Privacy-Preserving Prediction of Victim's Mortality and Their Need for Transportation to Health Facilities

Héber H. Arcolezi, Selene Cerna, Jean-Francois Couchot, Christophe Guyeux, Abdallah Makhoul

Privacy-Preserving Prediction of Victim's Mortality and Their Need for Transportation to Health Facilities

Héber H. Arcolezi*, Selene Cerna*, Jean-François Couchot, Christophe Guyeux, Abdallah Makhoul

***Abstract*—Emergency medical services (EMS) provide crucial prehospital care, such as in the case of cardiac arrest, where the victim requires immediate first-aid. For this reason, it is vital to improving EMS response time. This paper proposes a novel methodology based on machine learning (ML) techniques to predict both the victims' mortality and their need for transportation to health facilities using data gathered from the start of the emergency call until the Departmental Fire and Rescue Service of the Doubs (SDIS25) is notified. We first analyzed SDIS25 calls to find out associations between the call processing times and victims' mortality, and to measure the variables' importance. Next, we validated our proposed ML-based methodology, where mortality could be predicted with accuracy and area under the receiver operating characteristic curve (AUC) scores of 96.44% and 96.04% respectively, while the need for transportation achieved an accuracy and AUC scores of 73.62% and 78.91%, respectively. What is more, we found out that it was still possible to predict both targets perturbating the input data by applying *k*-anonymity and differential privacy techniques. In conclusion, the results showed the potential of ML for EMS, which can be used as a decision-support tool to early identify mortality and the use of resources (transportation) and, thus, help EMS to save more lives and avoid service disruptions.**

***Index Terms*—Emergency medical services, Decision support system, Privacy-preserving data mining, Machine learning, *k*-anonymity, Differential privacy, Pre-hospital emergency care.**

I. Introduction

A. Background

Emergency medical services (EMS) are a key component of healthcare systems around the world. An important measurement of their quality is their response time, which is measured from the time of the call to the time an ambulance arrives at the emergency scene. In fact, shorter ambulance response times are potential contributors to higher survival rates [1], [2] since every second is a matter of life. For instance, the response time also depends on how and by whom the call is processed in the EMS center [3]. For this reason, there is a need to optimize these services and take advantage of plenty of data gathered throughout the years in hospitals and EMS.

As reviewed in recent survey works [4]–[6], decision-support systems based on machine learning (ML) techniques have been proposed for application in emergency medicine. Indeed, in the context of this paper, for EMS, there are many interests in using ML methods for tasks such as: identifying possible medical conditions before arrival on emergency departments [7], to predict ambulances' demand to allow their reallocation [1], to predict ambulance response time [8], [9], to predict the ambulances' turnaround time in hospitals [10], to predict clinical outcomes [11], to early identify clinical conditions on emergency calls [12], to recognize and predict service disruptions [13], and so on.

In this paper, we analyze the case of the Departmental Fire and Rescue Service of the Doubs (SDIS25) in France, with data from 2015 until 2020. SDIS25 has 71 centers deployed in its territory and its service is activated through some emergency phone number (e.g., 18 or 112). Next, the call is treated by an *operator*, which collects the necessary information about the emergency (e.g., *victims*, location) and notifies the center(s) to deal with the intervention. Once the adapted armament is ready, the ambulance departs to the emergency scene and upon arrival, the crew provides first-aid treatment to the victim(s). After that, depending on the victim's status, the victim(s) is transported to some health facility (HF), e.g., hospital, private clinics, and so on. Lastly, the ambulance and its crew return to the SDIS25 center and are available again to attend other interventions.

Fig. 1 exemplifies the aforementioned workflow and the interval of interest we considered to predict whether the victim will die and if the ambulance will need to transport (or not) the victim to HFs. That is, we focus on the period comprising the time where the EMS call center's phone starts to ring until some center(s) is notified to handle the intervention or the call ends. Without loss of generality, 'Alert' can be either before or after the call ends. In this period, information about the call is recorded automatically together with the identity of the operator that treated the call. Besides, information about the intervention is manually recorded by the operator in a computerized system (e.g., location, observations). Lastly, the victim's information can be acquired during the call or not (e.g., unidentified victims in traffic accidents).

*These are co-first authors that contributed equally to this work.

This work was supported by the EIPHI-BFC Graduate School (contract "ANR-17-EURE-0002"), by the Region of Bourgogne Franche-Comté CADRAN Project, and by SDIS du Doubs, with the support of the French Ministry of Higher Education and Research (managed by the National Association of Research and Technology (ANRT) for the CIFRE thesis (N 2019/0372). The authors would also like to thank SDIS 25 Commander Guillaume Royer-Fey and Captain Céline Chevallier for their great collaborations and continuous feedback. All computations have been performed on the "Mésocentre de Calcul de Franche-Comté".

All authors are with Femto-ST Institute, University Bourgogne Franche-Comté, UBFC, CNRS, Belfort, France e-mail: {heber.hwang_arcolezi, selene_leya.cerna_nahuis, jean-francois.couchot, christophe.guyeux, abdallah.makhoul}@univ-fcomte.fr.

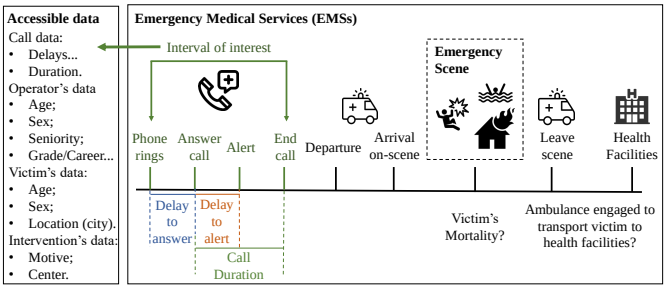

Fig. 1: General workflow of SDIS25 when dealing with an emergency. We consider as "interval of interest" the time between the phone starts to ring until some center(s) is notified (i.e., 'Alerted') or the call ends. With all data accessible within this interval, the objectives are to predict victim's mortality and their need for transportation to health facilities.

