

AI's Current Impact and Future Potential in Emergency Services: A Comprehensive Review and Analysis

Roxane Elias Mallouhy

Al Yamama University, Saudi Arabia

E-mail: r_mallouhy@yu.edu.sa

ORCID iD: <https://orcid.org/0009-0008-6433-310X>

Naoufal Sirri*

FEMTO-ST Institute, UMR 6174 CNRS, University of Franche-Comté, Belfort, France

E-mail: naoufal.sirri@univ-fcomte.fr

ORCID iD: <https://orcid.org/0000-0002-9900-337X>

*Corresponding author

Irum Nahvi

Prince Mohammad Bin Fahd University, Saudi Arabia

E-mail: inahvi@pmu.edu.sa

ORCID iD: <https://orcid.org/0000-0002-2895-8160>

Christophe Guyeux

FEMTO-ST Institute, UMR 6174 CNRS, University of Franche-Comté, Belfort, France

E-mail: christophe.guyeux@univ-fcomte.fr

ORCID iD: <https://orcid.org/0000-0003-0195-4378>

Received: 03 July 2024; Revised: 17 August 2024; Accepted: 20 September 2024; Published: 08 December 2024

Abstract: The dynamic force of artificial intelligence (AI) is reshaping our world, not in the distant future, but today. Its transformative potential, adaptability, and capacity to liberate human potential are becoming evident in a multitude of domains. AI's ability to process vast datasets, offer data-driven recommendations, and enhance decision-making processes underscores its pivotal role in addressing complex challenges. This article explores AI's current impact and its potential for further growth. It reviews 77 articles across diverse domains, highlighting AI's role in emergency services. Through an in-depth analysis of these studies, the paper provides a broad overview of the current state of AI in emergency services, identifying key trends, challenges, and future opportunities. By examining the methodologies, datasets, AI and deep learning techniques, feature selection processes, evaluation metrics, and prediction models used in each study, the paper aims to offer a thorough understanding of AI's role in this critical sector. This extensive body of knowledge is intended to be a valuable resource for researchers, practitioners, and policymakers. It supports the ongoing advancement of AI-driven emergency services, with the goal of saving lives, optimizing resource allocation, and enhancing response times in critical situations. Ultimately, this collaborative effort seeks to foster the development of more resilient and responsive emergency systems that can effectively mitigate risks and deliver timely aid to those in need. By advancing the capabilities of emergency response systems, AI enhances the precision and efficiency of critical interventions, ultimately leading to better outcomes and improved resilience in crisis situations.

Index Terms: Artificial Intelligence, Emergency Services, Machine Learning, Deep Learning, Data Analysis, Decision-Making, Healthcare, Public Safety, Data Insights, Innovative Technology.

1. Introduction

The past decade has witnessed a paradigm shift in the field of artificial intelligence (AI), fundamentally reshaping its capabilities and potential applications across diverse domains. This transformation is particularly evident within the

realm of emergency services, where AI has become a critical driver of progress. Some key factors have fueled this evolution: Firstly, ground-breaking advancements in AI research, both at the theoretical and algorithmic levels, have laid the foundation for this progress. The emergence of cutting-edge techniques such as gradient boosting, deep learning architectures including Long Short-Term Memory (LSTM) networks and transformers, and the development of large language models have propelled AI to unprecedented levels of performance and real-world utility within emergency services. For instance, LSTM networks have revolutionized the analysis of sequential data, enabling AI systems to effectively predict trends in emergencies like natural disasters or disease outbreaks. This critical capability allows emergency services to proactively prepare and deploy resources, potentially saving lives. Similarly, large language models like GPT (Generative Pre-trained Transformer) have empowered AI systems to interpret and generate human-like text, facilitating advancements in natural language understanding and generation. This proves invaluable for processing and comprehending emergency-related communications and reports, enabling more efficient and effective response efforts. Secondly, the exponential growth in computational power, driven by the widespread adoption of Graphics Processing Units (GPUs) and other high-performance computing technologies, has revolutionized the training and deployment of increasingly complex AI models within emergency services. The increase in processing capabilities has enabled practitioners to address previously daunting challenges and attain unmatched levels of accuracy and efficiency. For example, GPU-accelerated computing has significantly reduced the training time for deep learning models, allowing emergency response teams to rapidly develop and deploy AI-powered tools. These tools can assist with tasks like image recognition in disaster zones or real-time data analysis during medical emergencies, ultimately saving lives and minimizing damage. Finally, the continuous digitalization of society has generated and accumulated vast troves of data, providing the essential fuel for AI algorithms to learn and generalize effectively. The proliferation of sensors, connected devices, and online platforms has led to the creation of massive datasets encompassing diverse information sources relevant to emergency services, such as geospatial data, social media feeds, and medical records. By harnessing the power of AI to analyze and extract insights from these data streams, emergency responders gain unprecedented situational awareness. Furthermore, advancements in data processing techniques like federated learning and edge computing enable AI models to learn from distributed data sources while maintaining privacy and security, fostering collaborative emergency response efforts involving multiple stakeholders.

The transformative potential of artificial intelligence (AI) is increasingly evident as it dynamically reshapes various sectors, including healthcare, public safety, finance, agriculture, and transportation echoing the digitalization-driven advancements. Liberating human resources from routine tasks, AI enables professionals to focus on creative endeavors and strategic decision-making. Its proficiency in processing vast datasets and providing data-backed recommendations has revolutionized optimization processes across industries. For instance, in retail, AI optimizes prices for enhanced profitability [1], while in agriculture, it aids in disease detection and provides data-driven insights for superior crop management [2]. Healthcare benefits from increased patient care efficiency, with AI making strides in robotic surgery [3] and drug discovery [4]. Autonomous vehicles promise enhanced access and road safety [5], while customer service chatbots offer round-the-clock assistance and cost reduction [6]. Furthermore, traffic prediction aids drivers in selecting optimal routes and navigating through varying weather conditions [7]. Real estate price prediction becomes an invaluable tool for sound real estate decisions [8], while social media content moderation fosters a safer online environment [9]. Intelligent video surveillance enhances security through real-time detection [10], and customer segmentation refines marketing strategies by understanding customer needs and 60 optimizing pricing [11]. Finally, content and product recommendations enhance user experiences by proposing relevant products and content based on user interactions [12]. The breadth of AI applications underscores its adaptability and potential for metamorphosis, driven by advancements in AI research, computational power, and the continuous digitalization of society. This explosion in AI adoption is driven by groundbreaking techniques like deep learning architectures and large language models, as well as exponential growth in computational power and the proliferation of data from sensors, connected devices, and online platforms. As AI technology continues to evolve, its transformative impact on various sectors is poised to reach even greater heights in the years to come.

In the realm of emergency services, the powerfully AI has already been demonstrated and industrialized to a significant extent. AI systems are currently utilized for tasks such as predictive analysis of natural disasters and disease outbreaks, optimizing resource allocation, and real-time data analysis during medical emergencies. These applications have greatly enhanced the efficiency and effectiveness of emergency response efforts, ultimately saved lives and minimized damage. However, the potential of AI in the world of emergencies is far from fully realized. Many opportunities remain untapped and while strides have been made, there is still much to be done to fully leverage the capabilities of AI in emergency scenarios. Building upon these advancements, the objective of this article is to provide a comprehensive assessment of the progress made in the AI applications in different areas. By analyzing 77 distinct articles from the scientific literature, this paper aims to furnish a panoramic view of the current AI landscape within emergency services. It elucidates prevalent trends, confronts existing challenges, and underscores prospective opportunities within these crucial sectors. By rooting into the methodologies and materials utilized in each article, encompassing summaries, datasets, AI and deep learning techniques, feature selection methodologies, evaluation metrics, and prediction types, the aim is to provide a comprehensive grasp of the role of AI in emergency services. This repository of knowledge serves as an invaluable resource for researchers, practitioners, and decision-makers alike, facilitating the continual enhancement of AI-driven emergency services contributing to saving lives, optimizing resource allocation, and improving response times in critical situations. Ultimately, this collaborative effort leads to more resilient and responsive emergency systems,

capable of mitigating risks and providing timely assistance to those in need, thereby safeguarding communities and enhancing public safety on a global scale.

The remainder of this article is organized as follows: Section 2 reviews the literature on AI in emergency services, providing a foundation by summarizing key findings from 77 research papers that have shaped current understanding. Building on this, Section 3 outlines the materials and methods used in this study, detailing the data acquisition process, traditional statistical methods, AI techniques, feature selection, and evaluation methods employed to construct and assess the models. The outcomes of these methods are presented in Section 4, where the study's results are discussed supported by figures and tables. Section 5 then explores the broader implications of these results, examining the quality of predictions, the challenges encountered, and addressing important issues related to security and privacy. Finally, Section 6 concludes by synthesizing the key findings and suggesting directions for future research, emphasizing the ongoing need for advancements in AI-driven emergency services.

2. Literature Review

The research landscape presented in this paper consists of a comprehensive analysis of 77 distinct research papers, each dedicated to a wide range of subjects that fall within the domains of emergency services, fire management and ambulance dispatch. These papers collectively embody a diverse spectrum of primary and secondary objectives, each contributing to the advancement of critical aspects like emergency response, public safety, and resource allocation. This paper serves as a comprehensive summary of findings extracted from 77 distinct research papers, each delving into a diverse array of subjects. In their individual explorations, these research papers are driven by the pursuit of multiple primary and secondary objectives. The primary objectives predominantly revolve around the creation of predictive models, optimization strategies, or pioneering algorithms. For instance, certain papers are dedicated to predicting parameters such as ambulance demand, the risk of fire incidents, or strategies for efficient firefighter interventions. In parallel, others are dedicated to the incorporation of machine learning and deep learning methodologies within emergency services.

Meanwhile, the secondary objectives of these research papers are equally noteworthy. These objectives predominantly concentrate on elevating the efficiency, effectiveness, and sustainability of the respective services under investigation. By tackling these secondary goals, the research presented in this paper ultimately extends its contributions to the broader fields of emergency management and urban planning. Through the presentation of innovative solutions and the insights gleaned from data-driven approaches, these research papers offer valuable recommendations aimed at ameliorating the allocation of resources and streamlining response times. To further assist in understanding the breadth of these contributions, refer to Table S1, where both the primary and secondary goals outlined in these articles are thoughtfully summarized.

