

## Research Article

# Employing Deep Learning and Time Series Analysis to Tackle the Accuracy and Robustness of the Forecasting Problem

Haseeb Tariq,<sup>1</sup> Muhammad Kashif Hanif ,<sup>1</sup> Muhammad Umer Sarwar,<sup>1</sup> Sabeen Bari,<sup>2</sup> Muhammad Shahzad Sarfraz,<sup>3</sup> and Rozita Jamili Oskouei <sup>4</sup>

<sup>1</sup>Department of Computer Science, Government College University, Faisalabad, Pakistan

<sup>2</sup>Griffith College Dublin, Dublin, Ireland

<sup>3</sup>Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan

<sup>4</sup>Department of Computer Science and Information Technology, Mahdishahr Branch, Islamic Azad University, Mahdishahr, Iran

Correspondence should be addressed to Rozita Jamili Oskouei; rozita2020j@gmail.com

Received 3 February 2021; Revised 15 February 2021; Accepted 11 March 2021; Published 1 April 2021

Academic Editor: Shah Nazir

Copyright © 2021 Haseeb Tariq et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Crime is a bone of contention that can create a societal disturbance. Crime forecasting using time series is an efficient statistical tool for predicting rates of crime in many countries around the world. Crime data can be useful to determine the efficacy of crime prevention steps and the safety of cities and societies. However, it is a difficult task to predict the crime accurately because the number of crimes is increasing day by day. The objective of this study is to apply time series to predict the crime rate to facilitate practical crime prevention solutions. Machine learning can play an important role to better understand and analyze the future trend of violations. Different time-series forecasting models have been used to predict the crime. These forecasting models are trained to predict future violent crimes. The proposed approach outperforms other forecasting techniques for daily and monthly forecast.

## 1. Introduction

Urbanization is becoming a global trend [1]. As the city grows, different management challenges increase on a daily basis. Nowadays, crime is a problematic social matter. The crime rate in big cities is higher than in smaller localities. One of the main problems in many countries is the increase in the crime rate in urban areas. With the increasing amount of crimes, crime evaluation methods are needed to reduce the crime [2]. The criminal activities can be reduced by a distribution of patrol officers according to the crime rate. However, it is hard to predict future crimes accurately and efficiently.

Crimes can be categorized into different types such as violent and nonviolent crimes. A violent crime is a crime in which criminals threaten a targeted person. These crimes are considered more serious than nonviolent crimes [3]. A violent act is composed of different offenses such as homicide,

aggravated assault, battery, kidnapping, robbery, murder, and forcible rape [4, 5]. Violent crime may or may not happen with the weapon. Different countries also have distinct methods of recording and crime reporting.

The main challenge is to analyze the increasing volume of criminal data correctly and efficiently [6]. Mostly, security forces lack the tools and skills to recognize effective patterns in these enormous data. Data mining methods can be used to extract valuable information to enhance the efficiency of the city police and enable the officers to make better use of the confined resources. In addition, advanced analytic methods can be integrated with current planning tools. This can enable crime investigators to access huge databases without the need for training from data scientists.

Forecasting is used to project past and present events into the future. Forecasting techniques identify, model, and extrapolate the patterns found in historical data. Forecasting problems can be categorized into short, medium, and long

term based on prediction periods. The majority of forecasting problems use the time series data. A time series is a time-oriented sequence of observations. Time series analysis produces models that can help to understand the underlying causes by using the observed time series. Time series models use the statistical properties of the historical data to predict future patterns and trends [7].

Conventional time-series analysis models such as autoregressive integrated moving-average (ARIMA) [8] and machine learning models such as artificial neural network (ANN) are insufficient to tackle the forecasting problem of criminal data. Researchers have also employed the hybrid models to denoise the data and to find the linear and nonlinear patterns in the data to improve the performance of the forecasting [9, 10].

The objective of the current study is to evaluate the predictive capacity of the models for a short- and medium-term forecast for criminal data. This will help in optimal decision-making and resource management. This work compares different time-series analysis models and machine learning models, i.e., ARIMA, simple exponential smoothing (SES), Holt–Winters exponential smoothing (HW), and recurrent neural network (RNN), to predict the crime trends.

The rest of the paper is organized into different sections. Section 2 discusses different time series forecasting used in this work. Section 3 presents the related work. Section 4 describes the time-series forecasting methodology. Section 5 presents the experimental evaluation of the proposed technique. The outcomes are concluded in Section 6.

## 2. Time Series Forecasting

Structured series of data points listed at an equal-spaced time is called time series. Time series analysis can be separated into two parts. The first part is to obtain the structured underlying pattern of the ordered data. The second part narrates to fit a model for future prediction. The most challenging part that involves mathematical calculations is the fitting part of the time series. Time series can be used for univariate and multivariate analyses [11]. This section discusses different time-series forecasting models to predict future crimes.

**2.1. ARIMA Model.** ARIMA model is a widely used time-series forecasting model introduced by Box and Jenkins in 1970 [12]. ARIMA model is a general linear stochastic model which is the combination of autoregressive and moving-average models [13–15]. An autoregressive model uses a linear combination of past values to predict the variable of interest. The moving-average model uses the past predictions' errors similar to the regression model [16]. It carries a limited number of parameters such as  $(p, d, q)$ , where  $p$  represents the order of the AR model,  $d$  is the degree of differencing, and  $q$  is the moving-average model [17, 18].

$$y_t = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t, \quad (1)$$

where  $\varphi_1, \dots, \varphi_p$  are the parameters for the autoregressive model,  $\theta_1, \dots, \theta_q$  are the parameters for the moving-average model,  $y_{t-1}, \dots, y_{t-p}$  are defined as the past values (lags),  $e_t$  is the white noise, and  $y_t$  is the difference at degree  $d$  of the original series of time series.

**2.2. Exponential Smoothing Methods.** This time series forecasting is used for univariate data. The exponential smoothing technique processes smoothing parameters determined from past data. For prediction, new observations can have greater value than the previous observations [19]. Smoothing variables are determined by minimizing the mean absolute percentage error (MAPE) and root mean square error (RMSE).