B. Context of the problem

However, the collection of big data in real life, in particular medical data, leads to a major problem: the disclosure of personal and sensitive information [14]. More precisely, in our problem, there are two entities we are concerned with, namely, call center *operators* and *victims* with regard to privacy. Indeed, data breaches are all too common [15], which endanger users' privacy and can lead to substantial losses for companies under the General Data Protection Regulation [16]. In addition, ML models trained with raw data can also indirectly reveal sensitive information. For instance, in [17] the authors evaluate how some models can memorize sensitive information from the training data, and in [18], the authors investigate how ML models are susceptible to membership inference attacks. In the privacy-preserving data mining literature, there are few alternatives, e.g., objective perturbation [19], gradient perturbation [20], and input data perturbation [21]–[23], that can help to mitigate these problems. In this paper, we consider the latter *input perturbation* setting since it is independent of any ML algorithm in contrast with objective and gradient perturbations schemes and because others can utilize the sanitized data in other contexts (e.g., analysis).

In our case, with the raw dataset containing direct identifiers (e.g., names), one straightforward question as: "*Is there any operator linked with an increased ratio of victims' death*?" can be easily computed, which compromises the operators' privacy and can lead to social and/or economical damages. Similarly, one can easily access the reason for the intervention (e.g., cardiac arrest) and use this information to jeopardize the victims' privacy through discrimination in health insurance, for example. Besides, even by excluding direct identifiers, both victims' and operators' identities are still at risk of being retrieved. Indeed, attributes such as sex, age, and ZIP code (a.k.a. quasi-identifiers – QIDs) can be combined with public data to reidentify individuals [24].

C. Motivations and contributions

From 2016 to 2020, the SDIS25 has presented an average increase of about 15% in the number of EMS calls (i.e., emergency interventions), comparing to 2015, reaching the highest peak in 2018 with an increase of 25%. And just as calls increased, the number of fatalities also increased, on average by 7%, with a peak of 17% in 2018. Similarly, the number of travels to HFs grew on average by 15%, with a peak in 2018 of 26%. This implies that the resources of SDIS25 were engaged for longer, so there were fewer resources in the centers. And with the present scarcity of resources in its service, the SDIS25 is vulnerable to a major crisis (e.g., pandemics, natural disasters), which puts the population at risk. For this reason, an intelligent system is needed to allow them to predict the urgency of the intervention (victim's mortality) and plan the distribution of resources that will be engaged (victim's transportation).

This way, within the *interval of interest* in Fig. 1 and with all the accessible information about the call processing time; operators' and victims' personal data, and the intervention itself, the purpose of this paper is to contribute with a novel ML-based methodology to predict both victims mortality and their need for transportation to HFs, which can be used as a decision-support system by SDIS25 (and EMS in general). Indeed, when a center has received the alert to go to an intervention, the system could launch its predictions and, thus, allow EMS to better determine their resources to provide faster response time (and possibly avoid victim's mortality) and to know that the engaged armament will be in use for less time (by not going to a hospital, for example). In addition, we take into consideration both victims' and operators' privacy and, hence, we developed and assessed the effectiveness of two state-of-the-art privacy notions on sanitizing the datasets before training ML models. In summary, this paper makes the following contributions:

I **Variables' impact**. Extract mortality and transportation statistics from victims, discover and address privacy issues, and finally recognize the variables with the greatest impact for building predictive models. This would allow replication of the process in another EMS.

II **Prediction of victims' mortality**: Predict if the victim will die. This would allow to increase the level of urgency of the call and to improve the selection of adapted engaged resources.

III **Prediction of victims' transportation**: Predict the need to transport the victim from emergency scenes to HFs. This would allow knowing if the EMS resources will be engaged for a longer period of time, and as a result, to make a better decision to deploy its resources.

IV **Privacy-conscientious predictions**: In both items II and III, we take into consideration both victims' and operators' privacy. Indeed, we consider the privacy-preserving data mining setting named *input data perturbation* [21]–[23]. This allows EMS to sanitize the data before transmitting it to third parties in charge to build ML-based decision-support systems.

Paper's outline: The remainder of this paper is organized as follows. In Section II, we describe the material and methods used in this work, i.e., the collected data and its analysis; the ML techniques, the privacy models, and the proposed methodology. In Section III, we present the results of our

experiments and in Section IV our discussion with related work. Lastly, Section V shows our concluding remarks.

II. Material and Methods

A. Data collection

We used retrospective data recorded by SDIS25 from January 2015 to December 2020. All calls made to SDIS25 lines (e.g., 18 and 112) were considered. Calls that had an intervention identifier indicating that they were attended by SDIS25 centers were eligible for inclusion in the case there were *victims*. In total, after removing outlying observations, the primary dataset at our disposal comprises 177883 emergency calls with the following explanatory variables:

- Victim: age, sex, and center that assisted the victim; the city where the intervention occurred; distance between the center and the city; victim's mortality and victim's transportation to HFs.
- Operator: age, sex, grade, and seniority (experience time).
- Call/Intervention: hour, day, day of the week, month, year, delay time to answer the phone, call duration, delay time to diffuse the alert, and type of intervention. The latter is described by 3 variables: type of operation (aid to person, fire, etc), subtype of operation (an emergency, fire on the public road, fire in an individual room, etc), and the motive for departure (external hemorrhage, respiratory distress, etc). We also have the reason for departure. The difference between the *motive* and *reason* is that the first is what is recorded by the operator based on the call received (partial), and the second is what is recorded by the firefighters upon return from the mission (confirmed). This last one will not be considered during the modeling since the developed system will make predictions before the armament departs to the mission.

