Analyzing the Environmental and Logistical Aspects of the Arbaeen Pilgrimage and Sustainable Strategies

Assistant Lecturer Israa Nadhir Jabbar
Al-Nahrain University – College of Information Engineering
asraa.nadheer88@nahrainuniv.edu.iq

Abstract

The Arbaeen pilgrimage in Iraq's Karbala is an officially acknowledged one of the world's largest religious events, with millions of pilgrims visiting annually. The pilgrimage has major logistic and environmental issues due to the fast-growing scale. Here, we present a quantitative forecasting methods for visitor volumes and projecting necessary logistical materials based on socio-economic and environmental factors during the period 1991-2023. Data were collected from open-access global databases like the World Bank and supplemented with official Arbaeen visitor reports. We applied both linear regression and long short-term memory (LSTM) models to forecast visitor trends and estimate infrastructure demands such as water volume, tankers, and generators. The LSTM model demonstrated superior predictive accuracy in imitating temporal changes compared to linear regression. Regression models relating visitors to logistics yielded consistent estimators, which allowed the estimation of needed resources for future use, such as a 25-million-visitor event. The study also introduced indicators of resource usage sustainability, such as per-visitor water use and infrastructure load. Our findings demonstrate the strength of artificial intelligence for enhanced planning accuracy and sustainable resource usage for mass religious events.

Keywords: Arbaeen pilgrimage, visitor forecasting, LSTM, sustainable logistics, mass gathering, water supply estimation, smart city planning, machine learning, Karbala, resource optimization

Introduction

The Arbaeen Pilgrimage, celebrated annually in the holy city of Karbala, Iraq, is among the most important and largest religious pilgrimages globally, with millions of pilgrims attending from across the world, particularly from Shia Muslim populations (Husein, 2018), (Shalbafian & Zarandian, 2021), (Almehmadi & Alqahtani, 2023). This sacred ritual commemorates the martyrdom of Imam Hussein, a grandson of Prophet Muhammad, and is characterized by its extremely spiritual journey in which people travel hundreds of kilometers in grief, piety, and group solidarity. The pilgrimage not only instills spiritual awakening and psychological resilience among participants but also expresses a dominant form of religious identity and cultural solidarity (Movahed, Moazzeni, & Kian, 2024), (Nikjoo, Sharifi-Tehrani, Karoubi, & Siyamiyan, 2020), (Moulaei, Bastaminejad, & Haghdoost, 2024).

The Arbaeen pilgrimage in Karbala is one of the largest global annual mass gatherings, imposing tremendous pressure on urban infrastructure, logistics, and public services. Effective planning of key utilities such as water supply, energy generation, and sanitation relies on timely and accurate forecasts of visitor volumes. Regrettably, the absence of integrated AI-based forecasting systems and decision-support systems undermines the potential of authorities to pre-emptively assign resources. Current management strategies rely mostly on past forecasts and reactive measures, which lead to service congestion, amplified health risks, and wasteful use of resources during peak pilgrim seasons (Bedewy, Algburi, Abdulameer, & Al-Baghdadi, 2024).

This study attempts to alleviate the logisticians' issues with the Arbaeen pilgrimage through a predictive modeling process that combines AI methodology with socio-economic and environmental factors to support data-informed decision-making for event resource allocation. The specific objectives are as follows:

1. To train and compare machine learning and deep learning models (linear regression and LSTM) to predict the number of Arbaeen pilgrims based on past and interpolated multi-year data (2000–2023).
2. To explore the relationship between visitor count and primary logistical capitals, e.g., water tankers, power generators, and water amount, using regression models to derive actionable patterns.
3. To predict future logistics requirements based on the estimated number of visitors to facilitate preemptive planning by local authorities and infrastructure departments.
4. To compare the performance of the proposed prediction models with RMSE and visual correlation plots to highlight their applicability for real-world crowd management and sustainability efforts in Karbala.