**2.2.1. Simple Exponential Smoothing Method.** Simple exponential smoothing is the simplest method that is suitable for stationary series. It is a time-series forecasting approach for a single parameter without a trend and seasonality. SES models are generally based on the assumption that time series should be oscillating at a constant level or slowly changing over time [20]. This method requires little computation. Let  $z_1, z_2, z_3, \dots, z_n$  be a time series. Formally, SES can be computed as

$$\widehat{z}_{t+1} = \alpha z_t + (1 - \alpha) \widehat{z}_t, \quad (2)$$

where  $z_t$  is the actual known series of time period  $i$ ,  $\widehat{z}_t$  is the forecast value of  $\gamma$  for time period  $i$ ,  $\widehat{z}_{t+1}$  is the forecast value for time period  $i + 1$ , and  $\alpha$  is the smoothing constant. The prediction  $\widehat{z}_{t+1}$  is based on the weighted nearest observation  $z_t$ , its weight  $\alpha$ , the weight of the nearest prediction  $\widehat{z}_t$ , and the weight  $1 - \alpha$  [21].

**2.2.2. Holt–Winters Exponential Smoothing Method.** Holt–Winters exponential smoothing method was designed in 1960 by extending the exponential smoothing method. HW is applied when data are in the stationary form. For the calculation of the prediction measures, all the data values need to be in series. This method is suitable when data are with the trend and seasonality [22]. The basic equations applied in each update cycle for level  $L$ , trend  $b$ , seasonality  $S$ , and forecast  $F$  at time  $t$  are

$$\begin{aligned} L_t &= \alpha(x_t - S_{t-c}) + (1 - \alpha)(d_{t-1} + b_{t-1}), \\ b_t &= \beta(d_t - d_{t-1}) + (1 - \beta)b_{t-1}, \\ S_t &= \gamma(x_t - d_t) + (1 - \gamma)S_{t-s}, \\ F_{t+m} &= d_t + b_t m + S_{t-c+m}, \end{aligned} \quad (3)$$

where  $t = 1, 2, 3, \dots$   $L_t$  and  $b_t$  estimate the level of the series and the slope of the series at time  $t$ , respectively.

Exponential smoothing is not suitable for seasonal data including trends or cycles. However, the HW model uses a modified form of exponential smoothing. It applies three exponential smoothing formulae called triple exponential smoothing. First, the average is computed to give locals the average of the series. Second, the trend is smooth, and finally,

smooth each subseries seasonal estimates for each season separately. The exponential smoothing formula applies to a series of trend and constant seasonal elements using HW addition and multiplication methods. An additive method is applied when the season changes through the series are roughly unchanged. The multiplicative method is employed when changes are in proportional series [22]. This study is only applicable to the HW additive model.

**2.3. Recurrent Neural Network.** RNN is a type of ANN which has input, hidden, and output units. Generally, the RNN model has a unidirectional flow of information from input layers to hidden layers. It remembers end-to-end working of the model [23]. Figure 1 explains the RNN framework for modeling time-series observations. A directional loop can help to remember when to make a decision, what is an input of the current node, and what it had learned from the inputs received previously. Using the previous sequence samples may help understand the current sample. RNN can work well on time series because of its capability of remembering the previous input received using the internal memory. This can help to make the RNN forecast accurately.

Long short-term memory (LSTM) networks are modified versions of the RNN that can help to solve the short- and long-term dependencies which make it easier to remember previous data. LSTM networks are trained using back-propagation through time which helps to overcome the vanishing gradient problem. Traditional neural networks have neurons, while LSTM networks have memory blocks connected through sequential layers. Each module contains gates that can handle module status and outputs. The gated formation of the LSTM network manages its memory state. The use of neural networks reduces the need for extensive feature engineering and allows training of large datasets [25].

The difference between LSTM and RNN is an internal unit state which is also transmitted along with the hidden state. The LSTM block receives the input sequence and then uses a gate activation unit to decide if it is dynamic. This action creates a state change and adds information that conditionally passes through the block. Gates make blocks much better than the classic neurons and enable them to memorize current streams.

The weight of the gates can be learned during the training phase. The gating function controls the input, remembers the content in the internal state variables, and handles the output that makes the LSTM unit flexible. In LSTM cells, there are three types of gates, i.e., input, forget, and output (Figure 2). Each unit of LSTM has a cell which has a state  $c_t$  at time  $t$ . The cell read/modify action is controlled using the input gate  $i_t$ , forget gate  $f_t$ , and output gate  $o_t$ . At each time step, the LSTM unit receives the input from two external sources at each of the four terminals, i.e., the three gates and the input [26].

### 3. Related Work

This section discusses the popular existing techniques to predict crimes. However, these techniques can have

constraints. Specific algorithms can be chosen at the identification, feature, and modeling stages. These algorithms can identify and depict natural trends, models, and data relationships.

In recent years, ML algorithms have become increasingly popular and can be used for prediction. Researchers have analyzed the working of criminal activities by using these models in time series such as ARIMA, SES, HW, and RNN models by considering these accuracy metrics. Different researchers have worked on identifications of violations in different states of the United States by examining different datasets. Information such as the trend and seasonality of the crime was extracted to help people and peace enforcement agencies. The crime databases depend on places to identify violation hotspots. There exist a number of online map applications that can show the correct place of the crime and type of offense in any part of the city. Criminal sites can be identified precisely [27]. On the contrary, the historical data and present approaches primarily determine the criminal act [27]. Predictive police are working in Philadelphia where law enforcement agencies highlight and forecast crimes based on locations [28].

Marzan et al. [29] evaluated daily and weekly crime patterns using linear regression, multilayer perceptron, Gaussian processes, and sequential minimal optimization regression. They forecasted the outcome for 10 days and 10 weeks. History is the primary basis of crime forecasting. Cesario et al. [6] used autoregressive models to analyze and forecast crimes in selected regions of Chicago. They examined a number of crimes and violations over time and separated them into trends, seasonality, and random signals. They predicted the crime for one and two years. The downside is an analysis only for a specific area, and it is intended for length prediction.

