Assessing the Influence of Epidemics on Firefighter Services: A Factor Analysis Approach

Naoufal Sirri[1] [0000−0002−9900−337X] and Christophe Guyeux[1] [0000−0003−0195−4378]

FEMTO-ST Institute, UMR 6174 CNRS, University of Franche-Comté, Belfort, France, {naoufal.sirri,christophe.guyeux}@univ-fcomte.fr

Abstract. This study broadens the scope of research into various variable categories and their implications for firefighters. It is motivated by the observation that acute illnesses, while significant for health, are not typically the immediate focus of firefighting interventions. Therefore, the inquiry seeks to understand how these illnesses might impact emergency responses. The study particularly focuses on assessing the overall influence of epidemics on firefighting interventions, emphasizing their frequency. In France, for example, flu epidemics vary annually due to virus diversity, resulting in a considerable number of medical consultations, hospitalizations, and deaths within a short period according to a public health survey in 2023. This underscores the urgency and importance of the research. Over a nine-year period from 2015 to 2024, our approach involves meticulously preparing data, conducting thorough analyses, and utilizing the XGBoost predictive model, known for its efficiency and robustness. Through an iterative training process, features that enhance RMSE scores over a 24-hour horizon are selected, underscoring the vital role of epidemic-related variables across all timeframes. The key insight is that these variables impact firefighting interventions consistently and interdependently, starting with diarrhea, transitioning to flu influence by midday, and concluding with respiratory infections affecting interventions, indicating their significance for predicting outcomes throughout the duration. This nuanced understanding of the models associated with the presence or absence of acute illnesses offers practical insights for resource management and firefighting response optimization, potentially saving lives by preventing intervention failures during critical incidents. This study marks the beginning of a comprehensive exploration into variable families to better understand the factors influencing firefighting activities.

Keywords: Firefighters intervention · Epidemic impact · Feature selection · Intervention causes.

1 Introduction

The impact of epidemics on firefighting intervention frequency and nature is significant, particularly in critical areas like public health. In France, flu epidemics vary annually due to the diversity of circulating viruses, affecting different populations with varying degrees of severity. Typically, from December to March, the flu leads to 1 million consultations in primary care, over 20,000 hospitalizations, and approximately 9,000 deaths within a span of ten weeks. This seasonal flu places a considerable burden on the healthcare system and society, causing absenteeism and increased morbidity and mortality [1]. However, using epidemic surveillance data to detect specific events is complex. The spatial and temporal variability, coupled with the establishment of detection thresholds, necessitates thorough analysis. This study explores the potential correlation between acute illnesses and firefighting interventions, contributing to the development of more effective surveillance strategies.

In recent years, significant progress has been made in predicting epidemics and optimizing fire intervention strategies, thanks to advancements in artificial intelligence (AI) and its applications. These innovations have enhanced the accuracy and effectiveness of predictive models, leading to better emergency response and risk management. The study [2] developed advanced AI models, including a fine-tuned ResNet-50, to help radiologists accurately distinguish between pneumonia caused by COVID-19 and H1N1 influenza, achieving up to 96.7% accuracy. Similarly the study [3] investigated an improved genetic algorithm (IGA-BP-NN) to predict public opinion trends on influenza, thereby increasing prediction accuracy and providing effective early warning information. In another study, [4] predicted the incidence of infectious diarrhea using a stacking ensemble approach. By integrating four base models (artificial neural networks, LSTM, SVR, and XGBoost) they found that models using symptom surveillance data outperformed those using meteorological data. The ensemble model demonstrated superior performance compared to individual base models, underscoring its effectiveness in improving prediction accuracy. Additionally, [5] developed a model to predict diarrhea incidence using Nigeria's Demographic and Health Survey data and artificial neural networks, achieving high accuracy rates of 95.78% during training and 95.63% during testing. This model aids in proactive intervention strategies and enhances diarrhea prevention and control efforts, while also offering valuable insights for predicting similar future epidemics. In contrast, authors in [6] utilized explainable AI (XAI) and remote sensing data to identify key factors influencing standard mortality rates (SMR) for respiratory cancer across Italian provinces. By analyzing machine learning algorithms, they classified provinces based on SMR values and highlighted critical predictive factors, aiding in prioritizing interventions to reduce the risk of respiratory cancer. Nevertheless, there is a scarcity of studies focusing on the prediction of firefighter interventions. [7] utilized machine learning to forecast mortality and transportation needs based on emergency call data, achieving 96.44% accuracy for mortality and 73.62% for transportation. This demonstrates the potential of machine learning to enhance decision-making and resource allocation in Emergency Medical

Services (EMS). Similarly, [8] employed machine learning to predict ambulance response times while ensuring the privacy of location data, showing that data sanitization does not significantly affect prediction accuracy. Additionally, [9] applied machine learning to predict firefighter interventions during the COVID-19 pandemic using data from 2016 to 2021, highlighting a pivotal change in August 2020 due to the rise in coronavirus cases in France. Finally, in the same context, [10] examined the reliability of predictions related to firefighter interventions for optimizing resources and time. Same authors conducted three studies that analyzed the impact of specific factors on firefighter interventions, namely air quality, solar activity and river height, on firefighter interventions by [11], [12] and [13], respectively.