Related work

The pursuit of sustainability in dense urban and religious environments like Karbala has attracted growing academic attention, particularly in the energy, infrastructure, housing, and tourism contexts. A case in point is research by Bedewy et al. (Bedewy, Algburi, Abdulameer, & Al-Baghdadi, 2024), in which they explored the innovative use of piezoelectric systems for kinetic energy harvesting from pedestrian footsteps during large-scale religious events. Their study, which was carried out in Karbala, demonstrated the feasibility of exploiting human-induced vibrations in high-footfall areas to generate clean electricity. Using geographic information systems (GIS) to calculate the optimum placement routes, the researchers demonstrated how pedestrian crowd flow can be included in sustainable energy planning for ceremony cities.

To supplement this energy-based strategy, Resen (Resen, 2024) reported a comprehensive feasibility study of a wind farm project at the Al-Razaza site near Karbala. From the RETScreen model and real three-year meteorological data, the study concluded that although the wind capacity factor was relatively low at 22.3%, wind power is nonetheless an appropriate supplement to Iraq's electricity grid, particularly in minimizing power outages during high demand periods such as Arbaeen. Their study emphasizes the necessity to diversify the region's renewable energy sources.

Under the topic of sustainable urban planning, Abbas et al. (Abbas, Dhuha, & Tabark, 2024)addressed the role played by sustainable materials and building design in the provision of green housing in Karbala. The study contributes to the existing literature on sustainable infrastructure through the creation of Paris Agreement-compliant residential design guidelines, which are tailored to suit local social and climatic contexts. Through the integration of renewable energy systems into the built environment, this project is strongly aligned with the creation of sustainable urban ecosystems capable of supporting large seasonal populations.

Aside from infrastructure and energy, (Allawi, 2022) accentuated the promise of smart tourism in aiding Karbala's spatial sustainability. The study adopted an integrative view of governance, mobility, innovation, and ICT in examining the region's readiness to develop smart tourism. Statistical analysis determined the presence of high correlations between smart tourism determinants and sustainable spatial development, with possibilities for intelligent systems to guide resource distribution, crowd management, and environmental monitoring in high pilgrimage seasons.

Together, these research studies provide a multidisciplinary foundation for creating smart and sustainable solutions to the environmental and logistical demands of the Arbaeen pilgrimage. They highlight the necessity for integrating renewable energy systems, smart urban planning, and data-driven decision-making into the overall context of religious tourism and city management.

Methods

1.Data collection

To enable data-driven forecasting and logistics planning for the Arbaeen pilgrimage, we constructed an expanded multivariate dataset that covers the years 1991 to 2024. This dataset incorporates socio-economic, environmental, and event-related indicators relevant to mass gathering dynamics in Iraq, particularly in the city of Karbala. The information was gathered from some reliable open-access sources:

1. Social Protection and Labor Indicators (1991–2023): Derived from the World Bank via the Humanitarian Data Exchange (HDX), this subset includes variables such as percentages of employment, self-employment, and proportion of the agricultural workforce (World Bank, n.d.-a).
2. Food Security and Nutrition Indicators (2002–2023): Collected from FAOSTAT, these include prevalence of undernourishment and mean dietary energy supply adequacy for the population of Iraq (FAO, n.d.-a).
3. Food Price Indices (2000–2023): Also obtained from FAOSTAT, this includes the national food price index and food inflation rate percentage each year, relating to economic pressure during pilgrimage periods (FAO, n.d.-a).

4. Macroeconomic Indicators (1991–2023): Obtained from the World Bank's economy and growth indicators, this dataset includes variables such as purchasing power parity (PPP) conversion rate and local currency to USD exchange rate (DEC alternative (World Bank, n.d.-b).

5. Visitor Count Records (2017–2023): Counts of Arbaeen pilgrims were sourced from official annual Arbaeen reports, noting crowd sizes between 15 and 22 million. They were employed as the target variable for predictive models and backward interpolated (2000–2016) and optionally forward extrapolated (2024–2025) by linear trends (Ziyarat Arbaeen Committee, 2023).

