Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN : 2074-5621
https://jatuc.uoturath.edu.iq/

Statistical Prediction of Diabetes Prevalence in Iraq using a Hybrid ARIMA–LSTM Model

Ammar Kuti Nasser*

Department of *Psychological Counseling and Educational Guidance* , College of Basic Education , Mustansiriyah University , Baghdad, Iraq.

*Correspondence email: dr.ammar168.edbs@uomustansiriyah.edu.iq

KEYWORDS
Applied statistics; time-series analysis; ARIMA, LSTM; hybrid models; diabetes prevalence

ABSTRACT
This study concerns the emerging challenge of diabetes in Iraq and aims to forecast its prevalence with advanced statistical methodologies . Diabetes is a rising health problem in Iraq attributed to variables from obesity to hypertension and an outdated healthcare system. Conventional linear models fail to well illustrate the complicated and non-linear nature of the history of the disease, thus we use a two-stage method in this work. The linear trend is extracted from the data with ARIMA and then it is used to predict for non-linear analysis using LSTM. Data were gathered, refined and analyzed from 7 Iraqi governorates. The diabetes prevalence decline identified by using the ARIMA-LSTM hybrid model in this data suggests that the integrated model showed better prediction of diabetes prevalence than individual BPMs. The hybrid model was found to be the best way to reduce the prediction error, noise and incomplete data processing. Given the robustness of the model in various locations and sample sizes, it could be considered as a decision support instrument for health policy planning, early detection strategies, prevention and resource allocation. The paper suggests more in-depth data analysis and exploring other sociodemographic factors, as well as training healthcare software developers in advanced data science. This method may be suitable for other chronic diseases as well.

الكلمات المفتاحية
كلمة مفتاحية , كلمة مفتاحية, كلمة مفتاحية, كلمة مفتاحية , كلمة مفتاحية , كلمة مفتاحية

الملخص
يجب أن يكون الملخص فقرة واحدة مستقلة تلخّص المخطوطة بأكملها. من الأفضل أن يوفّر سياقًا موجزًا للدراسة، ثم يصف المنهج العلمي وبعض النتائج الرئيسية بشكل نوعي. وينتهي بجملة توضّح الآثار المترتبة على هذا البحث في مجاله. لا يجوز تضمين مراجع أو أشكال أو جداول في الملخص.

1. INTRODUCTION

Diabetes mellitus is among the most common and complicated chronic illnesses worldwide in several countries. It is a metabolic disorder in which the pancreas produces too little insulin, or the body cannot make efficient use of the insulin it does produce. This imbalance causes a sudden glucose spike, which can be extremely damaging to organs such as the nerves, kidneys and eyes. About 537 million people were thought to have diabetes worldwide in 2021. But they will need to account for this because is poised to affect a staggering 700 million and beyond people if effective strategies are not initiated [1].

Diabetes prevalence in Iraq is increasing, reflecting a major public health problem[2]. This syndrome is mainly a result of the representative biomedical conditions of high obesity prevalence, uncontrolled blood pressure and diverse blood glucose level. In addition, health care and influences are not equilibrate from one region to another inside the country: it also deteriorate patients condition. The complex medical mechanism of diabetes by many nonlinear factors, the integrating prediction without combination of linear statistical models is not easily made prediction accurately. To meet the need, in the present study we follow a hybrid approach that combines state-of-the-art statistical methods with deep learning procedures to detect both linear and non-linear patterns from medical data [3]. The advantage of the ARIMA model is that it can analyze and interpret linear trends in time series data, while the LSTM has a strong capability to analyse timetable feature when identifying long-term and short-term temporal dependencies, meanwhile capturing complex relationships underlying sequence data.Through the integration of the two methods, this ARIMA-LSTM hybrid model is intended to develop a more accurate and effective prediction technique with the purpose of correctly predicting diabetes spread in Iraq for enhancing national health preparedness and making better preventive decisions [4] [5] [6].This analysis includes data from seven of the key governorates in Iraq, which were selected to provide good geographical and health coverage throughout the country. Here, we are focusing on the biomedical variables themselves and their within-community associations. By using advanced statistics and AI techniques, this health data study sets out to address inequalities in local research whilst modelling accurate prediction of outcomes using high-quality clinical data. Lastly, this work aims to establish a scientific rationale for evidence-based decisions and preventive health practices in order to mitigate public health consequences in Iraq.