From here, we extracted 2 variables as our predictive targets: the victim's mortality and the need for transport to a hospital. Both will be developed in the following sections.

B. Data analysis and privacy awareness

Table I exhibits for each year the number (Nb.) of EMS calls, mortality, and transportation to HFs, and the augmentation (Growth in %) throughout years in comparison with 2015. The last row named 'Total' indicates the sum for Nb. and the average for Growth. One can notice that the total number of mortality and transportation to HFs represent 1.94% and 78.72% of the total number of EMS calls, respectively. Also, over the years, the number of EMS calls increased by 14.67% on average, which led to an average increase of 7.17% and 15.27% on the number of fatalities and transportation to HFs, respectively.

Fig. 2 presents statistics on call processing times by time intervals in seconds and their ratio of deaths as a percentage. One can observe that the ratio of deaths increases in the 3 time periods as time increases. This is more noticeable with 'Call duration' that reaches a rate of 28.57% of deaths when the calls lasted between 17 and 18 minutes (interval 1020-1080). Although, there were few calls (14), more cases of deaths

TABLE I: Summary statistics on the number (Nb.) of EMS calls, mortality, and transportation to HFs and the augmentation (Growth) in relation to the baseline year 2015. Total indicates the sum for Nb. and the average for Growth.

Year	EMS Calls		Mortality		Transport	
	Nb.	Growth (%)	Nb.	Growth (%)	Nb.	Growth (%)
2015	26417	–	544	–	20700	–
2016	28529	7.99	532	-2.21	22479	8.59
2017	31311	18.53	562	3.31	24990	20.72
2018	33150	25.49	638	17.28	26271	26.91
2019	29914	13.24	562	3.31	23370	12.90
2020	28562	8.12	621	14.15	22214	7.31
Total	177883	14.67	3459	7.17	140024	15.27

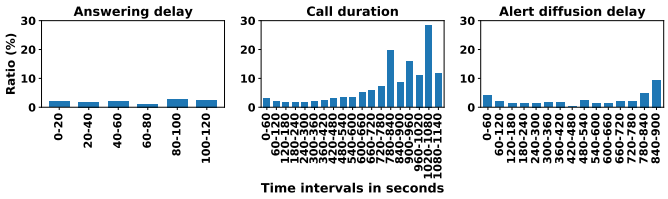

Fig. 2: Ratio of victims' death ($Nb.\ Deaths/Nb.\ Calls \times 100$) according to call center operators' delay time to answer the phone, the total call duration time, and the delay time taken to broadcast the alert, respectively, by time intervals in seconds.

appeared (4). Yet, we still kept these cases since they could influence the prediction of victim's mortality.

Furthermore, on analyzing the attributes *motive* and *reason* for departure, we remark that about 75% of the times both values were equal. This means, most of the time, the information was correctly acquired during the call, which helps SDIS25 to correctly define the armament needed to attend the victim, and could contribute to the accuracy of our ML models launched before the departure of the armament.

With regard to privacy, considering *victims*, by combining three available QIDs (sex, age, and city), one can find about 22000 cases with the trivial $k = 1$-anonymity level [24] (further explained in Subsection II-D). This means, in some cities with low population density, it would not be difficult to find out the person who needed help by knowing their sex and age. Similarly, combining four QIDs considering *operators* (sex, age, grade/career, and seniority) leads to a similar output with many unique rows. One exception is that there is a set of operators, and each row represents an *event* of who treated the emergency call. This reinforces the need for applying privacy-preserving techniques to protect the users' privacy.

C. Machine learning techniques

Given the easy and fast use of techniques based on decision trees with tabular data, we selected 2 remarkable representatives: Light Gradient Boosted Machine (LGBM) and Extreme Gradient Boosting (XGBoost). Also, given the large amount of data we have and the recognized potential of neural networks, we tested 2 types: Convolutional Neural Network (CNN) and Attentive Interpretable Tabular Learning (TabNET). These 4 techniques are described briefly in the following:

- LGBM [25] is a gradient boosting decision tree technique. It establishes a leaf-wise tree growth strategy to speed up computation and reduce memory consumption.
- XGBoost [26] is a robust gradient boosting decision tree technique and a pioneer in optimizing tree parallelization. Its main strategy is adding penalization to the model's complexity to avoid overfitting.
- CNN [27] is a type of deep neural network, applied mainly in the recognition and classification of images, but also texts and tabular data. Its 2 main layers are Conv and MaxPooling, where the first maps patterns and the second reduces the dimensionality of the drawn map.
- TabNET [28] is an interpretable canonical deep learning architecture for tabular data. It uses instance-wise feature selection strategy (attentive transformer) to improve nonlinear processing without falling into overfitting.

