A Hybrid Time-Series Prediction of the Greater Riyadh's Metropolitan Area Expansion

Faizah Alshammari
College of Computer Science and Engineering, Department of Computer Science and Artificial Intelligence, University of Jeddah, Saudi Arabia
falshammari.stu@uj.edu.sa (corresponding author)

Nahla Aljojo
College of Computer Science and Engineering, Department of Information System and Technology, University of Jeddah, Saudi Arabia
nmaljojo@uj.edu.sa

Araek Tashkandi
College of Computer Science and Engineering, Department of Information System and Technology, University of Jeddah, Saudi Arabia
astashkandi@uj.edu.sa

Abdullah Alghoson
College of Computer Science and Engineering, Department of Information System and Technology, University of Jeddah, Saudi Arabia
alghoson@uj.edu.sa

Ameen Banjar
College of Computer Science and Engineering, Department of Information System and Technology, University of Jeddah, Saudi Arabia
abanjar@uj.edu.sa

Nidhal K. El Abbadi
Computer and Engineering Techniques Department, College of Engineering and Techniques, Al-Mustqbal University, Iraq
nidhal.elabbadi@uokufa.edu.iq

Received: 31 August 2023 | Revised: 11 September 2023 | Accepted: 12 September2023

 DOI: https://doi.org/10.48084/etasr.6350

ABSTRACT

Riyadh is the most populous city in Saudi Arabia, with a population of over five million people. The governmental and economic centers of Saudi Arabia are located in the city. Due to the fact that the metropolitan region that surrounds Riyadh is continuously growing and expanding, appropriate planning is essential. To be able to formulate efficient plans, one needs access to trustworthy facts and information. Failing to have a clear picture of the future renders planning inefficient. Along with a hybrid time-series prediction of the expansion of the wider Riyadh metropolitan area, an urban growth forecasting model was constructed for the Riyadh region as part of this study. This model was used to make projections about the city's future population. This prediction was conducted with the application of Linear Regression (LR), Seasonal Auto-Regressive Integrated Moving Average (SARIMAX), and Auto-Regressive Integrated Moving Average (ARIMA). The dataset for this study consisted of satellite images of the region surrounding Riyadh that were acquired between 1992 and 2022. Mean Absolute Percentage Error (MAPE) was applied to measure the performance of the proposed hybrid models. The calculated MAPE vales are 2.0% for SARIMAX, 12% for LR, and 22% for ARIMA. As a consequence, the hybrid model's forecast for the future of the region suggests that the projections made regarding the expansion are keeping pace.

Keywords-urban-growth; ARIMA; SARIMAX; logistic regression

I. INTRODUCTION

In terms of landmass, the Kingdom of Saudi Arabia (KSA) is one of the largest countries in the Middle East and the world [1]. Due to the oil revolution that has occurred over the past six decades, the way of life in KSA has undergone a phenomenal transformation in all aspects. One of the things that occurred throughout the Renaissance was the transformation and development of urban areas, which resulted in the urbanization of many once-rural areas and villages [2]. Several million people live in Riyadh, which also serves as the political center of the KSA. It is considered one of the most important cities in the world that has undergone a massive urban revival and expansion, accompanied by growth and multiplication of its area [3]. Riyadh's expanse requires detailed planning, as it requires the construction of a new road network, new sewage and electrical systems, and additional public spaces and parks. When planning is not based on reliable information that touches on reality, it me problematic. The urbanization process, which is closely related to economic and social changes, has emerged as a primary source of concern in most cities around the globe. In 1950, metropolitan regions were home to barely 30% of the world's population, in 2014 it rose to 54%, and it is expected to reach 70%. The rate of urbanization in Asia and Africa is significantly higher compared to other regions of the world [4]. Most Saudi cities are under tremendous urbanization and it is expected that the urban population of the country will expand at a rapid rate. Urbanization is inescapable in developing countries [5]. Rapid and uncontrolled urban growth poses a threat to the possibility of sustainable development and, if not managed effectively, can result in the destruction of productive agricultural land, the overcrowding of habitats, the pollution of the air and water, water distribution problems, and increased levels of traffic congestion. Urbanization occurs generally on the outskirts of cities or in a linear fashion along motorways [6-7]. The term "urban sprawl" refers to the scattered growth around major thoroughfares or in the territory surrounding major cities. Due to advances in geospatial technology, there has been a considerable increase in modeling factors that contribute to urban sprawl, analyzing urban sprawl that has already occurred, and predicting urban sprawl that will occur in the future [8-9].

As urban expansion requires the development of plans for means of transportation, streets, bridges, water networks, power, and sanitation, government agencies in the Riyadh region are working to establish plans that respond to it. There is a shortage of precise information among government entities about the amount of expansion expected after some years. This study investigated the problem of insufficient data on the amount of future urban expansion to help government agencies in the process of making appropriate plans that are compatible with the anticipated expansion. This study used open satellite data to perform an in-depth analysis and determine the extent to which metropolitan areas and different land uses have changed. In addition to this, this study used Auto-Regressive Integrated Moving Average (ARIMA), Seasonal Auto-Regressive Moving Average (SARIMAX), and Linear Regression (LR) models to make projections on the future density of the urban city and forecast the urban development that will take place in the vicinity of Riyadh during the next decade. The results of these algorithms were compared using the Mean Absolute Percentage Error (MAPE).

