University of Baghdad
College of Engineering
JE
JOURNAL OF ENGINEERING

Journal of Engineering

journal homepage: www.jcoeng.edu.iq

Volume 31 Number 4 April 2025

Comparative Analysis of The Combined Model (Spatial and Temporal) and Regression Models for Predicting Murder Crime

Layth S. Ibrahim *, **Ghadeer Jasim Mohammed**

Department of Mathematics, College of Science, University of Baghdad, Baghdad, Iraq

ABSTRACT

This research dealt with the analysis of murder crime data in Iraq in its temporal and spatial dimensions, then it focused on building a new model with an algorithm that combines the characteristics associated with time and spatial series so that this model can predict more accurately than other models by comparing them with this model, which we called the Combined Regression model (CR), which consists of merging two models, the time series regression model with the spatial regression model, and making them one model that can analyze data in its temporal and spatial dimensions. Several models were used for comparison with the integrated model, namely Multiple Linear Regression (MLR), Decision Tree Regression (DTR), Random Forest Regression (RFR) and Neural Network Regression (NNR). The data used is about the monthly numbers of murder crimes for the police directorates in Baghdad and the governorates during the period from January 2015 to June 2023. The data was analyzed and then divided into two sets, a training and testing set, to perform these models in prediction. The accuracy of each modsl's performance was evaluated using two statistical measures: RMSE and R^2 in order to determine the best and most accurate performing model among the selected models. An important result was obtained in the comparison between these models, as the combined model obtained the most accurate performance than the other models, based on the values of the performance accuracy metrics for each model in relation to the data used in the murder crimes.

Keywords: Multiple Linear Regression (MLR), Random Forest Regression (RFR), Neural Network Regression (NNR), Combined model (spatial and temporal), Murder crime.

1. INTRODUCTION

Crime prediction plays an effective role in enhancing safety and reducing losses resulting from crimes. Recent studies have shown that crime prediction is closely related to urban development and the nature of people's lives **(Couch and Dennemann, 2000)**, which creates an urgent need for accurate crime prediction.

*Corresponding author
Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2025.04.02

Article received: 14/09/2024
Article revised: 13/01/2025
Article accepted: 04/02/2025
Article published: 01/04/2025

Here, efforts have focused on finding new and advanced models to achieve the necessary and required accuracy in prediction by knowing the most important economic, statistical and social characteristics of the population, such as educational level **(Ehrlich, 1975)**, ethnic and religious differences **(Braithwaite, 1989)**, and income level **(Patterson, 1991; Kennedy et al., 1998)**. It has also been proven that atmospheric data, such as weather data, are linked to crime **(Cohn, 1990; Ranson, 2014)**.
Criminological theories such as routine activity theory **(Cohen and Felson, 1979)** and rational choice theory **(Clarke and Cornish, 2014)** suggest that crime distribution is largely determined by time and space and as a result, this leads to the fact that temporal and spatial factors play an effective role in crime analysis **(Leong and Sung, 2015)**.
Modern models that have been extended, hybridized or combined help provide promising alternatives in accurate forecasting of different time and spatial series, in addition to traditional models such as multiple linear regression (MLR), decision tree regression (DTR), random forest regression (RFR) and neural network regression (NNR), which is a data analysis technique.
Data forecasting has received great attention from academic and scientific circles to provide future solutions based on historical data. The most important preventive measures in law enforcement management depend on accurate prediction of crime rates.
Many studies have emerged that focused on finding new accurate models suitable for the problem data to be forecasted, other than traditional data, either by extending, hybridizing or combining them. These studies include:
(Zhao and Tang, 2017) explored the presence of temporal and spatial correlations in crime and introduced a preliminary approach to model these correlations within a coherent TCP-based framework for crime prediction. Experimental results using real-world data confirm the effectiveness of the proposed framework.
In the study by **(Catlett et al., 2018)**, an approach leveraging spatial analysis and autoregressive models was introduced to automatically identify high-risk crime zones in urban environments and accurately forecast crime trends in each region. The final outcome of the algorithm is a spatiotemporal crime prediction model, comprising a collection of high-crime areas and corresponding crime predictors, where each predictor serves as a model for estimating the number of crimes expected in its designated area. Their experimental analysis using real-world data from a large area in Chicago showed that the proposed approach effectively predicts spatiotemporal crimes with high accuracy across different time horizons.
In the study **(Yi et al., 2018)**, the Clustered-CCRF model is presented, which is designed to effectively utilize spatial and temporal factors for integrated crime prediction. It has been observed that the crime rate in a particular area is influenced not only by its historical records but also by crime patterns in similar areas. They introduced two main predictors of crime: temporal autoregressive correlation and spatial feature-based correlation between areas, to effectively measure these patterns. They presented a tree-based clustering algorithm to identify regions with high similarity based on spatial features, aiming to improve the performance of their proposed model. Experiments using a real-world crime dataset demonstrated the superiority of their model over existing state-of-the-art methods.
In another study **(Kadar et al., 2019)**, developed machine learning models designed to predict spatial and temporal parameters were specifically adjusted for the unbalanced distribution of class classifications. These models were tested in a real-world setting using advanced predictors, including socioeconomic, geographic, temporal, weather, and crime variables, to achieve high accuracy. The results can assist decision-makers in law