3. Material and Methods

3.1. Data

In the domain of emergency prediction, the selection and acquisition of appropriate data sources are pivotal to the success of predictive modeling and analysis. This section provides an overview of the key data sources employed in this study. These sources encompass a diverse range of information, often collected from various origins, to create comprehensive datasets for training and evaluating AI algorithms. Some datasets from recent research that can be cited include:

- [13] leverages a dataset encompassing two distinct components. The first component comprises precise temporal records detailing firefighter interventions within the Doubs region of France, spanning the period from 2006 to 2017. The second component comprises a comprehensive set of external variables, encompassing meteorological conditions, traffic patterns, holidays, and various other factors potentially influencing the frequency of firefighting interventions.
- In [14], the dataset employed comprises data drawn from two primary sources, namely the National Emergency Department Information System (NEDIS) and Emergency Medical Service (EMS) run sheet. The NEDIS dataset comprises a substantial volume of records originating from emergency department visits, while the EMS run sheets data pertains to emergency medical service situations. For the purposes of model development, the study harnesses a sizable dataset, specifically 8, 981, 181 entries from the NEDIS dataset for development, and 2604 entries from the EMS run sheets dataset for testing.
- [15] amalgamates data from a diverse array of sources to fuel its analysis. The dataset encompasses 2543 historical fire incidents generously provided by the City of Atlanta Fire Rescue Department (AFRD), 32488 fire inspections, structural information concerning commercial properties procured from the CoStar Group, parcel data from Atlanta's Office of Buildings, and business license records obtained from the City of Atlanta's Office of Revenue, socioeconomic and demographic data from the U.S. Census Bureau, liquor license records, 2014 crime data sourced from the Atlanta Police Department, and Certificate of Occupancy (CO) data from the Atlanta Office of Buildings. The synergistic utilization of these datasets contributes substantively to the discovery of new insights and the development of a predictive model tailored for commercial fire risk estimation.

- The dataset used in [16] encompasses historical crime records and Twitter data sourced from the entirety of available Twitter users within the geographic area of interest, which is Chicago, Illinois.
- The dataset employed in [17] is derived from two distinct sources. The first comprises experimental burns specifically designed for the analysis of fire behavior, encompassing small-scale test fires. The second source comprises management burns executed with the primary objectives of reducing fuel hazards and enhancing habitat conditions. These datasets are characterized by 29 recorded burns that took place on flat terrain or slopes with inclines of less than 5%. These burns span four distinct fuel types, namely tall shrubland dominated by gorse (*Ulex europaeus*) in NW Portugal (comprising two fires), and low heathland featuring *Erica umbellata* and winged broom (*Chamaespartium tridentatum*) in northern and central Portugal (comprising 19 fires).

In summary, these data sources collectively contribute to a holistic understanding of emergency prediction. They enable the development of AI algorithms and predictive models that leverage diverse information streams to enhance the accuracy and effectiveness of emergency prediction systems. This section outlined the key data sources employed in the study, providing brief descriptions of each dataset's content and relevance. Researchers can refer to the resources in Table S2 to gain a comprehensive understanding of the data used in the research on emergency prediction.

3.2. Traditional Statistics

Statistical techniques serve as the cornerstone of AI, enabling the extraction of valuable insights from data. These methods encompass a broad spectrum of approaches, including, but not limited to, Descriptive Statistical Methods, Inferential Statistics, Regression Analysis, Computational Statistical Method, Time Series Analysis, Data Mining and Machine Learning, and many others. Statistical methods provide AI systems with the capability to make informed predictions, categorize information, and uncover patterns within extensive datasets. Machine learning algorithms, for instance, heavily rely on statistical principles to acquire knowledge from data and adapt to new information. In handling uncertainty, a fundamental aspect of AI, Bayesian inference and probability theory serve as indispensable tools. Furthermore, statistical methods are invaluable for evaluating the performance of AI models:

- **Descriptive Statistical Methods:** These methods are essential for data preprocessing, exploratory data analysis (EDA), and preparing data for further analysis or machine learning tasks. Here are some common descriptive statistical methods used in AI: central tendency measures (mean, average, median, mode), variability measures (range, variance, standard deviation, interquartile range), distribution characteristics (histograms, box plots, probability density functions), frequency analysis (frequency tables, charts), correlation/covariance, data visualization, summary statistics, and data preprocessing (data cleaning, scaling, normalization).
- **Inferential Statistics:** It is used to gain understanding and draw conclusions from data samples or entire populations. In the realm of machine learning and AI, this process is crucial for evaluating hypotheses, measuring model effectiveness, exploring how variables interact, and estimating uncertainty. It enables AI experts to make data-driven decisions and evaluate the reliability of their results, especially in cases involving large datasets and intricate algorithms. These techniques help determine statistical significance, set confidence intervals, and guide decisions in areas like choosing models, estimating parameters, and developing predictive models.
- **Regression Analysis:** Modeling the relationship between a dependent variable and one or more independent variables is done with this method. It aids in comprehending how, when one of the independent variables is changed while the other independent variables are kept constant, the dependent variable's typical value shifts. Logistic regression and linear regression are two popular forms of regression analysis.
- **Computational Statistical Methods:** These approaches include statistical data analysis through the application of computational techniques and algorithms. It involves data analysis using specialized software, re-sampling techniques, and simulations. This method works especially well when handling big datasets and intricate models.
- **Time Series Analysis:** In order to derive useful statistics and features from time-ordered sequence data, this statistical technique analyzes the data. In domains such as signal processing, meteorology, and economics, time series analysis is widely employed. It covers methods for handling trends, cyclical patterns, and seasonal fluctuations in the data.
- **Data Mining and Machine Learning:** Large datasets are analyzed through data mining in order to find important information, such as unknown correlations and hidden patterns. As a subset of data mining, machine learning employs algorithms to examine data, draw conclusions from the data, and make wise decision in light of the newfound knowledge. In predictive analytic, both machine learning and data mining are essential.

This empirical study delved into 77 articles, employing a diverse range of statistical methods as shown in Table S3. As indicated in the table's last column, "Others," a variety of techniques beyond those previously mentioned were employed. This encompasses approaches such as Bayesian statistics, multivariate analysis, survival analysis, and non-parametric statistics. The selection of these methods was carefully tailored to the specific research questions being explored and the nature of the data. After analyzing research in the context of our study, it's evident that most authors favored Data Mining and Machine Learning method. This choice is likely due to several factors:

- Data analysis: it entails the management and examination of large datasets sourced from diverse outlets such as emergency department, employing methodologies like data mining and machine learning to extract meaningfully insights.
- Pattern recognition: machine learning identifies complex patterns and trends, crucial for anticipating emergencies and optimizing resource management.
- Prediction: machine learning predicts future trends, such as the probability of fires in certain areas, peak hours for emergency calls, or the types of patients most likely to be admitted to hospitals. This allows responders to adequately prepare and optimize their intervention strategies.
- Decision-Making: data analysis guides informed decisions, optimizing available resources, whether in terms of deploying firefighters, managing patients in hospitals, or allocating medical resources in emergencies.

While on the other side, it has been found that there are few authors who have opted for the time series analysis method, and this choice is potentially due to several reasons:

- Nature of the data: data in these domains can be more dynamic and multidimensional, requiring more complex approaches than those offered by time series analysis.
- Model complexity: time series models may not effectively capture the complexity of observed phenomena, such as sudden variations or unexpected events in incident management.
- Adaptability: time series analysis methods may not be easily adaptable to the specific needs of emergency interventions or hospital management, requiring more flexible and personalized approaches.
- Priority: researchers may prioritize other methods more suited to their research goals or the specific characteristics of their data, such as machine learning for prediction or spatial analysis for incident mapping.
- Finally, it turned out that very few authors opted for computational statistics and inferential statistics methods, and this choice is potentially due to several reasons:
- Data characteristics: the nature of the data in these domains may not necessitate the application of computational or inferential statistics, as other methods such as data mining and machine learning might be more suited for managing large and diverse datasets.
- Priority: researchers may prioritize more pragmatic and operational approaches, such as using real-time data or predictive modeling, to address the immediate needs of emergency interventions.
- Complexity and resources: computational and inferential statistics can be complex and resource-intensive, requiring specialized knowledge, software, and computational power. Researchers may opt for simpler or more accessible methods to achieve their goals.

In summary, the integration of Data Mining and Machine Learning is essential for dissecting complex data and improving decision-making in domains requiring prompt interventions. While time series analysis, computational statistics, and inferential statistics offer valuable tools, their limited use in such contexts is often due to data complexity, urgency of situations, and specific operational needs.

3.3. Artificial Intelligence

Artificial Intelligence (AI) refers to the development of computer systems and software that can perform tasks that typically require human intelligence. These tasks include problem-solving, learning from experience, understanding natural language, recognizing patterns, and making decisions. AI systems are designed to mimic human cognitive functions, such as reasoning, problem-solving, perception, and language understanding [18]. AI encompasses a wide range of techniques and approaches, including rule-based systems, expert systems, machine learning, natural language processing, computer vision, and more. It is used in various domains, from healthcare and finance to autonomous vehicles and robotics [19]. Reviewed in this paper are multiple articles that have utilized deep learning and machine learning approaches such as: ANNs (Artificial Neural Networks), FFNNs (FeedForward Neural Networks), CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), LSTM (Long Short-Term Memory), or DNNs (Deep Neural Networks). These techniques are part of the broader field of deep learning and are known for their ability to model complex patterns and relationships in data, making them particularly useful for tasks like image recognition, natural language processing, and sequential data analysis:

- Artificial Neural Networks: ANNs are computational models inspired by the human brain. They consist of interconnected nodes or neurons organized in layers, including an input layer, hidden layers, and an output layer. ANNs are used for various tasks, including classification, regression, and pattern recognition [20].
- Feedforward Neural Networks: FFNN is a type of ANN where data flows in one direction, from the input layer through the hidden layers to the output layer. They are commonly used for tasks like image classification and text analysis.
- Convolutional Neural Networks: CNNs are designed for processing grid-like data, such as images. They use convolutional layers to automatically detect features within the data, making them highly effective for image recognition tasks [21].