D. Privacy models

In this work, we applied and compared two state-of-the-art privacy models to sanitize our datasets, which are briefly described in the following:

- *K*-anonymity (K-Anon) [24]: The K-Anon model requires that each released record to be indistinguishable from at least $k-1$ others. Intuitively, the larger k is, the better the privacy protection will be. On applying K-Anon, there is a difference between explicit identifiers (e.g., names), which are removed or masked to avoid direct re-identification; sensitive attributes (e.g., mortality), that might be preserved, and quasi-identifiers (e.g., age, sex), in which K-Anon seeks to ensure indistinguishability. We recall the definition of K-Anon in the following.
  Definition 1 (k-anonymity requirement [24]): Each release of data must ensure that every combination of values of QIDs can be indistinctly matched to at least *k* individuals.
- Differential privacy (DP) [29], [30]: DP, originally proposed in [29], is the current standard for data privacy. A formal definition of DP is given in the following:
  Definition 2 ((ϵ, δ)-Differential Privacy [30]): Given $\epsilon > 0$ and $0 \leq \delta < 1$, a randomized algorithm $\mathcal{A}$ is said to provide (ϵ, δ)-differential-privacy if, for all neighbouring datasets D_1 and D_2 that differ on the data of one user, and for all sets R of outputs:

$$\Pr[\mathcal{A}(D_1) \in R] \leq e^{\epsilon} \Pr[\mathcal{A}(D_2) \in R] + \delta.$$

  The additive δ on the right-side of Eq. (2) is interpreted as a probability of failure. Normally, a common choice for δ is to set it significantly smaller than $1/n$ where n is the number of users in the database [30].

E. Proposed ML-based methodology

This article proposes a new methodology based on ML techniques to predict the mortality of a victim and their need to be transported or not to an HF.

From our primary dataset, we modified the name of the victim's city (VIC_CITY) by a numerical identifier generated based on the proximity between cities, i.e., we took as references the centroid of the 3 main cities of the region, we calculated the distances with the centroid of the other 570 cities in the territory, and we assigned the identifiers in ascending order of distance. This way, it is easier to define the generalization of hierarchies (by masking) during sanitization.

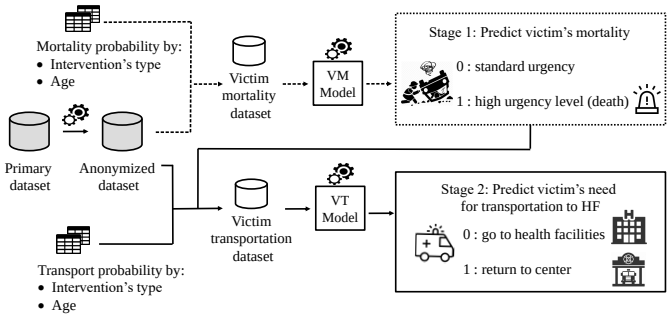

Fig. 3: Proposed ML-based methodology developed in 2 stages. Stage 1 predicts the victim's mortality, which will be used as a feature in stage 2 when predicting the need to transport the victim to an HF.

To increase the impact of the predictors, we added 6 other variables by combining the existing ones. The first 2 were:

- Age group (AGG_GROUP_AGE). The age of the victims was grouped into 8 categories. This variable was only used in the original dataset since sanitized datasets already group age in order to respect K-Anon (cf. Table III).
- Time between rings the phone until the alert is broadcast (AGG_INT_TM_TOTAL). It is the sum of delay time to answer the phone, call duration, and delay time to diffuse the alert.

Then we sort our dataset in ascending order by datetime and divide it into a training set (calls from 2015-2019) to build our models and testing set (calls from 2020) to validate the performance of the models. Finally, we add 4 more predictors:

- Probability of mortality by motive (AGG_MORT_PROBA_MOT). We took the training set and grouped the calls in which the victims died by intervention type (type, subtype, and motive of departure). The number of cases by intervention type is divided by the total number of deaths. This results in a probability, where the probabilities > 0 are stored in a temporary table. The testing set samples are completed by looking for their probability in the temporary table. If the intervention's type is not found, the sample will receive a probability of zero, and the table will be updated for the next sample. Otherwise, the sample would take the current probability, and again the table would be updated. This follows the actual process, where the death of the victim is confirmed only when the firefighters return from their mission.
- Probability of mortality by age (AGG_MORT_PROBA_AGE). The training set data that resulted in deaths are grouped by motive of departure and AGG_GROUP_AGE. In this way, the number of cases per category is divided between the total deaths to generate a probability and create a temporary table with the values > 0. The assignment of

probabilities to the testing set follows the same process of the variable described previously.

- Transport probability by motive (AGG_TRAN_PROBA_MOT). Its process is similar to the variable AGG_MORT_PROBA_MOT, with the difference that we exchanged deaths for non-transportation to HFs.
- Probability of transport by age (AGG_TRAN_PROBA_AGE). It follows the same process of AGG_MORT_PROBA_AGE, but with the "non-transportation to HFs" target.

The categorical variables such as: the grade and sex of the operator, the sex and city of the victim, the center that attended the victim, the intervention's type, the weekday and the month were converted using the LabelEnconder method of Scikit-Learn, and the remaining variables that are numerical maintained their original scale.

Fig. 3 presents the interaction of our 2 final models developed in 2 stages as explained in the following:

Stage 1: Prediction of victims' mortality. In this stage, the objective is to predict whether or not the victim is at great risk of dying if they are not treated quickly. This is represented by the binary target variable VICTIM_MORTALITY (0: alive, 1: dead). The used dataset is composed of all the explanatory variables of the primary dataset together with the modified ones and the new ones (AGG_GROUP_AGE, AGG_INT_TM_TOTAL, AGG_MORT_PROBA_MOT, AGG_MORT_PROBA_AGE). Since there is an imbalance in the number of samples between classes 0 and 1 (see Subsection II-B), we sought to compensate for the balance by using the Adaptive Synthetic sampling approach (ADASYN) and sample duplication, both only in the training set. In our preliminary experiments, duplication of samples in class 1 presented better results than ADASYN and, thus, duplication was used to construct the final models.

