# Optimizing Emergency Response with Machine Learning

Jackson Thetford
*Center for Engineering Leadership*
*Rice University*
Houston, Texas
j.thetford@rice.edu

Alan Salceda Monge
*Center for Engineering Leadership*
*Rice University*
Houston, Texas
alan.salceda@rice.edu

Diego Jimenez Vivaldi
*Center for Engineering Leadership*
*Rice University*
Houston, Texas
diego.jimenez@rice.edu

***Abstract*—This project presents a project that optimizes the emergency medical services in Allegheny County, Pennsylvania, using data science techniques in the analysis of the 911 dispatch data. It holds around 2 million entries of emergency calls with priority levels, geographical locations, and call descriptions. The research aims to enhance the efficiency of the EMS through the application of supervised and unsupervised machine learning models. The project includes cleaning and transformation of the dataset, where redundant columns were dropped, missing values were handled, and categorical variables were encoded. Supervised learning models were trained to classify calls based on their priority level with varying degrees of accuracy. Dimensionality reduction techniques, such as PCA, were applied to solve high-dimensionality data and enhance the interpretation and efficiency of the models.**

**The unsupervised clustering models, such as K-means and Gaussian Mixture Models were applied to the optimization of priority classification based on patterns in data. Overlaps of existing priority levels with each other suggest that these models make improvements in the system of priority classification. Techniques of visualization, such as PCA plots, were also deployed to validate cluster distributions and their efficiency. This research combines location, time analysis, and machine learning to solve a critical inefficiency problem in EMS response times in the county. The results show how streamlined priority classification and resource allocation can bring data-driven improvements in emergency response systems. The next steps include deploying the models and establishing priority levels to enhance real-world operations for EMS.**

## I. Introduction

When in the middle of an emergency, one's first instinct is usually to call for help. In a recent study by the American College of Emergency Physicians, about 76% of adults feel prepared to communicate with a 911 dispatcher [9]. However, in another study, it is shown that only 47.4% of people know how to check for a pulse in an unconscious person [10]. Due to this lack of information, some emergencies are not labeled or treated accordingly. This highlights that while 911 dispatchers are a crucial resource, their effectiveness can be limited by the caller's lack of basic emergency knowledge, potentially delaying emergency response.

Effective communication is critical in emergencies, as the way an incident is described influences the priority level assigned by dispatchers. In many cases, this can lead to issues due to inconsistencies caused by different levels of priority being assigned to similar incidents. To address this problem, an unsupervised machine learning model was created to help reduce the error in designating priority levels.

To do this, a dataset of 911 emergency medical service (EMS) calls made in Allegheny County in Pittsburgh, Pennsylvania, was used. The data set contained a list of incidents along with their description, location, and priority level. The priority levels ranged from E0 to F4, E0 being the highest priority and F4 the lowest. At first, the data needed to be cleaned and filtered by eliminating unnecessary columns and values that had no input. We then created a dictionary to group similar components, such as the priority levels and their descriptions. All of this data was then fed to the machine learning model to perform our unsupervised learning analysis.

## II. Literature Review

Understanding the geographical, demographic, and infrastructural context of Allegheny County is necessary for analyzing 911 emergency dispatch call data for the presented project. Located in Pennsylvania, Allegheny County is a major metropolitan area with a population of about 1.216 million people as of 2022 [2] according to the U.S. Census Bureau. In determining the Emergency Medical Services (EMS) response capacity the factors of population density, distribution of health facilities, and level of urbanization in the county become very relevant. The Data Commons platform extends this understanding by providing demographic and population statistics specific to Allegheny County [3]. Researchers better understand how population density within different areas of the county could create differences in the demand for emergency services by visualizing the concentration of the population.

The availability of health infrastructure, such as hospitals and clinics, impacts EMS operations. Several authors emphasize the existence of hospitals and clinics in Allegheny County that constitute important destinations for emergency medical transportation [1]. Thus, the number of, and the location of healthcare centers would directly relate to EMS route optimization, response time, and efficiency in delivering patient care. Given that the county offers a diverse range of urban, suburban, and rural areas, emergency medical services dispatch operations face unique challenges in both navigating through traffic and finding the best route to hospitals. Furthermore, the government of Allegheny County provides more insight into the county's public health and safety infrastructure. The official website of the county delineates the scope of public services, including EMS, in coordination with first response services during emergencies [4].