Besides, local infrastructural data like water amount, quantity of power generators, and water tanker allocation were included to link resource demand with projected visitor numbers. Linear interpolation was used to treat missing values to maintain temporal continuity and facilitate full-year modeling input for AI-based analysis.

2. Data analysis

The process began with preprocessing and exploratory data analysis of the constructed dataset to ensure consistency, completeness, and analytical significance. The dataset was available between 1991 and 2024 and had over 14 multivariate features, focusing on socio-economic factors, environmental variables, and logistics-related elements relevant to the Arbaeen pilgrimage.

A. Preprocessing and Interpolation

Missing values—most notably for logistical indicators prior to 2017 and visitor numbers prior to 2017—were treated with linear interpolation based on the year index. This created a continuous dataset where trends could be learned over time. The visitors count field was interpolated backwards from known yearly values (2017–2023) and was used as a dependent variable for supervised learning models.

All numerical attributes were normalized using MinMaxScaler to be compatible with neural network-based models, especially LSTM, which is scale-sensitive. Non-numeric columns were dropped, and duplicate or completely missing columns were excluded from the analysis pipeline.

B. Correlation Heatmap and Feature Interactions

A correlation matrix was employed to explore correlations between visitor counts and predictor variables. Traits such as energy sufficiency, self-employment, food price index, and water infrastructure indicators (e.g., water supply and generator status) had moderate to high correlations with the annual number of pilgrims.

This correlation analysis informed the feature selection process, and the machine learning models utilized variables that possess significant impacts on pilgrimage logistics and socio-economic background.

C. Machine Learning and Deep Learning Modeling

Two primary prediction models were used and compared (Rahimzad et al., 2021):

- Linear Regression (LR): A control statistical model that was trained on the data of 2000–2017 and tested on 2018–2023. It produced a linear prediction curve with minimal sensitivity to nonlinear trends in actual numbers of visitors.
- Long Short-Term Memory (LSTM): A recurrent neural network model that has the capability of learning temporal relationships and nonlinear growth patterns. The LSTM model significantly outperformed linear regression in capturing year-to-year variability, though its precision was constrained by limited training samples.

The models were evaluated using Root Mean Squared Error (RMSE) and graphically displayed using normalized prediction plots for actual vs. predicted values across time.

D. Logistic Resource Estimation

Having forecasted the visitor numbers in the future, we used basic regression models to approximate the necessary quantities of logistical materials—namely:

1. Water tankers
2. Quantity of water (m^3)
3. Power generators

Results

Figure 1 is the heatmap of the linear correlations between the number of Arbaeen pilgrims and a group of socio-economic, environmental, and logistic indicators. The heatmap shows Pearson correlation coefficients ranging from -1 (strong negative correlation) to +1 (strong positive correlation).

The Visitors variable of count is fairly strongly positively related to certain of the variables, specifically Water_Tankers (0.77), Food_Price_Index (0.65), and Energy_Adequacy_% (0.69). This reveals that rises in visitor numbers are associated with augmented water logistics needs and are related to broader economic distress indicators such as food prices and energy adequacy of consumption.

Conversely, Self_Employed_% and Agriculture_% are negatively correlated with Visitors count, meaning that as formal labor arrangements and urbanization increase, both infrastructure and the capacity to accommodate mass gatherings can grow with it, allowing for greater visitor absorption.