The specifications for the training process by technique is described in the following:

- LGBM: objective="binary", is_unbalance=True, importance_type="gain", boosting_type="gbdt", and the search space was n_estimators [50-1000], learning_rate [0.001-0.8], max_depth [1-10], and colsample_by_tree [0.5-1].
- XGBoost: objective="binary:logistic", boosting_type="gbtree", and the tunned hyperparameters were n_estimators [50-1000], learning_rate [0.001-0.5], max_depth [1-20], colsample_by_tree [0.2-1], and scale_pos_weight [20-60].
- CNN: we extracted 10% of the training samples for the validation set. In addition, 4 types of architectures were built using Keras library, their structure are presented in Table II. The main hyperparameters were loss="BinaryCrossentropy", optimizer=Adam, epochs=1000, monitor="val_loss", mode="min", patience=15, restore_best_weights=True. And the tunned hyperparameters were batch size [40-200], learning rate [0.0001-0.2], and the architecture's type [1,2,3,4].
- TabNET: we extracted 10% of the training test to be used as a validation set during the training process. The main specifications were optimizer="Adam", eval_metric="logloss", max_epochs=1000, and patience=15. The hyperparameters tunned were n_d [3-10], n_a [1-10], n_steps [2-10], n_independent [1-5], n_shared [1-5], gamma [1-2], lr [0.001-0.5], and batch size [800-2000].

TABLE II: Defined architectures for CNN with the number of neurons (*nn*), pool_size (*s*), and dropout rate (*r*) used.

Archi. 1	Archi. 2	Archi. 3	Archi. 4
Input	Input	Input	Input
BatchNormalization	BatchNormalization	BatchNormalization	BatchNormalization
Conv1D (nn=128)	Conv1D (nn=512)	Conv1D (nn=32)	Conv1D (nn=32)
MaxPooling1D (s=2)	MaxPooling1D (s=2)	MaxPooling (s=2)	MaxPooling (s=2)
Flatten	Flatten	Dropout (r=0.2)	Dropout (r=0.2)
Dense (nn=2)	Dense (nn=2)	Conv1D (nn=64)	Conv1D (nn=64)
Activation (sigmoid)	Activation (sigmoid)	MaxPooling (s=2)	MaxPooling (s=2)
		Flatten	Dropout (r=0.5)
		Dense (nn=2)	Conv1D (nn=128)
		Activation (sigmoid)	MaxPooling (s=2)
			Flatten
			Dense (nn=2)
			Activation (sigmoid)

Finally, to search hyperparameters for each technique, we used Bayesian optimization through the Hyperopt library with the Tree Parzen Estimator Suggest (tpe.suggest) and 500 iterations. The objective function employed to select the best victim mortality model (VM model) is a combination of Accuracy (ACC), Balanced Accuracy (BACC), Macro F1 score (MF1), and Recall of class 1 in the form of $\texttt{maximize}(ACC^2 \times Recall^2 \times BACC \times MF1)$. This is because we wanted to keep the models with high ACC and Recall without losing the balance between classes due to the class imbalance problem.

Stage 2: Prediction of victims' transportation. This stage predicts whether the victim will need or not to be transported to an HF. The target is represented by a binary variable VICTIM_TRANSPORTATION (0: needed, **1: did not need**). Given that most cases require (see Subsection II-B) a transfer to an HF (0), it would be more hard but advantageous to detect those calls that need the resources for less time (1), thus, recognize that the armament will return faster to the center and there will be more resources available ready to attend other interventions. The dataset used in this stage included as an explanatory variable the predicted victim mortality since if the prediction indicates a death, there will be a greater probability of going to an HF to save the victim's life. Therefore, the engaged armament will be unavailable for a longer time. Other predictors considered are all those defined in stage 1, and AGG_TRAN_PROBA_MOT and AGG_TRAN_PROBA_AGE. Unlike the previous phase, no duplication or generation of artificial samples was applied since no improvement was observed in the preliminary tests. Finally, we used the same search space and model selection process applied in stage 1 to choose the best victim transportation model (VT Model).

Experiments with sanitized data. In the context of this paper (i.e., of medical data), we used the differentially private data release mechanism proposed in [31], which produces truthful data output. More precisely, DP is ensured by sub-sampling, in which the sampling probability depends on ϵ, and data are released in a generalized form that also satisfies K-Anon (where k depends on ϵ and δ). In our experiments, we set $\epsilon = 1$, which is a common value in DP literature [30], [31] and we set $\delta = 10^{-6} \ll 1/n$. With these parameters, the differentially private training set was sub-sampled from

TABLE III: Final generalization hierarchy for each QID of each entity, namely, victim and call center operator.

Attribute	Final Generalization Hierarchy	
	k-Anonymity	Differential Privacy
Vic. Age	[0, 64[, [64, 101[	[0, 32[, ..., [96, 101[
Vic. Sex	M, F, NR	M, F, NR (not registered)
Vic. City	21***, 22***, 23***	212**, 222**, ..
Ope. Age	[22, 38[, ..., [54, 63[, $\geq$63	[22, 30[, [30, 38[, ... $\geq$62
Ope. Sex	F, M	F, M
Ope. Seniority	[0, 5760[, [5760, 10204[, $\geq$10204	*
Ope. Grade	*	*

149321 to 81538 samples and, besides DP guarantees, K-Anon is also satisfied with $k = 74$ [31]. Thus, for a fair comparison between the two privacy models, we also set $k = 74$ when applying the K-Anon model individually. In other words, we can thus evaluate the impact of having only a K-Anon guarantee in comparison with the case of DP that also satisfies K-Anon guarantees.