The safety and crime rates of various zones within Allegheny County are related to emergency response demands. Crime Grade [5], a governmental safety organization, rates various zones within Allegheny County as high violent zones, which is related to emergency response demands. Crime Grade ranks Allegheny County according to the relative safety of neighborhoods, identifying regions with higher risks of crime and violence [5]. That is important information for any EMS dispatch analysis because a higher crime rate zone might have more emergency calls due to violent incidents which could increase the burden of emergency response teams.

Higher-risk neighborhoods often require faster EMS response times to address life-threatening injuries from assaults, shootings, or accidents. This social context directly impacts the need for faster response times in specific areas. It also helps in focusing on neighborhoods with higher incidents of violence by providing priority to emergency calls. Crime Grade's high-risk area classification can be fed into machine learning models or predictive analytics for optimization of emergency response. This could help ensure that EMS resources are better allocated to high-demand areas, minimizing response delays in areas of high emergency call volume.

Emergency Medical Services form a critical component of public health infrastructure, providing immediate medical care in potentially life-threatening circumstances. The National EMS Office describes EMS as "a planned system with specified personnel, transport, and healthcare facilities that work together to provide immediate medical care [7]." This definition highlights the complexity involved in managing a response system with so many interconnected parts. This system's efficiency is often measured by response time. The response time is the time taken for an emergency response unit to reach the scene after receiving a call.

One of the primary datasets for this study is provided by the Western Pennsylvania Regional Data Center (WPRDC) and is titled "Allegheny County 911 Dispatches - EMS and Fire" [6]. The dataset includes all necessary features, such as time and location of emergency calls, dispatch times, and response times. Analyzing these dispatch records makes it possible to measure the performance of EMS units and find bottlenecks that slow response times. The WPRDC dataset will be the foundation for developing machine learning models and predictive algorithms to improve EMS response times in Allegheny County.

Another very critical component is the use of geospatial analysis and location-based routing. The University of Toronto's Ambulance Emergency Response Optimization (AERO) Initiative shows various ways in which emergency ambulance response can be optimized by using mathematical modeling, geospatial analysis, and predictive analytics [8]. AERO's approach includes predictive models to anticipate high-demand areas for emergencies and position EMS units closer to those locations. This preposition strategy reduces the distance that EMS vehicles have to travel and, as a result, shortens response times. All of these ideas—dynamic dispatching, ambulance reallocation, and demand prediction—are useful for the analysis of the 911 emergency call data in Allegheny County.

Further, the use of better routing algorithms can also improve emergency response. Traditional response systems usually rely on a fixed response zone; however, modern systems, such as those proposed by AERO, involve a dynamic allocation of EMS vehicles based on live information. This reduces dead time - time wasted by ambulances waiting for dispatch orders - and ensures that vehicles are efficiently positioned to respond to emergencies wherever they may arise.

An EMS response times analysis in Allegheny County identifies factors that delay response to emergencies. The supporting datasets and sources include: The WPRDC EMS and Fire dispatch dataset contains information on the location, priority, and timing of emergency calls [6]. A row in the dataset represents one dispatch event; therefore, temporal and spatial analysis can be performed on emergency calls. The analysis exposes strong trends, for example, the volume of calls, the time of day that sees the most emergencies, the quarter of the year in which the call was made, and EMS priorities with the fastest and slowest response times.

Selecting important input features is a key aspect of the development of predictive modeling for EMS response times. The main inputs are location, call type, time of day, and traffic conditions, which show a strong relation to the outcome variable, which is the response time. The focus of the AERO Initiative, geospatial optimization [8], shows how models can lower response times by allocating emergency resources close to high-demand areas. The same techniques could be applied to Allegheny County, where data from Crime Grade and the U.S. Census Bureau[2] can be applied to resource allocation prioritization towards the high-risk and densely populated areas. The area with multiple risk concerts is the city of Pittsburgh.

Another important concept is calculating the time from the dispatch call to the emergency scene. This concept can be understood as the difference between when a call is made and when the ambulance arrives. To lower this time as much as possible, this dataset from WPRDC needs analysis to identify where the biggest delays are. By grouping the data by priority level (E0, E1, E2, etc.), one can observe the types of emergencies that have the most delays. Analyzing this information provides insights into where process improvements are required.