2. LITERATURE REVIEW

Diabetes prevalence prediction using statistical or advanced artificial intelligence models: a literature review. Besides, many local and international studies have covered varying methods of predicting diabetes prevalence using the statistical model or more superior ones such as AI. Many of such works demonstrate the importance adopting a hybrid model, in which traditional time-series prediction models (i.e., ARIMA model) and artificial intelligence methods (such as Long Short Term Memory(LSTM) network), are combined.On the national level, Anwar Abdul Nasser (2009) at College of Medicine, University of Tikrit carried out a researchers on "ImmunoGenetics and diabetes in Iraq", it was one of big study immunological and genetic designed to elaboration on immunological responses and genetic factors for arise diabetes in Iraq. Prof Hassan Irfan Hussein, Institute for Life Sciences, University of Anbar advised us on this research and it provides some scientific clarifications about immunological aspects regarding the prevalence of the disease. In a related work, Al-Diwaniyah Teaching Hospital (2018) conducted an analytical study on blood specimens taken from 25 patients with diabetes in which interesting findings about immune markers were achieved and could play a role in the prediction modeling. Moreover, Enas Hussein Ali and her team conducted a systematic review of Iraqi researches on type 2 diabetes (2023) suggesting the use of hybrid models to increase prediction accuracy in light of such complexity for local datasets.Globally, Mandi et al. (2018) provided prediction model for complex time series data in diabetes disease management through LSTM neural network with higher level of accuracy than that of classic methods. Khan et al. (2023) compared ARIMA to deep learning models and highlighted that ARIMA model is unable to capture the non-linear patterns in data. These observations underscore the need for using sophisticated machine learning algorithms (Mustafa et al., 2025; Chachan et al., 2023; Suman, Singh, and Gulati, 2024) to learn complex temporal patterns from medical information.Al-Haddad et al. (2024) demonstrated that it was imperative to add the socioeconomic factors in the prediction model with regard of performance and evidence based decision-making in health. Another significant work, al Kilani (2021), analyzed the influence of obesity and food consumption on diabetes prevalence in Iraq through large-scale datasets and sophisticated statistical techniques. New evidence (Ansbro et al., 2022; Wang et al., 2025) has been provided researchers of the role of environmental and lifestyle determinants on diabetes burden in diverse ethnic groups.In summary, there is a growing evidence in literature which strongly suggests the need to establish an hybrid model by exploiting the best features of ARIMA and LSTM. In addition, it enables effective prediction of diabetes prevalence in Iraq with regard to both timing and accuracy and produces research products to inform strategic health policy planning while leading patients toward a better quality of life. The reviewed studies have also shown that considering the specific demographic, socioeconomic and geographical context is crucial to model disease prevalence in Iraq.

3. RESEARCH PROBLEM

Projections for the prevalence of diabetes in Iraq are increasingly challenging and difficult to make, especially when biomedical categories are taken as main core. Clinical information â€¦ In this paper, we focus on using computer assistance to extract useful information for enhancing the diagnosis of patients from clinical data which are usually highly non-linear and complicated in structure but very valuable. Also, it is common knowledge that medical records often come with seasonal and temporal variations, non-uniform quality and accuracy especially in areas where healthcare infrastructure has to struggle or even bear the brunt of instability.Based on these problems, the main research problem in this paper is that an integrated prediction model needs to be constructed which can overcome the drawbacks of the traditional methods. This model should integrate the linear time-series analysis of ARIMA model and the deep learning properties of Long Short-Term Memory (LSTM) network, which is effective at capturing nonlinear and complicated relations among biomedical variables. We would like to build models that are better calibrated and more flexible, incorporating this intricate relationship between medical variables and allowing informed decision-making in the management of health policy, preventive interventions as well as early diabetes screening programs. That is, the predominant issue is that our learned models are not rich enough for modelling the complexity in medical data and lead to quality distortion of prediction as well as tactical decision w.r.t. reality. A new hybrid model is needed to be filled the literature's gap that combines the strengths of contemporary statistical method and artificial intelligence (AI) methods which were designed based on Iraqs healthcare data peculiarities since its rates for diabetic patients remain increasing.