enforcement and enhance public decision-making processes, especially in sparsely populated areas.
In another study **(Sun et al., 2020)**, designed and implemented a comprehensive spatiotemporal deep learning framework called Crime Forecaster, that simultaneously captures both temporal frequency and spatial dependence across and within different regions. They developed a model to capture temporal dependencies using a closed recurrent network with diffusion convolution units, allowing simultaneous modeling of dependencies across regions. Experimental evaluations on two real-world datasets demonstrated the effectiveness of Crime Forecaster, outperforming the leading state-of-the-art algorithm by up to 21%. The Los Angeles crime dataset used in the study was collected over a ten-year period.
In the study conducted by **(Umair et al., 2020)**, focusing on predicting the behavior of criminal networks and turning it into actionable insights using natural language processing in Pakistan, they conducted a hotspot-based spatial analysis. Their approach achieved a maximum accuracy of 92% using the K-Nearest Neighbor (KNN) algorithm and 62% using the Random Forest algorithm. The results also indicated that robberies were the most common crime events.
In the study by **(Li et al., 2018),** seeks to analyze the basic characteristics of urban crimes in China by identifying crime data from original case records. By comparing the predicted crime patterns with actual observations, the validity of these basic characteristics and their governing principles is evaluated. First, a quantitative method based on Chinese descriptions was developed to transform unstructured case record information into a typical safety score. Second, the internal characteristics of the cases are analyzed, including the number of incidents, time of occurrence, and location. Third, a crime prediction model using the autoregressive integrated moving average (ARIMA) method is presented to predict crime trends over time. Experimental results show that the predictions are consistent with actual values and exhibit the same crime patterns.
In the study, a combined regression model is created, which is a model that combines two models, the time series regression model and the spatial regression model, into one model capable of analyzing data in both the temporal and spatial dimensions. For comparison, several combined model models are used, namely multiple Linear Regression, Decision Tree Regression, and Neural Network Regression with two measures: Root Mean Square Error (RMSE) and R-Squared Score (R^2).

2. METHODOLOGY

The methodology used includes the following:

2.1 Data Time Series Analysis Data

Examining the relationship between historical data and future data is referred to as time series analysis **(Alsuwaylimi, 2023)**. Time series data contains a large amount of information in the form of patterns **(Chen et al., 2023)**. This pattern information helps classify the behavior of observed data over a period of **(Hyndman and Athanasopoulos, 2018).**
Murder Crime data for Iraq were used for eight and a half years (102 months) for each governorate from December 2015 to June 2023.