The presented analysis of 911 emergency call data in Allegheny County is based on a wide source of information from contextual, operational, and methodological information. First, demographic and geographic information from the U.S. Census Bureau [2], Felt [1], and Data Commons [3] help ground this analysis in a population-level understanding of the areas and contexts in which emergency events occurred. Additionally, Crime Grade [5] informs about the level of violence and crime rate in explaining the demand for emergency response services. Operationally, results from the National EMS Office [7] indicate the need to enhance response times, an objective further supported by the geospatial and predictive modeling approaches of the AERO Initiative [8]. The EMS dispatch times from the main dataset of WPRDC [6] are essential, allowing for data-driven efficiency in emergency response. Taking the integration of the concepts from all these sources into consideration, this study would look to improve the performance of EMS through predictive modeling and dynamic resource allocation with geographic optimization. This paper, therefore, seeks to improve emergency response times, efficient prioritization, and optimal resource allocation.

## III. Methodology

This section describes the methods and models used.

### A. Data Collection and Preprocessing

The dataset used for this project was obtained from the Allegheny County 911 Dispatch records and contains nearly 2 million emergency call entries. The raw data included information about the given priority level of the incident, incident descriptions, geographical locations, and various other incident attributes. The initial preprocessing involved cleaning the dataset by:

- Removing irrelevant or redundant columns.
- Handling missing values through various techniques such as inference or deletion, based on the significance of the attribute.
- Converting categorical variables into numerical formats suitable for model training.
- Altering the scope of our model regarding the number of predictable classes analyzed due to limited sample sizes for certain priority levels.

Exploratory Data Analysis (EDA) was performed to give us info about the dataset, such as the distributions of priority levels, the recorded year and quarter of the calls, and the geographic distribution of the calls. A major part of this analysis was the visualizations, which included distributions, heatmaps, and temporal trends to understand the data better.

The data was also split into stratified testing and training sets for the supervised learning model. This allowed for a proportional representation of all the classes in our model training and prevented size bias.

### B. Research Hypothesis

Our research hypothesis is that emergency call descriptions, combined with location data, could be used to classify calls into priority levels (E0–E5 and F0–F5) with high accuracy, which will allow this process to be automated, removing the need for human dispatchers which will reduce human errors and speed up the emergency response process for those in need.

### C. Model Development

We tested several supervised machine learning models to classify the calls into predefined priority levels. As we are using supervised learning, we are using the “priority” level in the data set to be the dependent variable we are trying to predict. This makes the assumption that this variable is absolute and the data set is deterministic in regards to each priority level, or that there is no human error in the original dataset. The algorithms were explored and implemented, including:

- **K-Nearest Neighbors (KNN):** For non-linear decision boundaries in the feature space.
- **Random Forest:** This model was used for its ability to handle high-dimensional data and provide feature-importance insights.
- **XGBoost:** For its efficiency in handling large datasets and high performance in classification tasks.

Models were evaluated on metrics such as accuracy, precision, recall, and F1-score. Special emphasis was placed on recall to prioritize life-threatening situations appropriately.

However, after training our regression models, we found that our assumption of deterministic priorities was incorrect and there was an overlap or a “gray area” between priority levels. This will be further explained in a later section. Our solution to this was to run train unsupervised machine learning models that would enable priority prediction levels based on the absolute data we had and not the human-influenced data like priority level. This approach allowed the data itself to dictate groupings that better captured the nuanced characteristics of emergency calls. We performed PCA analysis to reduce dimensionality, identify the most significant features, and determine the best number of components and clusters for our models. The unsupervised algorithms we explored were:

- **K-Means Clustering:** An algorithm that assigns data points to a fixed number of clusters based on their similarity.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** A density-based algorithm that identifies clusters of varying shapes and sizes, chosen because of its ability to handle outliers as noise.
- **Gaussian Mixture Models (GMM):** A probabilistic clustering method that models data as multiple Gaussian

distributions.

These unsupervised techniques identified clusters and essentially redefined the priority levels of each EMS call. These clusters were compared with the original priority levels.

### D. Feature Engineering and Importance Analysis

Feature engineering was performed for both the supervised and unsupervised learning techniques implemented. For the supervised techniques, this included Term Frequency Inverse Document Frequency for text data, geospatial encoding, feature scaling, and various methods of encoding our categorical / non-numerical data. For the unsupervised techniques, the previously mentioned methods were used in addition to Principal Component Analysis (PCA) which helped reduce dimensionality and made the clustering algorithms simpler and easier to understand.