Moreover, researchers have analyzed the effectiveness and accuracy of algorithms for crime predictions and other potential applications for peace enforcement analysis such as identifying real crime locations, crime profiles, and discovering criminal trends. The most important component is the accuracy of creating new information (based on previous observations), which can reduce the crime rate. Borowik et al. [30] applied prophet forecasting and spectral analysis for real time series in Poland. The authors determined the weekly and annual seasonal patterns for long-period trends in selected sorts of events. There is still a considerable change in crime that cannot be taken by the expected model. It has commonly been assumed that anticipated levels are beneficial for more appropriate allocation of peace enforcement agencies [30].

Chen et al. [17] applied the ARIMA model for short-term forecasting on property crimes. They compared the forecasting results with simple exponential smoothing and Holt two-parameter exponential smoothing model. By the given data for 50 weeks of property crime, they forecasted one week ahead from the given observations using the ARIMA model [17]. However, they only compared straightforward techniques and measured the amount of crime over the whole city and not over districts or grid cells. Their approach used grid cells. The data also lacked historical information.

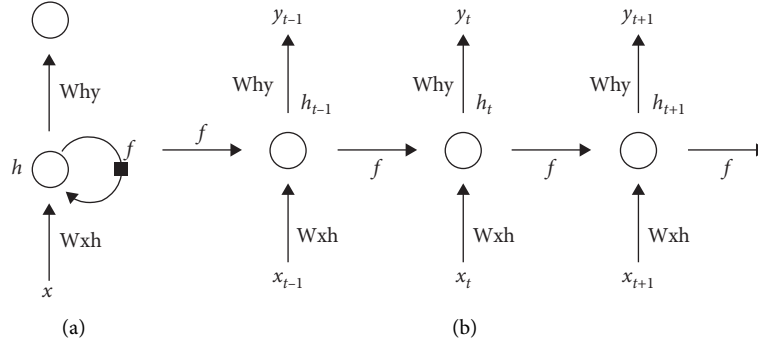


FIGURE 1: Example of (a) folded and (b) unfolded RNN [24].

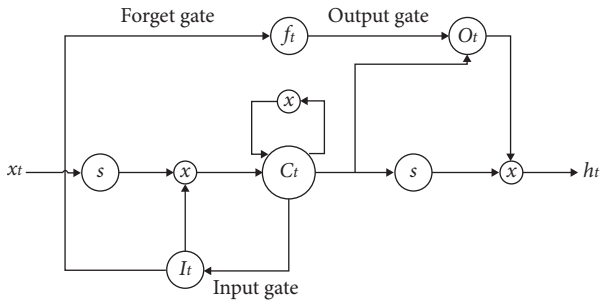


FIGURE 2: LSTM unit and its components [26].

Feng et al. [31] investigated crimes in Chicago, Philadelphia, and San Francisco by applying the Holt–Winters model. Firstly, the authors predicted the trend of crime in the next few years. After that, the category of crime was forecasted for time and location. For this, they collected multiple classes in a larger set and made an attribute selection. The outcomes showed that the tree classification models had performed better on classification tasks when compared with naive Bayesian methods and KNN. Holt–Winters model multiplicative seasonality provided good results when predicting the criminal tendency [31].

Singh [32] described a method to predict the crime for one week by taking 30 days' input of data using LSTM. He compared the performance of different models. Gated recurrent units have good crime prediction performance when compared with the traditional ARIMA model, artificial neural network, convolutional neural network, and RNN with its type [32]. Catlett et al. [33] proposed a predictive approach based on autoregressive and spatial analysis models to detect high-risk crime regions and forecasted crime trends.

Existing techniques have used different algorithms and strategies to forecast different types of data. Some techniques have been used for stationary data, while others are used for univariate data. Moreover, a majority of existing techniques are used to forecast a specific crime and focus on short-term prediction. In this study, the data are made stationary, and autocorrelation is used to find the correlation of lagged values. The proposed technique applies RNN along with LSTM to avoid the exploding gradient and vanishing gradient problems.

## 4. Methodology

This section describes the methods for data collection, preparation of the dataset, model testing, and training. The dataset is collected from the official website of Philadelphia crime [34] through the API. The dataset contains information on different kinds of violent crime from 2006 to 2016. The crime time and location information are used to forecast short- and medium-term crimes.

Figure 3 depicts the methodology used in this research for the violent crime dataset. First, data preprocessing is applied to transform raw data into clean data.

Data preprocessing includes removing unnecessary attributes, filling empty cells, and adding multiple related features. Data cleaning is employed to remove erroneous values. This is the most important and challenging part to achieve high accuracy. The features which contain more than 60% missing values are dropped since they are not helpful for further analysis. Moreover, outliers and duplicate values are filtered out. In the next step, data are standardized and normalized for further analysis.

Dimensionality reduction techniques reduce the high-dimensional data to low-dimensional data. In this study, we have applied the principal component analysis (PCA) method which provides linear mapping based on an eigenvector search. PCA provides different approaches to reduce the feature space dimensionality [35, 36]. In this study, the dataset is split into 70:30 ratio, i.e., 70% of the data is used for training, while 30% is used for the testing purpose.

The most important step in the workflow is choosing an appropriate model. Time series algorithms are used to predict the number of offenses that may occur in the next few years. Time series forecasting can be applied to time-dependent values. In this work, classical statistical methods are used along with machine learning techniques. Next, data exploration techniques are applied to understand the hidden insights of the dataset. Visualizations of a dataset are performed to find the trends and seasonality patterns in the data without transforming or changing the dataset.