Our research team tackles this problem by employing a novel approach: analyzing the impact of epidemics on firefighter responses. Utilizing comprehensive historical datasets that encompass data on acute illnesses and firefighter interventions, we explore this correlation. The main goal of this study is to evaluate how acute illnesses influence the frequency of firefighter interventions. To achieve this objective, we have developed a hypothesis and prediction framework as part of our research:

- Hypothesis: A rise in acute illness cases will initially correlate with a continuous increase in firefighter interventions.
  Prediction: The occurrence of any acute illness will systematically affect health, correlating with other illnesses, and resulting in a prolonged and consistent increase in firefighter interventions.

To ensure a systematic and organized approach to our research, we meticulously developed and followed a detailed methodological plan. In Section 2, we provide an exhaustive overview of the methods and materials used in our experiments, detailing our experimental approach. Section 3 presents the research process and the results obtained. In Section 4, we conduct a thorough analysis and critical reflection on the findings, addressing the initial research question and highlighting relevant implications and interpretations. The study concludes in Section 5, where we summarize the main conclusions, emphasize significant contributions, and propose potential directions for future research. This methodological framework enabled a systematic exploration of the correlation between epidemics and firefighting interventions while maintaining a comprehensive perspective throughout our investigation.

2 Methods

2.1 Data Preparation

Data Acquisition This study leverages a comprehensive dataset obtained from the Service d'Incendie et de Secours du Doubs (SDIS 25), France. The dataset encompasses 322,437 meticulously documented interventions spanning from January 1, 2015, to May 31, 2024. Each intervention record includes an identification code, time stamps (start and end), geographical coordinates, intervention type, and response duration. To enrich the intervention data and capture the broader context of each incident, we incorporated various contextual features. These features encompass holiday periods, meteorological parameters, air quality indices, solar activity metrics, lunar phases, river water levels (relevant due to firefighter responsibilities in flood rescue), epidemiological statistics and other pertinent factors. The process is detailed as follows:

- Fire Monitoring: earth observation data from NASA's VIIRS and MODIS satellites were continuously collected, to analyze fire propagation in specific areas, as documented in [14] and [15].
- Temporal Data: temporal details encompassing time, day, day of the week, day of the year, month, year, holidays, academic breaks sourced from [16], Ramadan observances, lockdowns, and curfews were integrated into the dataset.
- Air Quality: data on particulate matter ($PM_{2.5}$, PM_{10}), ozone (O_3), and nitrogen oxides (NO_2) were obtained from nearby air quality monitoring stations, as referenced in [17].
- Epidemiological Data: weekly epidemiological data on ailments like chickenpox, influenza, and acute diarrhea was incorporated from the Sentinelles network, as documented in [18].
- River water levels: hourly water level measurements of the initial forty rivers from the closest stations within the Doubs department were obtained from the governmental service "Hydroreel". This data was incorporated into the dataset for each corresponding hourly segment, as referenced in [19].
- Socioeconomic Data: variables associated with French league and Champions League football matches were included based on their potential impact on interventions, as referenced in [20].
- Astronomical Data: the [21] libraries were employed to calculate the spatial separation between the Earth, Moon, and Sun. Additionally, Astral [22] library was used to analyze moon phases, moonrise, sunrise, and sunset times to assess their potential influence on intervention occurrences. A boolean variable indicating "night" or "day" was derived based on sunrise and sunset data.
- Solar Activity: features including 10 cm radio flux, sunspot count, sunspot area, and X-ray emissions were included to assess their potential influence on interventions, as referenced in [23].
- The initial choice for meteorological data was Météo France [24] (the French public meteorological service). However, limitations arose due to restricted access to data from remote main stations and a three-hour sampling intervals, hindering both geographical and temporal accuracy. While Météo France bulletins provided insights into diverse weather risks, these limitations necessitated the inclusion of MeteoStat [25] as a complementary source. MeteoStat offers near-future forecasts, enhancing the capabilities of Météo France data.
  - High-Resolution Station Data: three meteorological stations within the Doubs department provided a comprehensive set of atmospheric data

from Météo France. These data points included: atmospheric pressure, cloud cover, humidity, barometric trends, dew point, temperature, average wind speed and direction tracked every 10 minutes, precipitation within the last hour, precipitation over the past three hours, horizontal visibility, gust speeds over a specified interval and prevailing climatic conditions.
- Weather Advisories: weather advisories obtained from Météo France were incorporated to enrich the meteorological dataset. These advisories covered various weather hazards (wind, rainfall, floods, storms, snowfall, freezing rain, heatwaves, and extreme cold) and were categorized using a color-coded system (green, yellow, orange, red) for improved clarity [26].
- High-Resolution Gridded Data: the MeteoStat API facilitated the acquisition of additional climatic variables from publicly available datasets. These variables included temperature, dew point, precipitation, snowfall, wind parameters (speed and direction), pressure, and humidity. Notably, high-resolution temperature data was retrieved from a comprehensive 11x11 grid network encompassing the entire Doubs department.