In [10], land use maps from aerial photographs, topographic maps, and satellite images for the period 1967-2019 for Vellore, India, were used to build regression models to calculate the density built at any distance from the city center or major roads in the space-time domain. The study also developed models based on seasonal ARIMA to predict future urban density for the year 2045 using the urban densities of the years 1967, 1991, and 2019, showing that more development is expected in suburban than urban areas. In [11], the built-up areas were extracted in 14 central cities of Hunan Province, using night-light remote sensing datasets (DMSP-OLS) from 1992 to 2018 to evaluate the spatiotemporal characteristics of the built-up areas in terms of area, expansion velocity, and main expansion direction. An inverse neural network (BP) and an ARIMA model were used to predict built-up areas from 2019 to 2026. The BP neural network-based urbanization model provided high prediction accuracy ($R = 0.966$) and estimated that the total area of urban built-up areas in Hunan Province will reach 2,463.80 km^2 by 2026. In [12], time-series satellite images were used to predict urban expansion using the ConvLSTM network, which can learn global spatiotemporal information without downsizing spatial feature maps. This network was applied to time-series satellite images, and the prediction results were compared with the Pix2pix and Dual GAN networks. Several multidate satellite images were used, representing the three largest cities in KSA, Riyadh, Jeddah, and Dammam. The evaluation results showed that the proposed model produced better prediction results in terms of mean square error, root mean square error, peak signal-to-noise ratio, structural similarity index, and overall classification accuracy. Moreover, the training time of this model was less than that of the Dual GAN architecture. According to [13], predicting the evolution of ecological space based on multiple scenarios is essential to have a sufficient grasp of the expansion of a region. In [14], a concept of change dynamics in land use and cover was presented to model future urban expansion. This study concluded that land use and land cover play a role in determining urban growth. Similarly, a mapping and prediction strategy for land cover changes was presented in [15], using machine learning approaches, to examine the stages of change that occur in small and medium cities. It is necessary to emphasize the influence that the rate of change has. This could also be within a scale level since the microscale assessment and forecast of urban thermal comfort zones are related to urban expansion [16].

In [17], the problem of agglomeration effects as spatially embedded social interactions was investigated by establishing urban scaling that extended beyond metropolitan areas. This was connected to [18], which showed that monitoring the rate of growth was an effective method for predicting land use and changes in land cover in the future. The findings of [19] demonstrate an improved regional pattern for land development, based on the growth of urban areas. In [20], a novel method was suggested to identify urban region land use, which could be used in the expansion decision process. The concept of eviction through suburbanization was highly important for monitoring the increasing levels of social disparity caused by the expansion of metropolitan areas [21].

Many studies have investigated the expansion of the Saudi Arabian population, but there are still some unanswered questions. As many studies focus on short-term population forecasts, longer-term projections are needed to help policymakers plan for the future. Additionally, there is a need for more in-depth research on demographic shifts in KSA based on age and gender. This could be useful for policymakers to understand the unique difficulties and possibilities of certain demographic subsets. Although many studies have investigated the demographic causes of KSA's booming population, more information is needed on the role of economic and social issues. For instance, how can alterations in economic or social policy impact population expansion? Most analyses of population growth in KSA focused on the country as a whole, but more work needs to be done to understand how this growth is distributed around the country. Pinpointing possible population growth hotspots will allow policymakers to allocate resources more efficiently. Furthermore, the investigation of migration patterns in KSA can help policymakers better manage the consequences of migration on population increase.

II. REGRESSION METHODS

The conceptualization of this study lies in the ongoing structural changes in KSA's native-born population, including higher life expectancy, higher share of working-age people, better puplic health, a more robust healthcare system, and higher rates of education and participation in the labor force [22]. This study took advantage of [22] to build the conceptual framework shown in Figure 1. The idea was developed through the formulation that as urban expansion requires planning, the growth of Riyadh city lies with transport, streets, bridges, water networks, power, sanitation, government agencies, and many other things. All these factors have a direct bearing on Riyadh's Metropolitan Area Expansion, indicating a causal relationship. In addition, this expansion is closely connected to the increase in population.

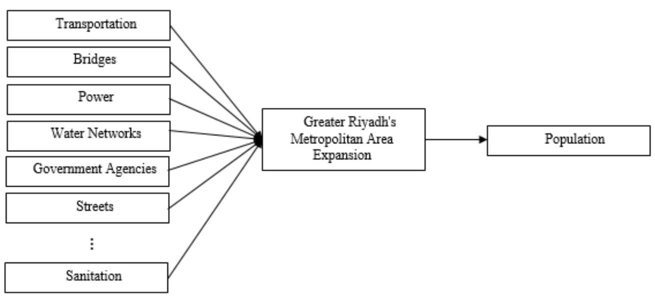

Fig. 1. The proposed conceptual framework.

