Time Series Crime Prediction Using a Federated Machine Learning Model

Mustafa Abdul Salam[1,2,*†], Sanaa Taha[3††] and Mohamed Ramadan[4†††]
[1] Artificial Intelligence Dept., Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
[2] Faculty of Computer Studies, Arab Open University, Cairo, Egypt
[3] Information Technology Dept., Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt
[4] Computer Science Dept., Faculty of Computers and Information, Egyptian E-Learning University, Cairo, Egypt

*Corresponding author's Email: mustafa.abdo@fci.bu.edu.eg

Abstract
Crime is a common social problem that affects the quality of life. As the number of crimes increases, it is necessary to build a model to predict the number of crimes that may occur in a given period, identify the characteristics of a person who may commit a particular crime, and identify places where a particular crime may occur. Data privacy is the main challenge that organizations face when building this type of predictive models. Federated learning (FL) is a promising approach that overcomes data security and privacy challenges, as it enables organizations to build a machine learning model based on distributed datasets without sharing raw data or violating data privacy. In this paper, a federated long short- term memory (LSTM) model is proposed and compared with a traditional LSTM model. Proposed model is developed using TensorFlow Federated (TFF) and the Keras API to predict the number of crimes. The proposed model is applied on the Boston crime dataset. The proposed model's parameters are fine tuned to obtain minimum loss and maximum accuracy. The proposed federated LSTM model is compared with the traditional LSTM model and found that the federated LSTM model achieved lower loss, better accuracy, and higher training time than the traditional LSTM model.

Keywords: *Federated Learning (FL), Deep Learning, Tensor- Flow Federated (TFF), Keras, Data Privacy, Long Short-Term Memory (LSTM).*

1. INTRODUCTION

1.1 Crimes

According to previous studies, crime is a commonsocial and economic problem that affects quality of life and lowers national and individual economic growth [1]. Thereis a theory that has shown that crime is predictable, as criminals tend to commit crimes that they have successfully committed before [2]. With an increase in crimes, countriesand law-enforcement agencies are continuously in search ofa crime prediction model to preemptively predict crime and enable its prevention. However, building this type of modelis a data privacy challenge as raw crime data are shared. Therefore, there is an urgent need for a new approach that enables countries and law enforcement agencies to build a crime predictive model without violating privacy.

1.2 Federated learning

Google introduced federated learning in 2016 as a new machine learning paradigm. The main purpose of federated learning is to build a collaborative machine learning model based on distributed datasets that preserve data privacy by training the collaborative model without sharing raw data [3], [4].
In federated machine learning, each client (i.e., data organization, data server, mobile device, or IoT device) participates in the learning process, as each client has their own dataset and local machine learning model. There is a centralized server in afederated environment that has a centralized machine learning model (global model), which aggregates the client's model parameters (model gradients). Each client locally trains on a dataset and shares the model parameters or weights with the global model. The global model makes several iterations to collect the distributed client model updates without sharingraw data [3], [4] as shown in Fig 1.

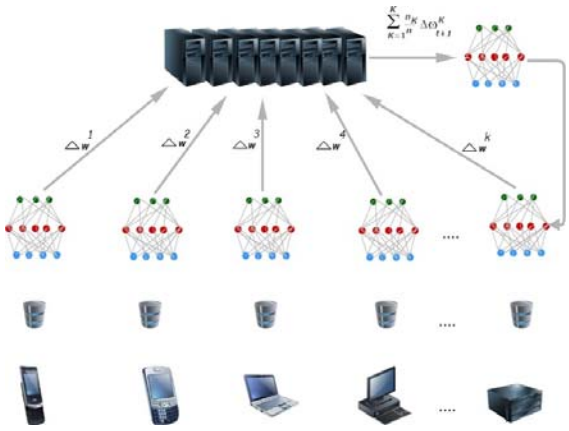


Fig. 1 The global model collection of local models updates.

Manuscript received April 5, 2022
Manuscript revised April 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.4.16

Federated learning can be applied in several disciplines, such as smart retail environments, sales, multiparty databases, and smart health care systems [5] for the following reasons:

- It does not require data transfer to a centralized server to train the model. Training occurs on each client locally, and then the model is updated on the global model.
- It preserves data privacy. This learning model implements several methodologies such as differential privacy, homomorphic encryption, and secure multipartycomputation (SMC).
- It enables a third party to be included in the training process as long as there are no privacy violations, and data are secured.
- It requires less computational power, as the training process is completed by each client. The centralized model's primary role is to collect gradient updates from the distributed models.
- It utilizes decentralized algorithms that may provide similar or better performance as centralized algorithms[4].

In environments where data privacy and data security are extremely important, it is highly recommended to use federatedmachine learning rather than traditional machine learning.

1.4 Long Short-Term Memory (LSTM)

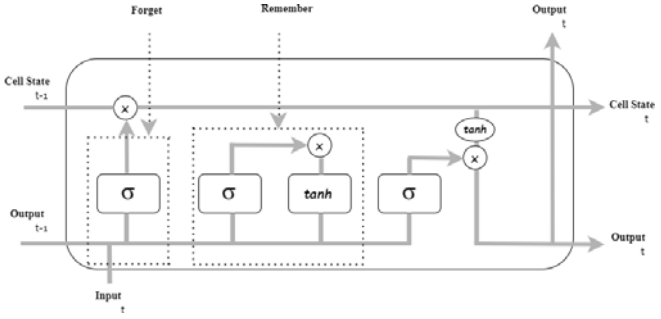


Fig. 2 LSTM architecture.