This study employs a comprehensive variable selection strategy. While acknowledging the potential inclusion of non-significant variables, this approach prioritizes the identification of all potential factors influencing firefighter interventions. This allows for a thorough exploration of correlations between epidemiological data and a wider range of parameters encompassing diverse risks, including accidents, fires, floods, and others.

Data Pre-processing During the data processing phase, missing values in certain meteorological datasets were addressed using linear interpolation. To prepare the data for our learning model, we utilized two techniques from the Scikit-learn library [27]. First, the "StandardScaler" method was employed to standardize numerical features, including influenza, visibility, varicella statistics, humidity, dew point, acute diarrhea, year, hour, gusts, wind speed and direction, temperature, river levels, cloud cover, precipitation, and lunar distance. This technique normalizes the distribution of values to have a mean of zero and a standard deviation of one.

Additionally, we applied the "TargetEncoder" technique [28] to handle categorical attributes such as the day of the week, month, year, holidays, events and barometric trend. This method converts these categorical variables by replacing each category with the mean of the target variable. The original target values, representing the count of interventions, were retained as discrete entities to accurately reflect the distribution of interventions.

Data Mining A thorough examination of the dataset proved essential for extracting relevant information for our study. On average, approximately 30,000 interventions were recorded annually, with a notable upward trend in intervention frequency over time. Regarding the variables associated with epidemiological data, Table 1 presents the key statistical measures of incidence rates per 100,000 inhabitants for each variable. This table provides a concise summary of central tendencies, variability, and distribution patterns observed across different categories within this domain. Initially, data on four epidemics within the department were collected, specifically influenza, diarrhea, chickenpox, and respiratory infections. All these variables are represented as continuous entities, predominantly demonstrating a right-skewed distribution.

Table 1: Data analysis of epidemics variables

Variable	Mean	Std	IQR	Max	Distinct values
ira_inc	330	858.4	42.7	7444	117
grippe_inc	294	885.4	0.0	6234	112
diarrhee_inc	373	752.8	435.1	5507	168
varicelle_inc	380	773.9	456.6	5231	135

Furthermore, the examination of these variables through time series analysis reveals a modest degree of similarity between two significant variables. To support this observation, we conducted a correlation analysis between these variables, yielding results that indicate substantial correlations. For instance, the variables "grippe_inc" and "diarrhee_inc" are correlated with a value of 0.39, suggesting that an acute flu may lead to diarrhea.

To conclude our analysis, we aimed to assess the influence of epidemics on the overall target variable through a correlation analysis. The results indicate a weak and statistically insignificant impact on the overall target variable. This provides a preliminary insight into the limited role of epidemics in shaping intervention models. Although notable correlations are absent, the possibility of a more complex relationship highlights our understanding of the nuanced interaction between epidemics and incidents requiring emergency services intervention.

2.2 Feature and Model Selection

Historically, it has been common practice to incorporate all available features from the training dataset, based on the assumption that maximizing the inclusion of information would yield an optimal model. However, advocating for a constraint on the number of considered features is driven by two main reasons. First, some variables may exhibit strong interdependencies, while others may contribute minimally to predictive capability, potentially leading to reduced model generalization or the introduction of redundant information. Second, including a large number of features can significantly increase computational complexity without corresponding improvements in model performance [29]. Therefore, using a more restricted set of features holds promise for achieving more efficient results.

In this study, we employed the "feature importance" method for feature selection, assigning scores to each variable in the dataset, with higher scores indicating greater relevance [30]. A threshold was established to retain the top 400 most relevant features. Various selection techniques were applied, including:

1. High Variance: preserved features with variances exceeding 0.5.
2. Pearson and Spearman correlation coefficients: filtered correlations with the target variable that had an absolute value of 0.4 or higher.
3. Chi-Square Selector: applied the chi-square test to evaluate the association between each feature and the target variable after normalizing the features using the 'Min-Max Scaler' function.
4. Extreme Gradient Boosting (XGBoost) [31]: utilized predefined hyperparameters (number of estimators = 100000, maximum depth = 7, early stopping after 10 rounds). The model was trained, and feature importance was computed or derived.
5. Light Gradient Boosting Model (LightGBM) [32]: used specified hyperparameters (number of estimators = 100000, maximum depth = 7, number of leaves = 2^7, objective function = regression, metric = RMSE, learning rate = 0.1, early stopping after 10 rounds), and extracted feature importance as previously mentioned.

The culmination of this selection process resulted in a refined list of features, prioritizing those consistently identified by multiple techniques. From an initial pool of 3912 features, approximately 10% were retained, resulting in a subset of 400 features used for model training.

In our model selection process, we chose for Extreme Gradient Boosting (XGBoost) [31]. XGBoost is renowned for its scalability, efficiency, flexibility, and speed, capable of handling complex datasets while achieving remarkable predictive accuracy. Its effectiveness lies in its integration of gradient boosting principles, enabling iterative improvement of weak learners' performance. Widely adopted across various fields, from finance to healthcare, XGBoost demonstrates versatility in managing regression, classification, and ranking tasks with exceptional competence. Its popularity is further underscored by its ability to handle missing data, perform feature selection, and adapt to different types of data.