- **Recurrent Neural Networks:** RNNs are specialized for sequential data, such as time series or natural language. They have feedback connections, allowing them to maintain a memory of previous inputs, making them suitable for tasks like text generation and speech recognition.
- **Long Short-Term Memory:** LSTMs are a type of RNN that addresses the vanishing gradient problem and can capture long-term dependencies in data. They are widely used in natural language processing, speech recognition, and time series analysis.
- **Deep Neural Networks:** DNNs are simply ANNs with many hidden layers. "Deep" refers to the number of layers, and they are used for various tasks where complex feature hierarchies are required [22].

The predominant choices among a large number of researchers, frequently adopted across the studies referenced in Table S4 and Table S5, offer valuable insights into the current trends and preferences within the research community. These choices influence future investigations and shape the trajectory of advancements in feature selection methodologies. Deep learning methods such as LSTM, CNNs, and DNN are commonly employed across a variety of studies. Time series analysis techniques, including SARIMA (Seasonal Autoregressive Integrated Moving Average), ARIMA (Autoregressive Integrated Moving Average), and exponential smoothing, are also utilized in multiple studies. Additionally, ensemble methods like Random Forest and XGB (Extreme Gradient Boosting) are widely applied. Other observed techniques include k-means clustering, NSGA-II (Non-dominated Sorting Genetic Algorithm II), logistic regression, and SVM (Support Vector Machines). Certain techniques within the realm of feature selection and predictive modeling have not received adequate attention or investigation in the research papers listed in the table. These areas may include novel methods of feature selection, emerging algorithms, or specific application domains where feature selection techniques could be further explored and refined. For instance, while k-means clustering is used in a few studies, the broader spectrum of clustering methods is not well-represented. GLMs (Generalized Linear Models), despite being a traditional statistical method, are notably absent. Only a few papers utilize survival analysis, indicating a potential area for further exploration. With the exception of Bayesian logistic regression, there's a lack of utilization of Bayesian methods across the papers. Similarly, the Naïve Bayesian Classifier, despite its simplicity and effectiveness in certain scenarios, is not widely explored in the listed studies. Critiques of the predominant choices observed in the research papers stem from various reasons, highlighting potential limitations and drawbacks. Many papers seem to focus on a single ML method without comparing it against alternative approaches, limiting the understanding of which method might be most effective for a given problem. Moreover, there is a bias towards deep learning methods, potentially neglecting simpler techniques that might be equally or more suitable for certain tasks. Heavy reliance on complex models like deep learning could raise concerns about overfitting, especially when dealing with limited datasets or noisy data. Additionally, the choice of ML methods should ideally be aligned with the specific characteristics and requirements of the problem domain, lacking clear justification in some cases. In dynamic emergency settings, advanced machine learning approaches such as transfer learning, reinforcement learning, and ensemble methods are pushing the limits of decision-making and predictive accuracy. Real-time data processing and administration is made possible by the combination of big data analytics and cloud computing, which is essential for prompt and well-informed answers. With the use of intelligent sensors and networked devices that offer real-time monitoring and early detection capabilities, the Internet of Things (IoT) is changing emergency services. Edge AI can be combined with IoT (Internet of Things) devices to monitor conditions in real-time, such as detecting hazardous conditions in fire-prone areas or providing real-time analytics during large public events to ensure public safety. [23] Innovative developments in human-computer interaction, such as augmented reality (AR) and virtual reality (VR), are revolutionizing training and operational support, while enhanced geospatial analysis through advanced Geographic Information Systems (GIS) provides better mapping and resource allocation. Predictive model integrity and sensitive data protection is guaranteed by the proliferation of cybersecurity measures. Furthermore, research into cutting-edge technologies like 5G networks and block chain holds promise for more developments in safe data storage and quick communication. Research can explore how Federated Learning can enhance collaboration between different emergency service providers (e.g., police, fire, and medical services) by securely sharing insights without sharing raw data. [24] SSL can be particularly useful for training models on diverse types of emergency data, such as video feeds from surveillance cameras, sensor data from smart buildings, or audio data from emergency calls, to improve prediction accuracy and response strategies. [25] Incorporating TinyML can enable rapid deployment of predictive analytics directly on-site, such as in disaster zones or remote areas, where immediate action is required and cloud connectivity may be limited. [26] Integrating XAI can improve the collaboration between AI systems and human operators in emergency services, ensuring that AI-driven decisions are transparent and justifiable, which is crucial for maintaining public trust and operational integrity. [27] Research could explore how automated systems, enhanced with AI and machine learning, can improve the accuracy and speed of incident detection, thereby improving overall response times and resource management. [28] AI-driven models that factor in dynamic elements like traffic, weather conditions, and real-time incident reports can be used to optimize resource distribution and improve response effectiveness. [29] as the technologies develop, it is a concern to the public trust and biases, and ethical issues as it integrates into emergency services.

In conclusion, while the prevalence of deep learning methods in the analyzed papers showcases their popularity and effectiveness in various applications, there's a need for a more balanced exploration of different ML techniques and a more critical evaluation of their suitability for specific contexts. Researchers should strive for more comparative analyses to better understand the strengths and weaknesses of different approaches.

3.4. Feature Selection

Feature selection has become a crucial component in the fields of machine learning and data analytics, acting as a key element for crafting effective and understandable models. Its importance stems from its ability to carefully select and isolate a relevant subset of features from a large dataset. The advantages of narrowing down the feature set are manifold, including:

- **Simplified Models:** Feature selection simplifies models by trimming down the number of features, making them more intelligible and interpretable. This simplification aids in comprehending the complex interactions between features and the outcome, offering insights into the processes involved.
- **Enhanced Model Performance:** It plays a critical role in improving the performance of machine learning models. By removing unnecessary, redundant, or noisy features, it reduces the risk of overfitting, where the model overly relies on training data and struggles to perform well with new data.
- **Lower Computational Demands:** Feature selection cuts down on data dimensionality, thereby decreasing the computational resources needed for training and using models. This is especially beneficial for large or complex datasets, aiding in the efficient creation and application of machine learning models.
- **Improved Data Preprocessing:** As a vital part of data preprocessing, feature selection helps in identifying and eliminating problematic features, such as those with missing values or high correlations. This process ensures data quality and integrity, forming a strong base for further modeling.
- **Better Interpretability:** Reducing the number of features also improves the interpretability of machine learning models. Enhanced interpretability allows for a deeper understanding of the model's reasoning, enabling more effective examination and validation of its predictions.

Feature selection is an essential method in data analytics and machine learning, offering a host of advantages that help create robust, understandable, and effective models. It is an essential tool for data scientists and machine learning specialists due to its ability to reduce model complexity, increase performance, decrease computational requirements, improve data preprocessing, and improve interpretability. This practice uses a variety of approaches, each with unique features and methods: regardless of machine learning algorithms, the Filter Method evaluates features based on intrinsic qualities like correlation with the target variable. Fisher's Score, Chi-Square tests, correlation coefficients, and variance thresholds are examples of common methods. On the other hand, Wrapper Methods assess portions of the features according to how well a specific machine learning model performs. This method applies techniques such as Recursive Feature Elimination, Forward Selection, and Backward Elimination to feature selection, treating it as a search problem.

Using embedded methods, the feature selection procedure is integrated into the model training stage. These techniques, which are specific to particular learning algorithms, include decision trees, regression models such as LASSO and RIDGE, and methods involving penalization functions. Conversely, by reducing the dimensions of the feature space, dimensional reduction techniques seek to reduce its size. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and t-SNE are notable techniques in this category.

Feature Learning involves algorithms that autonomously derive necessary representations from raw data for feature detection or classification. This approach is typically executed using deep learning models like Autoencoders, Convolutional Neural Networks, and other neural network frameworks. Moreover, machine Learning-Based Feature Selection chooses a subset of features for model training after evaluating each feature's importance. Methods in this category frequently make use of ensemble techniques that yield feature importance scores, such as Random Forest, XGBoost, and others. Each category signifies a distinct strategy for identifying the most pertinent features for model training, ranging from straightforward statistical methods to sophisticated machine-learning-based techniques. Table S6 shows the various methods of feature selection across the articles reviewed. The researchers can get critical information about the main practices and trends of feature selection in current research. This table might be also beneficial to researchers who focus on the identification of the most or less typical methods of feature selection used among different research. New procedures may supposedly be created of dealing with a problem by developing new manners and working out the ways for mitigation of arising issues, which were not covered by the traditional methods. Besides, simultaneous performance analysis of different feature selection methods among papers can unveil commonalities or disclose the most favorable strategies in certain research scenarios.

Performing the statistical analysis of the share of the feature selection techniques among 77 over-observed articles reveals many insights. First, it indicates that the Wrapper Methods are called by more than 50 percent of the articles (58.4%), thus indicating a high level of their popularity as an approach preferably used for feature selection. Supervised learning algorithms with the wrapper method that repeat the process of searching for the best feature subset until their performance is optimal is clearly an example of the researchers' recognition of the effectiveness of the method. The filter methods are also popular; 19 of 77 papers (24.7%) have used them to remove noise and for the ease of understanding which is important. Another technique cited in the review articles (13%) are the embedded techniques which is mostly used together with model training for feature selection. Yet, only 10.4% of articles apply Feature Selection based on Machine Learning and 5.2% on techniques for dimensionality reduction, which discloses certain spots to increase coverage. Surprisingly, only one article (which make up only 0.3% of overall papers) employs Feature Learning Methods, which implies that this technology hasn't been strongly implemented yet. Finally, it must be mentioned that close to a

quarter of the papers use a mixture of various techniques of the feature selection methods, so highlighting the fact that the hybrid methods are flexible and versatile. The fact here is, related to 25% of the articles there is a nonexistence of feature selection in the research, which suggest either the default method for setting up the experimental condition, or study is focused on other aspects. These empirical findings as well as the rich and diverse landscape of feature-selection techniques thus demonstrate the popular nature of feature-selection procedures in the field, offering researchers with the required insights for guiding future individual choices or studies.