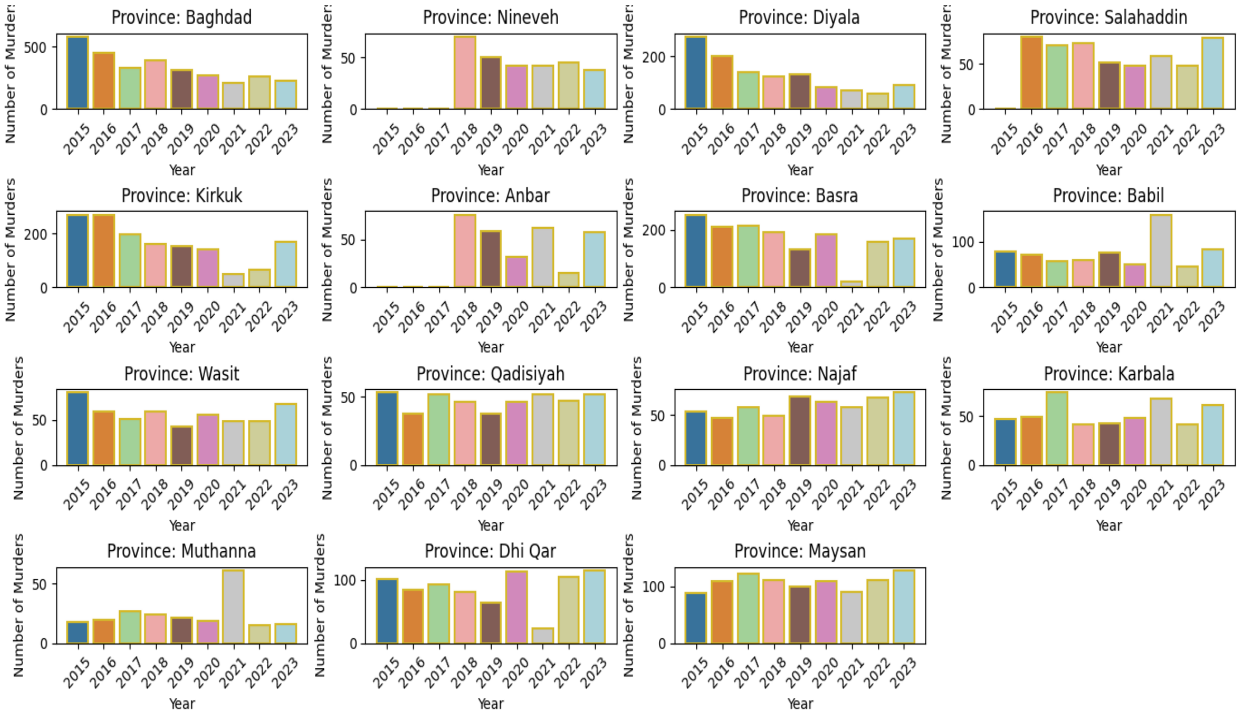

Figure 1. Number of actual murders recorded from 2015 to 2023

The general goal of statistical modeling is to develop a predictive equation that relates a standard variable to one or more predictor variables **(Albayati and Lateif, 2018).** The data include different spatial and temporal variables, such as the population density of each governorate, as well as the crime rate, which represents the number of murder crimes reported per month. Morder crime come in different types. The focus was on four types: premeditated murder, premeditated murder to wash away honor, assault leading to death, and manslaughter. These data were analyzed in the form of time periods representing the distribution of crime numbers from 2015 to 2023 for each governorate and then analyzed according to the four categories mentioned above, as shown in **Figs. 1 and 2**, respectively. The data were then analyzed in monthly and spatial time periods for each governorate, and our study will focus on the latter analysis.

From our analysis, we conclude that murders varied between governorates and decreased slightly in some Iraqi governorates during the past three years, 2021, 2022, and 2023, compared to 2015, 2016, and 2020. Data for the years 2015, 2016, and 2017 were zeroed for Anbar and Nineveh governorates, and data for Salah al-Din governorate were zeroed for the years 2015 and 2016 due to the unavailability of data during that period due to the presence of ISIS terrorist gangs in those governorates.

In **Fig. 2**, it was noticed that the rates of premeditated murder are high and appear in all governorates where murder rates are noticeably high, indicating the existence of a major and serious problem. As for honor crimes, they are present in most governorates, indicating the extent of society's need to address its cultural issues. It was also noted that aggression leading to loyalty is prophetically low, but it exists, and this data illustrates the violent conflicts present in society. Finally, manslaughter comes in second place in terms of rates, and although it is much lower than premeditated murder, it greatly affects the safety of society.

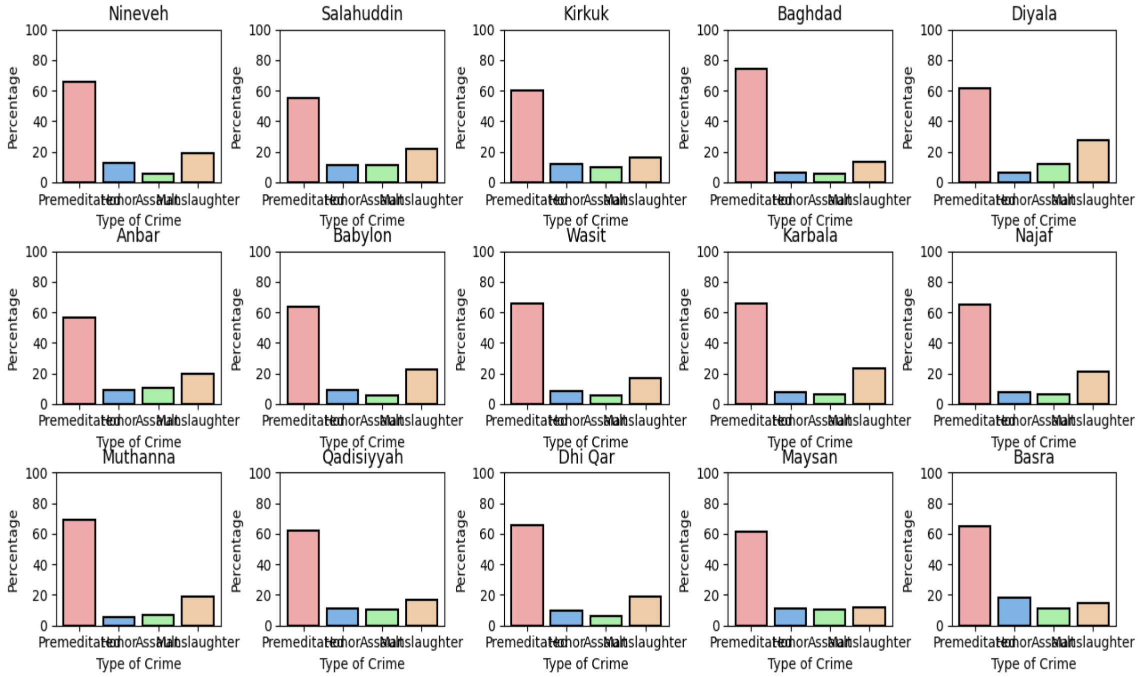

Figure 2. Percentages of selected types of murder crime for each governorate

2.2 Models

Types of regression were used as follows:

2.2.1 Time Series Regression (TSR)

Time series regression is a statistical technique **(Çalışır et al., 2022)** used to forecast future data based on historical data records. This method is referred to as dynamic autoregressive modeling

$$y_t = b_0 + \sum_{j=1}^{k} b_j x_{jt} + u_t \quad (1)$$

Time series analysis is generally suitable for studying the immediate and short-term effects of exposures **(Imai et al., 2015)**.