2.3 Methodologie Employed for the Prediction Tool

To implement our theoretical analysis in practice, we developed a comprehensive training and optimization pipeline (see Figure 1).

1. We initiated multiple training iterations, spanning durations from 1 to 24 hours, aiming to evaluate the influence of the epidemics up to 24 hours in advance.
2. Subsequently, we performed feature extraction as detailed in the "Feature Selection" section.

3. Iteratively, each variable selected in the previous phase was incorporated into the XGBoost model for training.
4. At each iteration, we divided the dataset into training, testing, and validation subsets. Initially, 20% of the data was allocated for testing, while the remaining 80% was used for training-validation. This allocation was later adjusted to 80% for training and 20% for validation.
5. We trained the XGBoost model using predefined hyperparameters (see Table 2) and implemented early stopping with "early_stopping_rounds=15". This approach halted the training process if no performance improvement was observed on the validation set for 15 consecutive rounds, considering the fixed number of 100,000 estimators in our model.
6. We assessed prediction scores using the RMSE metric (Equation 1) and compared them with previous scores. Variables demonstrating improvement were kept in the input list via a sequential feature selection method, and prior scores were adjusted accordingly.

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

7. Finally, we made a concerted effort to enhance prediction accuracy by experimenting with different hyperparameter combinations (see Table 3). This process included applying "early_stopping_rounds=25" together with a fixed set of parameters, including 100,000 estimators. By employing our iterative approach, we discovered the optimal hyperparameter configurations, identifying a set of values that significantly improved performance.

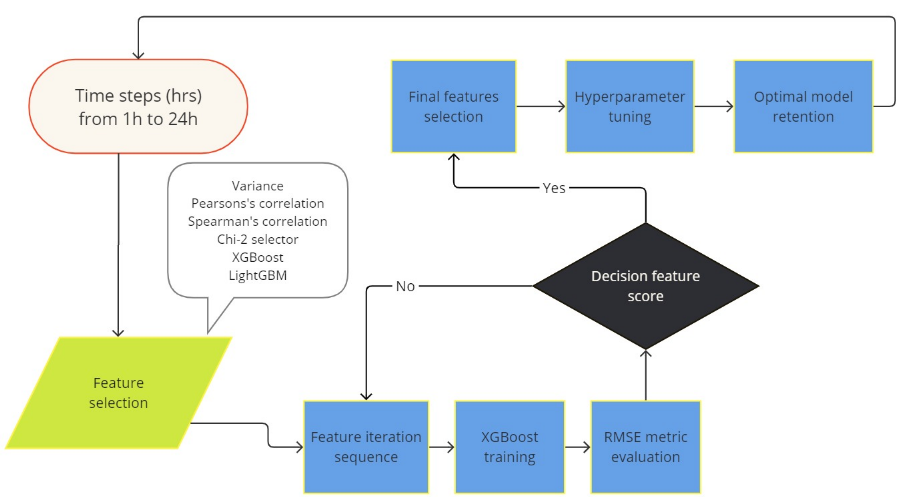


Fig. 1: Training pipeline representation model

Table 2: XGBoost default hyperparameters

Hyperparameter	Values
max_depth	7
min_child_weight	1
gamma	0
subsample	0.8
colsample_bytree	0.8
learning_rate	0.1

Table 3: Settings XGBoost hyperparameters

Hyperparameter	XGBoost
`max_depth`	[2, 14]
`min_child_weight`	[0, 14]
`gamma`	[0.0, 0.4]
`subsample`	[0.6, 0.9]
`colsample_bytree`	[0.6, 0.9]
`learning_rate`	[0.01, 0.009]

3 Results

Significant efforts have been invested in gathering, processing, and consolidating a comprehensive dataset, emphasizing the integration of information from various sources of past interventions. Particular attention was given to assessing the significance of each variable within this context. As highlighted earlier, the core objective of this study is to uncover valuable insights regarding the impact of epidemics on intervention forecasting. The previous section outlined a methodology for identifying key features. After training our model over a one-day period, we now present the outcomes of this feature selection process, focusing on the disease-related variables whose inclusion has enhanced prediction accuracy (refer to Table 4).

Table 4: Refining epidemic Analysis through Feature Selection techniques

Time horizon	Epidemics	Feature Selection Technique	Rank
5h	diarrhee_inc	Variance and XGBoost	147
7h	diarrhee_inc	Variance and LightGBM	153
11h	diarrhee_inc	Variance and LightGBM	143
15h	diarrhee_inc	Variance and LightGBM	131
	grippe_inc	Variance, XGBoost and LightGBM	144
16h	grippe_inc	Variance, XGBoost and LightGBM	183
	diarrhee_inc	Variance and LightGBM	202
17h	grippe_inc	Variance and LightGBM	155
19h	ira_inc	Variance, XGBoost and LightGBM	173

Additionally, we provide the classification outcomes generated by our model, which will be interpreted in the following section. Table 5 details the predictive performance of the XGBoost model concerning intervention counts, covering time horizons that include disease variables, with the most significant performances emphasized in bold. It also shows the RMSE metrics before and after the inclusion of epidemic variables, along with the percentage improvement. Figure 2 displays results from 300 samples aimed at predicting atypical intervention counts at the 16th hour. A noteworthy RMSE improvement of 1.24% is observed by incorporating a single acute disease variable at the

11th hour, compared to using other variables in earlier horizons. Figures 3(a) and 3(b) visually represent prediction accuracy, showing a maximum error range of 0 to 17 for the 16th hour before and after adding the epidemic variable, respectively. Since the XGBoost model produces decimal predictions (e.g., 7.25 interventions), the results were rounded to the nearest whole number (in this case, 7 interventions) for practical application.