In addition to the formulation of the urban expansion context, a prediction model was conceptualized to determine the growth of the Riyadh area using a hybrid time-series forecasting model. This concept lies in the accelerating growth of the Greater Riyadh Metropolitan area. An urban growth forecasting model should be associated with time series analysis because, over time, there are always updates in many directions. The prediction would be possible by involving the use of LR, SARIMAX, and ARIMA.

A. *Auto-Regressive Integrated Moving Average*

The term "Auto-Regressive" describes that the model uses previous values of the series to make predictions about future values. The term "Integrated" refers to the fact that the model takes into account the variation in values that occur between successive elements of the series to ensure that it remains stationary. The term "Moving Average" indicates that the model uses previous errors in forecasting to make predictions about future values [23]. ARIMA is frequently used in many time-series prediction problems. If the data are stationary, which means that statistical features of the data such as mean and variance do not change with time, it has the potential to be very useful. However, it is essential to keep in mind that ARIMA models are not suitable for all time-series data, and selecting the proper parameters (such as the order of the autoregressive and moving average terms) can be a difficult and iterative process [24]. ARIMA is defined by three order parameters, denoted by p, d, and q. The "Box-Jenkins" method is used to fit the data into an ARIMA model. An autoregressive AR(p) component uses previous values in a regression equation. This component is applied to series Y. The value of the autoregressive parameter p determines the total number of lags that are applied. For instance, the AR(2) or, more precisely, the ARIMA(2,0,0) model can be represented by:

$$Y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + e_t \quad (1)$$

where φ_1 and φ_2 are parameters for the model, and d represents the degree of difference in the integrated component $I(d)$. A series difference simply involves subtracting its current and previous values d times. Differencing is often used to stabilize the series when the stationarity assumption is not met. A moving average $MA(q)$ component represents the error of the model as a combination of previous error terms e_t. The order q determines the number of terms included in the model [24].

$$Y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t \quad (2)$$

Differencing, autoregressive, and moving average components make up a non-seasonal ARIMA model, which can be written as a linear equation:

$$Y_t = c + \theta_1 yd_{t-1} + \cdots + \theta_p yd_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \quad (3)$$

where yd is Y differenced d times and c is a constant. The particular values of the autoregressive and moving average parameters, in addition to the order of differencing, are often determined by a process known as parameter estimation [25]. This method can also be used to establish the order of differencing. During this stage of the process, the model is fitted to the historical data using the parameter values that produce the lowest amount of residual error. After the model parameters have been estimated, the model can be used to create projections of future values.

B. Seasonal Auto-Regressive Integrated Moving Average

SARIMAX was developed as an ARIMA modification to model and account for seasonality in the data and the effects of external or exogenous variables [26]. Similar to the ARIMA model, SARIMAX incorporates seasonal components in addition to the autoregressive, moving average, and differencing ones. The seasonal components reflect cyclical patterns and trends (daily, weekly, monthly, or yearly cycles) through the use of seasonal autoregressive, moving average, and differencing terms. Exogenous variables, outside of time series, are also included in SARIMAX alongside seasonal effects. The accuracy of forecasts can be improved by including explanatory variables such as weather patterns, economic indicators, and marketing efforts [10]. Moving average and autoregressive models work with stationary and linear data. However, the data are frequently nonstationary. Seasonal ARIMA (SARIMA) is a generalized form of the ARIMA model to handle seasonality in the data. An exogenous regressor term can also be used to incorporate external variables into the model. SARIMAX allows the user to include the effects of external variables in the model. Exogenous variables are variables that influence but are not influenced by a model. The SARIMAX model is represented as follows:

$$\begin{aligned} Y(t) &= \mu + \phi(1)Y(t-1) + \cdots + \phi(p)Y(t-p) \\ &-\theta(1)\varepsilon(t-1) - \cdots - \theta(q)e(t-q) \\ &+\varphi(1)Y(t-s) + \cdots + \varphi(P)Y(t-Ps) \\ &+\Theta(1)\varepsilon(t-s) + \cdots + \Theta(Q)\varepsilon(t-Qs) \\ &+\beta X(t) + e(t) \end{aligned} \quad (4)$$

where, $Y(t)$ is the value of the time series at time t, μ is the mean of the series, $\phi(1)$ through $\phi(p)$ are the autoregressive parameters of the model, $\theta(1)$ through $\theta(q)$ are the moving average parameters of the model, $\varphi(1)$ through $\varphi(P)$ are the seasonal autoregressive parameters, $\Theta(1)$ through $\Theta(Q)$ are the seasonal moving average parameters, $X(t)$ is the value of the exogenous variable(s) at time t, β is the coefficient(s) for the exogenous variable(s), s is the seasonal period, and $\varepsilon(t)$ is the error term at time t.