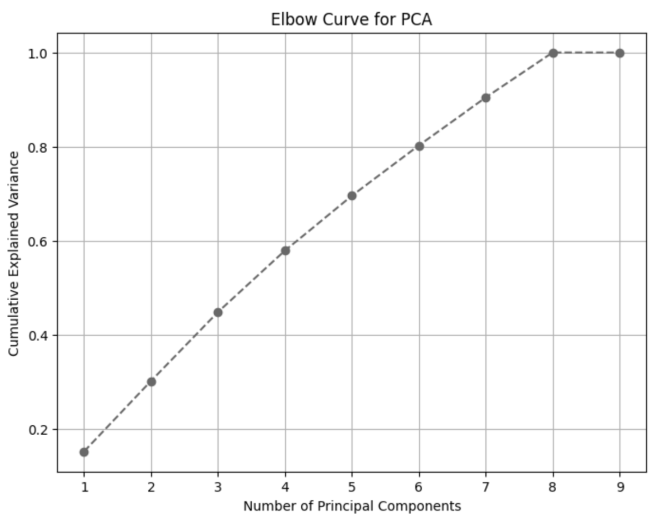


Fig. 1. PCA Elbow Curve

Feature importance analysis, using SHAP values and feature importance scores from Random Forest and XGBoost models, highlighted the most critical attributes influencing priority classification. Figure 2 shows the feature importance from the XGBoost model.

### E. Model Evaluation and Validation

Models were trained and validated using a stratified train-test split that balanced the representation of priority classes to the proportion of our original dataset. When relevant we performed cross-validation to help train our models to prevent overfitting. We used various metrics to analyze our model such as precision, recall, and f1-score for our supervised models, and silhouette score for our clustering models.

### F. Deployment and Impact

Our end goal is to create an efficient model to take in EMS call information and assign it a priority level, whether following the existing scale or making our own through new clustering, and ultimately increase the efficiency of Allegheny County's EMS response department. With further research and in future work, these improvements will be done by selecting the best-performing model and deploying it in the real world.

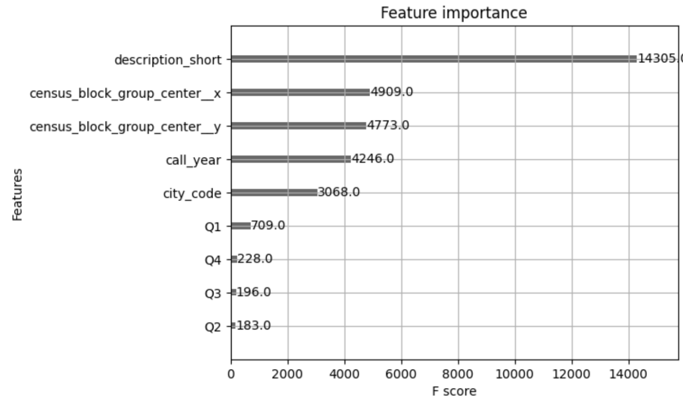


Fig. 2. XGBoost Feature Importance

## IV. Results and Discussion

### A. Supervised Learning Models

The supervised learning models, K-Nearest Neighbors (KNN), Random Forest, and Boost, were evaluated on their accuracy in classifying EMS call priorities. The models were also assessed on their prediction recall to minimize the number of false negatives. False negatives are when the model incorrectly classifies a true positive, which in our case could be potentially life-threatening in the case the model misclassifies a high, life-threatening priority call like an E0 or E1 as a lower priority. Each model had varying levels of success, but all models were within range of each other in terms of our evaluation metrics, and all performed sub-optimally.

- **KNN:** Achieved an accuracy of 59% with a macro-average F1-score of 0.65. This model had a weighted average recall of 59%. Being one of the simpler models, it struggled to differentiate between closely related priority levels, with significant misclassifications in intermediate categories (E2–E4).

- **Random Forest:** Provided similar results to KNN. It achieved a 59% accuracy, a macro F1-score of 0.67, and a weighted average recall of 59%. This model did well in the E5 and F4 categories, with an accuracy of 98% and 84%, respectively.

- **XGBoost:** Outperformed the other supervised models with an accuracy of 63% and a macro F1-score of 0.69. XGBoost, as expected, showed strong performance in categories with clear distinctions (edge categories like

E5 and F4) but struggled with closer priority levels (E2 and E3). XGBoost did even better than Random Forest in the E5 and F4 categories, with an accuracy of 99% and 96%, respectively.