Figure 4 illustrates the raw data visualization in the order of total crime occurrences from the observed data on a daily, monthly, and yearly basis, respectively. Crime data are plotted as a time series along the X-axis and the number of

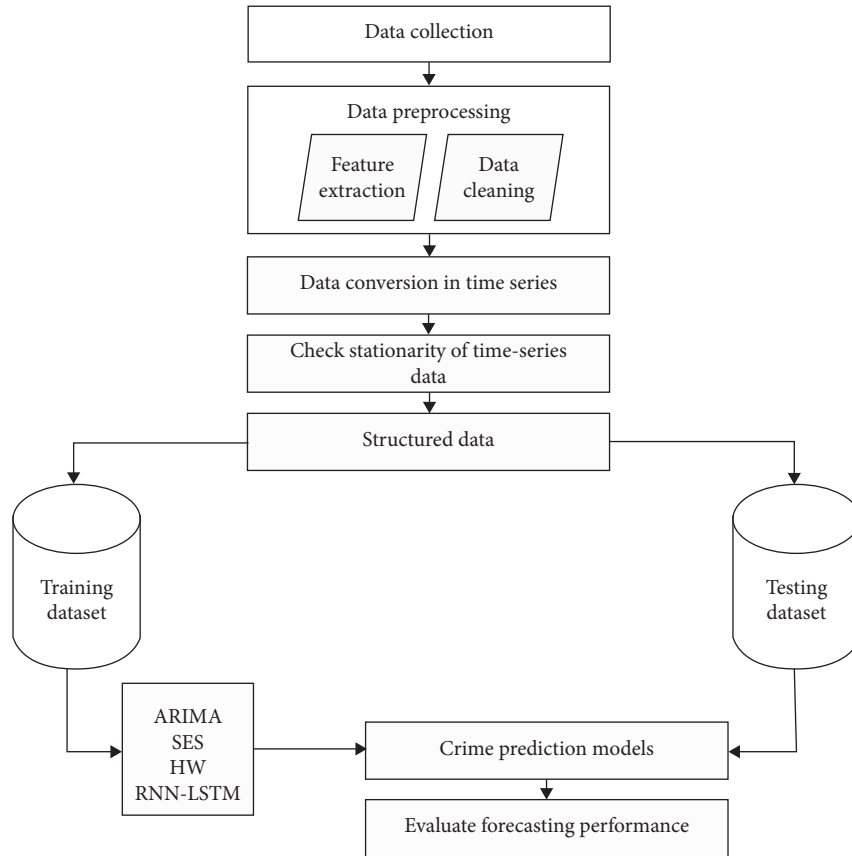


FIGURE 3: Proposed methodology.

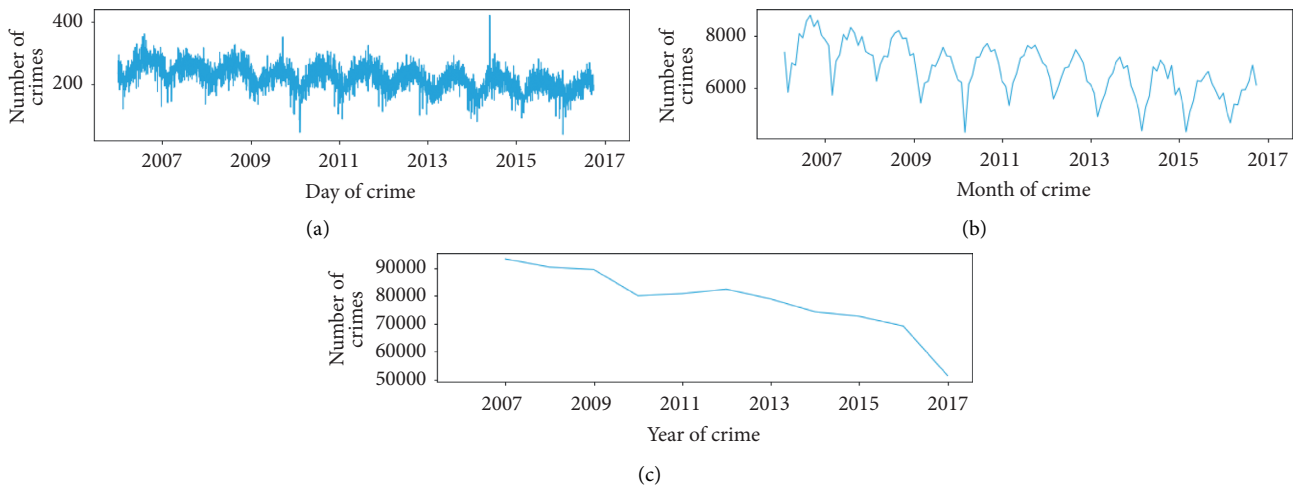


FIGURE 4: (a) Daily violent crimes, (b) monthly violent crimes, and (c) yearly violent crimes from 2006 to 2016.

crime occurrences on the Y-axis. The crime rate was gradually decreasing from 2007 to 2010. However, from 2011 to 2016, the violent crime was oscillating. Figure 4(c) shows the violent crime has a downward trend from the observed data by year (2006 to 2016). From this insight, it is hypothesized that the raw visualization of these data is distributed evenly over the days, but the trend is going down in monthly and yearly data. Crime occurrences by day, month, and year have a clear trend along with seasonal variations in

a dataset. There are many variations present in the daily data; thus, crime data are resampled in monthly data to apply the time series algorithms.

For time series analysis, data must be in a stationary form which means the series should be without trend and with constant variance, mean, and covariance over time. If the data are in the nonstationary form, then they are unpredictable and cannot be forecasted. Stationary data should have a constant mean, variance, and covariance over time.

The data exploration shows that crime data are nonstationary. Therefore, the data need to be converted into a stationarity form to forecast the crimes.

Data were made stationary using rolling statistics mean and augmented Dickey–Fuller (ADF). Rolling statistics mean can be applied to moving mean or moving standard deviation at any instant time  $t$ . This was applied for the  $t$ -statistic,  $p$  value, lags, and the number of observations. The differencing technique is employed through first-order differencing ( $d = 1$ ) on crime data to make data stationary on mean to remove the trend. The variance of the data should also be stationary to obtain reliable forecasts using different forecasting models. This test identifies whether the data consist of a unit root feature that has a severe impact on statistical inference. The unit root test determines the strength of a time series by trend. Many actual datasets are too complex to be captured by simple autoregressive models. Dickey–Fuller test is built on linear regression and is the easiest way to detect the unit root.