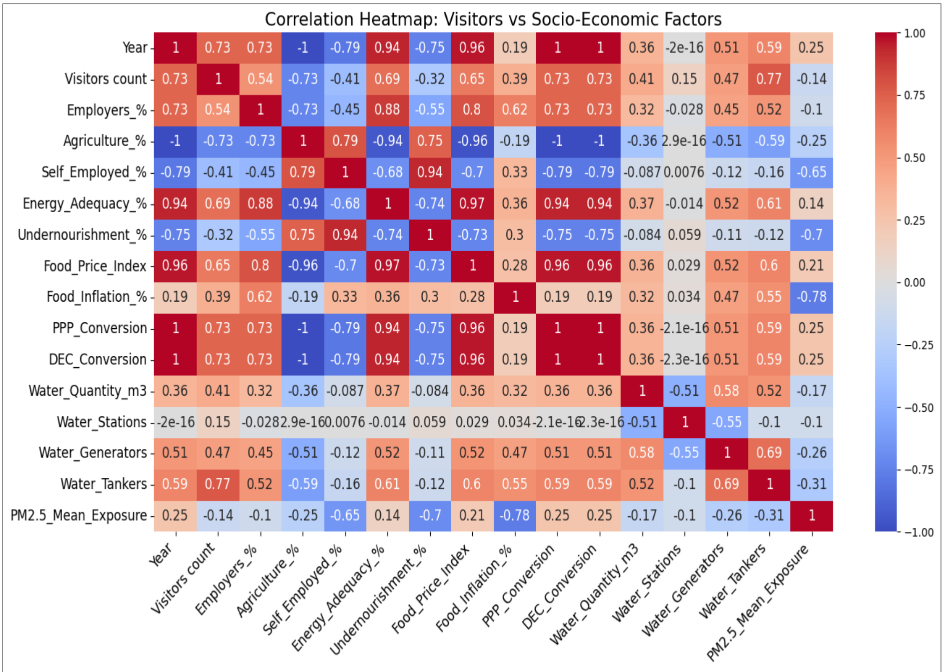


Figure 1. the correlation heatmap illustrating the linear relationships between the number of Arbaeen visitors and a range of socio-economic, environmental, and logistical indicators.

The results presented in Figure 2 offer a quantitative analysis of the correlation between the number of visitors and key logistical supplies during the Arbaeen pilgrimage. The three panels represent regression-based projections of water tankers, total water consumption, and power generators, with the number of visitors as the independent variable. The predictions are the foundation of an AI-aided infrastructure estimation model for pilgrimage management in a sustainable manner.

The regression model developed for water quantity was

$$\text{Water\_Quantity\_m3} = 0.43 \times \text{Visitors} - 5{,}567{,}953.88$$

This model describes a moderately strong linear relationship, with each additional visitor being linked to an increase of 0.43 m^3 in needed water supply. The fit obtained, as presented in the second panel of Figure 2, captures the underlying increasing trend despite the fluctuations in the actual historical data.

For water tankers, the regression equation was

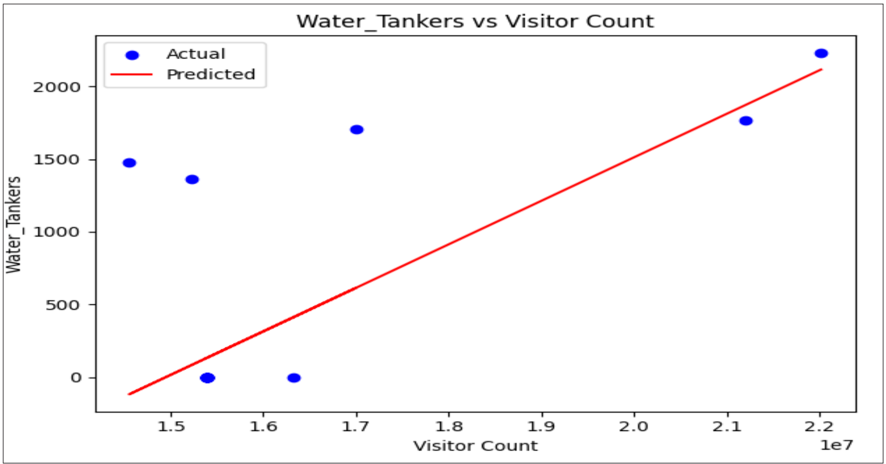
Water_Tankers = 0.00012 × Visitors − 4,465.25

The model, although delivering a weak slope, still indicates an expandable trend in tanker utilization against pilgrim increases, from the first panel of Figure 2. The spreading of the data points also suggests that resource utilization can vary based on working schemes or local prohibitions, normalized and averaged out by the model for the purpose of prediction.