LSTM as shown in Fig 2. is a special kind of RNN that is capable of handling long-term dependencies. It was introduced by Hochreiter & Schmidhuber (1997), and due to its capability to solve a large variety of problems, it has become a popular and widely used architecture. The main advantage of LSTM is that it is designed to remember information for a long period of time to prevent long-term dependency problems.
The LSTM architecture contains as following components:

- Forget Gate: This gate decides which information should be forgotten and which should be kept. Both current input and output information from previous hidden state is accepted at this gate. Then, the sigmoid function is applied to the information to obtain aninput between 0 and 1. Input values closer to 0 willbe forgotten.
- Input Gate: This gate accepts the current input information and the previous hidden state. The sigmoid function is applied to obtain values between 0 and1, and the tanh function is applied to obtain values between -1 and 1. Then, the two outputs are multipliedto represent the cell state.
- Output Gate: This gate decides what the next hidden state should be based on the output from the input gate.

2. MOTIVATION AND CONTRIBUTIONS

Federated machine learning overcomes the challenges faced by the traditional machine learning model as follows:

- To train the traditional model, all data sources must be moved to a centralized server to start the training process, which violates data privacy and data security rules, especially in military and health care organizations [6].
- In traditional learning, third-party companies are part of model building and training processes. They should understand, prepare, clean, restructure and reshape the data to be suitable for model training, which violates data privacyand data security rules, especially in military and health care organizations [6].
- Traditional machine learning requires a massive amount of historical data to train the model and to achieve acceptable accuracy (cold start) [6],[7].
- Traditional machine learning requires a large amount of time and high computational power to train the model and achieve acceptable accuracy, which may cause a delay for organizations, especially recently opened ones [6].
- Federated learning has the potential to overcome these issues, as it allows data servers to build andtrain their models locally, sharing only their model gradients without violating any data privacy rules [1].

The principal objective of this study is to compare a federated machine learning model and a nonfederated machine learning model by building federated and nonfederated LSTM models to forecast the numberof

crimes periodically and apply them to the same datasets. The comparison between the models is basedon loss, accuracy and training time.

3. RELATED WORK

Abdul Salam, M., Taha, S. and Ramadan, M.[6]. Proposed a federated machine learning model to predict whether a patient has COVID-19 from their chest X-ray images and descriptive data. The authors compared the federated model and the traditional model and found that the federated model resulted in better prediction accuracy, lower prediction loss, and higher training time than the traditional model.

Kim, S., Joshi, P., Kalsi, P.S. and Taheri, P.[8]. Proposed two crime prediction models by using the k-nearest neighbor and boosted decision tree algorithms. The authors implemented these two models on the Vancouver crime dataset and found that the models resulted in crime prediction accuracies between 39%and 44%, respectively.

Reier Forradellas, R.F., Náñez Alonso, S.L., Jorge-Vazquez, J. and Rodriguez, M.L.[9]. Proposed a crime prediction model according to communes. The authors applied their model by using the Python programming language, implementing the SEMMA model (sample, explore, modify, model, and assess), and applying their model on the Buenos Aires crime dataset.

Zhang, X., Liu, L., Xiao, L. and Ji, J.[10]. Proposed a comparison to assess the predictive power of several machine learning algorithms by applying them on the coastal city crime dataset from southeastern China. Algorithms includedin the comparison were such LSTM, KNN, random forest, support vector machine, naive Bayes, and convolutional neuralnetworks. The authors found that LSTM outperformed theother algorithms.

Wheeler, A.P. and Steenbeek, W.[11]. Introduced a machine learning model to spatially predict interpersonal robbery crimes in the city of Dallas, Texas by using a random forest (RF) algorithm. The authors found that the random forest (RF) algorithm provided more accurate predictions than kernel density estimation (KDE) and risk terrain modeling(RTM).

Bappee, F.K., Junior, A.S. and Matwin, S.[12]. Proposed a machine learning model to predict the relationship between criminal activity and geographical regions and applied themodel on the Nova Scotia (NS) crime dataset. The authors selected different categories of crime to identify hot points from crime hotspots.

Prabakaran, S. and Mitra, S.[13]. Proposed a survey for various data mining techniques used in crime analysis and prediction. The authors divided crimes into different types, such as fraud detection, violent crime, traffic violence, sexual assault and cybercrime. The authors mentioned general data mining techniques used to detect the crimes.

Ramasubbareddy, S., Srinivas, T.A.S., Govinda, K. and Manivannan, S.S.[14]. Proposed a crime prediction system (CPS) to predict and analyze the probability of crime occurrence by implementing an a priori algorithm in model building. The authors created a sample dataset from several city crime datasets and implemented the naive Bayesian algorithm and decision tree algorithm to predict crime in a certain area.

Chun, S.A., Avinash Paturu, V., Yuan, S., Pathak, R., Atluri, V. and R. Adam, N.[15]. Proposed a machine learning model to predict whether a criminal will commit a new crime in the future within a window of time. The authors found that the data pooling method outperformed multiclass crime prediction.

Nguyen, T.T., Hatua, A. and Sung, A.H.[16]. Proposeda machine learning model to predict the types of crimes that will occur based on location and time. The authors applied the model on the Portland Police Bureau (PPB) crime dataset and implemented several algorithms, such as support vector machines (SVMs), random forests, gradient boosting machines, and neural networks.

Hajela, G., Chawla, M. and Rasool, A.[17]. Proposed a spatiotemporal machine learning model coupled with 2-dimensional hot spot analysis to cluster and predict crime hotspots. The authors found that each crime showed geographical patterns such as weather and location and found that the model with hotspot analysis achieved better performance.

Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J. and Wang, F.[18]. Proposed a survey to summarize federated learning technologies, especially in the biomedical space. The authors mentioned the statistical challenges, system challenges, and privacy issues faced by federated learning and summarized the general solutions to these challenges.

Li, Q., He, B. and Song, D. [19]. Proposed model- contrastive federated learning (MOON) as a model to solve theproblem of nonindependently and nonidentically distributed (NON-IID) dataset. The authors found that the proposed model outperformed state-of-the-art models.