4. RESEARCH OBJECTIVES

In this work, we try to build a hybrid prediction model that incorporates the merits of ARIMA in analyzing linear time-series patterns and LSTM's in dealing with the non-linear and complex interaction among diabetes-related biomedical data. The model is intended to be applied for correct and optimal screening prediction of the distribution and spread spectrum of diabetes in Iraq when sufficient medical data are available at seven vicegerencies. Such forecasts seek to enhance the understanding of disease dynamics and prevention and treatment interventions.Secondary objectives are to compare model performance with each other and to that of sole biomechanics-predicted kinetic outputs versus full model outputs using robust accuracy metrics (e.g., MAE, RMSE) and establish the effect of biomedical covariate influences on predictive ability. The current study also aspires to provide practical evidence-based policy recommendations for health-care payers on how to improve the management of resources and support the scaling-up of data-driven and evidence-based health policies. Moreover, our work attempts to address the identified limitations in DM prediction studies within Iraq by considering model flexibility towards local medical data characteristics. It also underscores that more advanced artificial intelligence and statistical analysis should be considered to optimise patients' life quality and appropriate therapeutics concerning the Iraqi medical sectors.

5. METHODOLOGY

In this paper, a novel analytical solution that integrates sound statistical methods and state-of-the-art artificial intelligence (AI) methodologies is proposed on establishing the hybrid prediction model with data-borne characteristics for predicting diabetes prevalence based on biomedical variables. The research stages are then summarised in the subsequent main sections.

6. DATA COLLECTION AND PREPARATION

Published diabetes prevalence and secondary data sources such as the International Diabetes Federation (IDF), World Health Organization (WHO) and published local studies were used. The study was based on annual data covering the period 2014-2024 from seven Iraqi governorates. Linear interpolation was used to build a complete time series consisting of 77 observations. The columns of the dataset included important biomedical measures Average Systolic Blood Pressure (mmHg), Avg. Ongoing BG (mg/dL), BMI (kg/m²) and Obesity Rate (%) which were further studied to investigate their association with prediction errors. The cleaned data was normalized using

Min-Max Scaling and partitioned into 80% for training and 20% for testing in order to ensure that the predictive model perform well. [2], [3], [4].

7. EVALUATION AND ANALYSIS

Models were evaluated in a robust, unbiased manner using multiple splits. The model performance was evaluated using classical statistical accuracy indices including MAE, RMSE and the coefficient of determination (R^2). These numbers provide a general sense of what we expect with respect to prediction errors and model fit quality [11].Finally we performed statistical analysis to compare forecasting accuracy of ARIMA, LSTM and hybrid approach -ARIMA-LSTM with statistically significant difference found in favor of the latter. In addition, the influence of different biomedical variables on model performance was illustrated which indicated that the hybrid model can be more flexible in dealing with some complex relationships existing in medical data [12]. Sensitivity tests were carried out to show the data noise and model variability of robustness, indicating hybrid model is a stronger tolerance ability and stability [13].This thorough testing can provide confidence that the presented modeling framework provides reliable and plausible predictions which is necessary for informed health policy planning for diabetes care in Iraq [14].