2.2.2 Spatial Linear Regression (SLR)

Spatial linear regression models can be considered a generalization of standard linear regression models, where spatial autocorrelation is explicitly incorporated and handled by spatial models **(Durbin, 1960)**.

$$\mathrm{y_i} = \mathrm{b_0} + \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{b_i x_i} + \epsilon \quad (2)$$

Spatial regression models, usually with a linear additive specification, define the relationship among areal units exogenously by using weights that represent the spatial structure and the interaction patterns **(Chi and Zhu, 2008).**

2.2.3 Neural Network Regression (NNR)

NNR refers to a subfield of artificial intelligence that mimics the structure of the brain, enabling computers to understand concepts and make decisions like humans **(Alnuaimi and Albaldawi, 2024).** It is a neural network consisting of two or more layers, and NNR is employed to predict time series data **(Zhang, 2012; Mohamed et al., 2016).**

$$y_t = g(y_{t-1}, y_{t-2}, \ldots . y_{t-p}) + e_t \tag{3}$$

where e_t represents the error at time t. The sigmoid function is expressed as follows, noting that it is a biased function:

$$sigmoid(x) = \frac{1}{1+e^{-y}} \tag{4}$$

2.2.4 Decision Tree Regression (DTR)

DT is a rule-based classifier that organizes data based on the values of its attributes. Each node in the tree represents an input feature, while each branch corresponds to a value of that feature **(Obaid and Saleh, 2024)**. A decision tree repeatedly partitions data based on features, with the goal of maximizing information gain (for classification) or minimizing variance (for regression) at each node. A decision tree is a supervised machine learning method used to address classification and regression problems by repeatedly partitioning data according to specific parameters **(Schneider et al., 2010).**

2.2.5 Random Forest Regression (RFR)

Random Forest is a supervised learning algorithm that employs ensemble methods to address both regression and classification problems **(Alnuaimi and Albaldawi, 2024)**. It ensembles several Decision Trees classifiers combined to obtain accurate and robust predictors with overall improvement in outcomes **(Mohammed and Hussein, 2022)**. It is a combination of tree predictions $\{h(x, \Theta_i), i = 1, 2, \ldots\}$, where each tree depends on the values of the random vector $\{ \Theta_i\}$, which are independently sampled and follow the same distribution across all trees in the forest **(Dai et al., 2018; Breiman, 2001).** For a given independent variable, each decision tree provides an opinion on choosing the best outcome **(Dai et al., 2018).** The generalization error of any Decision Trees $h(x)$ is considered.

$$h(x) = E_{X;Y} \left(Y - h(X)\right)^2 \tag{5}$$

Where X is the input vector and Y is the output vector. The expected value of Random Forest Regression is equal to the average value of i Decision Trees $h(x, \Theta_k)$.

2.2.6 Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a widely used statistical technique for establishing relationships between two or more variables. n a multivariate context, the regression model can be expanded to relate Y to a set of p explanatory variables x_1 , x_2 , … , x_p **(Jobson, 1991)**.

$$y_i = b_0 + \sum_{i=1}^{p} b_i x_i + \epsilon_i \tag{6}$$

2.3 Evaluation Metrics

The performance was evaluated using:

2.3.1 Root Mean Square Error (RMSE)

RMSE is one of the most widely used measures for assessing the quality of forecasts **(Gilroy et al., 1990),** which explains the difference between the true value and the predicted value by applying the following formula:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\|Y_i - \hat{Y}_i\|}{N}} \tag{7}$$

2.3.2 R-Squared Score (R^2)

R^2 is a statistical metric that represents the proportion of variance explained by the model **(Rights and Sterba, 2018)**.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{8}$$

3. RESULTS AND DISCUSSION

3.1 Building The Combined Model

To build the combined model, we combine represented by temporal variables and represented by spatial variables to have an algorithm that combines both temporal and spatial properties. The combined model can be represented using the following equation:

$$Y_{it} = b_0 + b_1 X_{1it} + b_2 X_{2it} + \cdots + b_n X_{nit} + \gamma Z_t + u_i + \epsilon_{it} \tag{9}$$

where, Y_{it} is the dependent variable for individual i at time t. X_1, $X_2, \dots, X_n$ are the spatial independent variables, Z_t is the temporal, γ is the coefficient that represents the effect of the time variable Z_t on the dependent variable Y_{it} , u_i the individual-level random effect and ϵ_{it} is the error.

3.2 Comparison with Other Models

It is worth noting that determining which model is an ideal model must depend on the nature of the data and the purpose of analyzing it, and the performance of each model to be most appropriate for the specific task.