Table III exhibits the final generalization approach for each QID we considered of each entity (Victim – Vic. and Operator – Ope.) and privacy model (K-Anon and DP). The symbol $*$ in Table III indicates full suppression for an attribute or masking of a digit (for Vic. City). Also, we highlight that both privacy models were applied only in the *training set* and, hence, the *testing set* was transformed using the final generalization hierarchies.

Finally, with the sanitized datasets, we selected only the ML technique that performed the best with original data among the four we evaluated (i.e., LGBM, XGBoost, TabNET, and CNN). Similar to the 2-stage methodology previously explained, we used Bayesian optimization with 500 iterations for the first VM Model, and its predictions were also used as input to the subsequent VT Model for 500 iterations too. Lastly, models were evaluated with the aforementioned classification metrics.

III. Results

Table IV exhibits for each dataset (original and sanitized ones) the metrics results considering each experimented ML technique calculated for each target and all samples from 2020. We also included the metrics: Area Under the Receiver Operating Characteristic Curve (AUC) and the Precision of class 1. From Table IV, one can notice that for ML models trained with original data, XGBoost consistently outperformed all the other models in both binary classification tasks, with results highlighted in bold. This is why for sanitized datasets, we only present results with XGBoost.

In general, one can notice that the developed models reached high scores when predicting victims' mortality, with ACC, BACC, and AUC scores higher than 88% getting to 96%. Also, as an imbalanced classification problem, for mortality, all models achieved M-F1 scores of about $70\% \sim 75\%$ since they had lower precision scores $30\% \sim 36\%$ for the minority class (victim will die – 1) while achieving a good recall of about $80\% \sim 85\%$. On the other hand, the results for classifying if the ambulance will return to the EMS center (positive class – 1) or will transport the victim to HFs (negative class – 0) were not as good as for the victims' mortality target. In this case, all models achieved intermediate scores of about $70\% \sim 78\%$ for ACC, BACC, and AUC. Besides, the M-F1 metric decreased to about $65\% \sim 68\%$ with lower recall values ranging from $66\% \sim 71\%$ and higher precision scores $40\% \sim 44\%$ in comparison with predicting mortality.

Moreover, one can notice some minor decrease in the performance scores when comparing the XGBoost model trained with original and sanitized datasets. Indeed, although some features (the QIDs) suffered transformation and/or suppression, models were still able to classify both targets as good as with the original dataset, while providing privacy guarantees for both victims and operators. These results suggest that some patterns were kept even with the transformed features. To analyze this behavior, Fig. 4 illustrates the Pearson correlation coefficient ranging from -1 (negative correlation) to +1 (positive correlation), considering all the variables of the 3 types of dataset. We highlight that in the case of sanitized datasets we transformed the variable names used as QIDs following the acronym 'K' and 'DP' (e.g., K_OPE_AGE). One can notice that in the 3 datasets, few of the initial variables have a slight impact (VIC_SEX/K_VIC_SEX/DP_VIC_SEX, INT_OPERATION_TYPE, INT_OPERATION_SUBTYPE, and INT_MOTIVE_DEPARTURE). However, by combining some of these (see Subsection II-E) and generating the aggregates (AGG prefix) we obtain variables with more correlation (AGG_MORT_PROBA_AGE, AGG_MORT_PROBA_MOT, AGG_TRAN_PROBA_AGE, and AGG_TRAN_PROBA_MOT) even with sanitization.

In addition, Fig. 5 illustrates the 12 features with the highest impact by target, dataset, and best prediction model generated by the XGBoost "Gain" feature importance algorithm. This is based on the relative contribution of each feature to improve the accuracy in the division of a branch. Thus, from Fig. 5, we can confirm that the aggregate variables generated from probabilities and the type of intervention (type, subtype, and motive) have a great impact on the creation of both models. In the case of victim's mortality, age and sex are more important than the alert diffusion time and duration of the call. In addition to that, we discovered that the time and day of the week also have an impact on the model. In the case of non-transportation of the victim to an HF, the greatest impact is generated by the mortality of the victim and, to a lesser degree, the sex, the hour of the call, and the center from which the ambulance departed. Although the age of the victim did not show influence when it was sanitized, it did show an effect with the original data. Moreover, in our case, operators' personal data did not show much importance for any ML model, in this way, for upcoming works, we consider not using such predictors as there would be a need for preserving their privacy. We believe that our analysis of feature importance can provide an overall idea of what other EMS and research groups might consider with the intention of replicating this work.

IV. Discussion and Related Work

When reviewing the related works, we find that some EMS are exploiting their data with ML techniques [1], [4]–[9],

TABLE IV: Results of the best models by ML technique with and without privacy-preserving sanitization for 2020.

Dataset	ML Technique	Victim's Mortality (%)						Victim's Transportation (%)					
		ACC	BACC	AUC	MF1	Recall	Precision	ACC	BACC	AUC	MF1	Recall	Precision
Original	LGBM	96.07	90.83	95.60	73	85	34	73.53	70.99	73.78	67	66	44
	XGBoost	**96.44**	**90.86**	**96.04**	**75**	**85**	**36**	**73.62**	**72.63**	**78.91**	**68**	**71**	**44**
	CNN	95.97	89.44	94.63	73	83	33	70.42	70.30	75.76	65	70	40
	TabNET	95.39	90.24	94.97	71	85	30	70.98	70.26	74.17	65	69	41
K-Anon	XGBoost	96.26	91.00	95.50	74	86	35	73.62	72.52	78.70	68	71	44
DP	XGBoost	96.04	90.18	95.89	73	84	34	73.22	72.35	78.53	68	71	44