8. MODEL DEVELOPMENT

An ARIMA model was developed in order to explore linear trend within the epidemiological data of diabetes prevalence. The ARIMA model is used to fit non-stationary series that includes differencing, autoregressive and moving average components . A Long short term memory (LSTM) neural network was proposed to model the non-linear and complex temporal dynamics in biomedical data. The LSTM is specifically designed to capture longer term dependencies in sequences through the use of gated memory blocks that regulate the flow of information [4] [5].The hybrid ARIMA--LSTM model is thus developed by using both methods effectively. Here, ARIMA represents the linear part and seasonality and LSTM represents residual non-linearity pattern. The current model is a fused one in order to enhance accuracy and reliability of predicting diabetes prevalence in the Iraqi population [6] [10].

9. APPLIED PROCEDURES

Hybrid ARIMA–LSTM model output was used to conduct spatial and spatiotemporal visualization in the form of maps across Iraq with respect to diabetes prevalence. This implementation phase endeavored to sample the different governorate trends and patterns in Iraq[15]. ConclusionsBased on the prediction findings, some suggestions were made for public health planning and effective prevention [16]. The approaches demonstrate an association between model results and individual health policy decisions that may improve the management of diabetes and its population health effect [17].This comprehensive approach provides locally relevant information that is evidenced-based and has public health policy and practice implications in Iraq. The statistical analyses were performed using MATLAB R2023a software [19].

10. THE ARIMA MODEL (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

The ARIMA model is one of the classical and robust statistical method for analyzing and prediction of non-stationary time series data. It consists of three aspects: Autoregression (AR), differencing for stationarity (I) and Moving Average (MA) in the sense of modeling the residual noise [4]. The general expression for an ARIMA(p,d,q) model is given as:

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t \qquad (1)$$

where:
Y_t: the value of the time series at time t;
B: the backshift operator that shifts Y_t to Y_{t-1};
ϕ: autoregressive coefficients;
θ: moving average coefficients;
ε_t: white noise error term with zero mean and constant variance.

Expanding the equation gives the more detailed form:

Expanding the equation gives the more detailed form:

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (2)$$

Differencing $(1 - B)^d$is the art of detrending or deseasonalising a time series such that it may have better prediction capacity for the model.

The model parameters ϕ, θ are generally learned by statistical procedures like Maximum Likelihood Estimation (MLE). The adequacy of the model is assessed based on some criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) .

is the decomposition of the time series to de-trend or de-seasonalize it which will then make predictions with model easier.

11. AUTOREGRESSIVE (AR) AND MOVING AVERAGE (MA) MODELS

The ARIMA model is based on two basic functionalities: Autoregressive (AR) and Moving Average (MA) models.

The Autoregressive model of order p, indicated as AR(p), models a time series to be linearly dependent on its past p terms and noise term [4]:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \qquad (3)$$

where:
c: is the model constant;
ϕ_i: are the autoregressive coefficients;
ε_t: represents the white noise error term.

The Moving Average model, MA(q) depends on past forecast errors to model the time series and can be mathematically written as:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (4)$$

where μ is the average for the series and θ_i are coefficients of moving averages.

For the nonstationary time series, The ARMA model forms a combination of previous two models AR and MA which explains relationships among prior estimations of the series and prior residuals to enhance modeling/forecasting precision[4].

12. THE ARMA MODEL (AUTOREGRESSIVE MOVING AVERAGE)

The ARMA(p,q) model combines autoregressive (AR) and moving average (MA) for a stationary time series. This very dualism makes it possible to effectively capture dependencies based on previous observations and past forecast errors in the same equation.

The ARMA(p,q) model in general is as [4]:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \qquad (5)$$

Where:

- y_t represents the observed value at time t,
- c is a constant term,
- ϕ_i are the autoregressive coefficients,
- θ_j are the moving average coefficients,
- ϵ_t denotes the white noise error term.

It is the starting point in the process of dealing with a stationary time series and serves as crest glue for ARIMA when differencing requires to be carried out. Although it is good at capturing short-range linear dependencies, Kurignach PDF may not address nonlinear or long-term dependencies in complex biomedical data sets.

13. THE SARIMA MODEL (SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

The SARIMA model is a more general and the seasonal version of the traditional ARIMA framework and has become increasingly popular as it can accommodate seasonality often observed in real-world time series data such as for diabetes prevalence in this paper.