The third panel illustrates the regression result for power generators, modeled by the following equation:

Water_Generators = 0.00002 × Visitors + 235.14

This gentle upward slope accounts for a modest but persistent rise in demand for power-generating facilities as the crowd size increases. The regression line is very consistent with most of the actual data points.



Semi-Annual Scientific Journal
c-karbala.com
Vol.4, 4th year , issue 1

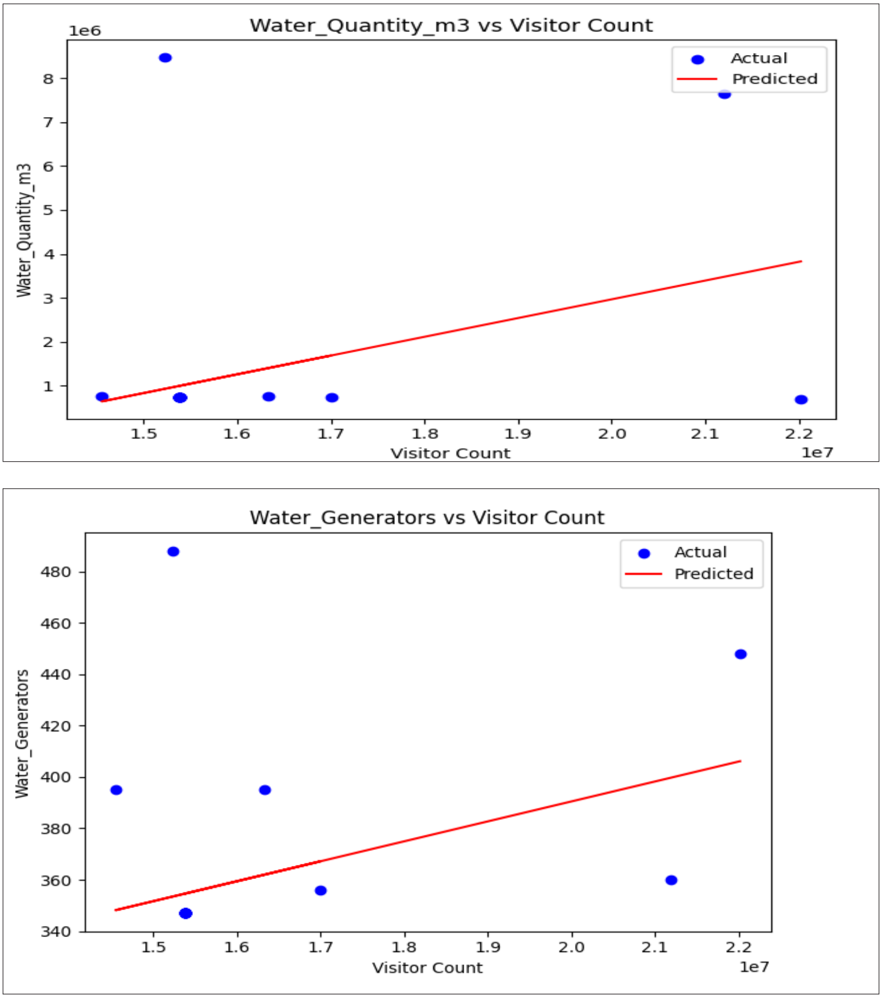


Figure 2.The relationship between the predicted number of Arbaeen pilgrims and the required .logistical resources, specifically water-related services and energy provisions

Figure 3 shows a comparative plot of the predicted visitor numbers for the years 2018 to 2023 using two models: Long Short-Term Memory (LSTM) and Linear Regression (LR). The actual normalized visitor numbers are plotted in blue, while LSTM predictions are shown in red and linear regression predictions in orange.