Table 1. Comparison between the characteristics of the models (MLR, DTR, NNR**,** CR and RFR)

Property	MLR	DTR	NNR	CR	RFR
Complexity	Low	Medium	High	High	Medium
Handling non-linearity	No	Yes	Yes	Yes	Yes
Ease of Understanding	Easy	Medium	Difficult	Complex	Medium
Interpretability	High	High	Low	Medium	Medium
Adaptability to Large Data	Low	Medium	High	Medium	Medium
Avoiding Overfitting	Relatively easy	Challenging without pruning	Requires careful parameter tuning	Requires careful tuning	Reduces overfitting by averaging

Optimal Applications	Sales, market analysis	Medicine, risk analysis	Patterns, finance, images	Crimes, epidemics, environment	Complex datasets with non-linear relationships

Table 2. Root Mean Square Error (RMSE) accuracy performance calculation for models (MLR, DTR, RFR, NNR, Combined Model).

Directorate	MLR	DTR	RFR	NNR	Combined Model
Baghdad	19.7	21.2	16.9	19	16.2
Nineveh	20.7	22.6	19.5	20.6	17.1
Diyala	20.6	23	18.2	24	15.8
Salahaddin	21.8	24.6	19.4	21.2	17
Kirkuk	23.4	25.6	19.4	21.8	16.2
Anbar	24.7	26.7	20.2	22.9	17.5
Basra	25.1	29.1	20.1	22.2	16.1
Babil	25.5	28.8	20	23.3	15.6
Wasit	22.6	29.8	20.2	23.8	15.4
Qadisiyah	27.9	31.8	20.4	25.3	16.2
Najaf	25	32.6	21.4	25.6	15.8
Karbala	21.5	27.4	22.3	25.2	15.5
Muthanna	26.6	24.9	17.1	21.4	16.7
Dhi Qar	31.8	25.5	32.6	28.6	16.9
Maysan	29	34.3	19.5	24.8	16.1

From the two **Tables 2 and 3**, an important scenario result was obtained in the comparison between these models, as the combined model obtained the most accurate performance than the other models, based on the values of performance accuracy measures in forecasting for each model and each governorate. **Fig. 3** shows the number of actual and expected Murder Crime by months for the Iraqi governorates for the year 2023 using the models Multiple Linear Regression (MLR), Decision Tree Regression (DTR), Random Forest Regression (RFR), Neural Network Regression (NNR) and the combined model (spatial and temporal).

Table 3. The R^2 accuracy performance calculation for models (MLR, DTR, RFR, NNR, Combined Model).

Directorate	MLR	DTR	RFR	NNR	Combined Model
Baghdad	0.75	0.65	0.85	0.80	0.95
Nineveh	0.83	0.73	0.87	0.76	0.91
Diyala	0.77	0.75	0.92	0.81	0.96
Salahaddin	0.73	0.67	0.81	0.74	0.89
Kirkuk	0.63	0.61	0.80	0.79	0.88
Anbar	0.71	0.65	0.86	0.70	0.90
Basra	0.75	0.69	0.88	0.77	0.93
Babil	0.70	0.73	0.90	0.75	0.97
Wasit	0.75	0.78	0.89	0.80	0.94
Qadisiyah	0.74	0.69	0.86	0.77	0.89
Najaf	0.60	0.63	0.81	0.74	0.85
Karbala	0.66	0.67	0.82	0.76	0.87
Muthanna	0.62	0.75	0.79	0.72	0.88
Dhi Qar	0.69	0.72	0.80	0.73	0.83
Maysan	0.70	0.75	0. 81	0.80	0.85