The first half of the equation is the same as the ARIMA and represents the auto-regressive and moving average components of the model ($-\theta(q)\varepsilon(t-q)$). This section of the equation begins with the term $-(q)(t-q)$ and continues until it ends. The seasonal patterns and trends in the data are captured by the following two portions of the equation, which include the seasonal autoregressive and moving average elements, and the seasonal error factors [10, 26]. The effect of extraneous factors on the time series is taken into account in the very last component of the equation, which consists of the exogenous variable(s) and the related coefficient(s). Before the model can be used to create forecasts, the particular values of the parameters ϕ, θ, φ, Θ, β and the order of differencing d and e seasonal differencing D must be estimated from the historical data through a process known as parameter estimation. This must be done before the model can be put into use.

C. Linear Regression

Regression is a statistical analysis method to determine the relationship between variables. A dependent variable and one or more independent variables can be modeled using LR. The purpose of LR is to locate the line or hyperplane that provides the most accurate description of the linear connection that exists between the variables [27]. In the basic regression, as presented in (5), there is just one independent variable X, and one dependent variable Y. The objective is to determine the equation of a line that provides the best possible fit to the data [28].

$$Y = a + bX \quad (5)$$

where b represents the slope of the line and a represents the intercept. Simple LR and multiple LR are both types of linear analysis. The error term accounts for the discrepancies between the predicted and the actual values. LR is used to model and describe the relationship between variables and predict future values of the dependent variable in many different domains, including economics, finance, biology, psychology, and engineering.

III. RESEARCH METHODOLOGY

Figure 2 presents the methodological flow of this study.

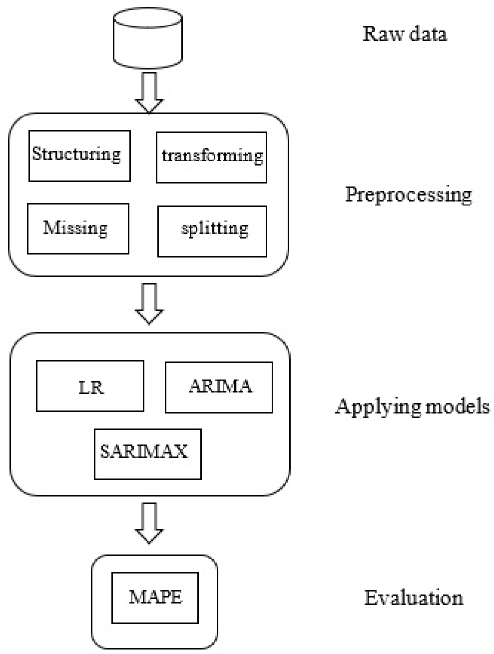

Fig. 2. The proposed methodology flow.

An experimental analysis was carried out to predict urban sprawl and formulate the findings to apply them in wider applications. Based on the methodological flow, after obtaining the dataset for the region of Riyadh from satellite images, the data were subjected to some form of preprocessing. Then the models were used. In the last step, the MAPE measure was used to evaluate the results of the models. The proposed framework used many different datasets, some of which were geographical. These datasets included satellite photos, population data, and land use data. At first, pre-processing was performed to extract key features from the datasets. These key characteristics can then be used to train and test machine learning models.

A. Dataset Generation

Geographic images of Riyadh City and the surrounding area from 1992 to 2022 were used with ArcGIS to obtain historical data on its urban development. The dataset was provided through the Google Earth Engine using the Google Collab program. This dataset was used along with data from 2015 to 2022 from the Copernicus Sentinel-2 satellite. The Geemap, Rasterio, and JSON Python libraries were used to process geospatial images, along with CV2 libraries to process images generated by the Google Earth Engine. Additionally, Matplotlib, NumPy, shutil, and time libraries were used. An object was used to represent the channels in the RGB images and clarify the raw dataset. This object had a set of bands that corresponded to the specified satellite (B4, B3, B2). The scale of the images was determined at 40 m. Data from the specified geographical area were available for a period of eight years, starting in 2015 through 2022. The years 2015 through 2022 were counted as seven years, and the dates January 1, 2022, through June 30, 2022, were added to generate a picture of the eighth year. The mask cloud filter was selected by working with the band QA60, to filter cloud satellite images from the clouds and fog and to ensure that the photographs contain less than 20% cloud cover and fog. When working with photos that have less than 20% blur, the QA60 band was used for filtering, and then these images were divided by 10,000 for normalization. The satellite images were saved in TIFF format. A copy of the satellite images was then created in RGB format to use them in the natural colors that can be seen by the human eye. The generated RGB photos were cut into 150×150 to produce more than 10,000 pieces per image per year, to strike a balance between simplicity and accuracy.

B. Units

The issue posed by the massive urbanization of Riyadh involves thinking about how the geographical coordinates of previous years can be used to gain an understanding of urban growth and make projections about how it will develop in the years to come. Following that, identifying the geographic data to be acquired is the next step in the data collection process. After that, the target or relevant data based on the objective or task was picked. In addition, the unprocessed data were categorized into a variety of forms and dimensions. At first, areas were categorized as either urban or non-urban based on the qualities of the images. The next step was to determine how many hectares are covered by urban and non-urban areas, respectively. A detection program was used to track urban and non-urban areas and determine annual changes. After data organization, the data were transformed into an appropriate form suitable for analysis. The properties of the images were used to make the initial classifications of the locations, which were then broken down further into urban and non-urban regions. Then, the land of the urban and non-urban areas was calculated in hectares, keeping track of the annual change in area size. The images were transformed into data (integer and float). After transforming, data were taken from three columns, Year, Area_by_Hectares, and Increasing rate, as shown in Figure 3 and Table I. Figure 3 shows the entire Riyadh region, including urban and non-urban areas. Non-urban areas are represented in gray, while urban areas are represented in red. The x-axis provides information regarding the length of the region's borders measured in kilometers. Figure 4 shows the range of the data after transformation.