In this study, the ADF test is applied to the raw crime data. We applied the difference of lag 1 on the raw crime data where the series is not having a longer trend. Moreover, a difference of lag 12 on raw data is applied to see the trend by removing the seasonality. A double-differencing technique is used in which the series is differenced by lag 12 and then differenced by lag 1. This gives us a double-difference series where there is no trend and seasonality.

The data should also be stationary on the variance to obtain reliable forecasts using ARIMA, SES, and HW models. Therefore, the logarithm is taken to transform the data to make them stationary on variance and to evaluate the influence of seasonality. The resultant integrated part of ARIMA, SES, and HW is equal to one as the first difference makes the series stationary.

Autocorrelation function (ACF) describes the correlation between lagged values of any series at different times [37]. ACF depicts the relationship between the present and past values of the series. It considers time series components such as seasonality, trend, cyclic, and residual to find correlations.

Next, ACF is plotted on stationary crime data to identify the presence of AR and MA components in the residuals (Figure 5). ACF values are shown on a vertical axis which ranges from  $-1$  to  $1$ . The horizontal axis illustrates the size of the lag between the elements of the time series. Daily and monthly ACF sample patterns determine the summarized model processes. The lag refers to the correspondence order. In the daily ACF plot at lag 0, the correlation is 1. The reason is the data are correlated with themselves. At a lag of 1, the correlation is approximately  $-0.4$ . There are enough dotted horizontal lines present that conclude residuals are not random. There is a seasonal component present in the residuals at the lag of 12, and information available can be extracted by AR and MA models.

The prediction methods were applied on univariate data that require the least number of observations prior to starting the models such as ARIMA, SES, HW, and RNN-LSTM. For the parametrized ARIMA model, there are three distinct integers ( $p, d, q$ ), where  $p$  is for the AR model,  $q$  is

for the MA model, and  $d$  is for an integrated part. In the ARIMA, a model fits the importance of the parameters with a certain number of parameters and tests. This means whether the parameters are expressed in unit roots (null hypothesis) or not (alternative hypothesis). The standards such as  $t$ -statistics and  $P$  value are used to evaluate the importance of the parameters considered for the model [38]. We have used ( $p = 1, q = 1, d = 0$ ) to fit the ARIMA model based on ACF results. In SES, the values of the data series are analyzed without trends and seasonality. The stationary data have been used to apply on SES. On the contrary, HW data values are forecasted with trends as well as seasonality. There occur some significant jumps in a few successive time points. After applying SES and HW, the amplitude of fluctuations varies based on the nature of data [39].

Lastly, LSTM is employed along with the RNN as the building unit or extension of the RNN. LSTM can read, write, and delete information or can retain information in its memory. RNN is applied along with LSTM to avoid the exploding gradient and vanishing gradient problem. RNN uses short-term memory where LSTM is working like a gated cell in the form of sigmoid ranging from 0 to 1 which can help backpropagation and keep the gradient steep, so the training is short and accuracy is high. RNN is used to handle the sequence-dependent variables of daily and monthly violent crimes. The normalization technique is applied to the data to make them uniform. LSTM for regression with time steps is applied on the violent crime in which the previous time step is taken in the series as the input to forecast the output at the next time step. This process is applied by setting the columns to be time-step dimension and changing the values of dimension back to 1. In this method, mapping is applied by finding the end of the data pattern, checking the limits of sequence, and gathering input and output parts of the pattern. Then, reshaping is done by taking the current time ( $x = t$ ) which is going to predict the value at the next time in the sequence ( $y = t + 1$ ). The network is trained with 100 epochs, 1 batch size, and 2 verbose.

## 5. Results and Discussion

The time-series prediction techniques have been applied and compared to evaluate the effectiveness and efficiency. In order to perform regression tasks and their validation, the crime data are divided into training and testing data. This study is conducted by using a univariate data structure where UCR\_General is the variable used against Dispatch\_Date\_Time. UCR\_General is the criminal code that is classified into violent crime and property crime.

There exist several ways to measure the accuracy of the forecasting method. For the regression problem, MAPE (equation (4)) and RMSE (equation (5)) are used as error metric measurements. Both MAPE and RMSE are used to evaluate modeling capabilities as well as predictive ability. Any forecast with the MAPE value  $\leq 10\%$  is observed as highly accurate. The value between 10 and 20% is considered good, while 21–50% is supposed to be reasonable. The value greater than 50% is considered inaccurate forecasting [40]. In this study, the MAPE value obtained is less than 10%. The

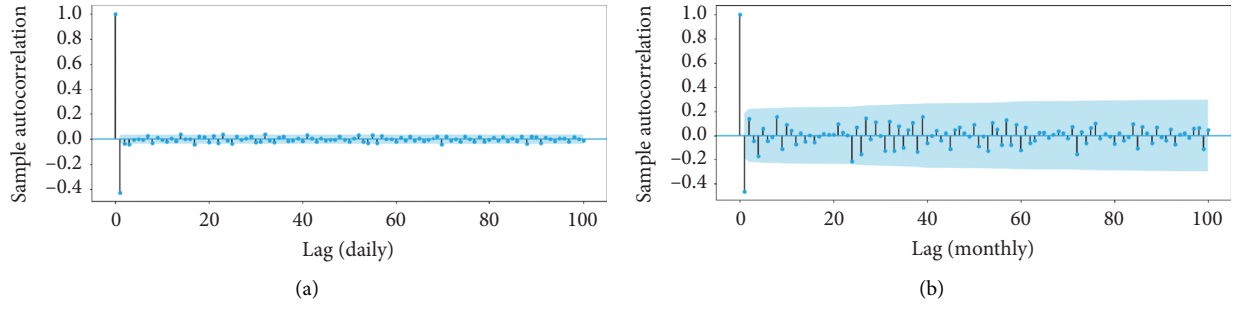


FIGURE 5: (a) Daily ACF and (b) monthly ACF in time series.

