score	support
0	0.91	0.99	0.95	130
1	0.00	0.00	0.00	12
accuracy			0.91	142
macro avg	0.46	0.50	0.48	142
weighted avg	0.84	0.91	0.87	142

Fig 13. Classification report for RF and XG Boost

	precision	recall	f1-score	support
0	0.91	0.98	0.95	130
1	0.00	0.00	0.00	12
accuracy			0.90	142
macro avg	0.46	0.49	0.47	142
weighted avg	0.84	0.90	0.87	142

Fig 14. Classification report for SVM

	precision	recall	f1-score	support
0	0.97	0.68	0.80	130
1	0.18	0.75	0.29	12
accuracy			0.69	142
macro avg	0.57	0.72	0.55	142
weighted avg	0.90	0.69	0.76	142

Fig 15. Classification report for CNN

	precision	recall	f1-score	support
0	0.95	0.64	0.76	130
1	0.15	0.67	0.24	12
accuracy			0.64	142
macro avg	0.55	0.65	0.50	142
weighted avg	0.89	0.64	0.72	142

Fig 16. Classification report for LSTM

5 Conclusion

In the present work, the machine learning powered 911 Call Analyzer presents a viable option for automating emergency call processing and ranking crucial incidents. Its ability to precisely identify critical circumstances underscores its potential to improve public safety outcomes by streamlining emergency response systems and cutting response times. To further refine and implement this essential tool in practical circumstances, it will be essential to solve limitations on data and standardize evaluation measures going forward.

6 References

1. Trujillo A, Orellana M, Acosta MI. “Design of emergency call record support system applying natural language processing techniques.” 2019.

2. Orellana M, Trujillo A, Acosta M-I. “A methodology to predict emergency call highpriority: Case study” ECU-911 2020.

3. Abi Kanaan M, Couchot J-F, Guyeux C, Laiymani D, Atechian T, Darazi R. “A methodology for emergency calls severity prediction: From preprocessing to BERT-based classifiers 2023.

4. Tenney I, Das D, Pavlick E. “BERT rediscovers the classical NLP pipeline”. 2019.

5. Blomberg SN, Folke F, Ersbøll AK, Christensen HC, Torp-Pedersen C, Sayre MR, Counts CR,
Lippert FK. “Machine learning as a supportive tool to recognize cardiac arrest in emergency calls” 2019.

6. Rui Zhong, Xiaocen Dong, Rongheng Lin, Hua Zou "An Incremental Identification Method for Fraud Phone Calls Based on Broad Learning System" 2019.

7. Marianne Abi Kanaan, Jean-François Couchot, Christophe Guyeux, David Laiymani, Talar Atechian, Rony Darazi "Combining a multi-feature neural network with multi-task learning for emergency calls severity prediction" 2024.

9. K. Lee, J. K. Kim, M. W. Park, L. Kim, and K. F.
Hsiao, "A Situation- Based Dialogue Classification Model for Emergency Calls"2017.

10. T. Yasuda, Y. Yamada, F. Hamatsu, and T. Hamagami, "A call triage support system for emergency medical service using multiple random forests," 2017.

8. Krishna Kumar, Kapil Chaturvedi "An Audio Classification Approach using Feature extraction neural network classification Approach." 2020.