TABLE I. DATA AFTER TRANSFORMATION

Year	Area_by_Hectares	Increasing_rate
1992	56636	11140
1993	67777	14396
1994	82173	811
1995	82985	1673
1996	84658	1236

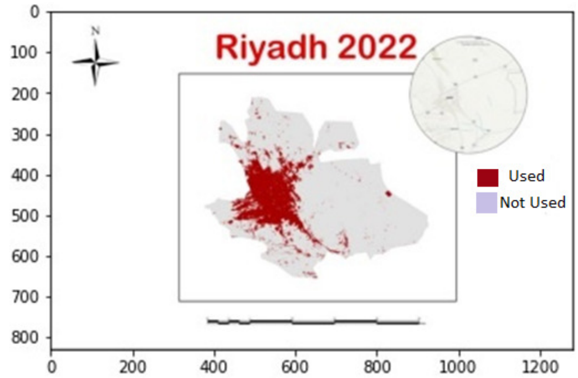

Fig. 3. Raw data before transformation.

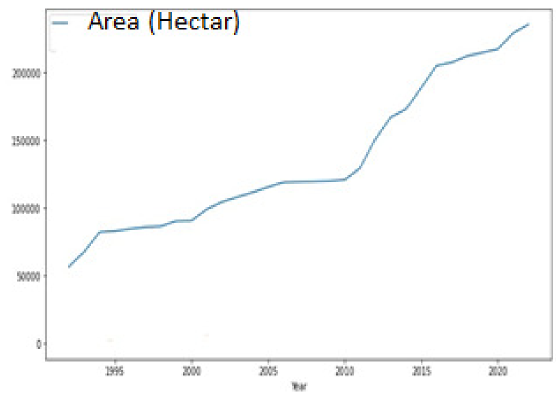

Fig. 4. Range of data after transformation.

Any missing values were calculated based on the arithmetic mean of the sum of the preceding and the following year.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

Before loading the preprocessed data, they were split into a 70:30 ratio for training and testing, respectively. The models were trained with the help of the first portion and then evaluated with the help of the second.

A. Experimental Analysis

The experiment was performed using the Python programming language. The MAPE measure was used to evaluate the models [29], as it expresses the precision by:

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{6}$$

The ARIMA model was run using the following parameters: p = 1, d = 0, and q = 2. The SARIMAX method was run using p = 4, d = 2, and q = 3, however, since the seasonal order was used, the values (0,0,0,0) were employed.

B. Results

A MAPE score of 0% represents a perfect prediction, which means that the projected values exactly match the values that really occurred. When the MAPE is 100%, the projected values are, on average, twice as far apart from the actual values as the actual values themselves are. A more accurate prediction model would have a lower MAPE. The errors from the three models are presented in Table II. The result shows that SARIMAX has the lowest MAPE of 2.0%. MAPE implies that the average absolute percentage difference between the projected and actual values of a dataset is very negligible and the accuracy of a prediction model, represented as a percentage, is within the allowed range. It is essential to take into account the fact that the interpretation in this setting falls within a lower value range of below 3% for one model, however, while using LR, it rose up to 12%, which is still considered to be low.

TABLE II. MAPE COMPARISON BETWEEN ARIMA, SARIMAX, AND LR

ARIMA	SARIMAX	LR
2.2%	2.0%	12.4%

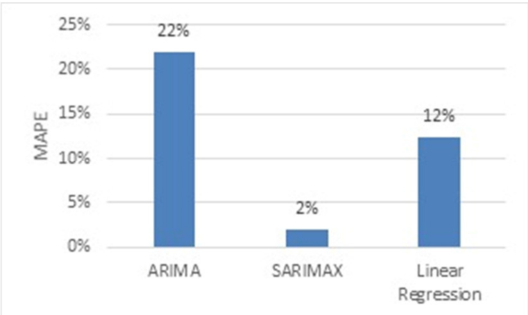

Fig. 5. MAPE of the selected algorithms

SARIMAX was used to predict the urban expansion of the region surrounding Riyadh during the next ten years. This prediction indicates that the region surrounding Riyadh will see tremendous growth, as shown in Table III and Figure 6. In Figure 6, the blue line represents the actual area that has been occupied by the Riyadh region over the past thirty years (1992-2022), while the orange one represents the area that is expected to be occupied in the next 10 years (2023-2032).