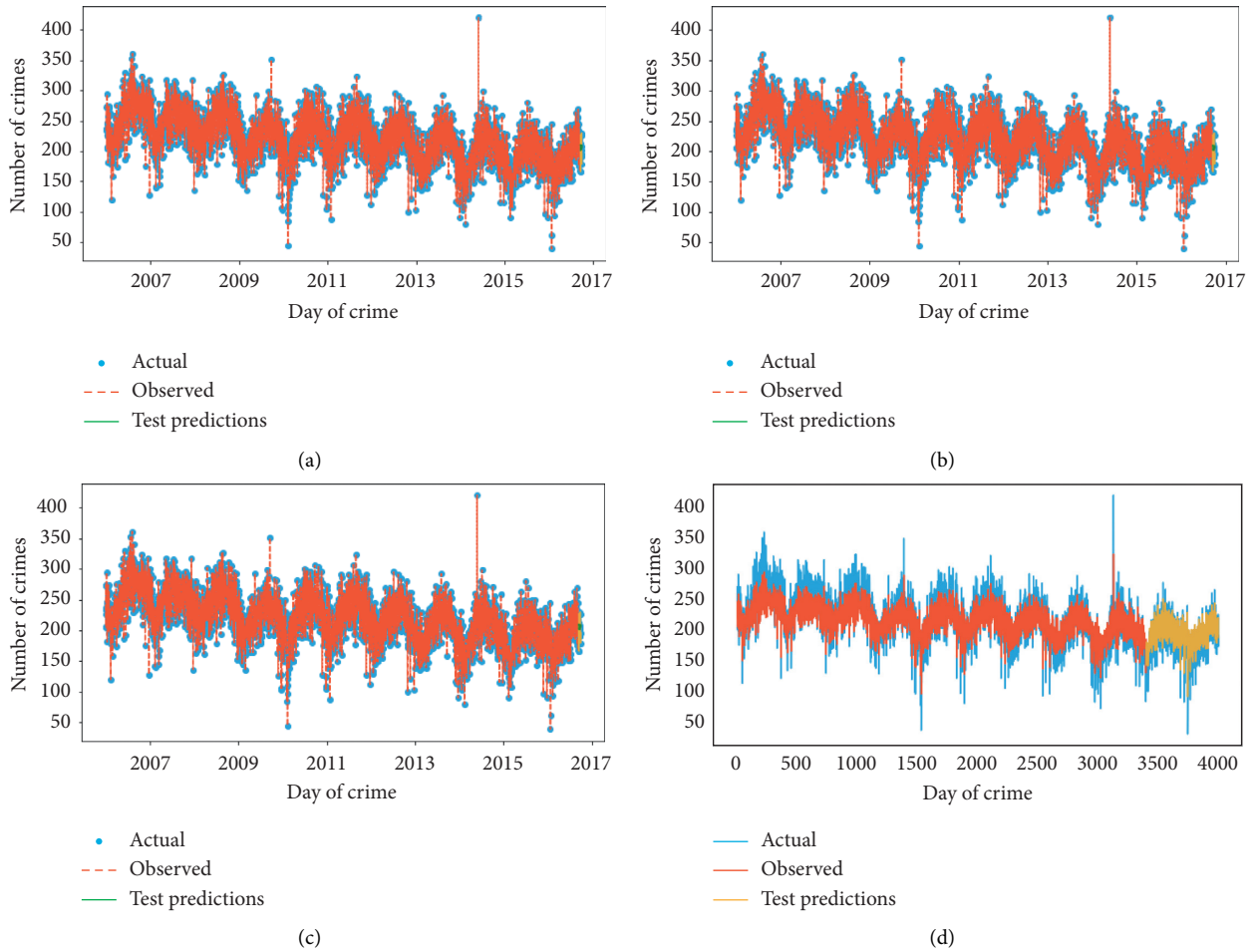


FIGURE 6: Actual and observed values for daily violent crimes by different models. (a) ARIMA model. (b) HW model. (c) SES model. (d) RNN model.

RMSE value can range from 0 to  $+\infty$ , where 0 is the best value and indicates no difference between the values of the modeled and observed data.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{a_t - \hat{a}_t}{a_t} \right|, \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (a_t - \hat{a}_t)^2}. \quad (5)$$

The violent crime data are analyzed in different ranges and periods for different models used in this study