Zhang, Weishan, et al. [20]. Proposed a novel dynamic fusion-based federated learning approach for image analysis to detect COVID-19 infections. The authors compared theproposed model with GhostNet, ResNet50, and ResNet101, and found that the proposed approach provided better accuracy than the other methods.

4. MATERIALS AND METHODS

This section addresses the applied tools and methodologiesused to build the federated and traditional models to forecast crime time series. In this paper, TensorFlow with Keras API was used to build the federated and traditional models using the following steps:

4.1 Crime Traditional LSTM Model

Algorithm 1 Crime Traditional LSTM Model

```
Variables Initializing
crime series ← read from buston crime csv()
crime series ← drop unused columns()
while Data Still Stationary do
    crime series ← get data next difference()
    crime series ← get adfuller p value()
end while
series supervised ← transforming to supervised()
crime series scaled ← MinMaxScaler()
train data, test data ← splitting crime data()
train dataset ← Create tensors dataset(train data)
train dataset ← shifting by window
train dataset ← flatten dataset(train data)
train dataset ← shuffling dataset(train data)
train dataset ← batching dataset(train data)
test dataset ← Create tensors dataset(test data)
test dataset ← shifting by window(test data)
test dataset ← flatten dataset(test data)
test dataset ← shuffling dataset(test data)
test dataset ← batching dataset(test data)
test dataset ← prefetching dataset(test data)
lstm model ← Create keras model()
compile keras model(optimizer, loss function)
for round in clients no do
    history ← Train The Model(train dataset)
    accuracy list ← Save Accuracy Mertics(history)
end for
Display And Plot The Results(accuracy list)
```

4.2 Crime Federated LSTM Model

Algorithm 2 Crime Federated LSTM Model

```
Variables Initializing
crime series ← read from buston crime csv()
crime series ← drop unused columns()
while Data Still Stationary do
    crime series ← get data next difference()
    crime series ← get adfuller p value()
end while
series supervised ← transforming to supervised()
crime series scaled ← MinMaxScaler()
train data, test data ← splitting crime data()
train dataset ← Create tensors dataset(train data)
train dataset ← shifting by window()
train dataset ← flatten dataset(train data)
train dataset ← shuffling dataset(train data)
train dataset ← batching dataset(train data)
test dataset ← Create tensors dataset(test data)
test dataset ← shifting by window(test data)
test dataset ← flatten dataset(test data)
test dataset ← shuffling dataset(test data)
test dataset ← batching dataset(test data)
test dataset ← prefetching dataset(test data)
federated dataset ← repeating dataset()
lstm model ← Create keras model()
federated lstm model ← from keras model()
iterative process ← federated average process()
for round in clients no do
    history ← Train The Model(federated dataset)
    accuracy list ← Save Accuracy Mertics(history)
end for
Display And Plot The Results(accuracy list)
```

5. THE PROPOSED MODEL

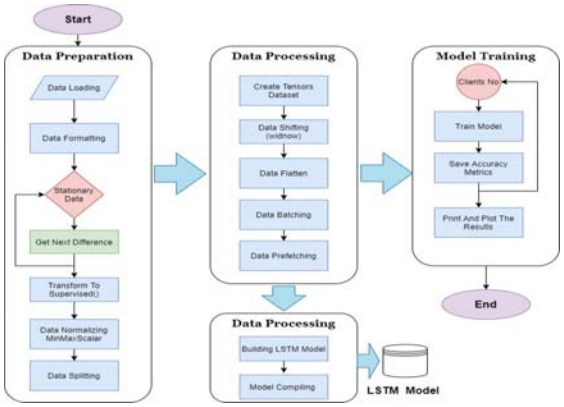


Fig. 3 The proposed traditional LSTM model for crime time seriesforecasting.

The proposed traditional LSTM model for crime time series forecasting.

5.1 Crime Traditional LSTM Model:

As shown in Fig. 3,the proposed LSTM model building steps are:

- Data Loading
  The data are loaded by using the pandas
  read_csv() API.
- Data Formatting

Unused columns are removed and date columns are formatted.

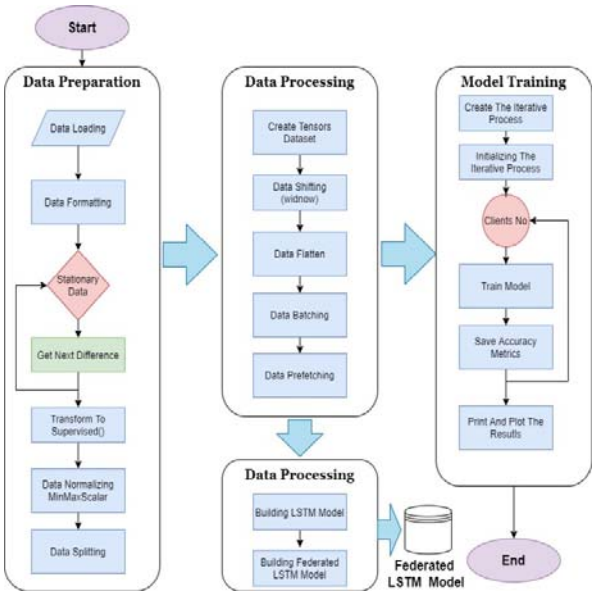
- Data Checking
  An adfuller test is used to check whether the data are stationary.
- Obtain Next Difference
  The next data difference is obtained.
- Transform to Supervised
  The data are transformed to supervised data to enhancethe model accuracy.
- Data Splitting
  The data are split into testing and training sets.
- Create Keras Tensor Dataset
  The Keras dataset is created by using the from_tensor_slices API.
- Data Shuffling
  The data are shuffled to avoid obtaining the same results.
- Data Flattening
  The ndarray dataset is flattened to 1 darray dataset.
- Data Batching
  Data are grouped into batches to enhance their performance.
- Data Prefetching
  Data are cached in memory for better performance.
- Create LSTM Deep Learning Model
  The sequential deep learning model is built using the Keras API.
- Model Compiling
  The model optimizer and loss function are identified.
- Model Training
  The model performance is evaluated by printing and plotting the evaluation metrics.