TABLE III. THE FORECASTED AREA FOR TEN YEARS

Year	Area_by_Hectar
2023	246622
2024	267250
2025	284135
2026	290721
2027	305200
2028	312984
2029	314570
2030	317563
2031	319888
2032	322039

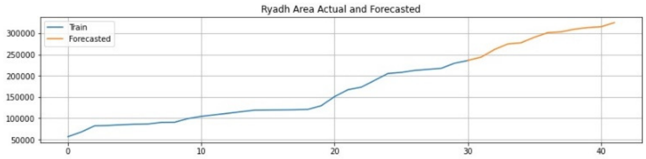

Fig. 6. The forecasted area for ten years.

V. DISCUSSION

This study used Google Earth Engine [30], and its results indicate that the capital city of KSA, located in the central part of the country, will continue to grow over the next several decades. Although the region surrounding Riyadh is the most extensive in KSA, comprising an area of more than 404,000 km^2, it is not the most populous city in the country. According to an estimate, the population of the area is projected to be approximately 8.3 million people. Throughout the last few years, the region has seen remarkable development in a variety of fields, including infrastructure, real estate, and tourism. The government has been making large investments in several different infrastructure projects, such as the Riyadh Metro, which will serve as the basis of the city's public transportation system, and the King Salman Energy Park, which will serve as a key hub for the energy industry. Both projects are expected to yield major fruits in the relatively near future. These are some of the most vital details regarding the expansion.

In addition, there is another component of the development that lies ahead. There has been an increase in the number of real estate construction projects in the previous several years. As part of these developments, new residential and commercial properties have been constructed throughout the region. As a direct result of this factor, the real estate market is currently through a period of expansion and prices have been consistently increasing. Tourism is another primary focus for the region and one of the primary objectives of the government is to establish Riyadh as a premier destination for visitors from both within and outside the country. This is yet another component that will contribute to the expansion, and it is also one of the most essential elements for the area. In addition to its many historical and cultural attractions, such as the Masmak Fortress and the National Museum of Saudi Arabia, this region is also known for its vibrant shopping and dining scenes.

According to the findings of this study, it is projected that the Riyadh region will continue to experience growth and development in a range of industries over the next few years. This growth and development is expected to continue for at least the next few years, and the city is expected to become a smart city where even traffic flow will be computerized [31]. This is because the Saudi government is putting a strong emphasis on diversifying the economy and reducing the country's dependence on the revenue from oil sales. The rapid growth of urban areas is a major challenge for many cities around the world [32]. Due to the increase in economic activity and population, there have been major modifications in land use patterns and urban landscapes in recent decades [33]. Therefore, urban planners must find ways to manage and accommodate expanding cities and, along with policymakers, anticipate probable changes in population and land use patterns. Urban growth forecasting is a crucial part of urban planning. In recent years, machine learning techniques have gained popularity in urban growth models due to their ability to produce extremely accurate estimates [34]. Riyadh is rapidly becoming one of the most populous urban areas, as its population has exploded during the recent decades, causing major changes in building and transportation habits in the area. Therefore, it is important to create predictions about Riyadh's urban growth to guide the city's development and ensure that it meets the needs of its residents.

VI. CONCLUSION

Utilizing the Integrated Regressive Moving Average (ARIMA), the Automatic Seasonal Integrated Regressive Moving Average (SARIMAX), and the Linear Regression (LR), this study performed urban growth forecasts in the region surrounding Riyadh for the next ten years. To do this, we analyzed satellite photos of the Riyadh region taken between 1992 and 2022. We utilized the Mean Absolute Percentage Error (MAPE), to analyze the models that were used so that we could choose the model that was most accurate for making realistic predictions. According to the findings, the SARIMAX algorithm was the most effective out of the three algorithms, with a MAPE value of 2.0%. On the other hand, the ARIMA algorithm was the least effective, with a MAPE value of 22%. We can observe that the precision of the developed model really high and produces very good outcomes in comparison to other efforts that are in a similar vein. At the conclusion of the experiment, we used the SARIMAX algorithm to make a prediction regarding the rate of urbanization in the region surrounding Riyadh. According to the projection, the total land area inside the Riyadh region will amount to 322,039 hectares by the year 2032. We propose that as a future endeavour, the data set could be expanded to encompass all regions of Saudi Arabia. Additionally, forecasting can be utilized to generate forecasts for the future in other areas such as population and power consumption.

REFERENCES

[1] L. Bursztyn, A. L. González, and D. Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, vol. 110, no. 10, pp. 2997–3029, Oct. 2020, https://doi.org/10.1257/aer.20180975.

[2] O. H. Sayed and Y. S. Masrahi, "Climatology and phytogeography of Saudi Arabia. A review," *Arid Land Research and Management*, vol. 37, no. 3, pp. 311–368, Jul. 2023, https://doi.org/10.1080/15324982.2023.2169846.

[3] M. D. Alotaibi *et al.*, "Assessing the response of five tree species to air pollution in Riyadh City, Saudi Arabia, for potential green belt application," *Environmental Science and Pollution Research*, vol. 27, no. 23, pp. 29156–29170, Aug. 2020, https://doi.org/10.1007/s11356-020-09226-w.

[4] United Nations, Department of Economic and Social Affairs, Population Division, *World Urbanization Prospects - The 2018 Revision*. New York, NY, USA: United Nations, 2019.