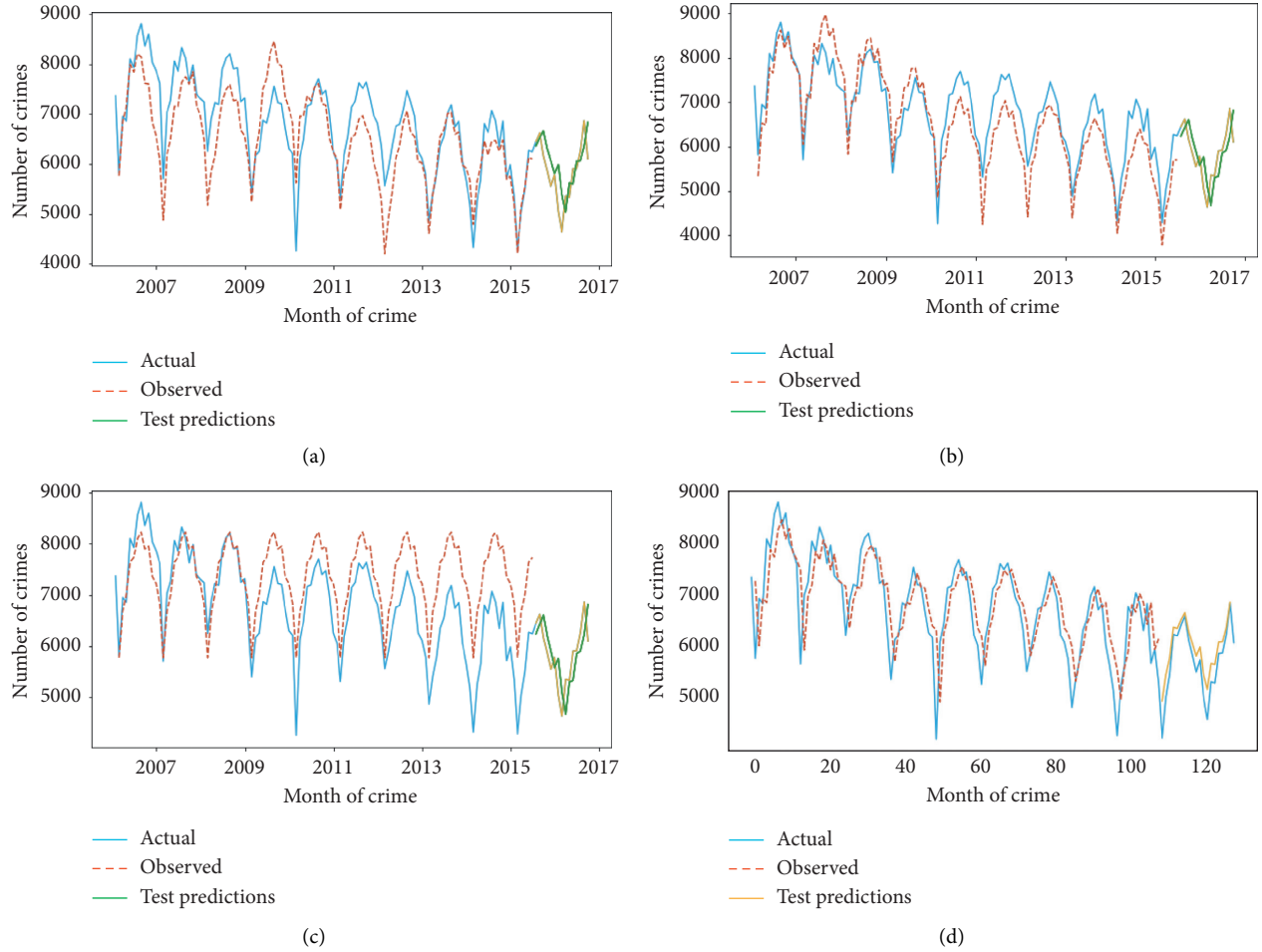


FIGURE 7: Actual and observed values for monthly violent crimes by different models. (a) ARIMA model. (b) HW model. (c) SES model. (d) RNN model.

TABLE 1: Error metric measurements: a performance comparison of daily and monthly violent crimes.

Model	Daily violent crime		Monthly violent crime	
	RMSE	MAPE	RMSE	MAPE
ARIMA model	201.96	81.84	583.09	6.90
HW model	40.60	17.39	536.51	7.20
SES model	39.60	16.37	973.01	13.28
RNN-LSTM model	13.42	4.75	18.42	6.03

(Figures 6 and 7). Each graph represents the number of crime events related to a particular aspect. The trends depict the actual and expected values for daily and monthly crimes by using different time series models. Figure 6 shows the fluctuated series obtained for crimes through a different model. This figure demonstrates the original and predicted values for daily violent crimes by using different models. There are a number of offenses in differing amounts and intervals of time. The violent crime increases in the middle of the day and descends in the evening of the day. There is a downward trend component in daily crimes from 2013 to 2016 period.

Figure 7 depicts the result of the monthly violent crime using different models. There is a series of offenses to different extents and time frames. Violent incident headlines are made on a regular basis in Philadelphia, and the violent crime spikes in summer [28]. The crimes ascend in months of summer (June, July, and August) and descend in months of winter. There is a downward trend component in monthly crimes that have come down around 2016.

Table 1 provides more details about error metrics for daily and monthly crimes. RNN-LSTM has much better performance than the ARIMA, SES, and HW for daily and monthly crime forecast. LSTM using RNN has a higher forecasting accuracy and the lower gap from other models between training and testing errors. The proposed method is useful and can be easily applied to the time-series regression problems.

## 6. Conclusion

The purpose of this study was to develop a time series model using statistical model experimentation and predict the daily and monthly violent crime in Philadelphia. This study performs the comparative analysis of predictive models