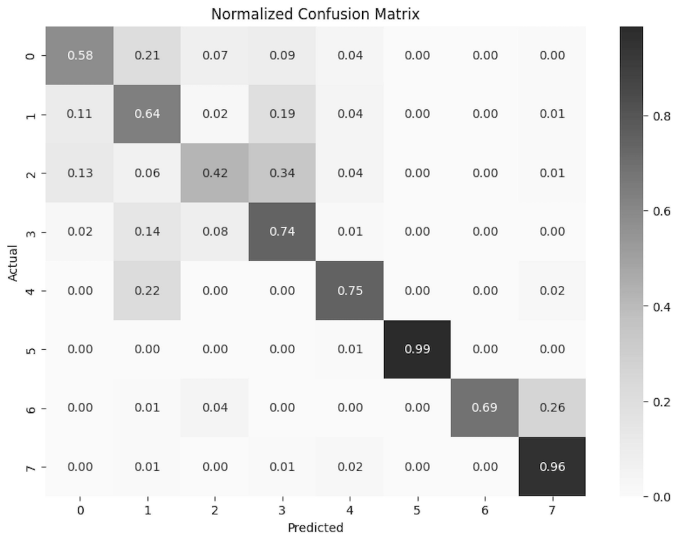


Fig. 3. XGBoost Confusion Matrix

Interestingly, the models did very well at predicting E5 priority EMS calls, with the worst-performing model being KNN, with an accuracy of 98% for this priority level. Also, even though the models show low accuracy, it should be noted that there is a clear distribution of inaccurate predictions close to the true value. For example, when predicting an E0 priority call, the most common incorrect prediction was E1, followed by E2. There were very few "bad" guesses, such as a model guessing E5 for a true E0 call. We can also see the feature importance from the XGBoost model shown below, where the random forest feature plot was nearly identical.

Overall, the models did better than guessing. However, they could not completely define the categories they were given. This is most likely due to the inherent overlap and subjectivity in the dataset's priority labels, which is likely due to human error and ambiguous definitions in the original data.

### B. Unsupervised Learning Models

To address the aforementioned challenges, unsupervised learning techniques were implemented to redefine priority levels based on inherent data characteristics rather than subjective labels. Principal Component Analysis (PCA) was used to reduce dimensionality, and clustering algorithms such as K-Means, DBSCAN, and Gaussian Mixture Models (GMM) were explored.

- **K-Means Clustering:** This method provided the most interpretable results, where four very isolated clusters were found. Using 3D PCA visualizations, the clusters were visualized, but it was later discovered that the clusters were only representative of the quarter (temporal) that the call was received, which is not indicative of a call priority by itself. Another K-means clustering model was made without considering the quarter, and a messier but more realistic clustering model was found. After further optimization and feature selection, this model showed four quasi-isolated clusters without excessive overlap that will allow the prediction of EMS calls into four new priority levels. One possible issue with this application of K-means clustering is to further research the rigid assumptions, such as spherical distribution and an equal number of points in each cluster.

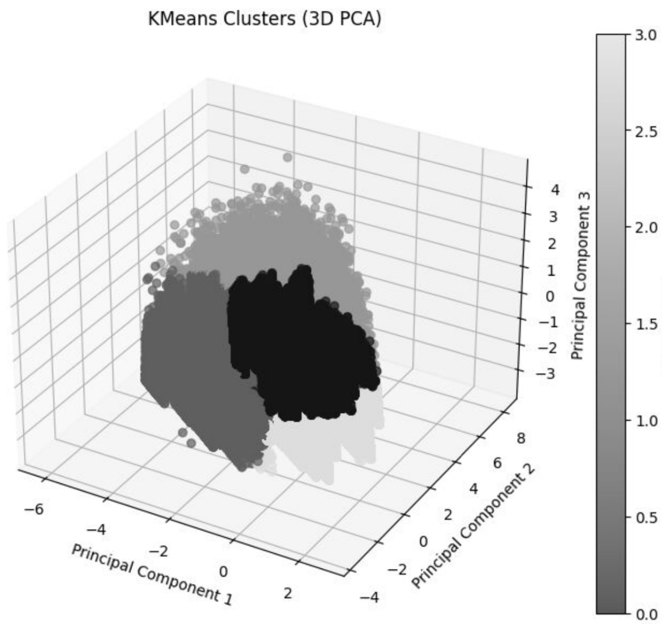


Fig. 4. 3D Kmeans Clustering without Quarter Feature

- **DBSCAN and GMM:** These methods were used for their flexibility in handling outliers and probabilistic clusters. However, they did not result in meaningful groupings for this dataset, likely due to the high dimensionality and sparsity of some features.

### C. Discussion

The limited performance of supervised learning models shows the difficulty of using human-labeled data to train classification models for critical applications like EMS call prioritization. The subjective decision-making in the labeling process likely contributed to overlapping categories and non-trivial error rates.