[5] M. Arshad, K. M. Khedher, E. M. Eid, and Y. A. Aina, "Evaluation of the urban heat island over Abha-Khamis Mushait tourist resort due to rapid urbanisation in Asir, Saudi Arabia," *Urban Climate*, vol. 36, Mar. 2021, Art. no. 100772, https://doi.org/10.1016/j.uclim.2021.100772.

[6] S. Awasthi, "'Hyper'-Urbanisation and migration: A security threat," *Cities*, vol. 108, Jan. 2021, Art. no. 102965, https://doi.org/10.1016/j.cities.2020.102965.

[7] S. Dinda, K. Das, N. D. Chatterjee, and S. Ghosh, "Integration of GIS and statistical approach in mapping of urban sprawl and predicting future growth in Midnapore town, India," *Modeling Earth Systems and Environment*, vol. 5, no. 1, pp. 331–352, Mar. 2019, https://doi.org/10.1007/s40808-018-0536-8.

[8] S. Ghosh, "A city growth and land-use/land-cover change: a case study of Bhopal, India," *Modeling Earth Systems and Environment*, vol. 5, no. 4, pp. 1569–1578, Dec. 2019, https://doi.org/10.1007/s40808-019-00605-y.

[9] K. T. Deribew, "Spatiotemporal analysis of urban growth on forest and agricultural land using geospatial techniques and Shannon entropy method in the satellite town of Ethiopia, the western fringe of Addis Ababa city," *Ecological Processes*, vol. 9, no. 1, Sep. 2020, Art. no. 46, https://doi.org/10.1186/s13717-020-00248-3.

[10] K. Malarvizhi, S. V. Kumar, and P. Porchelvan, "Urban sprawl modelling and prediction using regression and Seasonal ARIMA: a case study for Vellore, India," *Modeling Earth Systems and Environment*, vol. 8, no. 2, pp. 1597–1615, Jun. 2022, https://doi.org/10.1007/s40808-021-01170-z.

[11] L. Xu, X. Liu, D. Tong, Z. Liu, L. Yin, and W. Zheng, "Forecasting Urban Land Use Change Based on Cellular Automata and the PLUS Model," *Land*, vol. 11, no. 5, 2022, https://doi.org/10.3390/land11050652.

[12] W. Boulila, H. Ghandorh, M. A. Khan, F. Ahmed, and J. Ahmad, "A novel CNN-LSTM-based approach to predict urban expansion," *Ecological Informatics*, vol. 64, Sep. 2021, Art. no. 101325, https://doi.org/10.1016/j.ecoinf.2021.101325.

[13] X. Liu, X. Wang, K. Chen, and D. Li, "Simulation and prediction of multi-scenario evolution of ecological space based on FLUS model: A case study of the Yangtze River Economic Belt, China," *Journal of Geographical Sciences*, vol. 33, no. 2, pp. 373–391, Feb. 2023, https://doi.org/10.1007/s11442-023-2087-9.

[14] B. Mahdavi Estalkhsari, P. Mohammad, and A. Karimi, "Land Use and Land Cover Change Dynamics and Modeling Future Urban Growth Using Cellular Automata Model Over Isfahan Metropolitan Area of Iran," in *Ecological Footprints of Climate Change: Adaptive Approaches and Sustainability*, U. Chatterjee, A. O. Akanwa, S. Kumar, S. K. Singh, and A. Dutta Roy, Eds. Cham, Switzerland: Springer International Publishing, 2022, pp. 495–516.

[15] M. Shrestha, C. Mitra, M. Rahman, and L. Marzen, "Mapping and Predicting Land Cover Changes of Small and Medium Size Cities in Alabama Using Machine Learning Techniques," *Remote Sensing*, vol. 15, no. 1, 2023, https://doi.org/10.3390/rs15010106.

[16] A. E. AlDousari, A.-A. Kafy, M. Saha, Md. A. Fattah, A. Bakshi, and Z. A. Rahaman, "Summertime Microscale Assessment and Prediction of Urban Thermal Comfort Zone Using Remote-Sensing Techniques for Kuwait," *Earth Systems and Environment*, vol. 7, no. 2, pp. 435–456, Jun. 2023, https://doi.org/10.1007/s41748-023-00340-6.

[17] D. Strumsky, L. Bettencourt, and J. Lobo, "Agglomeration effects as spatially embedded social interactions: identifying urban scaling beyond metropolitan areas," *Environment and Planning B: Urban Analytics and City Science*, Jan. 2023, Art. no. 23998083221148200, https://doi.org/10.1177/23998083221148198.

[18] H. B. Akdeniz, N. S. Sag, and S. Inam, "Analysis of land use/land cover changes and prediction of future changes with land change modeler: Case of Belek, Turkey," *Environmental Monitoring and Assessment*, vol. 195, no. 1, Nov. 2022, Art. no. 135, https://doi.org/10.1007/s10661-022-10746-w.