3.5. Evaluation Method

Evaluation metrics play a crucial role in assessing the performance, effectiveness, and quality of Artificial Intelligence (AI) models and systems. In the realm of AI, which encompasses machine learning, deep learning, natural language processing, and more, the choice of appropriate evaluation metrics is essential for understanding how well a model or system is performing its intended tasks. These metrics provide quantifiable measures of various aspects of AI models, such as accuracy, precision, recall, F1 score, mean squared error, and many others, depending on the specific task and application. The selection of the right metrics depends on the nature of the problem being solved, whether it's classification, regression, language generation, image recognition, or recommendation systems (see Table S7). Here are some definitions of these metrics:

- RMSE (Root Mean Square Error) (1): RMSE measures the average magnitude of the errors between predicted and actual values, emphasizing larger errors. It is commonly used in regression analysis [30].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1)$$

- MSE (Mean Squared Error) (2): MSE is a measure of the average squared difference between predicted and actual values, often used in regression analysis.

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \quad (2)$$

- MAE (Mean Absolute Error) (3): MAE calculates the average absolute difference between predicted and actual values, providing a more interpretable measure of error [31].

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (3)$$

- MAPE (Mean Absolute Percentage Error) (4): MAPE expresses errors as a percentage of the actual values, providing a relative measure of accuracy [32].

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (4)$$

- R (Correlation Coefficient) (5): The correlation coefficient (R) measures the strength and direction of a linear relationship between two variables [33].

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5)$$

- Accuracy (Classification Accuracy) (6): Accuracy measures the proportion of correctly classified instances in a classification problem [34].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

- Precision (7): Precision is the ratio of true positive predictions to the total positive predictions. It quantifies the accuracy of positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

- Recall (Sensitivity) (8): Recall, also known as sensitivity, measures the proportion of true positives among all actual positive instances. It quantifies the model's ability to capture all relevant instances.

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \quad (8)$$

- Specificity (9): Specificity measures the proportion of true negatives among all actual negative instances. It quantifies the model's ability to correctly identify negatives.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

- AUC-ROC (Area Under the Receiver Operating Characteristic Curve): AUC-ROC quantifies the overall performance of a binary classification model by measuring the area under the ROC curve [36].
- F1 Score (10): The F1 score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance in binary classification [35].

$$\text{F1 Score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \quad (10)$$

- Tsec (Time in seconds): Tsec represents a time duration in seconds, commonly used for measuring computational time or response time in various applications.
- NBINT (Number of Exact Predictions): NBINT denotes the number of Exact Predictions.
- ACC0E (Accuracy with Margin of Error Zero) (A0): ACC0E represents the accuracy score in a classification problem where predictions are considered accurate only when they match the actual values exactly, with no margin of error or discrepancy.
- ACC1E (Accuracy with Margin of Error One) (A1): ACC1E denotes the accuracy score in a classification scenario where predictions are considered accurate if they are within a margin of error of one. In other words, predictions are counted as correct if they are either an exact match or off by a single unit.
- ACC2E (Accuracy with Margin of Error Two) (A2): ACC2E signifies the accuracy score in a classification context where predictions are considered accurate if they fall within a margin of error of two. This means that predictions are counted as correct if they match the actual values exactly or differ by no more than two units.

After analyzing the chosen metrics in research conducted in this context, it is evident that most authors have opted for RMSE and MAE metrics. This choice is attributed to the nature of the interventions by firefighters, incident management, and hospital management, where the target variables are often continuous, such as response time or hospital bed occupancy. To evaluate the accuracy of these predictions, metrics such as RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) are widely employed. RMSE, sensitive to the magnitude of errors, is favored when the consequences of prediction errors are significant, as in emergency firefighter interventions. Conversely, MAE, insensitive to outliers, is useful in environments where data may be noisy or contain outliers, such as in emergency situations. These metrics provide a robust way to assess the accuracy of regression models, essential for informed decision-making in these critical contexts. However, there is a lack of usage of metrics such as AUC, precision, specificity, and F1-score in the research conducted, as these metrics may not be suitable for the data used in their studies. Authors may have used other more appropriate evaluation metrics for their specific study. For example, in some cases, it may be more important to evaluate model performance in terms of response time or cost rather than precision or specificity. In other cases, application-specific metrics may be used, such as the distance between prediction and the actual location of an incident.

4. Results

4.1. Online/Offline Predictions

AI's terms "online" and "offline" predictions relate to two different ways that predictive models can function. Online predictions, sometimes referred to as real-time or dynamic predictions, entail data processing and dynamic prediction generation in response to newly available data. This strategy is essential in situations where quick decisions must be made, involving real-time fraud detection systems, autonomous cars, and stock trading algorithms. Online prediction models are highly responsive, but they frequently come at the expense of high computational demands and the requirement for a strong infrastructure to manage data streams. These models are constantly updating and improving their predictions in light of the most recent data. On the other hand, static data sets are used for offline predictions, in which the model processes and examines data that has been gathered over time. This method is frequently applied in situations like research studies, marketing analyses, and strategic business planning where there is no need for quick decisions. Since offline models have no need to be concerned about producing results right away, they can use more sophisticated and computationally demanding algorithms. The price to be paid, though, is that in quickly evolving environments, there is a chance that the predictions will lose their significance before they are finished.

Many AI systems in use today combine aspects of predictions made offline and online. An AI system might, for example, use online predictions for quick responses to changing conditions and offline predictions for long-term planning and strategic decision-making. By combining the best features of both approaches, this hybrid strategy enables

organizations to guarantee both responsiveness and depth in their predictive capabilities.

In the empirical study outlined in this paper, the analysis extends to 77 articles examining the usage of online, offline, or both types of predictive models across various domains. Particularly Table S8, provides a comprehensive breakdown for each article, detailing the specific type of prediction methodology employed. This meticulous documentation offers valuable insights into the prevalent trends and preferences within the realm of predictive modeling research. By categorizing the methodologies utilized in each study, researchers and practitioners can gain a nuanced understanding of how different sectors leverage online and offline predictions, or a combination thereof, to address specific challenges and objectives. Such granular information not only enriches the discourse surrounding predictive modeling but also aids in the formulation of informed strategies for future research endeavors and practical implementations across diverse domains

4.2. Further Details

In order to assess prediction results, metrics are commonly employed to measure the performance and reliability of algorithms used during the training phase. Following our exposition of various metrics used in previous research, as listed in (see Table S7), it is evident that the most frequently utilized metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), MSE (Mean Squared Error), F1 Score, Accuracy, Area Under the Receiver Operating Characteristic Curve (AUROC), Recall (Sensitivity), Specificity and Precision. These metrics have been identified as the most relevant for evaluating the performance of each research study. Among these metrics, RMSE, MAE, and MSE have proven particularly significant.

- Table 1: Presents the prediction results of firefighter intervention obtained using LSTM and XGBoost models for all years (2006-2017) in the study [13]. It is clear that the most compelling results were achieved with the XGBoost model, which exhibited RMSE values of 1.61 and MAE values of 1.16. The LSTM model also demonstrated impressive performance, with an RMSE of 1.59 and an MAE of 1.13.

Table 1. Results of firefighter intervention prediction using LSTM and XGBoost

Year	LSTM					XGBoost				
	RMSE	MAE	ACC0E	ACC1E	ACC2E	RMSE	MAE	ACC0E	ACC1E	ACC2E
2006	1.60	1.13	28.28%	73.04%	90.99%	1.61	1.16	25.55%	73.27%	90.86%
2007	1.63	1.19	27.27%	70.91%	89.06%	1.66	1.20	26.19%	71.48%	88.83%
2008	1.59	1.16	26.83%	71.68%	90.28%	1.64	1.22	24.45%	69.94%	89.55%
2009	2.28	1.49	22.72%	62.28%	83.00%	2.39	1.58	21.59%	59.04%	80.36%
2010	2.32	1.49	22.17%	61.96%	81.92%	2.22	1.51	22.65%	60.82%	81.50%
2011	2.49	1.68	21.05%	57.54%	78.92%	2.55	1.69	21.07%	58.16%	78.93%
2012	2.06	1.53	21.30%	58.26%	81.11%	2.08	1.55	21.16%	58.03%	80.02%
2013	2.05	1.53	21.15%	58.81%	80.58%	2.06	1.54	20.91%	58.68%	80.22%
2014	2.04	1.52	21.26%	59.10%	81.17%	2.06	1.52	21.47%	59.37%	81.00%
2015	2.09	1.58	21.14%	56.41%	79.49%	2.09	1.56	21.51%	57.70%	79.48%
2016	2.64	1.71	18.94%	53.91%	77.51%	2.58	1.67	19.16%	55.49%	78.42%
2017	2.26	1.69	19.90%	54.63%	76.80%	2.27	1.68	20.83%	55.59%	76.94%

- Table 2: Illustrates the values of MAE, Mean Relative Error (MRE), and RMSE for a three-day traffic dataset in the study [37]. Notably, the most outstanding results of traffic Accident Risk prediction were obtained through the utilization of deep learning techniques, with an RMSE of 0.034, an MSE of 0.001, and an MAE of 0.014.

Table 2. Comparison accuracy scores with others deep learning methods on GTSRB dataset

Method	MAE	MSE	RMSE
Lasso	0.046	0.006	0.076
SVR	0.066	0.006	0.075
DTR	0.021	0.004	0.058
ARIMA	0.058	0.049	0.169
TARPMML	0.014	0.001	0.034

- Table 3: Showcases MAE and RMSE values for the years 2017-2018 in the study. It is evident that the XGBoost model exhibited the most convincing performance, with an RMSE of 2.28 and an MAE of 1.68. The Gradient Boost and AdaBoost models also delivered robust results of firefighter intervention prediction, with RMSE values of 2.29 and 2.35, and MAE values of 1.69 and 1.71, respectively.