Table 5: Emphasizing classification and outcomes:

Time horizon	**Epidemics**	**RMSE pre-sel**	**RMSE post-sel**	**Improvement**
5h	diarrhee_inc	3.4267	3.4098	0.49%
7h	diarrhee_inc	3.4886	3.4661	0.64%
11h	diarrhee_inc	3.7655	3.7192	**1.24%**
15h	diarrhee_inc	3.4597	3.4351	0.71%
	grippe_inc	3.4212	3.4048	0.48%
16h	grippe_inc	2.3940	2.3832	0.45%
	diarrhee_inc	2.3716	**2.3605**	0.47%
17h	grippe_inc	3.4842	3.4709	0.38%
19h	ira_inc	3.6242	3.6226	0.04%

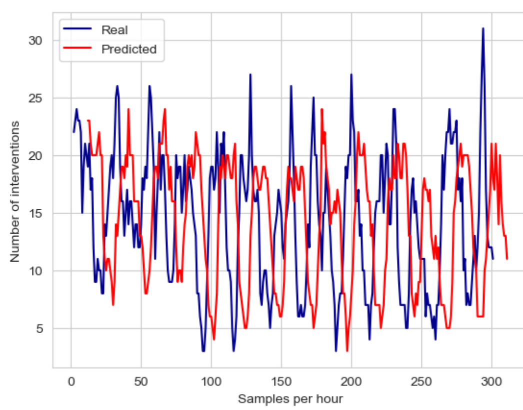
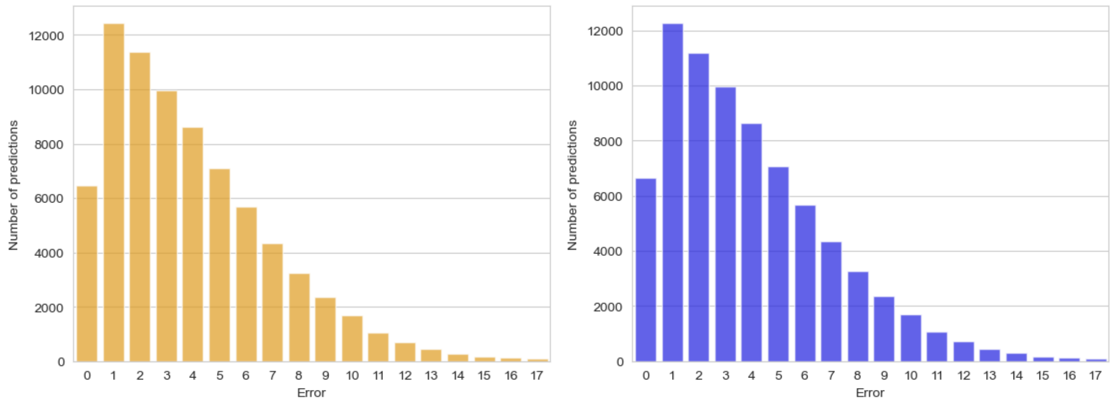


Fig. 2: Prediction for the 16th hour

Notably, XGBoost achieved its most impressive RMSE score of 2.3605 at the 16th hour, significantly influenced by incorporating two acute disease variables, specifically diarrhea and influenza. The scores consistently improved across the initial time horizons due to the selection of the single variable of diarrhea. Midway through the time horizons, the addition of both diarrhea and influenza variables consistently enhanced the scores for two hours. Subsequently, the influenza variable continued to improve the score for the remainder of the day until the selection of the last respiratory infection



(a) Prediction's error at the 16th hour pre-selection (b) Prediction's error at the 16th hour post-selection

Fig. 3: Emphasizing from predictive modeling

Table 6: Best hyperparameter for the 11th hour

Hyperparameter	**Values**
max_depth	8
min_child_weight	10
gamma	0
subsample	0.8
colsample_bytree	0.8
learning_rate	0.008

Table 7: Best hyperparameter for the 16th hour

Hyperparameter	**Values**
max_depth	6
min_child_weight	9
gamma	0.2
subsample	0.7
colsample_bytree	0.8
learning_rate	0.1

variable which in turn improved the score. Following this, the XGBoost model underwent thorough training across all horizons using a meticulous grid search procedure for hyperparameter tuning. This method systematically explores various combinations of attributes to find the optimal solution. Tables 6 and 7 detail the hyperparameters identified and used for the best RMSE score at the sixteenth hour and the highest percentage improvement at the eleventh hour, respectively, ensuring the model achieves its optimal performance.