[19] Y. Gao, Z. Shen, Y. Liu, C. Yu, L. Cui, and C. Song, "Optimization of differentiated regional land development patterns based on urban expansion simulation—A case in China," *Growth and Change*, vol. 54, no. 1, pp. 45–73, 2023, https://doi.org/10.1111/grow.12637.

[20] T. He, Y. Lu, W. Yue, W. Xiao, X. Shen, and Z. Shan, "A new approach to peri-urban area land use efficiency identification using multi-source datasets: A case study in 36 Chinese metropolitan areas," *Applied Geography*, vol. 150, Jan. 2023, Art. no. 102826, https://doi.org/10.1016/j.apgeog.2022.102826.

[21] D. Q. Rutan, P. Hepburn, and M. Desmond, "The Suburbanization of Eviction: Increasing Displacement and Inequality Within American Suburbs," *RSF: The Russell Sage Foundation Journal of the Social*

Sciences, vol. 9, no. 1, pp. 104–125, Feb. 2023, https://doi.org/10.7758/RSF.2023.9.1.05.

[22] A. A. Salam, "Ageing in Saudi Arabia: new dimensions and intervention strategies," *Scientific Reports*, vol. 13, no. 1, Mar. 2023, Art. no. 4035, https://doi.org/10.1038/s41598-022-25639-8.

[23] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*. Cham, Switzerland: Springer International Publishing, 2017.

[24] F. Marandi and S. M. T. Fatemi Ghomi, "Time series forecasting and analysis of municipal solid waste generation in Tehran city," in *2016 12th International Conference on Industrial Engineering (ICIE)*, Tehran, Iran, Jan. 2016, pp. 14–18, https://doi.org/10.1109/INDUSENG.2016.7519343.

[25] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results in Physics*, vol. 27, Aug. 2021, Art. no. 104509, https://doi.org/10.1016/j.rinp.2021.104509.

[26] P. Manigandan *et al.*, "Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models," *Energies*, vol. 14, no. 19, 2021, https://doi.org/10.3390/en14196021.

[27] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, https://doi.org/10.38094/jastt1457.

[28] A. R. Ramadhan, M. Choi, Y. Chung, and J. Choi, "An Empirical Study of Segmented Linear Regression Search in LevelDB," *Electronics*, vol. 12, no. 4, 2023, https://doi.org/10.3390/electronics12041018.

[29] E. Vivas, H. Allende-Cid, and R. Salas, "A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score," *Entropy*, vol. 22, no. 12, 2020, https://doi.org/10.3390/e22121412.

[30] M. Ahmadi and M. Ghamary Asl, "Monitoring urban growth in Google Earth Engine from 1991 to 2021 and predicting in 2041 using CA-MARKOV and geometry: case study—Tehran," *Arabian Journal of Geosciences*, vol. 16, no. 2, Jan. 2023, Art. no. 107, https://doi.org/10.1007/s12517-022-11089-z.

[31] S. M. Abdullah *et al.*, "Optimizing Traffic Flow in Smart Cities: Soft GRU-Based Recurrent Neural Networks for Enhanced Congestion Prediction Using Deep Learning," *Sustainability*, vol. 15, no. 7, 2023, https://doi.org/10.3390/su15075949.

[32] U. Iftikhar, K. Asrar, M. Waqas, and S. A. Ali, "Evaluating the Performance Parameters of Cryptographic Algorithms for IOT-based Devices," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7867–7874, Dec. 2021, https://doi.org/10.48084/etasr.4263.

[33] N. B. Serradj, A. D. K. Ali, and M. E. A. Ghernaout, "A Contribution to the Thermal Field Evaluation at the Tool-Part Interface for the Optimization of Machining Conditions," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7750–7756, Dec. 2021, https://doi.org/10.48084/etasr.4235.

[34] M. U. Farooq, A. Ahmed, S. M. Khan, and M. B. Nawaz, "Estimation of Traffic Occupancy using Image Segmentation," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7291–7295, Aug. 2021, https://doi.org/10.48084/etasr.4218.

AUTHORS PROFILE

Faizah Alshammari is with the College of Computer Science and Engineering, Department of Computer Science and Artificial Intelligence, University of Jeddah, Jeddah, Saudi Arabia and has great attention to various research endevours.

Nahla Al Jojo is with College of Computer Science and Engineering, Department of Information System and Technology, University of Jed-dah, Jeddah, Saudi Arabia. Her research interest is in the area of artifical intelligence.

Araek Tashkandi is currently with the College of Computer Science and Engineering, Department of Information System and Technology, University of Jed-dah, Jeddah, Saudi Arabia. He has published many researsh papers.

Abdullah Alghoson is with the College of Computer Science and Engineering, Department of Information System and Technology, University of Jeddah, Jeddah, Saudi Arabia. His research interest is on the area of machine learning.

Ameen Banjar is currently with the College of Computer Science and Engineering, Information system and Technology Department, University of Jeddah, Jeddah, Saudi Arabia. His research interest is in the area of machine learning.

Nidhal K. El Abbadi is currently with the Computer and engineering techniques Department, College of engineering and tech-niques, Al-Mustqbal University, Babylon, Iraq. His research interest is in the area of Machine learning.