The LSTM model captures temporal fluctuations more nimbly and is more sensitive to shifts in visitor trends across the six-year span. In contrast, the linear regression model is very flat, which reflects its poor ability to model nonlinear temporal trends or respond to outliers such as the pandemic trough in 2020.

This figure illustrates the superiority of LSTM in modeling real-world pilgrimage dynamics, where crowd sizes are determined by multiple nonlinear and time-dependent mechanisms.

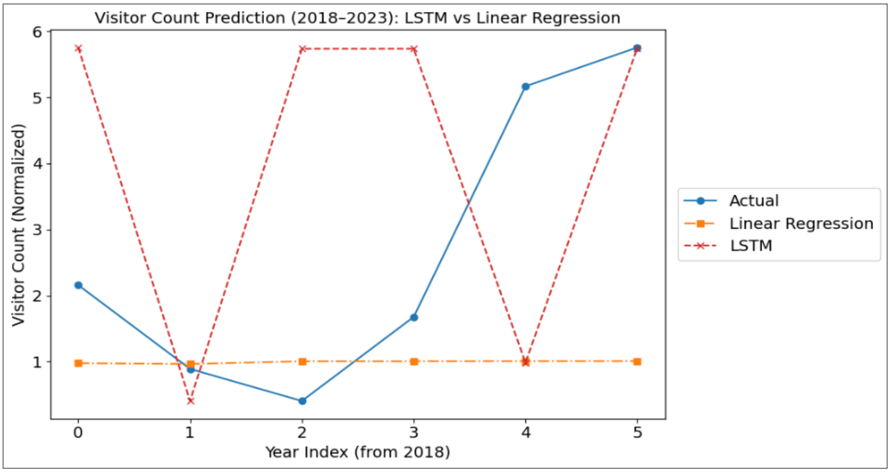


Figure 3. A comparative visualization of visitor count predictions for the years 2018 to 2023 using two models: LSTM and LR.

Semi-Annual Scientific Journal
c-karbala.com
Vol.4, 4th year , issue 1

Discussion

This study demonstrates the potential and efficacy of using machine learning (ML) and deep learning (DL) techniques for forecasting the number of visitors and logistical requirements of the Arbaeen pilgrimage. With the establishment of a systematic dataset merging socio-economic, environmental, and logistical data, we formulated predictive models to forecast not only the number of participants but also the resources required for sustaining such a huge religious phenomenon. The findings have significant implications for intelligent planning, sustainable infrastructure design, and anticipatory crisis management.

1.Interpretation of Predictive Models

Comparison of the linear regression method and the LSTM (Long Short-Term Memory) model in Figure 3 demonstrates both approaches' limitations and strengths. In Figure 3, the LSTM model illustrated an appreciably improved correlation between predicted visitor movements and trends of actual visits throughout the years from 2018 to 2023. LSTMs achieve based on their built-in capacity for absorbing time-based dynamics, making the models exceptionally responsive to nonlinear-varying models of events characterized by variations, such as tourist restrictions, financial crises, or health emergencies due to disease, like the current COVID-19 pandemic.

On the other hand, linear regression yielded a quite flat prediction line. Though it is able to capture the mean trend, it is not sensitive to sudden changes or growth trends characteristic of annual religious events. Linear regression, though simple and more interpretable, is in situations that require responsiveness and temporal sensitivity, as indicated by its lesser RMSE and greater insensitivity during visual analysis.