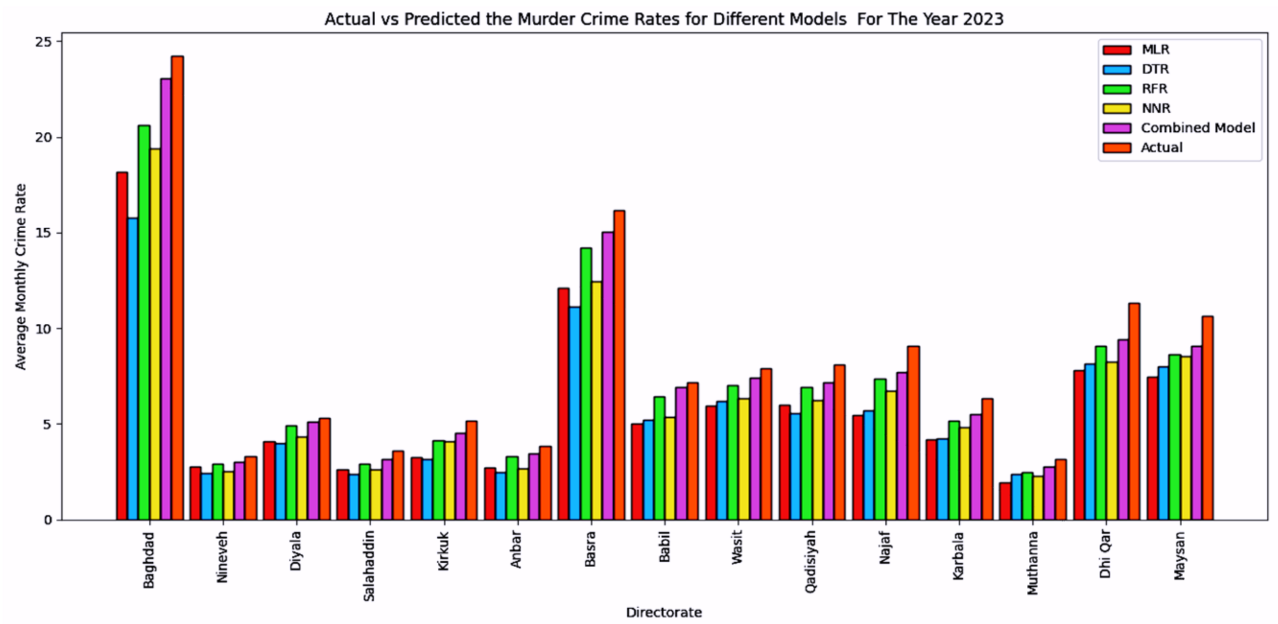

Figure 3. Projected the murder crime rates using regression models for the year 2023

Fig. 3 shows tables of the number of actual murders recorded in 2023 in the police directorates in Baghdad and the governorates, compared with the number of expected murders in 2023 using the combined model.

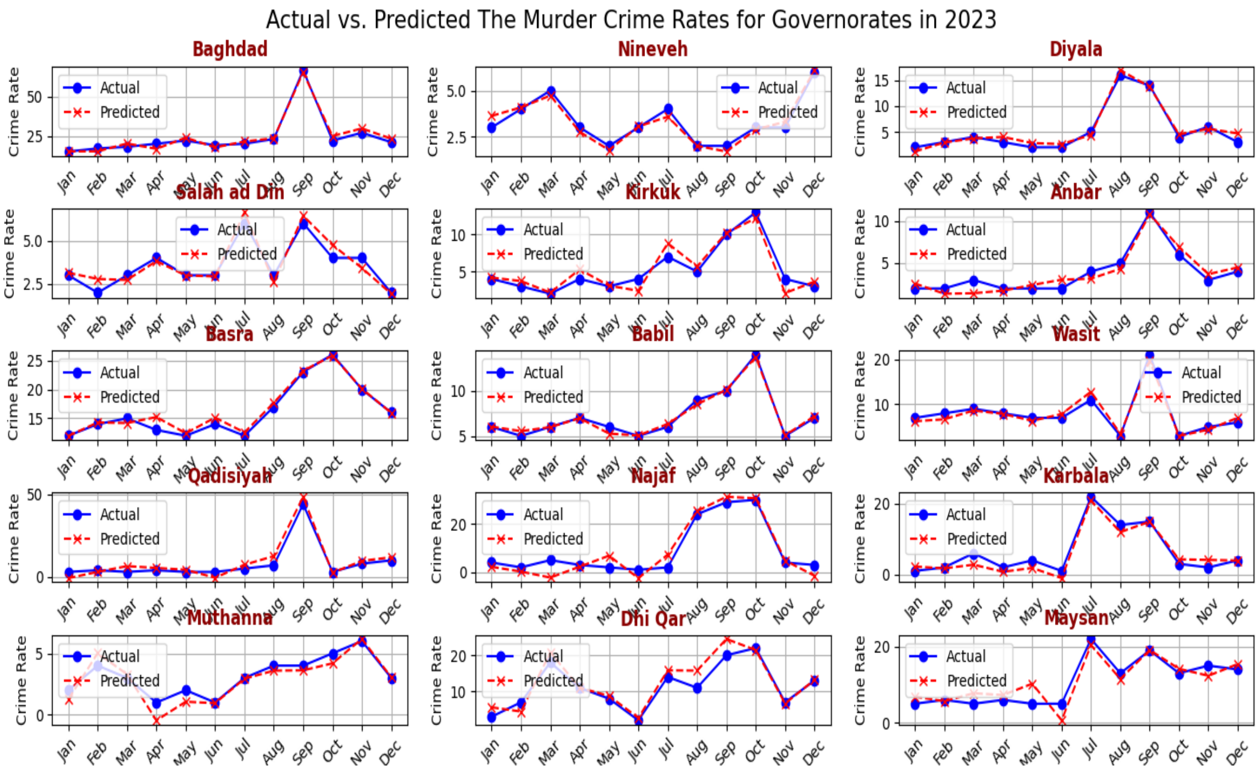

Figure 4. Actual Vs. Predicted Crime Rates for Govemorates by month using the combined model (spatial and temporal).

4. CONCLUSIONS

The most important results that we reached are through the two tables above (2 and 3), which represent performance accuracy measures, which are (RMSE) and R^2 respectively. We found that the combined model gave distinctive results that outperformed all regression models with which it was compared, namely, Multiple Linear Regression (MLR), Decision Tree Regression (DTR), Random Forest Regression (RFR) and Neural Network Regression (NNR). However, the results of Random Forest Regression (RFR) were good compared to the rest of the models in terms of performance accuracy, as it was very close to the combined model.

Finally, traditional models in time series and spatial data at the same time may give better performance in some other data, but innovative or combined models that were processed to have common characteristics and features gave distinctive and promising results.

In the future, a new model can be built either by hybridizing two traditional models, expanding the traditional model, or combining other models to address the problems of traditional models in the accuracy of performance in forecasting the data used.