based on RMSE and MAPE values. RNN-LSTM has achieved better results than the other models with the values of RMSE 4.75 and MAPE 13.42. The RNN-LSTM model has a higher forecasting accuracy and a lower gap from other models between training and testing errors. These results can help law enforcement agencies in decision-making. In the future, we are interested to develop specific recommendations or targeted crime prevention strategies for different crime prevention models. Moreover, we will perform the scalability analysis and implement the proposed method for different datasets.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Q. Wang, G. Jin, X. Zhao, Y. Feng, and J. Huang, "CSAN: a neural network benchmark model for crime forecasting in spatio-temporal scale," *Knowledge-Based Systems*, vol. 189, Article ID 105120, 2020.
- [2] H. Borrión, P. Ekblom, D. Alrajeh et al., "The problem with crime problem-solving: towards a second generation pop?" *The British Journal of Criminology*, vol. 60, no. 1, pp. 219–240, 2020.
- [3] "The texas crime report for 2017," 2017, <https://www.dps.texas.gov/crimereports/17/citCh2.pdf>.
- [4] "Violent crime - wikipedia," 2016, <https://en.wikipedia.org/wiki/Violentcrime>.
- [5] "Fbi-violent crime," 2011, <https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/violent-crime/violent-crime>.
- [6] E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models," in *Proceedings of the IEEE 14th International Conference on Dependable, Autonomic and Secure Computing*, pp. 795–802, IEEE, Sydney, Australia, 2016.
- [7] C. L. J. Douglas, C. Montgomery, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, Wiley Online Library, Hoboken, NJ, USA, 2015.
- [8] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the arima model," in *Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106–112, IEEE, 2014.
- [9] H. Yu, L. J. Ming, R. Sumei, and Z. Shuping, "A hybrid model for financial time series forecasting-integration of ewt, arima with the improved abc optimized elm," *IEEE Access*, vol. 8, pp. 84 501–584 518, 2020.
- [10] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [11] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017.
- [12] D. Peter and P. Silvia, "Arima vs. arimax—which approach is better to analyze and forecast macroeconomic time series," in *Proceedings of the 30th International Conference Mathematical Methods in Economics*, pp. 136–140, Karviná, Czech Republic, 2012.
- [13] M. Ahmad, A. Khan, M. Mazzara, and S. Distefano, "Seeking optimum system settings for physical activity recognition on smartwatches," in *In Advances in Computer Vision*, K. Arai and S. Kapoor, Eds., pp. 220–233, Springer International Publishing, Cham, Switzerland, 2020.
- [14] M. Ahmad, M. Alqarni, A. Khan et al., "Smartwatch-based legitimate user identification for cloud-based secure services," *Mobile Information Systems*, vol. 2018, Article ID 5107024, 14 pages, 2018.
- [15] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, A. Ali, and A. Tufail, "Extended sammon projection and wavelet kernel extreme learning machine for gait-based legitimate user identification," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pp. 1216–1219, ACM, New York, NY, USA, 2019.
- [16] C. D. R. Rodriguez, D. M. Gomez, and M. A. M. Rey, "Forecasting time series from clustering by a memetic differential fuzzy approach: an application to crime prediction," in *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, IEEE, Honolulu, HI, USA, 2017.
- [17] P. Chen, H. Yuan, and X. Shu, "Forecasting crime using the arima model," in *Proceedings of the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 5, pp. 627–630, IEEE, Jinan, China, 2008.
- [18] H. B. Hwang and H. Ang, "A simple neural network for arma time series," *Omega*, vol. 29, no. 4, pp. 319–333, 2001.
- [19] P. S. Kalekar, "Time series forecasting using holt-winters exponential smoothing," *Kanwal Rekhi School of Information Technology*, vol. 4329008, no. 13, 2004.
- [20] E. Ostertagova and O. Ostertag, "Forecasting using simple exponential smoothing method," *Acta Electrotechnica et Informatica*, vol. 12, no. 3, p. 62, 2012.
- [21] E. Ostertagová and O. Ostertag, "The simple exponential smoothing model," *Proceedings of Conference*, Technical University of Košice, in *Proceedings of the 4th International Conference on Modelling of Mechanical and Mechatronic Systems*, pp. 380–384, Technical University of Košice, Košice, Slovak Republic, 2011.
- [22] G. Tirkes, C. Guray, and N. Celebi, "Demand forecasting: a comparison between the holt-winters, trend analysis and decomposition models/predvidanje potraznje: usporedba između holt-winters modela, analize trenda i modela dekompozicije," *Tehnicki Vjesnik-Technical Gazette*, vol. 24, no. S2, pp. 503–510, 2017.
- [23] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *Ieee Access*, vol. 5, pp. 21 954–21 961, 2017.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [25] A. Stec and D. Klabjan, "Forecasting crime with deep learning," 2018, <http://arxiv.org/abs/1806.01486>.
- [26] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep lstm-rnn," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2727–2740, 2019.
- [27] T. Almanie, R. Mirza, and E. Lor, "Crime prediction based on crime types and using spatial and temporal criminal hot-spots," 2015, <http://arxiv.org/abs/1508.02050>.
- [28] J. Ratcliffe, R. Taylor, A. Askey, J. Grasso, and R. Fisher, *The Philadelphia Predictive Policing Experiment: Impacts of Police*

- Cars Assigned to High Crime Grids*, Center for Crime Science, Temple University, Philadelphia, PA, USA, 2017.
- [29] C. S. Marzan, M. J. C. Baculo, R. de Dios Bulos, and C. Ruiz, "Time series analysis and crime pattern forecasting of city crime data," in *Proceedings of the International Conference on Algorithms, Computing and Systems*, pp. 113–118, ACM, Beijing, China, 2017.
- [30] G. Borowik, Z. M. Wawrzyniak, and P. Cichosz, "Time series analysis for crime forecasting," in *Proceedings of the 2018 26th International Conference on Systems Engineering (ICSEng)*, pp. 1–10, IEEE, Sydney, Australia, 2018.
- [31] M. Feng, J. Zheng, Y. Han, J. Ren, and Q. Liu, "Big data analytics and mining for crime data analysis, visualization and prediction," in *Proceedings of the International Conference on Brain Inspired Cognitive Systems*, pp. 605–614, Springer, Guangzhou, China, 2018.
- [32] P. Singh, *Time series forecasting on crime data in amsterdam for a software company*, Ph.D. dissertation, NOVA Information Management School, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Lisbon, Portugal, 2018.
- [33] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "Spatio-temporal crime predictions in smart cities: a data-driven approach and experiments," *Pervasive and Mobile Computing*, vol. 53, pp. 62–74, 2019.
- [34] Metadata catalog – phila.Gov," <https://metadata.phila.gov/home/datasetdetails/5543868920583086178c4f8e/representationdetails/570e7621c03327dc14f4b68d/>.
- [35] S. Deegalla, H. Boström, and K. Walgama, "Choice of dimensionality reduction methods for feature and classifier fusion with nearest neighbor classifiers," in *Proceedings of the 15th International Conference on Information Fusion (FUSION)*, pp. 875–881, IEEE, Chicago, IL, USA, 2012.
- [36] S. Ahmadkhani and P. Adibi, "Face recognition using supervised probabilistic principal component analysis mixture model in dimensionality reduction without loss framework," *IET Computer Vision*, vol. 10, no. 3, pp. 193–201, 2016.
- [37] J. D. Cryer, *Time Series Analysis*, vol. 286, Duxbury Press, Boston, MA, USA, 1986.
- [38] Y. S. Triana and A. Retnowardhani, "Enhance interval width of crime forecasting with arima model-fuzzy alpha cut," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1193–1201, 2019.
- [39] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, Melbourne, Australia, 2018.
- [40] J. J. M. Moreno, A. P. Pol, A. S. Abad, and B. C. Blasco, "Using the r-mape index as a resistant measure of forecast accuracy," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013.