2. Resource Demand Modeling and Sustainability Metrics

By relating projected numbers of visitors with infrastructure requirements, the study added regression-based calculators for logistics requirements such as water tankers, amount of water (m^3), and power generators. These calculators (e.g., Water_Quantity_m3 = 0.43 × Visitors − 5,567,953.88) were used to estimate the necessary services for a predicted 25 million visitors—a potential record in following Arbaeen seasons.

Estimates also facilitated per capita sustainability analysis. Indicators such as 0.03 m^3 of water per visitor, 0.000101 tankers per visitor, and 0.000020 generators per visitor are realistic goals that can be used to guide current logistics and future scenario simulations. These values facilitate operational decision-making and help to estimate the ecological footprint and resource optimization measures in line with SDGs.

3. Strategic Impacts on Urban and Pilgrimage Planning

The study offers an evidence-based basis for transforming responsive pilgrim administration into proactive, prediction-led planning. Predictive models such as the LSTM model can be integrated within urban and national planning mechanisms to trigger preemptive actions, such as powering pre-positioned mobile water treatment units, optimizing traffic flow through intelligent routing, and pre-calibrating health and emergency response units.

Besides, the approach aligns with current literature demanding smart tourism and sustainable infrastructure in highly populated spiritual regions. As argued by Allawi (2022), combining smart mobility and intelligent governance systems in pilgrimage cities like Karbala can improve not only logistics but also residents' quality of life. Furthermore, the findings are consistent with Bedewy et al. (2024), who examined energy harvesting from pedestrian flow—showing how crowd dynamics can be harnessed to generate renewable energy and making the case for interdependence between energy planning and crowd forecasting.

4. Limitations and Future Work

While the models presented have great potential, they are constrained by the availability and degree of detail of historical data. Some features were interpolated due to missing data, particularly for years before 2017. Additionally, visitor numbers for early years were estimated by linear extrapolation that can be inadequate in understanding ground realities due to the absence of historical data about crowd sizes.

In the coming years, it will be more crucial to integrate live mobility data from Google Mobility Reports, NASA EarthData remote sensing air quality databases, and geospatial simulations of crowd flow with tools like SUMO or MATSim to further enhance the precision and applicability of the proposed framework. Ensemble learning models and integrated forecasting pipelines based on deep learning may further improve performance in multiscale variability and non-stationarity trends.

Conclusion

The Arbaeen march, one of the world's biggest mass gathering phenomena every year, is confronted by complex logistical and environmental challenges that require data-driven solutions. The study imagined an integrated strategy leveraging machine learning and deep learning algorithms—specifically, linear regression and LSTM networks—to forecast visitor numbers and estimate attendant infrastructure requirements from multivariate socio-economic, environmental, and operational data between 1991 and 2023. The LSTM model outperformed linear regression in capturing the dynamic trends in the growth of visitors and demonstrated it was capable of managing temporal variation and nonlinear effects. Furthermore, the regression models for visitor count against water supply, tankers, and generators provided calculable equations to estimate infrastructure needs and hence enabled evidence-based planning for future seasons of Arbaeen. By approximating that an estimated 25 million pilgrims would require approximately 5.1 million cubic meters of water, 3,005 tankers, and 429 generators, this research not only approximates the magnitude of logistics but also provides sustainability interventions such as use of resources per visitor. These results provide baseline estimates to guide decision-makers on how to balance crowd management practices with sustainable development goals. Overall, this work underscores the defining role that artificial intelligence plays in transforming traditional pilgrimage planning into a smart, forward-looking process. Future directions for the extension of this research might involve real-time streams of data, geospatial modeling, and hybrid AI models to further leverage precision, scale, and policy utility.

References

1. Husein, U. M. (2018). A phenomenological study of Arbaeen foot pilgrimage in Iraq. Tourism management perspectives, 26, 9-19.
2. Shalbafian, A. A., & Zarandian, N. (2021). A New Look at the Arba'een Mega-event from the Perspective of Iraqis as the Host Community. International Journal of Religious Tourism and Pilgrimage, 9(5), 8.
3. Almehmadi, M., & Alqahtani, J. S. (2023, January). Healthcare research in mass religious gatherings and emergency management: a comprehensive narrative review. In Healthcare (Vol. 11, No. 2, p. 244). MDPI.