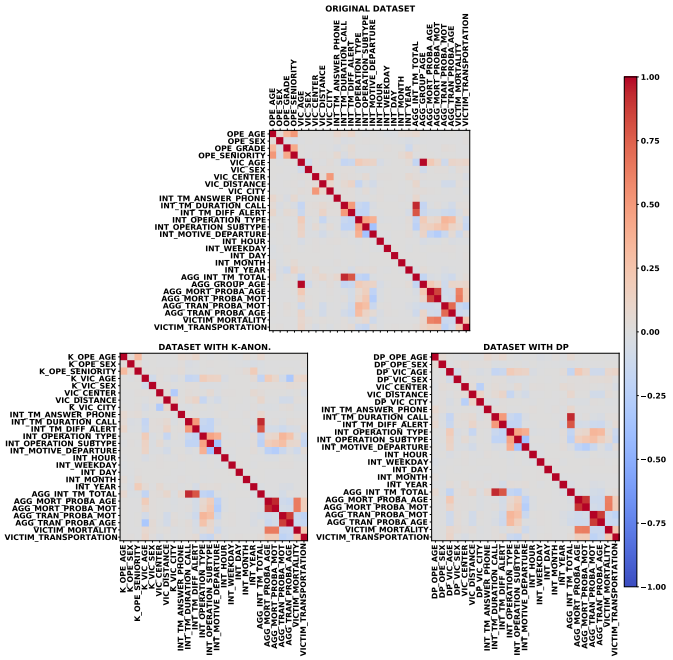

Fig. 4: Pearson correlation between all predictors and targets with the 3 kinds of datasets: original, K-Anon, and DP.

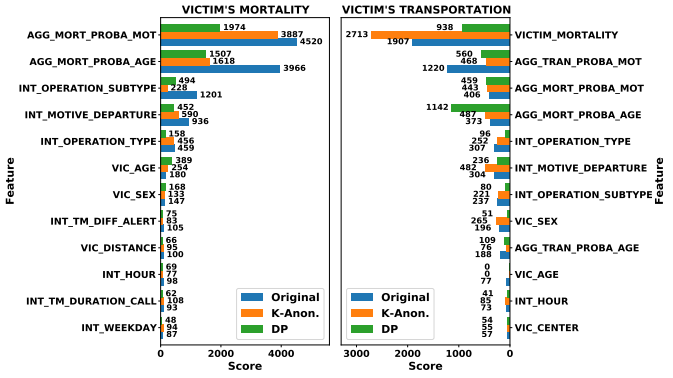

Fig. 5: Feature importance from the best XGBoost models, considering the type "Gain" as score and the first 12 variables.

[11]–[13]. For instance, in [1], the authors identified that shorter response time is associated with a higher survival rate and predicted the demand of ambulances to allow their reallocation. In [12], the authors used an ML framework to early recognize cardiac arrest by voice records in emergency calls. The latter consider an interval of interest similar to ours (Fig. 1), with the difference that we use tabular data recorded in the last 5 years, and where our first objective is to predict the victim's mortality. Since this prediction is done before dispatching ambulances, EMS could activate proactive decisions to mobilize their armament and provide quicker response times. In addition, we also propose to use the prediction of the victim's mortality as an input to a second model (cf. Fig. 3), which predicts if the ambulance will return to the EMS center and, as consequence, the crew will be engaged for less time and more resources will be available. Related to this point, few works were found in the literature. For instance, [10] analyzes and predicts the spending time of an ambulance in a hospital to complete the transfer of the victim and return to its center. If these times are large or the workload is heavy, it can lead to breakdowns in the EMS [13].

Although the collection of medical data allows investigations to propose improved ML-based decision-support tools, on the other hand, there is a problem with the disclosure of personal and sensitive information. Currently, data breaches are all too common [15], which in our context, would jeopardize both the victims' and operators' privacy. In addition, data breaches are not the only issue since ML models can also leak private information in an indirect way [17], [18]. To solve these issues, in the privacy-preserving data mining literature, some works propose to train ML models with sanitized data [22], [23], which is also known as input perturbation [21].

We also adopt *input perturbation* because it allows using any ML and post-processing techniques in contrast with gradient [20] or objective perturbation [19]. Besides, sanitizing the training dataset as we performed in this paper is in accordance with real-world applications where EMS would use and/or share only sanitized data with trusted third parties to train and develop ML-based decision support systems. This way, because each sample in the dataset is made private, data are protected from data leakage and are more difficult to reconstruct when the ML model receives attacks, for example. As shown in the results, it was still possible to achieve similar scores than when training ML models with the original data, which suggests that some patterns were kept even with the transformed QID features (cf. Fig. 4 and Fig. 5). More specifically, while the results with the k-anonymous dataset approximates the ones with original data, the results with DP decreased more. This could be due to DP applying both subsampling of the training dataset as well as the generalization and/or suppression of QIDs to satisfy K-Anon.

V. Conclusions and Perspectives

This paper proposes a new ML-based methodology to predict both the victims' mortality and their need for transporta-

tion to health facilities (HFs) using data available at the start of the emergency call until an EMS center is notified. The results indicate that mortality could be predicted with both accuracy and AUC scores as high as 96%. On the other hand, the need for transportation to HFs could be predicted with intermediate accuracy and AUC scores of about $70\% \sim 78\%$. These results showed the potential of ML for EMS, which can be used as a decision-support tool to early identify mortality and the use of resources (transportation) and, thus, help EMS to save more lives and avoid service disruptions. In addition, we also took into consideration both *victims'* and call center *operators'* privacy when training ML models. In this case, even with sanitized datasets (*k*-anonymous and differentially private), it was still possible to predict the intended targets. With these findings, EMS may consider using and/or sharing privatized datasets to avoid possible data leakages, membership inference attacks, and data reconstruction attacks [15], [17], [18].