Table 3. Results of firefighter intervention prediction using boosting techniques

Year	Technique	TSec	MAE	RMSE	A0(%)	A1(%)	A2(%)
2017	AdaBoost	0.91	1.71	2.35	20.88	55.79	75.38
	GradientB	0.16	1.69	2.29	20.55	55.66	76.85
	XGBoost	0.24	1.68	2.28	20.76	56.47	76.91
2018	AdaBoost	0.91	1.85	2.60	19.20	53.04	74.02
	GradientB	0.06	1.81	2.50	18.78	52.68	74.98
	XGBoost	0.24	1.80	2.50	19.02	53.08	75.34

- Table 4: Presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for the year 2018, broken down by region, for the results of firefighter intervention prediction in the context of the study [38]. The data in this table were generated using both original and anonymized datasets, with comparisons between normalized and non-normalized data. Notably, the analysis reveals that the most favorable outcome was achieved with the XGBoost model, yielding an RMSE of 2.68 and an MAE of 1.91.

Table 4. Results of firefighter intervention prediction using XGBoost

Model	Normalized ratio		Non-normalized ratio	
	MAE	RMSE	MAE	RMSE
Baseline (mean)	-		2.5556	3.3237
Original	-		1.8552	2.5821
$f = 0.20, \epsilon_{\infty} = 4.39$	1.8666	2.5963	2.1748	2.8822
$f = 0.40, \epsilon_{\infty} = 2.77$	1.9271	2.7194	2.7436	3.6736
$f = 0.60, \epsilon_{\infty} = 1.69$	1.9151	2.6848	4.2475	4.9567
$f = 0.80, \epsilon_{\infty} = 0.81$	1.9403	2.7002	7.8542	8.4985

- Table 5: Illustrates the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for the year 2017, specifically pertaining to the LSTM model's predictive performance. The model's task involved forecasting the number of interventions for both the subsequent one and three hours, utilizing various time steps of 3, 12, and 24 hours, as part of the study [39]. It becomes evident that the most favorable outcome achieved by this model was an RMSE of 2.28 and an MAE of 1.69.

Table 5. Results of firefighter intervention prediction using LSTM

Time steps (hrs)	Predicting next 1h			Predicting next 3h		
	RMSE	MAE	NBINT	RMSE	MAE	NBINT
3h	2.2864	1.6996	1820	4.4295	3.3394	314
12h	2.2901	1.6934	1835	4.2834	3.2496	302
24h	2.2828	1.7006	1813	4.2706	3.2142	328

- Table 6: Provides a visual representation of the performance evaluation of the XGBoost and MLP models for the year 2019, as outlined in the study by [40]. The primary objective of these models was to forecast incidents of public service breakdowns on a monthly basis. The results indicate that the XGBoost model achieved the most favorable outcomes, with an average RMSE of 2.56 and an average MAE of 2.02. On the other hand, the MLP model yielded an average RMSE of 2.68 and an average MAE of 2.07. These findings underscore the predictive capabilities of the XGBoost model as it outperformed the MLP model in forecasting public service breakdowns.
- Table 7: Presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for four distinct models: XGBoost, LGBM (Light Gradient Boosting Machine), MLP (Multi-Layer Perceptron), and LASSO. These models were employed to predict ambulance response times using both original and sanitized location data, as examined in the study by [41]. The results reveal that among these models, XGBoost performed the most admirably, achieving an RMSE of 5.53 and an MAE of 3.42. Similarly, LGBM demonstrated competitive performance with an RMSE of 5.54 and an MAE of 3.38. MLP achieved an RMSE of 5.59 and an MAE of 3.56, while LASSO produced an RMSE of 5.65 and an MAE of 3.47. These findings highlight the varying predictive capabilities of these models when applied to ambulance response time forecasting, with XGBoost and LGBM exhibiting slightly superior performance in comparison to MLP and LASSO.

Table 6. Results of prediction public service breakdowns by month using MLP and XGBoost

Month	Baseline		XGBoost		MLP		BRR		SVM		RF	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
January	3.4641	2.7097	2.2930	1.9032	2.5209	2.0323	2.8455	2.3548	3.5058	2.6774	2.7591	2.1290
February	3.5355	2.9286	2.0529	1.6429	2.2520	1.7857	3.2950	2.7857	2.4128	1.8214	2.2678	1.8571
March	3.6056	2.9355	2.0791	1.6774	2.2433	1.7419	2.3071	1.8387	2.5400	2.0645	2.2860	1.8065
April	3.9115	3.2333	2.6013	1.9000	2.6141	2.0333	2.9665	2.3333	3.0768	2.4000	2.8925	2.2333
Mai	4.1697	2.9354	3.6544	2.8387	3.6323	2.6774	3.8646	2.8065	3.9595	2.8387	3.9226	2.8710
June	5.1897	4.1333	3.0332	2.2667	3.3166	2.4000	4.5753	3.8000	5.0133	3.8667	3.7283	2.9000
July	4.9416	3.9677	3.4827	2.8387	4.0081	3.2258	5.0609	4.0645	5.2976	4.2581	4.5791	3.6774
August	3.8813	3.2580	2.4561	2.0968	2.4822	2.0323	2.7474	2.2581	2.6396	2.2581	2.4363	2.0645
September	3.2352	2.5333	2.5949	2.000	2.1909	1.6667	2.7447	2.2000	2.3381	1.8667	2.6394	2.1667
October	2.9182	2.3226	2.0320	1.5484	2.1627	1.5806	2.4626	1.8710	2.3280	1.8065	2.2361	1.7097
November	2.3664	1.8000	2.2583	1.8333	2.3594	1.9667	2.6708	2.2667	2.5298	2.0667	2.2876	1.8333
December	4.0040	3.2581	2.1850	1.7419	2.3827	1.8065	2.8793	2.2258	2.6881	2.0000	2.2504	1.6451
Average	3.7686	3.0013	2.5602	2.0240	2.6804	2.0791	3.2016	2.5671	3.1941	2.4937	2.8571	2.2411

Table 7. Results predict ambulance response times using MLP, LASSO, LGBM, and XGBoost

Data	Metric	XGBoost	LGBM	MLP	LASSO
Original	RMSE	5.5398	5.5427	5.5916	5.6511
	MAE	3.4286	3.3880	3.5623	3.4760
	MAPE	30.114	29.476	31.867	30.260
	R^2	0.3412	0.3405	0.3289	0.3145
$\epsilon = 0.005493$	RMSE	5.5547	5.5544	5.6401	5.6596
	MAE	3.4515	3.3915	3.5773	3.4960
	MAPE	30.432	29.628	32.307	30.571
	R^2	0.3377	0.3378	0.3172	0.3124
$\epsilon = 0.002747$	RMSE	5.5617	5.5536	5.6959	5.6636
	MAE	3.4430	3.4628	3.6357	3.4991
	MAPE	30.364	30.688	32.687	30.606
	R^2	0.3360	0.3379	0.3036	0.3115
$\epsilon = 0.001155$	RMSE	5.5788	5.5867	5.8184	5.6671
	MAE	3.4803	3.4991	3.8550	3.5094
	MAPE	31.097	31.327	35.704	30.835
	R^2	0.3319	0.3300	0.2733	0.3106
$\epsilon = 0.000866$	RMSE	5.5892	5.5885	5.8575	5.6716
	MAE	3.5033	3.4702	3.8736	3.5134
	MAPE	31.515	30.964	35.810	30.907
	R^2	0.3295	0.3296	0.2635	0.3095
$\epsilon = 0.000693$	RMSE	5.5962	5.5978	6.0463	5.6717
	MAE	3.5119	3.5087	3.9704	3.5171
	MAPE	31.638	31.543	36.122	31.007
	R^2	0.3278	0.3274	0.2153	0.3095

- Table 8: Presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for four distinct models: AR (AutoRegressive), MA (Moving Average), ARIMA (AutoRegressive Integrated Moving Average), and Prophet. These models were applied to predict the number of firefighter interventions spanning the entire duration of the study [42], encompassing the years 2006 to 2017. The results reveal that among these models, AR exhibited RMSE and MAE values of 1.95 and 1.99, respectively. MA demonstrated competitive performance with RMSE and MAE values of 1.86 and 1.34, respectively. ARIMA outperformed the others, yielding an RMSE of 1.30 and an MAE of 1.01, while Prophet produced an RMSE of 1.63 and an MAE of 1.31. These findings underscore the varying predictive capabilities of these models when applied to the dataset spanning the study period, with ARIMA emerging as the most accurate in terms of both RMSE and MAE, followed by AR, MA, and Prophet.

Table 8. Results of firefighter intervention prediction (MAE (left) and RMSE (right)) using AR, MA, ARIMA, and prophet

Year	AR		MA		ARIMA		Prophet	
2006	1.481	2.046	1.349	1.86	1.018	1.307	2.55	3.53
2007	1.601	2.064	1.429	1.924	1.376	1.822	1.95	2.97
2008	1.496	1.952	1.385	1.868	1.263	1.644	1.31	1.63
2009	2.374	3.35	1.854	2.787	1.414	1.904	3.25	5.69
2010	2.161	3.058	1.847	2.716	1.169	2.154	2.00	2.23
2011	2.574	3.676	2.09	2.922	1.699	2.247	6.00	11.00
2012	1.99	2.5	1.84	2.415	1.642	2.031	2.44	2.87
2013	1.972	2.478	1.83	2.392	1.682	2.14	2.33	2.66
2014	2.04	2.545	1.874	2.451	1.504	1.843	2.44	2.90
2015	2.145	2.678	1.939	2.525	1.829	2.43	2.73	3.15
2016	2.223	3.137	2.026	2.807	1.898	2.31	2.45	2.99
2017	2.359	2.917	2.111	2.738	1.941	2.544	2.86	3.49
2018	2.552	3.216	2.24	2.929	2.075	2.742	1.85	2.60

- Table 9: Presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values for two models: XGBoost and LightGBM. These models were applied to predict the number of firefighter interventions before the merging of datasets and after feature selection, as part of the study [43]. The results indicate that among these models, XGBoost achieved an RMSE of 1.31 and an MAE of 0.11, while LightGBM demonstrated competitive performance with an RMSE of 1.23 and an MAE of 0.20. These findings highlight the varying predictive capabilities of XGBoost and LightGBM when applied to firefighter intervention prediction, with LightGBM exhibiting slightly superior performance in terms of both RMSE and MAE compared to XGBoost.