4 Discussion

The objective of this study was to assess the influence of acute disease levels on the prediction of fire department interventions over a nine-year span from 2015 to 2024. Our methodological approach was meticulously planned, encompassing careful data preparation and thorough analysis to guide our experimental decisions. This included feature selection using various statistical techniques and machine learning methods. XGBoost was chosen as the predictive model for its robustness in handling outliers, making it particularly suitable for continuous variables related to acute diseases (see Table 1). We employed an iterative training pipeline, systematically selecting features that improved the RMSE score over a 24-hour horizon (see Figure 1). The findings, presented in Table 5, underscore the importance of epidemic-related variables, especially in the early horizons and over extended periods. Specifically, at the 11th hour, the inclusion of the "diarrhea" variable resulted in the most significant RMSE improvement of 1.24%, enhancing the score from the 5th hour up to 10 hours. Additionally, at the 16th hour, the combination of "diarrhea and influenza" variables yielded a promising RMSE score of 2.3605. This combination maintained its effectiveness for 2 hours, after which the "influenza" variable continued to improve the score independently. Towards the end of the day, the "acute respiratory infection" variable took over, further enhancing the score on its own.

This finding demonstrates a consistent and progressive influence of epidemic variables across all time periods, underscoring their enhanced value for both short-term and long-term forecasting. Our results indicate that certain epidemic-related factors significantly affect all horizons due to various temporal and atmospheric factors. Our analysis suggests a continuous association of these three diseases throughout the day. It seems that diarrhea symptoms may be more severe upon waking due to overnight dehydration, leading to urgent medical needs in the morning and triggering emergency calls. As the day progresses, the symptoms of acute illnesses like diarrhea and influenza can intensify, requiring more medical intervention. This can be attributed to the highly contagious nature of influenza, which spreads rapidly in workplaces or schools, causing acute symptoms. By the end of the day, fatigue and exhaustion might exacerbate the symptoms of acute respiratory infections, or individuals might experience a general decline in health, particularly the elderly or those with pre-existing conditions, leading to an increase in emergency calls. These findings suggest that other variables may have delayed effects or that other factors influence the behavior of epidemics affecting fire department interventions. Seasonal weather conditions such as humidity and precipitation, along with pollutants like fine particulate matter ($PM_{2.5}$, PM_{10}) and nitrogen dioxide (NO_2), can impact air quality. This could explain the immediate and prolonged impact of these variables. In conclusion, the significant potential impact of acute diseases on health should not be underestimated, as their effects can have serious consequences, consistently prompting fire department interventions.

Additionally, Figure 4 demonstrates that the correlation between variables from different categories and the epidemic characteristic is modest, particularly for non-epidemic factors such as precipitation, humidity, and pollutants, across all time horizons. This suggests an indirect causal relationship, indicating that severe weather conditions or poor air quality might indirectly impact public health. Further analysis, depicted in Figure 5, shows a systematic increase in forecast accuracy when a variable from the epidemic category is included, particularly during the cold season. This finding underscores the value of incorporating epidemic-related variables during this period to enhance prediction accuracy.

A thorough analysis, depicted in Figure 3, demonstrates a reduction in prediction errors, thereby confirming the validity of our results. Specifically, the figure shows a decrease in errors, with an increase in zero errors and a reduction in one and two errors after incorporating the two variables "diarrhea" and "flu," compared to the errors observed before their inclusion. This study provides deeper insights into how epidemics affect firefighter interventions. Nonetheless, certain limitations, such as the exclusion of other boosting models or machine learning methods for comparative purposes and the sole reliance on grid search for hyperparameter optimization, have constrained our findings. Future research will focus on exploring additional variables related to other diseases, examining the combined impact of acute diseases and meteorological