5.2 Crime Federated LSTM Model:

As shown in Fig. 4, the proposed LSTM model building steps are:

- Data Loading
  Data are loaded by using the pandas `read_csv()` API.
- Data Formatting
  Unused columns are removed, and date columns areformatted.
- Data Checking
  An adfuller test is used to check whether the data arestationary.

Fig. 4. The proposed federated LSTM model for crime time series forecasting.



- Obtain Next Difference
  The next data difference is obtained.
- Transform to Supervised
  The data are transformed to supervised data to enhancethe model accuracy.
- Data Splitting
  The data are split into testing and training sets.
- Create Keras Tensor Dataset
  The Keras dataset is created by using the `from_tensor_slices` API.
- Data Shuffling
  Data are shuffled to avoid obtaining the same results.
- Data Flattening
  The ndarray dataset is flattened to 1 darray dataset.
- Data Batching
  Data are grouped into batches to enhance their performance.
- Data Prefetching
  Data are cached in memory for better performance.
- Creating Federated Data
  Data are repeated to simulate the number of

clients.

- Create LSTM Deep Learning Model
  The sequential deep learning model is built using theKeras API.
- Create Federated Learning Model
  The deep learning model is created by using the
  `from_keras_model` Keras API .
- Create a Federated Average Process
  Local model gradients are generated, and updates are sent to the global model.
- Model Initialization and Training
  The iterative process is initiated, and training is started.
- Model Training
  The model performance is evaluated by printing and plotting the evaluation metrics.

5.3. Hardware Specifications

Our experiments were conducted using the hardware specified in Table 1.

TABLE 1: HARDWARE SPECIFICATIONS FOR THE MACHINE USEDDURING THE EXPERIMENTS

Criteria	Specification
CPU	Intel Core i7-6700HQ
GPU	NVIDIA GeForce GTX 950 M (4GB DDR3)
Storage	256GB SSD + 1000GB HDD
RAM	16GB DDR3 L, 2133 MHz

6. RESULTS

After applying the two proposed models, traditional LSTMand federated LSTM, on the Boston crime dataset, we found that:

- The proposed federated model resulted in a lower prediction loss than the proposed traditional model, as shown in Fig. 5, Fig. 6, Table 2 and Table 3.
- The proposed federated model resulted in a lower prediction MAE, MSE, RMSE and log_coch_error than the proposed traditional model, as shown in Fig. 5, Fig. 6, Table 2 and Table 3.
- The proposed federated model resulted in a slightly higher prediction RMSLE, cosine_similarity and MAPE than the proposed traditional model, as shown in Fig. 5, Fig. 6, Table 2 and Table 3.

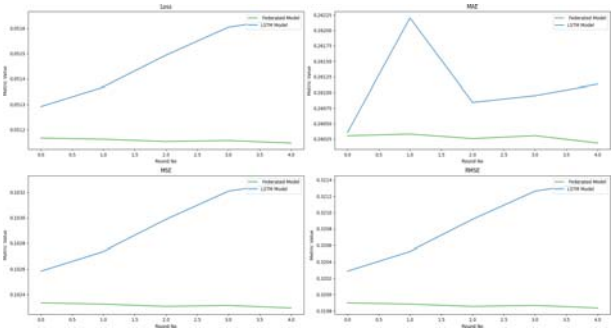


Fig. 5 Comparison of the loss, MAE, MSE, RMSE metrics of the federatedand traditional models.

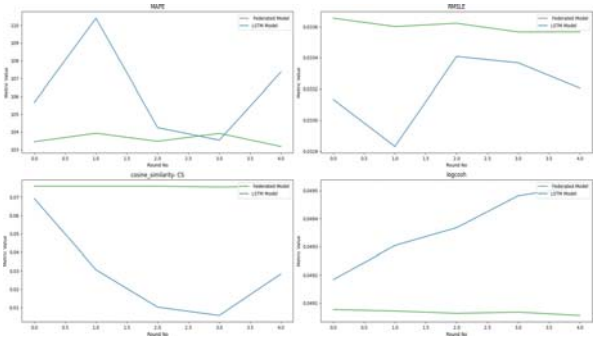


Fig. 6 Comparison of MAPE, RMSLE, cosine similarity, and logcoshMetrics comparison of the federated and traditional models.

TABLE 2: CRIME LSTM MODEL METRICS

R	MAE	MSE	RMSE	MAPE	RMSLE	Cosine Similarity	LogCosh Error	Loss
1	0.260351	0.102584	0.320287	0.033138	0.069239	0.0491839	105.663	0.051291823
2	1	0.102738	0.320527	0.03285	0.030647	0.0493043	110.408	0.051368788
3	0.102988	2	0.320917	0.033408	0.010215	0.0493678	104.246	0.051493943
4	0.10321	0.321263	3	0.033367	0.005675	0.049482	103.536	0.051604897
5	0.103284	0.321378	0.0332041	4	0.028376	0.0495162	107.363	0.051641814

TABLE 3: CRIME FEDERATED LSTM MODEL METRICS

R	MAE	MSE	RMSE	MAPE	RMSLE	CosineSimilarity	LogCoshError	Loss
1	0.260303	0.102335	0.319898	0.033654	0.076049	0.0490769	103.447	0.051167395
2	1	0.102325	0.319883	0.033600	0.076049	0.0490717	103.927	0.051162444
3	0.102307	2	0.319855	0.033621	0.076049	0.0490635	103.47	0.051153574
4	0.102315	0.319867	3	0.033565	0.075595	0.0490677	103.919	0.051157355
5	0.102295	0.319835	0.033566	4	0.076049	0.0490558	103.178	0.051147364

7. DISCUSSION

7.1 Datasets

In this work, the Boston crime dataset is used:

- This dataset contains crime incident reports providedby the Boston Police Department

(BPD).