Table 9. Results of firefighter intervention prediction using XGBoost and LGBM

2*Dataset	Before merging the datasets				After Feature Selection			
	XGBoost		LightGBM		XGBoost		LightGBM	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Childbirth	0.479	1.523	0.632	1.452	0.434	1.611	0.403	1.2638
Drown	0.118	2.453	0.327	2.327	0.223	3.244	0.206	2.68
Wasp	0.322	1.313	0.471	1.332	0.417	1.851	0.403	1.274
Brawl	0.865	2.221	1.263	2.437	1.398	2.915	1.452	2.725
Fire on public road	1.577	3.688	2.049	4.042	1.509	4.911	0.926	3.186
Suicide	1.491	3.464	2.465	4.367	2.302	4.708	1.096	3.469
Water-flood	1.331	3.731	1.445	3.487	1.589	4.487	1.085	3.235
Public road accident	5.503	13.307	6.721	11.37	6.382	11.719	3.505	8.893
Traffic accident	5.406	10.689	6.348	11.029	6.567	11.806	3.332	9.423
Witness	3.967	6.402	4.775	6.932	4.292	6.76	2.258	5.64
Heating	3.247	10.546	3.507	9.485	4.14	16.908	4.768	17.234
Fire	4.435	14.07	4.848	11.374	4.93	22.406	5.036	28.863
Help for people	5.973	9.812	6.934	10.177	5.792	9.828	3.225	8.144
Emergency aid to people	6.914	11.85	8.772	12.846	8.087	13.43	4.227	10.864

Table 10. Results of the need for critical care prediction using deep learning-based models

Predictive model	AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	F1_score [95% CI]
AI + ESI	0.923 [0.920-0.926]	0.799 [0.795-0.803]	0.857 [0.854-0.859]	0.439 [0.433-0.445]	0.968 [0.966-0.969]	0.567 [0.562-0.571]
AI + KTAS	0.909 [0.906-0.912]	0.799 [0.795-0.803]	0.859 [0.856-0.862]	0.442 [0.436-0.449]	0.968 [0.906-0.912]	0.569 [0.565-0.575]
AI only	0.867 [0.864-0.871]	0.799 [0.795-0.803]	0.768 [0.764-0.772]	0.324 [0.318-0.331]	0.965 [0.963-0.968]	0.461 [0.454-0.467]
ESI only	0.839 [0.831-0.846]	0.357 [0.348-0.365]	0.991 [0.988-0.994]	0.851 [0.843-0.859]	0.917 [0.913-0.921]	0.503 [0.492-0.514]
KTAS only	0.824 [0.815-0.832]	0.376 [0.368-0.384]	0.971 [0.966-0.975]	0.642 [0.631-0.653]	0.918 [0.912-0.922]	0.474 [0.462-0.486]
NEWS (5)	0.741 [0.734-0.748]	0.310 [0.299-0.322]	0.976 [0.971-0.980]	0.647 [0.635-0.659]	0.910 [0.902-0.918]	0.419 [0.407-0.431]
NEWS (3)	0.696 [0.691-0.699]	0.288 [0.275-0.301]	0.936 [0.929-0.942]	0.387 [0.375-0.397]	0.904 [0.898-0.910]	0.330 [0.316-0.343]

- Table 10: presents the values of AUC (Area Under the Curve), F1 Score, Sensitivity, and Specificity for deep learning-based models. These models were employed to predict the need for critical care in prehospital emergency medical services, as examined in the study [14]. The results reveal that these models achieved notable performance, with an AUC of 0.92, an F1 Score of 0.56, a Sensitivity of 0.79, and a Specificity of 0.99. These findings underscore the efficacy of the deep learning-based models.
- Table 11: presents the Area Under the Curve (AUC) values for the Random Forest and Support Vector Machine (SVM) models. These models were employed to predict fire risk in the context of the study [15]. The results indicate that among these models, SVM achieved an AUC of 0.80, while Random Forest outperformed the others with an AUC of 0.82. These findings highlight the varying predictive capabilities of these models in assessing fire risk, with Random Forest exhibiting slightly superior performance compared to SVM.

Table 11. Results of the fire risk prediction using svm and random forest

Training window	Testing AUC of the following year	
	Random Forest	SVM
2011-2012	0.7624	0.7614
2011-2013	0.8030	0.7914
2011-2014	0.8246	0.8079

- Table 12: displays the values of F1 Score, Accuracy, Precision, and Recall for models based on Natural Language Processing (NLP) techniques. These models were applied to predict the peak in firefighter interventions resulting from rare incidents, as investigated in the study. The results reveal that these models achieved notable performance, with an F1 Score of 0.89, an Accuracy of 0.65, a Precision of 0.87, and a Recall of 0.90. These findings underscore the effectiveness of NLP-based models in predicting peaks in firefighter interventions stemming from infrequent incidents.

Table 12. Results of the peak in firefighter interventions prediction using nlp techniques

Technique	Input	F1_score	Accuracy	Balanced accuracy	Precision	Recall
CNN	Bulletin text	0.84	0.56	0.78	0.83	0.85
LSTM	-	0.85	0.56	0.80	0.84	0.86
FlauBERT	-	0.87	0.59	0.82	0.86	0.89
CamemBERT	-	0.89	0.65	0.84	0.87	0.90

- Table 13: presents the F1 Score values for models based on the deep learning approach using Convolutional Neural Networks (CNN). These models were applied to predict the severity of traffic accidents in the context of the study [21]. The results indicate that these models achieved a remarkable F1 Score of 0.87. This finding highlights the effectiveness of the deep learning-based approach, specifically utilizing CNN, in accurately predicting the severity of traffic accidents.

Table 13. Results of traffic accident risk prediction using deep learning techniques

Depth	Structures of TASP-CNN	Micro_F1score
Depth-4	conv2d → flatten → dense → dense	0.86
Depth-5	conv2d → conv2d → flatten → dense → dense	0.87
Depth-6	conv2d → conv2d → conv2d → flatten → dense → dense	0.78
Depth-7	conv2d → conv2d → conv2d → conv2d → flatten → dense → dense	0.76

5. Discussion

The empirical analysis of the 77 articles reveals several key insights into the performance and applicability of various machine learning models across different tasks in emergency services, fire risk management, and related domains. A notable trend is the widespread use of a diverse array of machine learning models in the reviewed articles, including but not limited to k-Nearest Neighbors (k-NN), logistic regression, support vector machines (SVM), hazard-based models, decision trees, neural networks, and ensemble methods such as Random Forests and Gradient Boosting Machines (GBM). Additionally, more sophisticated approaches like Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and reinforcement learning are also frequently employed. These models are typically evaluated using a consistent set of metrics, including accuracy, precision, recall, F1-score, AUC-ROC, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), providing a solid basis for comparative analysis.

The comparison of these models provides valuable insights into their relative performance across different tasks. Models that achieve higher accuracy or lower RMSE in predicting emergency scenarios are identified as particularly effective in terms of predictive capability. In cases where results align closely across studies, the emphasis shifts to discussing additional benefits, such as enhanced computational efficiency, broader applicability across multiple domains, or the use of innovative methodologies that might offer practical advantages. Furthermore, the study highlights the varied methodological approaches across the reviewed articles. While some studies utilize advanced techniques like reinforcement learning or ensemble methods to optimize predictions, this analysis also underscores the effectiveness of alternative or novel approaches, particularly when they yield comparable or superior results. The contextual application of findings is explored, demonstrating how different models can be adapted for various emergency response scenarios or public safety applications, thereby broadening the utility of the research.

5.1. Security and Privacy

At times, it becomes imperative to secure sensitive personal data or safeguard the privacy of the population before embarking on predictive tasks. Consequently, numerous research endeavors have been undertaken to effectively address this challenge (see Table S9). Several key studies in this domain are noteworthy:

- In Study [38], the authors encountered the challenge of safeguarding sensitive data, particularly the location of firefighter interventions, which demands the utmost care in handling. To mitigate the risk of data leaks, they opted to anonymize this information using the Differential Privacy (DP) technique, a constructionally secure method (refer to Fig. 1). This research primarily focused on predicting the number of firefighter interventions in specific localities while adhering to the core concept of DP. A localized approach to differential privacy was employed to anonymize location data.

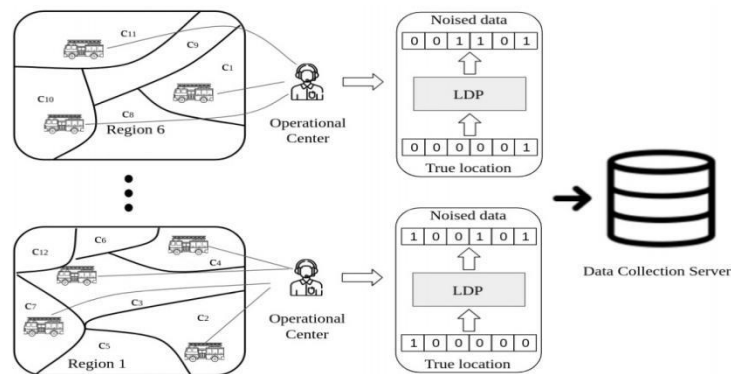


Fig.1. The DP approach applied

- In the study [41], the authors build upon the same concept as in the previous study. Here, they successfully predict ambulance response times while preserving the precise location of the emergency incident, be it in an urban or rural area. This is particularly crucial as such location data can be sensitive, potentially revealing who received medical attention and for what reason. To address this concern, the authors introduce the "input perturbation" parameter within the machine learning framework to ensure data privacy. In this study, the concept of "geo-indistinguishability" is applied to sanitize each emergency location data point, representing an advanced formal notion rooted in differential privacy.