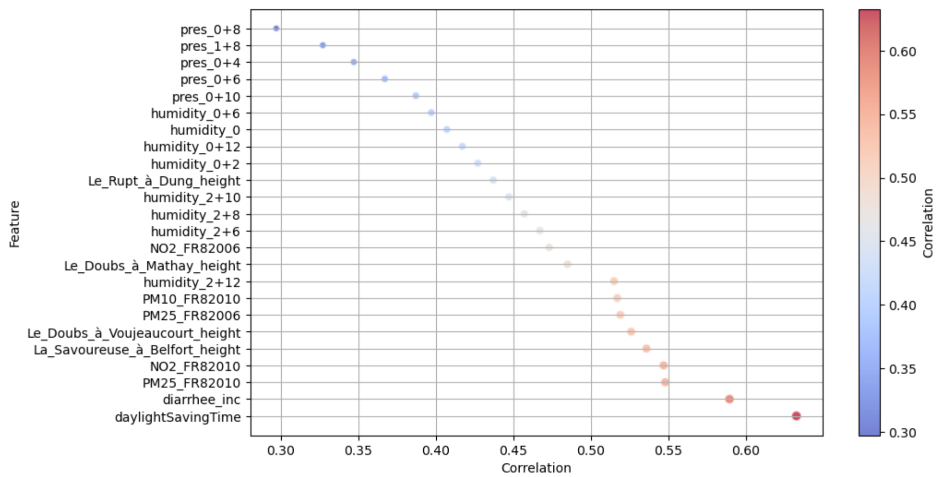


Fig. 4: Correlation with the disease "grippe_inc" at the 16th hour

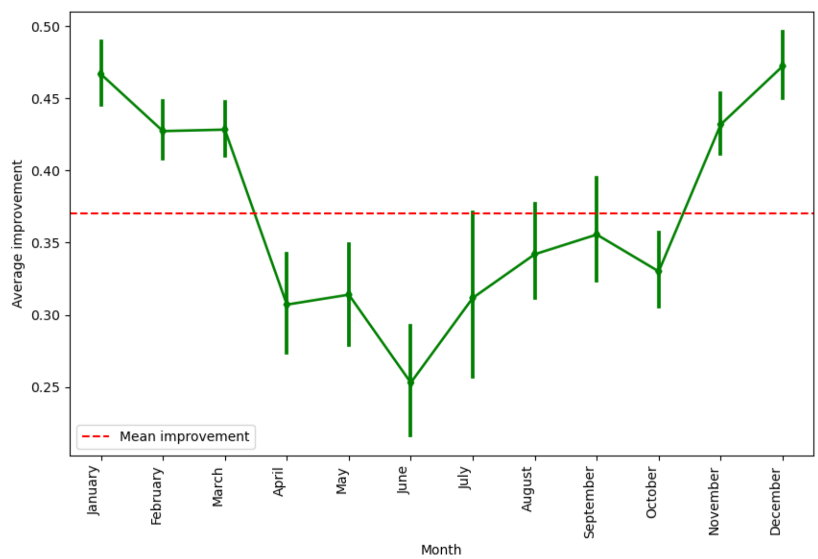


Fig. 5: Improvement statistics per month

conditions, employing alternative models, and adopting more advanced sequential optimization techniques, such as Bayesian optimization.

5 Conclusion

This study delves deeper into a series of articles focused on examining various categories of variables and their implications for fire intervention strategies. Concentrating on the prediction of future intervention numbers, which is crucial for firefighting services worldwide, this paper explores the influence of epidemics on emergency service forecasts. Utilizing a comprehensive nine-year dataset from the Fire and Rescue Service (SDIS 25) of the Doubs region in France, the research investigates how the predictive model identifies significant trends associated with acute disease variables at different times of the day. The results indicate that these variables can have a lasting impact from early time horizons over an extended period. It was noted that diseases are interconnected, with interventions starting due to the influence of acute diarrhea at the 5th hour, followed by interventions halfway through the horizons caused by the impact of both acute influenza and diarrhea, and concluding with interventions at the 19th hour due to the influence of acute respiratory infections. While this approach proves to be valuable for resource management and optimizing response times, ongoing improvements are necessary to ensure seamless integration into the decision-making processes of fire services.

Continuing our research, we will explore the subtle effects of various other variable categories (such as fire monitoring, alert bulletins, events, etc.) on forecasts over different time frames. Our aim is to derive valuable insights from these areas and pinpoint the most influential variables according to the types of interventions and the prediction timeframes. By utilizing this information, we intend to enhance our forecasting accuracy for both short-term and long-term scenarios through the careful selection of models and hyperparameters that are most pertinent to specific variables. Additionally, we will emphasize the integration of geolocated variables to refine predictions based on factors such as population density and forest area. Our primary goal is to bolster predictive capabilities, allowing for the proactive implementation of operational strategies that effectively address the demand for fire service interventions.

References

1. Public Health France. https://www.santepubliquefrance.fr/les-actualites/2023/poids-et-impact-de-la-grippe-saisonniere-en-france-metropolitaine-bilan-des-epidemies-de-2011-2012-a-2021-2022. Accessed on 15 June 2024.
2. Sotoudeh, Houman and Tabatabaei, Mohsen and Tasorian, Baharak and Tavakol, Kamran and Sotoudeh, Ehsan and Moini, Abdol Latif, *Artificial intelligence empowers radiologists to differentiate pneumonia induced by COVID-19 versus influenza viruses*, *Acta Informatica Medica*, vol. 28, pp. 190, 2020, Publisher: The Academy of Medical Sciences of Bosnia and Herzegovina.
3. Zhang, Jie and Yuan, Chao, *Analysis and Management of Flu Disease Public Opinion Based on Machine Learning*, *Journal of Medical Imaging and Health Informatics*, vol. 11, p. 1791–1797, 2021, Publisher: American Scientific Publishers.
4. Wang, Pengyu and Zhang, Wangjian and Wang, Hui and Shi, Congxing and Li, Zhiqiang and Wang, Dahu and Luo, Lei and Du, Zhicheng and Hao, Yuantao, *Predicting the incidence of infectious diarrhea with symptom surveillance data using a stacking-based ensembled model*, *BMC Infectious Diseases*, vol. 24, pp. 265, 2024, Publisher: Springer.