- This dataset includes crime reports from June 14, 2015, to September 3, 2018.
- This dataset contains 319,073 incident reports. This dataset is downloaded from https://www.kaggle.com/AnalyzeBoston/crimes-in-boston.
- The number of clients is 10, to stimulate the federatedenvironment.
- Table 4: shows the columns description and action taken for the model training process.
-

TABLE 4: CRIME INCIDENT REPORTS IN BOSTON.

Column Name	Column Description	Data Type	Contains Null	Action
INCIDENT_NUMBER	incident id unique	number	no	
OFFENSE_CODE	offense code id	number	yes	removed
OFFENSE_CODE_GROUP	offense description	string	yes	removed
OFFENSE_DESCRIPTION	offense description	string	yes	removed
DISTRICT	district code	string	yes	removed
REPORTING_AREA	reporting area code	number	yes	removed
SHOOTING	is it a shooting incident (Y,N)	string	yes	removed
OCCURRED_ON_DATE	incident date	date	yes	
YEAR	incident year	number	yes	
MONTH	incident month	number	yes	
DAY_OF_WEEK	incident day description	string	yes	removed
HOUR	incident hour	number	yes	removed
UCR_PART	uniform crime reporting	string	yes	removed
STREET	incident street	string	yes	removed
Lat	incident latitude	float	yes	removed
Long	incident longitude	float	yes	removed
Location	incident location (latitude, longitude)	float	yes	removed

- To create a time-series dataset, the data should be grouped to represent the number of incidents per date.
- After removing the unused columns and grouping data by date, the dataset now contains the (`incident_date`, `incident_no` and `incident_day_no`) columns.
- Table 5: contains the column description after removing unused columns and grouping the data to represent the number of crime incidents per date.

TABLE 5: NUMBER OF CRIME INCIDENTS PER DATE IN THE CITY OF BOSTON.

Column Name	Column Description	Data Type	Contains Null
INCIDENT_DATE	incident date	date	no
INCIDENT_DAY_NO	the incident date after converting it to number	number	no
INCIDENT_NUMBER	number of incident	number	no

- After preparing the dataset, the data are plotted to determine whether they are stationary, as shown in Fig. 7.

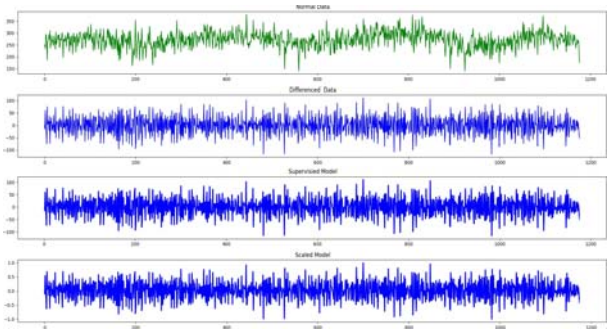


Fig. 7 Crime dataset analysis.

- The data density is plotted to determine the type ofdata distribution, as shown in Fig. 8.

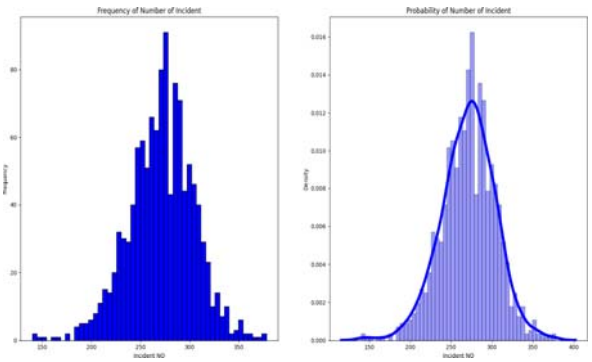


Fig. 8 Crime dataset density.

8. RESULTS DISCUSSION

The model parameters were modified multiple times to achieve the maximum accuracy and minimum loss. These modifications:

- Model optimizer
- Window size
- Batch size
- Split ratio
- Number of rounds

8.1 Model Optimizer

- In this study, the optimizer parameters were changed multiple times to determine the efficacy of changing the optimizer on the performance of the proposed models.
- Table 6: shows the model parameter values during the model training process.
- After applying the SGD optimizer model to the crime dataset, we found the accuracy metrics, as shown in Fig. 9.
- After applying the Adam optimizer model on the crime dataset, we found the accuracy metrics, as shown in Fig. 10.

Window size	60	60
Dataset split ratio	80% training - 20% testing	80% training - 20% testing
Round number	5	5

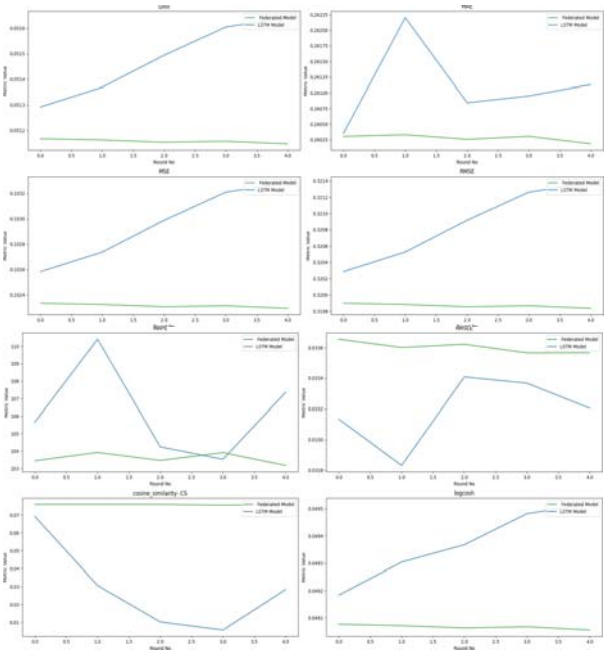


Fig. 9 SGD optimizer loss comparison.