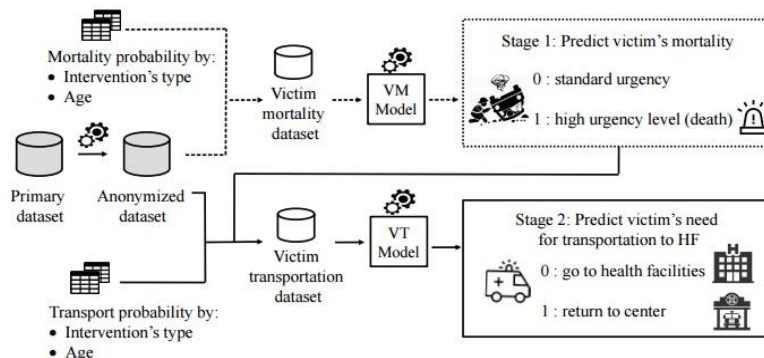


Fig.2. The DP and k-anonymity approach applied

- In Study [44], the authors introduced a novel methodology grounded in machine learning (ML) techniques. This methodology serves the dual purpose of predicting both the mortality of victims and their requirements for transportation to healthcare facilities. Furthermore, they observed that it remained feasible to predict these two sensitive targets by applying the k-anonymity and differential privacy techniques to perturb the input data. (Refer to Figure 2) for an illustration of these techniques).

5.2. Quality of the predictions

Various algorithms have been identified for predicting the central subject of "A survey on emergency prediction" through several recent research endeavors. However, certain algorithms have demonstrated superior performance and reliability in prediction tasks. Presented below is a table summarizing the top-performing algorithms based on the metrics employed:

Table 14. Algorithm performance

Type of Prediction	Algorithm	Result
Traffic Accident Risk	LSTM	RMSE: 0.034, MAE: 0.014
Critical Care Need	Deep Learning	AUC: 0.92, Sensit: 0.79, Specif: 0.99
Firemen Interventions	LSTM	RMSE: 0.015, MAE: 0.011
Fire Risk	Random Forest	AUC: 0.82
Victim Mortality	XGBoost	AUC: 0.96, F1 Score: 0.75, Recall: 0.85
Peak in Firefighter Intervention	NLP	F1 Score: 0.89, Precision: 0.87, Recall: 0.90

5.3. Limitations of Approaches

While the transformative potential of AI in emergency services is undeniable, the approaches reviewed in this article are not without their limitations. Many AI methodologies, despite their sophistication, can struggle with real-time adaptability due to the dynamic nature of emergencies, where situations can evolve rapidly and unpredictably. Additionally, AI models heavily reliant on vast datasets may face challenges related to data quality, availability, and bias. The dependence on historical data for training can lead to models that may not generalize well to novel or rare emergency scenarios. Moreover, the implementation of AI in high-stakes environments like emergency services raises significant ethical and practical concerns, including the transparency of decision-making processes, the accountability of AI-driven actions, and the potential for reduced human oversight. These challenges must be carefully considered and addressed to fully harness the benefits of AI in critical domains.

6. Conclusions

This extensive review of various research articles has provided valuable insights into the landscape of deep learning and machine learning techniques employed across diverse emergency domains. These articles collectively revealed a multitude of primary and secondary goals, each tailored to address specific challenges and opportunities within their respective fields. This review, therefore, serves as a critical resource, laying the foundation for developing more robust, efficient, and novel AI systems that can be applied across various domains, particularly in critical fields such as emergency services. The prevalence of deep learning and machine learning techniques throughout the reviewed articles not only showcases the adaptability and effectiveness of these methods in solving complex problems but also highlights their transformative potential across multiple domains. Researchers have harnessed the power of neural networks, support vector machines, decision trees, and various other algorithms to accomplish their objectives. Their innovative applications, ranging from natural language processing to image recognition, underscore the versatility of these technologies, further emphasizing their potential to revolutionize emergency response systems. Additionally, the pivotal role that databases played in this research, acting as the backbone for data acquisition and model training, cannot be overstated. The exploration of diverse data sources, both structured and unstructured, highlights the crucial importance of data quality and diversity in achieving robust results, ultimately enhancing the reliability and applicability of AI-driven solutions.

Moreover, this study provides a comprehensive understanding of how AI, particularly deep learning and machine learning, is being leveraged across different domains to address both common and unique challenges. The rigorous application of statistical methods and metrics to evaluate model performance further underscores the meticulous approach taken by researchers. By employing precision, recall, F1-score, and other measures, they were able to gauge the predictive accuracy and generalization capability of their models, thereby providing valuable insights into the strengths and limitations of different approaches. Furthermore, the systematic use of feature selection was a common practice, with researchers identifying the most influential variables to optimize model performance. Techniques such as feature engineering and dimensionality reduction were skillfully employed, enhancing both model efficiency and interpretability, which are crucial for practical applications. However, findings from this review suggest that while AI has made significant advancements in improving decision-making and operational efficiency, particularly in emergency services, there remain considerable challenges. These challenges are not limited to technical aspects, such as data quality and model interpretability, but also extend to broader ethical considerations. The differentiation between online and offline

predictions emerged as a significant consideration, with some studies focusing on real-time, online predictions to enable timely decision-making, while others concentrated on batch processing for large-scale data analysis. This distinction underscores the need for tailored AI approaches depending on the specific requirements of the emergency scenarios being addressed.

Security and privacy concerns were also pervasive throughout the reviewed articles, underscoring the growing importance of safeguarding sensitive information. The consistent need to balance predictive accuracy with data protection emerged as a recurring theme, highlighting the vital role of ethical AI practices in this domain.

While the quality of predictions varied across studies, reflecting the nuances and complexities of individual research contexts, the review has highlighted both the remarkable achievements and the ongoing challenges within the field. Addressing these challenges is essential for the continued evolution of AI technologies, ensuring that AI systems are not only powerful but also responsible, equitable, and trustworthy in their application. Ultimately, this review underscores the critical importance of continued innovation and interdisciplinary collaboration in the development of AI technologies. By addressing the challenges identified, the AI community can ensure the evolution of systems that are both effective and ethically sound, capable of making significant contributions not only to emergency services but also to broader societal needs. The insights provided here offer a valuable foundation for future research and development, guiding efforts to create AI systems that are as transformative as they are responsible.

7. List of References Utilizing ML Methods in all the Articles

Table S10 likely summarizes key information from these 77 articles related to their use of ML methods.

The references of 77 articles in the table specifically used techniques related to machine learning. ML encompasses various algorithms and approaches to allow computers to learn from data and make predictions.

Acknowledgment

This work has been achieved in the frame of the EIPHI Graduate school (contract "ANR-17-EURE-0002").

References

- [1] Ajiga David Iyanuoluwa, Ndubuisi Ndubuisi Leonard, Asuzu Onyeka Franca, et al, "AI-driven predictive analytics in retail: a review of emerging trends and customer engagement strategies", *International Journal of Management & Entrepreneurship Research*, 2024, vol. 6, no 2, p. 307-321. doi.org/10.51594/ijmer.v6i2.772.
- [2] Aafar Abbas, Bibi Nabila, Naqvi Rizwan Ali, et al, "Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations", *Frontiers in Plant Science*, 2024, vol. 15, p. 1356260. doi.org/10.3389/fpls.2024.1356260.
- [3] Chatterjee Swastika, Das Soumyajit, Ganguly Karabi, et al, "Advancements in robotic surgery: innovations, challenges and future prospects", *Journal of Robotic Surgery*, 2024, vol. 18, no 1, p. 28. doi.org/10.1007/s11701-023-01801-w.
- [4] Jiménez-Luna José, Grisoni Francesca, Weskamp Nils, et al, "Artificial intelligence in drug discovery: recent advances and future perspectives", *Expert opinion on drug discovery*, 2021, vol. 16, no 9, p. 949-959. doi.org/10.1080/17460441.2021.1909567.
- [5] Michałowska Maria, Ogłowski Mariusz, "Autonomous vehicles and road safety", 17th International Conference on Transport Systems Telematics, TST 2017, Katowice–Ustroń, Poland, April 5-8, 2017, Selected Papers 17. Springer International Publishing, 2017. p. 191-202. doi.org/10.1007/978-3-319-66251-0_16.
- [6] Chong Terrence, Yu Ting, Keeling Debbie Isobel, et al, "AI-chatbots on the services frontline addressing the challenges and opportunities of agency", *Journal of Retailing and Consumer Services*, 2021, vol. 63, p. 102735. doi.org/10.1016/j.jretconser.2021.102735.
- [7] Dikshit Srishti, Atiq Areeba, Shahid Mohammad, et al, "The use of artificial intelligence to optimize the routing of vehicles and reduce traffic congestion in urban areas", *EAI Endorsed Transactions on Energy Web*, 2023, vol. 10. doi.org/10.4108/ew.4613.
- [8] Baldominos Alejandro, Blanco Iván, Moreno Antonio José, et al, "Identifying real estate opportunities using machine learning", *Applied sciences*, 2018, vol. 8, no 11, p. 2321. doi.org/10.3390/app8112321.
- [9] Wang Sai, "Factors related to user perceptions of artificial intelligence (AI)-based content moderation on social media", *Computers in Human Behavior*, 2023, vol. 149, p. 107971. doi.org/10.1016/j.chb.2023.107971.
- [10] Rajan Arathy, Binu, V. P, "Enhancement and security in surveillance video system", *The 2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*. IEEE, 2016. p. 1-5. doi.org/10.1109/ICNGIS.2016.7854056.
- [11] Chintalapati Srikrishna, Pandey Shivendra Kumar, "Artificial intelligence in marketing: A systematic literature review", *International Journal of Market Research*, 2022, vol. 64, no 1, p. 38-68. doi.org/10.1177/14707853211018428.
- [12] Raji Mustafa Ayobami, Olodo Hameedat Bukola, Oke Timothy Tolulope, et al, "E-commerce and consumer behavior: A review of AI-powered personalization and market trends", *GSC Advanced Research and Reviews*, 2024, vol. 18, no 3, p. 066-077. doi.org/10.30574/gscarr.2024.18.3.0090.
- [13] Cerna Selene, Guyeux Christophe, Arcolezi Héber, et al, "A comparison of LSTM and XGBoost for predicting firemen interventions", *World Conference on Information Systems and Technologies*. Cham: Springer International Publishing, 2020. p. 424-434. doi.org/10.1007/978-3-030-45691-7_39.
- [14] Kang Da-Young, Cho Kyung-Jae, Kwon Oyeon, et al, "Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services", *Scandinavian journal of trauma, resuscitation and emergency medicine*, 2020, vol. 28, p. 1-8. doi.org/10.1186/s13049-020-0713-4.