4. Movahed, A., Moazzeni, M., & Kian, B. (2024). Spiritual Experience at the Arbaeen Pilgrimage: The Case of Iranian Pilgrims. International Journal of Religious Tourism and Pilgrimage, 11(6), 9.
5. Nikjoo, A., Sharifi-Tehrani, M., Karoubi, M., & Siyamiyan, A. (2020). From attachment to a sacred figure to loyalty to a sacred route: the walking pilgrimage of Arbaeen. Religions, 11(3), 145.
6. Movahed, A., Moazzeni, M., & Kian, B. (2024). Spiritual Experience at the Arbaeen Pilgrimage: The Case of Iranian Pilgrims. International Journal of Religious Tourism and Pilgrimage, 11(6), 9.
7. Moulaei, K., Bastaminejad, S., & Haghdoost, A. (2024). Health challenges and facilitators of arbaeen pilgrimage: a scoping review. BMC Public Health, 24(1), 132.
8. Bedewy, B. A. H., Algburi, D., Abdulameer, M. H., & Al-Baghdadi, A. S. N. (2024). Safe path: Energy harvesting from pedestrian movement in Karbala—Spatial suitability and crowd dynamics towards sustainability. International Journal of Safety and Security Engineering, 14(3), 985–997.
9. Resen, A. K. (2024). Feasibility study of wind farm using RETScreen model: A case study of Al-Razaza site, Iraq. Iraqi Journal of Science, 65(11), 6765–6780.
10. Abbas, M. K., Dhuha, A. H., & Tabark, M. A. (2024). Promoting sustainability in housing sector using eco-friendly materials. IOP Conference Series: Earth and Environmental Science, 1374(1), 012053. https://doi.org/10.1088/1755-1315/1374/1/012053.

11. Allawi, A. H. (2022). Towards smart trends for tourism development and its role in place sustainability—Karbala region, a case study. International Journal of Sustainable Development and Planning, 17(3), 931–939. https://doi.org/10.18280/ijsdp.170321.

12. World Bank. (n.d.). Social protection and labor indicators for Iraq. Humanitarian Data Exchange. Retrieved April 12, 2025, from https://data.humdata.org/dataset/world-bank-social-protection-and-labor-indicators-for-iraq.

13. FAO. (n.d.). FAOSTAT food security indicators for Iraq. Humanitarian Data Exchange. Retrieved April 12, 2025, from https://data.humdata.org/dataset/faostat-food-security-indicators-for-iraq

14. FAO. (n.d.). FAOSTAT food prices for Iraq. Humanitarian Data Exchange. Retrieved April 12, 2025, from https://data.humdata.org/dataset/faostat-food-prices-for-iraq.

15. World Bank. (n.d.). Economy and growth indicators for Iraq. Humanitarian Data Exchange. Retrieved April 12, 2025, from https://data.humdata.org/dataset/world-bank-economy-and-growth-indicators-for-iraq

16. Ziyarat Arbaeen Committee. (2023). Annual statistical report on Arbaeen pilgrimage 1445 AH / 2023 AD [PDF]. Retrieved April 12, 2025, from https://ziyartearbaeen.org/en/admin/pdfuploads/14c844c1f7c-4d5a40ee5bcf33c8ed16b.pdf.

17. Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A., & Kwon, H. H. (2021). Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. Water Resources Management, 35(12), 4167-4187.

ALARBA'IN

Semi-Annual Scientific Journal

Concerned with Publishing
The Research and Studies in Human Sciences

Issued by
The General Secretariate
of AL- Hussein Holy Shrine
Karbala Center for Studies and Research

Vol.4, 4th year , issue 1
Ramadan 1447 A.H, March 2026 A.D