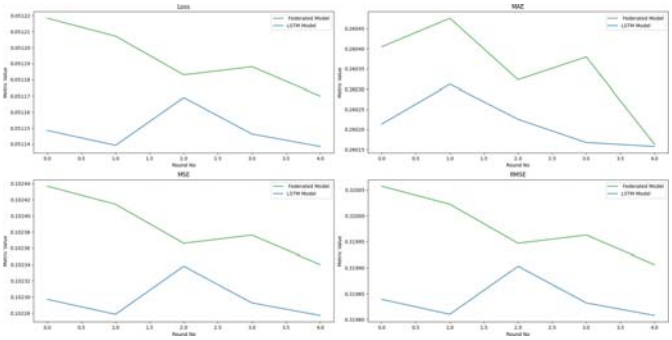


Fig. 10 Adam optimizer loss comparison.

- After performing the model training by applying the SGD and Adam optimizer, we found that:
  - In the federated LSTM model, the SGD optimizer resulted in lower loss and accuracy metrics than the Adam optimizer, as shown in Fig. 11.
  - In the traditional LSTM model, the SGD optimizer resulted in lower loss and accuracy metrics than the Adam optimizer, as shown in Fig. 11.

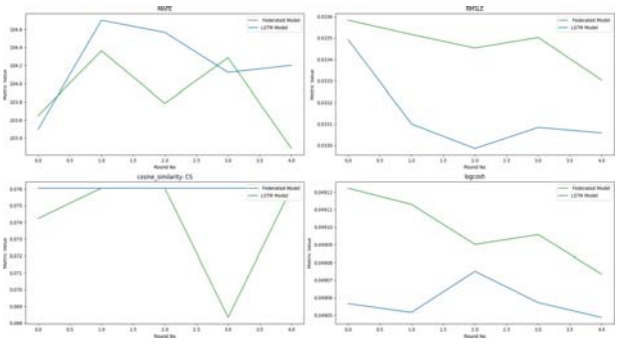


TABLE 6: MODEL PARAMETER VALUES DURING THE TRAINING PROCESS.

KPI Name	LSTM Model	Federated LSTM Model
Loss	Huber	Huber
Optimizer	SGD, ADAM	SGD, ADAM
Metrics	All	All
Data size	319,073 samples	319,073 samples
Batch size	4	4

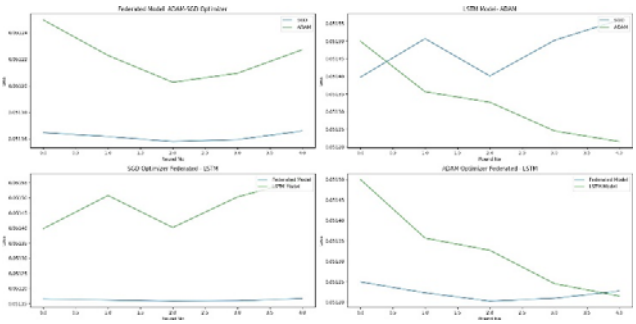


Fig. 11 SGD vs. Adam optimizer loss comparison.

8.2 Window size:

- In this study, the window size parameter was changed multiple times to determine the efficacy of changing the window size on the performance of the proposed models.
- Table 7: shows the model parameter values during the model training process.

TABLE 7: MODEL PARAMETER VALUES DURING THE TRAINING PROCESS.

KPI Name	LSTM Model	Federated LSTM Model
Loss	Huber	Huber
Optimizer	SGD	SGD
Metrics	All	All
Data size	319,073 samples	319,073 samples
Batch size	4	4
Window size	60,100,150,200	60,100,150,200
Dataset split ratio	80% training - 20% testing	80% training - 20% testing
Round number	5	5

- After performing the model training, with different values of window size, we found that In the federated LSTM model, a lower
  - Window size resulted in a lower prediction loss, as shown in Fig. 12.
  - In the traditional LSTM model, a lower window size resulted in a lower prediction loss, as shown in Fig. 12.

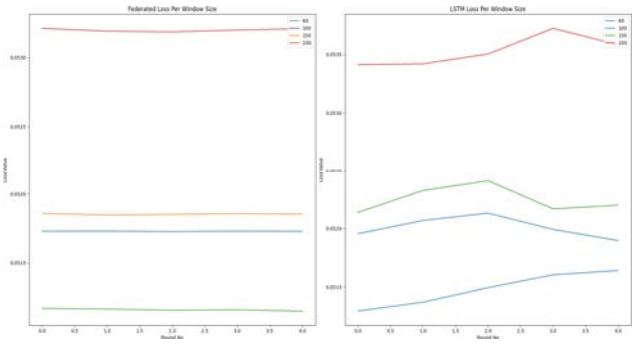


Fig. 12 Window size loss comparison.

8.3 Batch size:

- In this study, the batch size parameter was changed multiple times to determine the efficacy of changing batch size on the performance of the proposed models.
- Table 8: contains the model parameter values during the model training process.
- After performing the model training, with different values of window size, we found that
  - In the federated LSTM model, a lower batch size resulted in a slightly lower prediction loss, as shown in Fig. 13.
  - In the traditional LSTM model, a higher batch size resulted in a lower prediction loss, as shown in Fig. 13.