SARIMA adds seasonal terms to the autoregressive, differencing and moving average terms in order to handle repeating patterns over constant periods. This is particularly beneficial if there are seasonal patterns within the data, for example perspiringactivity over a year [21].

The SARIMA model is usually denoted as $SARIMA(p,d,q)(P,D,Q)_s$, where:

- p, d, q are the orders of the non-seasonal AR, differencing, and MA components respectively.
- P, D, Q are the seasonal orders of the AR, differencing, and MA components respectively.
- s is the length of the seasonal cycle (e.g., 12 for monthly data).

Mathematically, the model can be expressed as:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t \qquad (6)$$

Where B is the backshift operator, and ϕ,θ and their related ones are polynomials of non-seasonal and seasonal components.

The SARIMA model is able to adequately model seasonal behavior as well non-stationary trends, resulting in more accurate forecasts when seasonality dominates both.

14. THE LSTM MODEL (LONG SHORT-TERM MEMORY NETWORK)

LSTM model is a specific type of RNNs (Recurrent Neural Networks) which was created to address the shortcoming of traditional neural networks that do not take long range dependencies into account. LSTM networks are particularly useful for modeling sequential data that have non-linear and complex interactions like time sensitive series patterns especially as observed in biomedical datasets relevant to diabetes[22] .

LSTM works through the aid if memory cells that maintain informational states, update state and also remove them based on a set of gates i.e input gate, forget get and output gate. An LSTM cell can be mathematically expressed via the following equations:

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (7)$$

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (8)$$

Candidate cell state:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \qquad (9)$$

Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \qquad (10)$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \qquad (11)$$

Hidden state:

$$h_t = o_t * \tanh(C_t) \qquad (12)$$

where:

- x_t is the input vector at time t,
- h_{t-1} and h_t are the previous and current hidden states,
- C_t is the cell state at time t,
- W_f, W_i, W_C, W_o are weight matrices,
- b_f, b_i, b_C, b_o are bias vectors, and
- σ and tanh denote the sigmoid and hyperbolic tangent activation functions, respectively.

One of the advantages with LSTM networks is due to its effective sequence modeling that benefits in a robust representation learning from complex sequences, including biomedical data with nonlinear temporal dependencies like diabetes prevalence trends.

15. THE HYBRID ARIMA–LSTM MODEL

The hybrid ARIMA–LSTM model is prepared to take advantage of the complementary advantages of ARIMA and LSTM in better prediction with complicated time series data such as medical (biomedical) data on diabetes prevalence. The ARIMA part well models linear trends and seasonality in a time series, while the LSTM-part captures nonlinear and long term dependences that are not possible to address by ARIMA on its own[12] [13].

The methodology of hybrid modeling consists of the following main steps:

1. Fit the ARIMA model to original series data to remove both linear as well as seasonal effects.

2. Take the Residuals of your ARIMA model to extract the non linear part.

3. Fit the residual data to an LSTM model to capture the nonlinear relationships.

Finally, add the ARIMA predictions to the LSTM predictions of residuals to get the final forecast:

$$\widehat{Y}_t = \hat{Y}_t^{ARIMA} + \hat{Y}_t^{LSTM} \qquad (13)$$

Where:

- $\hat{Y}_t^{ARIMA}$ is the forecast from the ARIMA model,
- $\hat{Y}_t^{LSTM}$ is the forecast of residuals from the LSTM model,
- $\widehat{Y}_t$ is the combined hybrid forecast.

This hybrid model has shown high accuracy and stability in the presence of complex, noisy biomedical data while either model alone performs considerably worse. The hybrid model enhances generalization by comb The hybrid integration between ARIMA and LSTM models is summarized in the following steps:

Step 1: The ARIMA model was trained on the original diabetes prevalence data Y_t for the period 2014–2024.