- [15] Madaio Michael, Chen Shang-Tse, Haimson Oliver, et al, "Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. p. 185-194. doi.org/10.1145/2939672.2939682.
- [16] Gerber Matthew, "Predicting crime using Twitter and kernel density estimation", Decision Support Systems, 2014, vol. 61, p. 115-125. doi.org/10.1016/j.dss.2014.02.003.
- [17] Fernandes Paulo Martins, "Fire spread prediction in shrub fuels in Portugal", Forest ecology and management, 2001, vol. 144, no 1-3, p. 67-74. doi.org/10.1016/S0378-1127(00)00363-7.
- [18] Kühl Niklas, Schemmer Max, Goutier Marc, "Artificial intelligence and machine learning", Electronic Markets, 2022, vol. 32, no 4, p. 2235-2244. https://doi.org/10.1007/s12525-022-00598-0.
- [19] Gamal Donia, Alfonse Marco, El-Horbaty El-Sayed, et al, "Analysis of machine learning algorithms for opinion mining in different domains", Machine Learning and Knowledge Extraction, 2018, vol. 1, no 1, p. 224-234. doi.org/10.3390/make1010014.
- [20] Grekousis George, "Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis", Computers, Environment and Urban Systems, 2019, vol. 74, p. 244-256. doi.org/10.1016/j.compenvurbsys.2018.10.008.
- [21] Zheng Ming, Li Tong, Zhu Rui, et al. Traffic accident's severity prediction: A deep-learning approach-based CNN network. IEEE Access, 2019, vol. 7, p. 39897-39910. doi.org/10.1109/ACCESS.2019.2903319.
- [22] Liu Song, Liu Xiaojie, Lyu Qing, et al, "Comprehensive system based on a DNN and LSTM for predicting sinter composition", Applied Soft Computing, 2020, vol. 95, p. 106574. doi.org/10.1016/j.asoc.2020.106574.
- [23] Myagmar-Ochir Yanjinkham, Kim Woosong, "A survey of video surveillance systems in smart city", Electronics, 2023, vol. 12, no 17, p. 3567. doi.org/10.3390/electronics12173567.
- [24] Rahman Anchur, Hasan Kamrul, Kundu Dipanjali, et al, "On the ICN-IoT with federated learning integration of communication: Concepts, security-privacy issues, applications, and future perspectives", Future Generation Computer Systems, 2023, vol. 138, p. 61-88. doi.org/10.1016/j.future.2022.08.004.
- [25] Yan Wei Qi, "Introduction to intelligent surveillance: surveillance data capture, transmission, and analytics", Springer, 2019.
- [26] Costa Daniel, Peixoto João Paulo, Jesus Thiago, et al, "A survey of emergencies management systems in smart cities", IEEE Access, 2022, vol. 10, p. 61843-61872. doi.org/10.1109/ACCESS.2022.3180033.
- [27] Došilović Filip Karlo, Brčić Mario, Hlupić Nikica, "Explainable artificial intelligence: A survey", The 41st International convention on information and communication technology, electronics and microelectronics (MIPRO). IEEE, 2018. p. 0210-0215. doi.org/10.23919/MIPRO.2018.8400040.
- [28] Liu Chen, Chen Jiming, Liu Haoyu, et al, "Towards Efficient Traffic Incident Detection via Explicit Edge-Level Incident Modeling", IEEE Internet of Things Journal, 2024. doi.org/10.1109/JIOT.2024.3371482.
- [29] Ahmed Abdulaziz, Ashour Omar, Ali Haneen, et al, "An integrated optimization and machine learning approach to predict the admission status of emergency patients", Expert Systems with Applications, 2022, vol. 202, p. 117314. doi.org/10.1016/j.eswa.2022.117314.
- [30] Willmott Cort, MATSUURA, Kenji, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance", Climate research, 2005, vol. 30, no 1, p. 79-82. doi:10.3354/cr030079.
- [31] Hyndman Rob, Koehler Anne, "Another look at measures of forecast accuracy", International journal of forecasting, 2006, vol. 22, no 4, p. 679-688. doi.org/10.1016/j.ijforecast.2006.03.001.
- [32] Wheelwright Steven, Makridakis Spyros, Hyndman Rob, "Forecasting: methods and applications", John Wiley & Sons, 1998.
- [33] Pearson Karl, "VII. Note on regression and inheritance in the case of two parents", proceedings of the royal society of London, 1895, vol. 58, no 347-352, p. 240-242. doi.org/10.1098/rspl.1895.0041.
- [34] Sokolova Marina, Lapalme Guy, "A systematic analysis of performance measures for classification tasks", Information processing & management, 2009, vol. 45, no 4, p. 427-437. doi.org/10.1016/j.ipm.2009.03.002.
- [35] Powers David, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", arXiv preprint arXiv:2010.16061, 2020. doi.org/10.48550/arXiv.2010.16061.
- [36] Hanley James, Mcneil Barbara, "The meaning and use of the area under a receiver operating characteristic (ROC) curve", Radiology, 1982, vol. 143, no 1, p. 29-36. doi.org/10.1148/radiology.143.1.7063747.
- [37] Ren Honglei, Song You, Wang Jingwen, et al, "A deep learning approach to the citywide traffic accident risk prediction", The 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018. p. 3346-3351. doi.org/10.1109/ITSC.2018.8569437.
- [38] Arcolezi Héber, Couchot Jean-François, Cerna Selene, et al, "Forecasting the number of firefighter interventions per region with local-differential-privacy-based data", Computers & Security, 2020, vol. 96, p. 101888. doi.org/10.1016/j.cose.2020.101888.
- [39] Nahuis Selene, Arcolezi Héber, et al, "Long short-term memory for predicting firemen interventions", The 6th International Conference on Control, Decision and Information Technologies (CoDIT). Ieee, 2019. p. 1132-1137. doi.org/10.1109/CoDIT.2019.8820671.
- [40] Cerna Selene, Royer Guillaume, et al, "Predicting fire brigades operational breakdowns: A real case study", Mathematics, 2020, vol. 8, no 8, p. 1383. doi.org/10.3390/math8081383.
- [41] Arcolezi Héber, Cerna Selene, et al, "Preserving geo-indistinguishability of the emergency scene to predict ambulance response time", Mathematical and Computational Applications, 2021, vol. 26, no 3, p. 56. doi.org/10.3390/mca26030056.
- [42] Elias Mallouhy Roxane, Abou Jaoude Chady, et al, "Time series forecasting for the number of firefighters interventions", The International Conference on Advanced Information Networking and Applications. Cham: Springer International Publishing, 2021. p. 39-50. doi.org/10.1007/978-3-030-75100-5_4.
- [43] Abou Jaoude Chady, et al, "Machine learning for predicting firefighters' interventions per type of mission", The 8th International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, 2022. p. 1196-1200. doi.org/10.1109/CoDIT55151.2022.9804035.
- [44] Arcolezi, Héber, Cerna Selene, Couchot Jean-François, et al, "Privacy-preserving prediction of victim's mortality and their need for transportation to health facilities", IEEE Transactions on Industrial Informatics, 2021, vol. 18, no 8, p. 5592-5599. doi.org/10.1109/TII.2021.3123588.

Authors' Profiles



Roxane E. Mallouhy is an assistant professor with expertise in Computer science and machine Learning. With 8 years of experience in academic and professional careers, Roxane Elias Mallouhy has contributed significantly to the field through various publications in conferences and journals. She has published extensively in reputable journals and conferences. Currently, Roxane Elias Mallouhy is focused on applying her expertise in various domains of machine learning, aiming to develop her professional career and enhance her knowledge. She is excited to contribute to the scholarly discourse in machine learning and is committed to advancing the field through rigorous research and collaboration.



Sirri. Naoufal, having obtained his Master's degree in Advanced Computing and Applications from the University of Bourgogne Franche-Comté in France, pursued his studies as a PhD student in Computer Science in the same university. With a strong expertise and experience in software development and data science, both professionally and academically, he stands out with his sharp skills in these fields. Currently, he is dedicated to in-depth research as part of his doctoral thesis, focusing on predicting firefighter interventions. This innovative topic demonstrates his commitment to applying his expertise to address concrete societal challenges.



Nahvi. Irum holds a Master's degree in Computer Applications from the University of Kashmir, India. Presently, she serves as a Lecturer at Prince Mohammad Bin Fahd University, KSA, bringing with her 13 years of teaching expertise. With some publications in high-impact factor conferences and journal, her enthusiasm lies in furthering research and collaboration in the realm of machine learning.



Guyeux. Christophe research interests of Pr. Guyeux are in the areas of interdisciplinary sciences and complex systems. He applies techniques from mathematics and/or computer science to solve scientific questions raised in biology, environment, or computer science fields. Mathematical ingredients used to face complexity are discrete dynamical systems, probability and statistics, and numerical analysis, while artificial intelligence, optimization methods, graph theory, big data, and distributed computing are the computer science part of his research. Current areas of application in computer science include machine learning, information security (pseudorandom number generation, symmetric ciphers, hash functions, and information hiding), and wireless sensor networks (security issues, data survivability and aggregation, prognostic and health management based on WSN). In multidisciplinary

sciences, he contributes to solve issues raised by colleagues in biology (microbiology, phylogeny, bioinformatics and computational biology, genomics) and environment (data analysis, clustering, etc.).

How to cite this paper: Roxane Elias Mallouhy, Naoufal Sirri, Irum Nahvi, Christophe Guyeux, "AI's Current Impact and Future Potential in Emergency Services: A Comprehensive Review and Analysis", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.6, pp.1-19, 2024. DOI:10.5815/ijisa.2024.06.01