Acknowledgment

We would like to thank the Iraqi Ministry of Interior and its Criminal Statistics Department for providing access to the data set and to everyone who helped us complete this research.

Credit Authorship Contribution Statement

Layth S. Ibrahaim: Investigation, Methodology, Formal analysis, and Writing – original draft
Ghadeer Jasim Mohammed: Supervision and proofreading

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

REFERENCES

Albayati, A. H. and Lateif, R. H., 2018. Statistical analysis of mortality and morbidity due to traffic accidents in Iraq. *Journal of Engineering*, 24(1), pp. 20-40. https://doi.org/10.31026/j.eng.2018.01.02

Alnuaimi, A. F. and T. H. Albaldawi, 2024. An overview of machine learning classification techniques. In *BIO Web of Conferences. EDP Sciences*. https://doi.org/10.1051/bioconf/20249700133

Alsuwaylimi, A. A., 2023. Comparison of arima, ann and hybrid arima-ann models for time series forecasting. *Information Science Letters An International Journal*, 12(2), pp. 1003-1016. http://dx.doi.org/10.18576/isl/120238

Braithwaite, J., 1989. Crime, shame and reintegration. Cambridge Press.

Breiman, L., 2001. Machine Learning. Kluwer Academic Publishers, 45. pp. 5-32. https://doi.org/10.1023/A:1010933404324

Çalışır, K., Krczal, G. and Uslu, V.V., 2022. Small RNA-seq dataset of wild type and 16C Nicotiana benthamiana leaves sprayed with naked dsRNA using the high-pressure spraying technique. *Data in Brief*, *45*, p.108706. https://doi.org/10.1016/j.dib.2022.108706

Catlett, C., Cesario, E.; Talia, D. and Vinci, A., 2018 A data-driven approach for spatio-temporal crime predictions in smart cities. *In 2018 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, June*. pp.17-24. https://doi.org/10.1109/SMARTCOMP.2018.00069

Chen, Z., Ma, M., Li, T., Wang, H., and Li, C., 2023. Long sequence time-series forecasting with deep learning: A survey. Information Fusion, 97, P. 101819. https://doi.org/10.1016/j.inffus.2023.101819

Chi, Guangqing and Jun Zhu, 2008. Spatial regression models for demographic analysis. Population Research and Policy Review, 27 , pp.17-42. https://doi.org/10.1007/s11113-007-9051-8

Cohen, L. E.; Felson, M., 2003. Social change and crime rate trends: A routine activity approach. Crime: Critical Concepts in Sociology, 1, 316. ISBN-0-415- 25265-2.

Couch, C.; Dennemann, A., 2000. Urban regeneration and sustainable development in Britain: The example of the Liverpool Ropewalks Partnership. *Cities*, 17(2), pp.137-147. https://doi.org/10.1016/S0264-2751(00)00008-1

Clarke, R. V., 2014. Introduction to the transaction edition. The reasoning criminal: Rational choice perspectives on offending, ix-xvi. ISBN-978-1-4128-5275-3

Cohn, E. G. 1990. Weather and crime. *The British Journal of Criminology*, 30(1), pp. 51-64. https://doi.org/10.1093/oxfordjournals.bjc.a047980

Dai, B., Gu, C., Zhao, E., and Qin, X., 2018. Statistical model optimized random forest regression model for concrete dam deformation monitoring. tructural Control and Health Monitoring 25 (6), P.2170. https://doi.org/10.1002/stc.2170

Durbin, James, 1960. Estimation of parameters in time-series regression models. Journal of the Royal Statistical Society Series B: *Statistical Methodology*, 22, pp. 139-153. https://doi.org/10.1111/j.2517-6161.1960.tb00361.
Ehrlich, I., 1975. On the relation between education and crime. *In Education, income, and human behavior*. NBER. ISBN: 0-07-010068-3

Gilroy, E., Hirsch, R. M. and Cohn, T. A., 1990. Mean square error of regression-based constituent transport estimates. *Water Resources Research*. 26(9). pp. 2069-2077. https://doi.org/10.1029/WR026i009p02069

Hussein, M. A., 2022. Performance analysis of different machine learning models for intrusion detection systems*. Journal of Engineering*, 28(5), pp. 61-91. https://doi.org/10.31026/j.eng.2022.05.05

Hyndman, R. J. and G. Athanasopoulos, 2018. *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. https://www.otexts.org/fpp.

Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P. and Hashizume, M., 2015. Time series regression model for infectious disease and weather. *Environmental Research*, 142, pp. 319-327. https://doi.org/10.1016/j.envres.2015.06.040

Kadar, C., Maculan, R and Feuerriegel, S., 2019. Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. *Decision Support Systems*, 119, pp. 107-117. https://doi.org/10.1016/j.dss.2019.03.001

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., Lochner, K. and Gupta, V., 1998. Social capital, income inequality, and firearm violent crime. *Social Science and Medicine*, 47(1), pp. 7-17 . https://doi.org/10.1016/S0277-9536(98)00097-5

Leong, K., and Sung, A., 2015. A review of spatio-temporal pattern analysis approaches on crime analysis. *International E-Journal of Criminal Sciences*, (9).