Lastly, some limitations and prospective directions of this paper are described in the following. First, on sanitizing the datasets, there is a clear difference in the type of privacy we provided for each *entity*. On the one hand, because *victims* were unique in our dataset, DP and K-Anon provided *user-level* [30] privacy. On the other hand, there is a unique set of *operators* that treated many emergency calls and, thus, DP and K-Anon provided *event-level* [30] privacy. Investigating a uniform notion of privacy for both entities is an intended direction. Also, we considered an ideal case where the information of all victims in the testing set was acquired during the call. However, this may not always occur in real life, e.g., when someone activates EMS for unidentified victims. One future direction would be evaluating our models with randomly excluded data from victims (i.e., sex and age) to assess the ML models' robustness. Other future directions would be working with the *observations* registered by operators during calls in text format, which could be treated with natural language processing techniques, as well as voice registers.

REFERENCES

[1] A. Y. Chen, T.-Y. Lu, M. H.-M. Ma, and W.-Z. Sun, "Demand forecast using data analytics for the preallocation of ambulances," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1178–1187, Jul. 2016.

[2] J. P. Byrne *et al.*, "Association between emergency medical service response time and motor vehicle crash mortality in the united states," *JAMA Surgery*, vol. 154, no. 4, p. 286, Apr. 2019.

[3] C. Penn, T. Koole, and R. Nattrass, "When seconds count: A study of communication variables in the opening segment of emergency calls," *Journal of Health Psychology*, vol. 22, no. 10, pp. 1256–1264, Feb. 2016.

[4] N. Shafaf and H. Malek, "Applications of machine learning approaches in emergency medicine; a review article," *Archives of Academic Emergency Medicine*, vol. 7, no. 1, Jul. 2019.

[5] K. J. W. Tang, C. K. E. Ang, T. Constantinides, V. Rajinikanth, U. R. Acharya, and K. H. Cheong, "Artificial intelligence and machine learning in emergency medicine," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 156–172, Jan. 2021.

[6] J. Stewart *et al.*, "Applications of machine learning to undifferentiated chest pain in the emergency department: A systematic review," *PLOS ONE*, vol. 16, no. 8, p. e0252612, Aug. 2021.

[7] D.-Y. Kang *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*, vol. 28, no. 1, Mar. 2020.

[8] H. Arcolezi, S. Cerna, C. Guyeux, and J.-F. Couchot, "Preserving geo-indistinguishability of the emergency scene to predict ambulance response time," *Mathematical and Computational Applications*, vol. 26, no. 3, p. 56, Aug. 2021.

[9] X. Lian, S. Melancon, J.-R. Presta, A. Reevesman, B. Spiering, and D. Woodbridge, "Scalable real-time prediction and analysis of san francisco fire department response times," in *2019 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*. IEEE, Aug. 2019.

[10] S. Cerna, H. H. Arcolezi, C. Guyeux, G. Royer-Fey, and C. Chevallier, "Machine learning-based forecasting of firemen ambulances' turnaround time in hospitals, considering the COVID-19 impact," *Applied Soft Computing*, vol. 109, p. 107561, Sep. 2021.

[11] J. myoung Kwon *et al.*, "Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes," *Resuscitation*, vol. 139, pp. 84–91, Jun. 2019.

[12] S. N. Blomberg *et al.*, "Machine learning as a supportive tool to recognize cardiac arrest in emergency calls," *Resuscitation*, vol. 138, pp. 322–329, May 2019.

[13] S. Cerna, C. Guyeux, G. Royer, C. Chevallier, and G. Plumerel, "Predicting fire brigades operational breakdowns: A real case study," *Mathematics*, vol. 8, no. 8, p. 1383, Aug. 2020.

[14] G. Kaissis *et al.*, "End-to-end privacy preserving deep learning on multi-institutional medical imaging," *Nature Machine Intelligence*, vol. 3, no. 6, pp. 473–484, May 2021.

[15] D. McCandless *et al.*, "World's biggest data breaches & hacks," https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/, jan 2021, online; accessed 11 March 2021.

[16] European-Commission, "2018 reform of EU data protection rules," https://gdpr-info.eu/, 2018, online; accessed 10 April 2020.

[17] C. Song, T. Ristenpart, and V. Shmatikov, "Machine learning models that remember too much," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Oct. 2017.

[18] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, May 2017.

[19] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization." *Journal of Machine Learning Research*, vol. 12, no. 3, 2011.

[20] M. Abadi *et al.*, "Deep learning with differential privacy," ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 308–318.

[21] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, Oct. 2008.

[22] M. K. Paul, M. R. Islam, and A. S. Sattar, "An efficient perturbation approach for multivariate data in sensitive and reliable data mining," *Journal of Information Security and Applications*, vol. 62, p. 102954, Nov. 2021.

[23] M. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil, "Efficient data perturbation for privacy preserving and accurate data stream mining," *Pervasive and Mobile Computing*, vol. 48, pp. 1–19, Aug. 2018.

[24] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, Oct. 2002.

[25] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017.

[26] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016.

[27] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[28] S. O. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," *arXiv preprint arXiv:1908.07442*, 2019.

[29] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*. Springer Berlin Heidelberg, 2006, pp. 265–284.

[30] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[31] R. Bild, K. A. Kuhn, and F. Prasser, "SafePub: A truthful data anonymization algorithm with strong privacy guarantees," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 1, pp. 67–87, Jan. 2018.