TABLE 8: MODEL PARAMETER VALUES DURING THE TRAINING PROCESS.

KPI Name	LSTM Model	Federated LSTM Model
Loss	Huber	Huber
Optimizer	SGD	SGD
Metrics	All	All
Data size	319,073 samples	319,073 samples
Batch size	4,8,16,32	4,8,16,32
Window size	60	60
Dataset split ratio	80% training - 20%testing	80% training - 20%testing
Round number	5	5

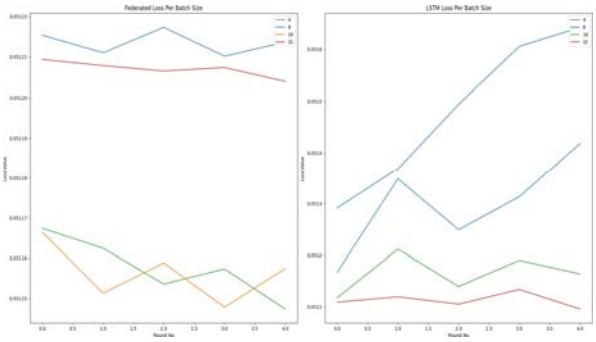
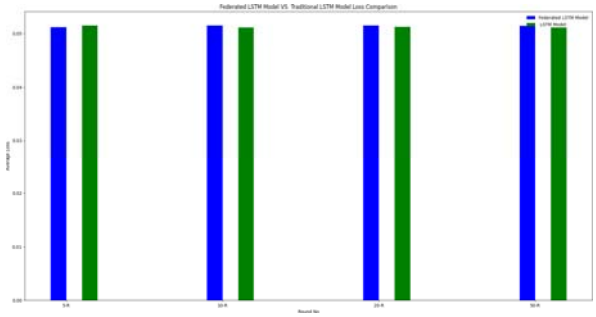


Fig. 13 Batch size loss comparison.

8.4 Split ratio:

- In this study, the split ratio parameter was changed multiple times to determine the efficacy of changing the split ratio on the performance of the proposed models.
- Table 9: shows the model parameter values during the model training process.
- After performing the model training, with different ranges of split ratios, we found that
  - In the federated LSTM model, a higher

Manuscript received April 5, 2022
Manuscript revised April 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.4.1



split ratio resulted in a lower prediction loss, as shown in Fig. 14.

- In the traditional LSTM model, a higher split ratio resulted in a lower prediction loss, as shown in Fig. 14.

TABLE 9: MODEL PARAMETER VALUES DURING THE TRAININGPROCESS.

KPI Name	LSTM Model	Federated LSTM Model
Loss	Huber	Huber
Optimizer	SGD	SGD
Metrics	All	All
Data size	319,073 samples	319,073 samples
Batch size	4	4
Window size	60	60
Data split ratio	80% training - 20% testing 60% training - 40% testing 50% training - 50% testing	80% training - 20% testing 60% training - 40% testing 50% training - 50% testing
Round number	5	5

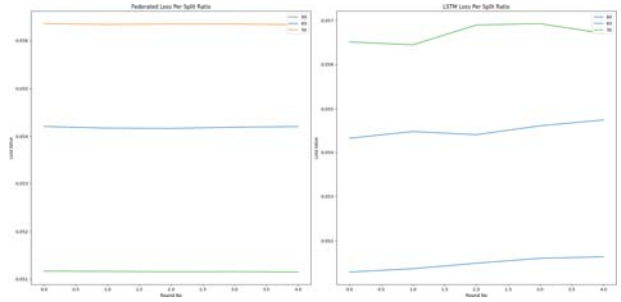


Fig. 14. Split ratio loss comparison.

8.5 Number of rounds:

- In this study, the number of rounds parameter was changed multiple times, to determine the efficacy of changing the number of rounds onthe performance of the proposed models.
- Table X shows the model parameter values during the model training process.

TABLE 10: MODEL PARAMETER VALUES DURING THE TRAININGPROCESS.

KPI Name	LSTM Model	Federated LSTM Model
Loss	Huber	Huber
Optimizer	SGD	SGD
Metrics	All	All
Data size	319,073 samples	319,073 samples
Batch size	4	4
Window size	60	60
Data split ratio Round number	80% training - 20%testing 5,10,20,50	80% training - 20%testing 5,10,20,50

- After performing the model training with different number of rounds, we found that
  - In the federated LSTM model, a higher number of rounds resulted in a lower prediction loss, as shown in Fig. 15.
  - In the traditional LSTM model, a higher number of rounds resulted in a lower prediction loss, as shown in Fig. 15.

Fig. 15. Number of rounds loss comparison.

9. CONCLUSION

In this paper, we compared federated machine learning withtraditional machine learning by:

- Introducing a federated LSTM machine learning model and traditional LSTM machine learning model to forecast the number of crimes periodically.
- Applying the two proposed models, traditional LSTM and federated LSTM, on the Boston crime dataset, wefound that
  - The proposed federated model had a lowerprediction loss, MAE, MSE, RMSE, and log_coch_error than the proposed traditional model.
  - The proposed federated model had a slightly higher prediction RMSLE, cosine_similarity and MAPE than the proposed traditional model.
- Conducting several attempts to determine which model parameters had an effect on the model performance, we found that
  - Model optimizer
    - In the federated LSTM model, the SGD optimizer resulted in lower loss and accuracy metrics than the Adam optimizer.
    - In the traditional LSTM model, the SGD optimizer resulted in lower loss and accuracy metrics than the Adam optimizer.
  - Window size
    - In the federated LSTM model, a lower window size resulted in a

lower predictionloss.
- In the traditional LSTM model, a lower window size resulted in lower prediction loss.