Unsupervised learning offered a promising alternative by letting the data dictate the groupings. The K-Means clustering model highlighted the potential to redefine the priority levels more consistently and objectively. However, these results come with some rigid assumptions and would need to be further validated with stakeholders to ensure its practicality.

Several key areas for improvement were identified:

1) **Data Quality:** Subjectivity and potential errors in the priority labeling should be improved.

2) **Feature Engineering:** Additional context-specific features, such as dispatcher notes, could be used to enhance model performance.

3) **Model Refinement:** Exploring hybrid approaches, such as combining clustering methods with supervised models or ensemble supervised learning methods.

The supervised learning models show the importance of data collection, processing, and feature engineering, and how data is often the bottleneck for the effectiveness of a model. This use of data also proves the information theory law that no process can result in a net gain of information, and often processes and transformations lead to a loss of information. In this case, the transformation of the scene that led to the EMS call, to the EMS call operator logging the call, to the final data used to make our models, led to inaccuracies and uncertainties to the fault of our model. This was somewhat resolved by the unsupervised learning models, but still most likely do not represent the truth with 100% accuracy.

Overall the models developed in this study show the potential to significantly enhance EMS call prioritization and resource allocation efficiency; however, for the models derived in this paper, there is still significant room for improvement in every step of the data science life cycle.

## V. Conclusion

Our analysis of the model's performance using a supervised learning technique revealed several insights. Firstly, when looking at the classification report and XGBoost used to determine the accuracy of our model, we can see that our model had an average accuracy of 60%. This means that, when it came to classifying the priority levels, the model had issues, especially for the categories E0, E2, and E3. In the confusion matrix it is shown that, even though the model predicted many of the values wrong, it was not off by far. After looking at these results, it was determined that human error was involved in assigning priority levels to the incidents. This raised the question of: is a supervised learning model the best way to go about doing this.

In the end, after determining that the supervised model was not optimal, an unsupervised model was used to reevaluate the priority categorization system. A PCA analysis was performed and it showed that the optimal number of clusters was 4 which was further supported by the elbow curve. This led to the belief that many of the priority levels currently used by 911 dispatchers had much more in common than was originally intended. Taking all this into account, having 8 different priority levels is not efficient and EMS dispatching should consider implementing a more streamlined approach.

## References


[1] Felt, "Hospitals and Clinics in Allegheny County, Pennsylvania," *Felt*, [Online]. Available: https://felt.com/explore/hospitals-clinics-allegheny-county-pennsylvania.

[2] U.S. Census Bureau, "Allegheny County, Pennsylvania," *QuickFacts*, [Online]. Available: https://www.census.gov/quickfacts/fact/table/alleghenycountypennsylvania/PST045222.

[3] Data Commons, "Allegheny County, Pennsylvania," *Data Commons*, [Online]. Available: https://datacommons.org/place/geoId/42003?utm$_{m}edium = exploremprop = countpopt = Personhl = en.$

[4] Allegheny County, "Home," *Allegheny County*, [Online]. Available: https://www.alleghenycounty.us/Home.

[5] CrimeGrade, "The Safest and Most Dangerous Places in Allegheny County, PA," [Online]. Available: https://crimegrade.org/safest-places-in-allegheny-county-pa/.

[6] Western Pennsylvania Regional Data Center, "Allegheny County 911 Dispatches - EMS and Fire," *WPRDC*, [Online]. Available: https://data.wprdc.org/dataset/allegheny-county-911-dispatches-ems-and-fire.

[7] U.S. Department of Transportation, "What is EMS?," *EMS.gov*, [Online]. Available: https://www.ems.gov/what-is-ems/.

[8] University of Toronto, "Ambulance Emergency Response Optimization (AERO)," *Centre for Global Engineering (CGEN)*, [Online]. Available: https://cgen.utoronto.ca/research-initiatives/ambulance-emergency-response-optimization-aero/.

[9] "Nearly Half of U.S. Adults are Unprepared to Help in a Medical Emergency," *www.emergencyphysicians.org*. https://www.emergencyphysicians.org/press-releases/2023/1-18-23-nearly-half-of-u.s.-adults-are-unprepared-to-help-in-a-medical-emergency

[10] J. Bahr, H. Klingler, W. Panzer, H. Rode, and D. Kettler, "Skills of lay people in checking the carotid pulse," *Resuscitation*, vol. 35, no. 1, pp. 23–26, Aug. 1997, doi: https://doi.org/10.1016/s0300-9572(96)01092-1.