5. Abubakar, Ismaila Rimi and Olatunji, Sunday Olusanya, *Computational intelligence-based model for diarrhea prediction using Demographic and Health Survey data*, *Soft Computing*, vol. 24, pp. 5357–5366, 2020, Publisher: Springer.
6. Romano, Donato and Novielli, Pierfrancesco and Diacono, Domenico and Cilli, Roberto and Pantaleo, Ester and Amoroso, Nicola and Bellantuono, Loredana and Monaco, Alfonso and Bellotti, Roberto and Tangaro, Sabina, *Insights from Explainable Artificial Intelligence of Pollution and Socioeconomic Influences for Respiratory Cancer Mortality in Italy*, *Journal of Personalized Medicine*, vol. 14, pp. 430, 2024, Publisher: MDPI.
7. Arcolezi, Héber H and Cerna, Selene and Couchot, Jean-François and Guyeux, Christophe and Makhoul, Abdallah, *Privacy-preserving prediction of victim's mortality and their need for transportation to health facilities*. In: *IEEE Transactions on Industrial Informatics*, vol. 18, pp. 5592–5599, 2021, Publisher: IEEE.
8. Arcolezi, Héber H and Cerna, Selene and Guyeux, Christophe and Couchot, Jean-François, *Preserving geo-indistinguishability of the emergency scene to predict ambulance response time*. In: *Mathematical and Computational Applications*, vol. 26, no. 3, pp. 56, 2021, Publisher: MDPI.
9. Elias Mallouhy, Roxane and Guyeux, Christophe and Jaoude, Chady Abou and Makhoul, Abdallah, *Anomalies and breakpoint detection for a dataset of firefighters' operations during the COVID-19 period in France*. In: *World Conference on Information Systems and Technologies*, Springer, 2022, pp. 3–12.
10. Sirri, Guyeux, *Firefighter Intervention Predictive Modeling: Reliability Assessment*. In: *International Conference on Circuit, Systems and Communication*, IEEE , 2024. Accepted for presentation on 28 June 2024.
11. Sirri, Guyeux, *Air Quality Impact on Firefighter Interventions: Factors Analysis*. In: *The 7th International Conference on Big Data and Internet of Things*, Springer, 2024. Accepted for presentation on 17 April 2024.
12. Sirri, Guyeux, *Solar activity Impact on Firefighter Interventions: Factors Analysis*. In: *The 5th International Conference on Deep Learning and Applications*, Springer, 2024. Accepted for presentation on 11 July 2024.
13. Sirri, Guyeux, *River Levels Affecting Firefighter Interventions: Factor Analysis*. In: *The 28th International Database Engineered Applications Symposium*, Springer, 2024. Proposed for presentation on 28 August 2024.
14. NASA. https://lance.modaps.eosdis.nasa.gov/viirs/. Accessed on 15 June 2024.
15. MODIS. https://lance.modaps.eosdis.nasa.gov/modis/. Accessed on 15 June 2024.
16. Ministry of National Education. http://www.education.gouv.fr/pid25058/le-calendrier-scolaire.html. Accessed on 15 June 2024.
17. ATMO-BFC. https://www.atmo-bfc.org/accueil. Accessed on 15 June 2024.
18. The Sentinel Network. https://www.sentiweb.fr/?page=table. Accessed on 15 June 2024.
19. Ministry of Ecological Transition. http://www.hydro.eaufrance.fr/. Accessed on 15 June 2024.
20. Soccer. https://www.footendirect.com/. Accessed on 15 June 2024.
21. Skyfield. https://github.com/skyfielders/python-skyfield. Accessed on 15 June 2024.
22. Astral. https://pypi.org/project/astral/0.5/. Accessed on 15 June 2024.
23. NASA. https://www.swpc.noaa.gov/. Accessed on 15 June 2024.
24. Météo-France. https://www.ecologie.gouv.fr/. Accessed on 15 June 2024.
25. Meteo-Stat. https://pypi.org/project/meteostat/. Accessed on 15 June 2024.
26. Vigilance-France. https://vigilance.meteofrance.fr/fr. Accessed on 15 June 2024.

27. Pedregosa, Fabian and Varoquaux, Gael and Gramfort, Alexandre and Michel, Vincent and Thirion, Bertrand and Grisel, Olivier and Blondel, Mathieu and Prettenhofer, Peter and Weiss, Ron and Dubourg, Vincent and others. *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, 12, 2011, pp. 2825–2830.
28. Target Encoder. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.TargetEncoder.html. Accessed on 15 June 2024.
29. Garreta, Raúl and Moncecchi, Guillermo. *Learning scikit-learn: machine learning in Python: experience the benefits of machine learning techniques by applying them to real-world problems using Python and the open source scikit-learn library.* (No Title), 2013.
30. Zien, Alexander and Krämer, Nicole and Sonnenburg, Sören and Rätsch, Gunnar. *The feature importance ranking measure.* In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II*, pp. 694–709, 2009, Springer.
31. Chen, Tianqi and Guestrin, Carlos. *Xgboost: A scalable tree boosting system.* In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
32. Ke, Guolin and Meng, Qi and Finley, Thomas and Wang, Taifeng and Chen, Wei and Ma, Weidong and Ye, Qiwei and Liu, Tie-Yan. *Lightgbm: A highly efficient gradient boosting decision tree.* In: *Advances in neural information processing systems*, volume 30, 2017.