- Batch size
  - In the federated LSTM model, a lower batch size resulted in a slightly low pre- diction loss.
  - In the traditional LSTM model, a higher batch size resulted in a lower prediction loss.
- Split ratio
  - In the federated LSTM model, a higher split ratio resulted in a lower prediction loss.
  - In the traditional LSTM model, a higher split ratio resulted in a lower prediction loss.
- Number of rounds
  - In the federated LSTM model, a higher number of rounds resulted in a low pre- diction loss.
  - In the traditional LSTM model, a higher number of rounds resulted in a low prediction loss.

10. FUTURE WORK

Swarm intelligence algorithms will be used in the future to optimize the proposed federated model for global optimization and reduce the communication overhead.

REFERENCES

[1] Kim, S., Joshi, P., Kalsi, P.S. and Taheri, P. , 2018, November. Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEM-CON) (pp. 415-420). IEEE.

[2] Ivan, N., Ahishakiye, E., Omulo, E.O. and Taremwa, D. ,, 2017. Crime Prediction Using Decision Tree (J48) Classification Algorithm.

[3] Zhang, Weishan, et al. Dynamic fusion-based federated learning for COVID-19 detection. IEEE Internet of Things Journal (2021).

[4] Lian, Xiangru, et al. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. arXiv preprint arXiv:1705.09056 (2017).

[5] Yang, Qiang, et al. Federated machine learning: Concept and applica-tions. ACM Transactions on Intelligent Systems and Technology (TIST) 10.2 (2019): 1-19.

[6] Abdul Salam, M., Taha, S. and Ramadan, M.,2021. COVID-19 detection using federated machine learning. Plos one, 16(6), p.e0252573.

[7] Li, Tian, et al. Federated learning: Challenges, methods, and future directions IEEE Signal Processing Magazine 37.3 (2020): 50-60.

[8] Kim, S., Joshi, P., Kalsi, P.S. and Taheri, P., 2018, November. Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEM-CON) (pp. 415-420). IEEE.

[9] Reier Forradellas, R.F., N´a˜nez Alonso, S.L., Jorge-Vazquez, J. and Rodriguez, M.L., 2021. Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction. Social Sciences, 10(1), p.4.

[10] Zhang, X., Liu, L., Xiao, L. and Ji, J., 2020. Comparison of machine learning algorithms for predicting crime hotspots. IEEE Access, 8, pp.181302-181310.

[11] Wheeler, A.P. and Steenbeek, W., 2021. Mapping the risk terrain forcrime using machine learning. Journal of Quantitative Criminology, 37(2), pp.445-480.

[12] Bappee, F.K., Junior, A.S. and Matwin, S., 2018, May. Predicting crime using spatial features. In Canadian Conference on Artificial Intelligence (pp. 367-373). Springer, Cham.

[13] Prabakaran, S. and Mitra, S., 2018, April. Survey of analysis of crime detection techniques using data mining and machine learning. In Journal of Physics: Conference Series (Vol. 1000, No. 1, p. 012046). OP Publishing.

[14] Ramasubbareddy, S., Srinivas, T.A.S., Govinda, K. and Manivannan, S.S., 2020. Crime prediction system. Innovations in Computer Science and Engineering, pp.127-134.

[15] Chun, S.A., Avinash Paturu, V., Yuan, S., Pathak, R., Atluri, V. and R. Adam, N., 2019, June. Crime prediction model using deep neural networks. In Proceedings of the 20th Annual International Conference on Digital Government Research (pp. 512-514).

[16] Nguyen, T.T., Hatua, A. and Sung, A.H., 2017. Building a learning machine classifier with inadequate data for crime prediction. Journal of Advances in Information Technology Vol, 8(2).

[17] Hajela, G., Chawla, M. and Rasool, A., 2020. A clustering based hotspot identification approach for crime prediction. Procedia Computer Science, 167, pp.1462-1470.

[18] Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J. and Wang, F., 2021.Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5(1), pp.1-19.

[19] Li, Q., He, B. and Song, D., 2021. Model-Contrastive Federated Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10713-10722).

[20] Zhang, Weishan, et al. "Dynamic fusion-based federated learning for COVID-19 detection." IEEE Internet of Things Journal (2021).

Mustafa Abdul Salam was born on 1st November, 1981 in Sharkia, Egypt. He received the BS from Faculty of Computers Informatics, Zigzag University, Egypt in 2003 with grade very good with honor, and obtains master degree in information system from faculty of computers and information, Menufia university, Egypt in 2009 specializing in Hybrid Machine Learning and Bio Inspired Optimization algorithms. He obtained his Ph.D. degree in information system from faculty of computers and information, Cairo University, Egypt. He is currently a Lecturer in Scientific Computing and AI department, Faculty of Computers and Information, Benha University, Egypt. He has worked on a number of research topics. 8Mustafa has contributed more than 30+ technical papers in the areas of neural networks, support vector machines, optimization, and time series prediction, and extreme learning machine, hybrid CI models in international journals, international conferences, local journals and local conferences. His majors are Machine Learning, Big Data, Stream Data Mining, and Deep Learning.

11. **Sanaa Taha** received her B.Sc. (2001) and M.Sc. (2005) degrees from the Department of Information Technology, Faculty of Computers and Information, Cairo University, Egypt, and a Ph.D. degree (2013) in Electrical and Computer Engineering from the University of Waterloo, Canada. She is currently an assistant professor in the Department of Information Technology, Faculty of Computer and Information, Cairo University, Cairo, Egypt. Her research interests include wireless network security, mobile networks security, mobility management, and applied cryptography.

Mohamed Ramadan was born on 1st 25October, 1983 in Qalyubia, Egypt. He received the B.Sc. from Faculty of Electronic Engineering Menofia University, Egypt in 2005 with grade very good with honor.