Mohamed, S. R., Burhan, A. M. M. and Hadi, A. M. A., 2016. Calculating the transport density index from some of the productivity indicators for railway lines by using neural networks. *Journal of Engineering*, 22(9), pp. 1-19. https://doi.org/10.31026/j.eng.2016.09.14

Obaid, M. M., and Saleh, M. H., 2024. Efficient intrusion detection through the fusion of ai algorithms and feature selection methods. *Journal of Engineering*, 30(07), pp. 184-201 . https://doi.org/10.31026/j.eng.2024.07.11

Patterson, E. B., 1991. Poverty, income inequality, and community crime rates. *Criminology*, 29(4), pp. 755-776 .https://doi.org/10.1111/j.1745-9125.1991.tb01087

Ranson, M., 2014. Crime, weather, and climate change. Journal of environmental economics and management, 67(3), pp. 274-302. https://doi.org/10.1016/j.jeem.2013.11.008

Rights, J.D. and Sterba, S.K., 2018. A framework of R-squared measures for single-level and multilevel regression mixture models. *Psychological Methods*, *23*(3), p.434.https://doi/10.1037/met0000139

Schneider, A., Hommel, G. and Blettner, M., 2010. Linear Regression Analysis. Part 14 of a Series on Evaluation of Scientific Publications. 107(44). *Deutsches Ärzteblatt International*. pp. 776-782. https://doi.org/10.3238%2Farztebl.2010.0776

Sun, J., Yue, M., Lin, Z., Yang, X., Nocera, L., Kahn, G. and Shahabi, C., 2020. Crime forecaster: Crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies. In Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*: European Conference, ECML PKDD 2020 Part V. Ghent, Belgium: Springer International Publishing*, pp. 52-67

Jobson, J. D, 1991. Multiple Linear Regression. In: Applied Multivariate Data Analysis. Springer Texts in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-3-030-67670-4_4

Yi, F., Yu, Z., Zhuang, F., Zhang, X., and Xiong, H., 2018. An integrated model for crime prediction using temporal and spatial factors. *In 2018 IEEE International Conference on Data Mining (ICDM). IEEE, November.* pp. 1386-1391. https://doi.org/10.1109/ICDM.2018.00190

Zhang, G.P. 2012. Neural networks for time-series forecasting. In: Handbook of Natural Computing. *Springer, Berlin, Heidelberg*. pp. 461–477. https://doi.org/10.1007/978-3-540-92910-9_14

Zhao, X. and J. Tang., 2017. Modeling temporal-spatial correlations for crime prediction. *In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management., November.* pp. 497-506. https://doi.org/10.1145/3132847.3133024

التحليل المقارن للنموذج المشترك (المكاني والزماني) ونماذج الانحدار للتنبؤ بجرائم القتل

ليث سالم ابراهيم*, غدير جاسم محمد

قسم الرياضيات، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

تناولت هذه الدراسة تحليل بيانات جريمة القتل في العراق بابعادها الزمانية والمكانية . ثم تناولت بناء نموذج جديد بخوارزمية تجمع الصفات التي ترتبط فيها السلاسل الزمنية والمكانية بحيث يكون هذا النموذج قادر على التنبؤ بدقة اداء اكبر من النماذج الاخرى عن طريق مقارنتها بهذا النموذج الذي اطلقنا عليه نموذج الانحدار المدمج و يتكون من دمج نموذجين هما نموذج انحدار السلاسل الزمنية مع نموذج الانحدار المكاني وجعلها نموذج واحد يكون قادر على تحليل البيانات بابعادها الزمانية والمكانية. استخدمنا عدة نماذج لكي يتم مقارنتها بالنموذج المدمج وهي الانحدار الخطي المتعدد ، انحدار شجرة القرار ، انحدار الغابات العشوائية و انحدار الشبكة العصبية . اما البيانات المستخدمة في دراستنا هي بيانات أعداد جرائم القتل الشهرية لمديريات الشرطة في بغداد والمحافظات خلال الفترة من كانون الثاني 2015 الى حزيران 2023. وقد تم تحليل البيانات ومن ثم تقسيمها الى مجموعتين مجموعة تدريب ومجموعة اختبار ، لأداء هذه النماذج في التنبؤ. ومن ثم قمنا بتقييم دقة أداء كل نموذج باستخدام اثنين من المقاييس الإحصائية: ($RMSER^2$,)، وذلك لتحديد النموذج الأفضل والأدق أداءً بين النماذج المختارة. تم الحصول على نتيجة سيناريو مهمة في المقارنة بين هذه النماذج، إذ حصل النموذج المدمج على الأداء الأكثر دقة من النماذج الاخرى ، وذلك بالاعتماد على قيم مقاييس دقة الأداء لكل نموذج فيما يتعلق بالبيانات المستخدمة في جريمة القتل.

الكلمات المفتاحية: الانحدار الخطي المتعدد (MLR)، انحدار الغابة العشوائية (RFR)، انحدار الشبكة العصبية (NNR)، النموذج المركب (المكاني والزماني)، جريمة القتل.