Step 2: Generation of ARIMA forecasts: $\hat{Y}_t^{ARIMA} = f_{ARIMA}(Y_{t-p}, \ldots, Y_{t-1})$

Step 3: Extraction of ARIMA residuals by calculating the difference between the original values and the predicted values: $e_t = Y_t - \hat{Y}_t^{ARIMA}$

Step 4: The LSTM model was trained on the residual data e_t to capture the nonlinear patterns.

Step 5: Generation of LSTM forecasts for the nonlinear component: $\hat{Y}_t^{LSTM} = f_{LSTM}(e_{t-n}, \ldots, e_{t-1})$

Step 6: Integration of both predictions using Equation 13 to obtain the final hybrid forecast.

In this final combined forecast, both the linear trends (extracted by ARIMA) and the nonlinear dynamics (extracted by LSTM) are integrated, thereby achieving superior prediction accuracy for diabetes prevalence rates across the seven Iraqi governorates.

ining the temporal linearity of ARIMA with the non-linearity pattern recognition of

LSTM, thus particularly accommodated for LTC trend prediction such as diabetes [6] [23]

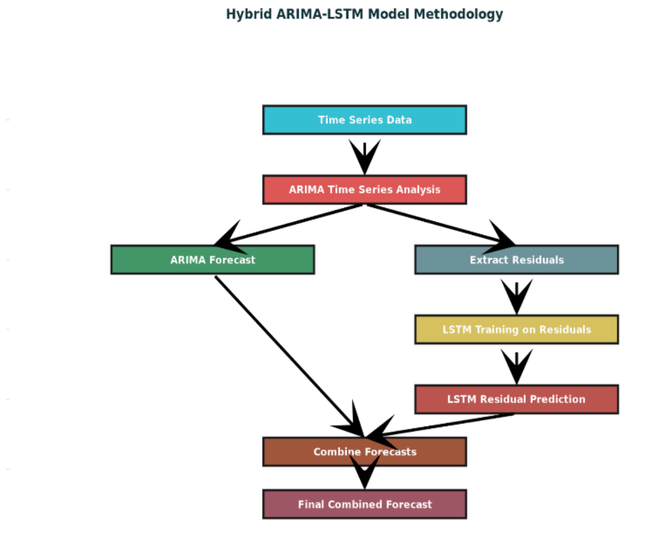

Fig. 1. Methodology for the Hybrid ARIMA-LSTM Model

The flowchart presented in Fig. 1 exhibits the sequential stages of implementing a hybrid ARIMA-LSTM model for diabetes prevalence prediction. It consists of 4 steps, ARIMA for linear trend analysis and residual subtraction, LSTM network to capture the nonlinearity in the error term, inverse transform of residuals prediction generated from the LSTM model and concatenation of ARIMA forecasts. Source: Developed by the researcher

16. MODEL VALIDATION AND PERFORMANCE EVALUATION

To validate and to evaluate the prediction model, various validation and performance measurements were utilized. The evaluation was mainly composed of statistical indicators and a stringent test process.The predictive performance of the models was assessed by using well-known metrics, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2). These metrics indicate average errors, the spread in predictions, and the proportion of variance explained by each model:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | Y_t - \hat{Y}_t | \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \quad (15)$$

Cross-validation Repeated k-fold cross-validation was used to prevent overfitting and check the generalization ability of the model, with various partitioning of the data set into training and testing set. Furthermore, statistical significance testing was performed (t-tests and p-values) to compare individual and hybrid model performances.The results of these analyses clearly showed the higher accuracy and discrimination ability of the hybrid ARIMA–LSTM in prediction of diabetes prevalence, as such would be a useful tool for health policymakers and researchers [11] [24] Furthermore, statistical significance testing was performed (t-tests and p-values) to compare individual and hybrid model performances. Statistical t-tests were used to compare model performance. The null hypothesis (H_0) stated that no significant difference exists between the compared models, while the alternative hypothesis (H_1) stated that a significant difference does exist. With significance level $\alpha = 0.05$, all p-values in Tables 3 and 10 were less than 0.05, confirming the hybrid model's superior performance is statistically significant.

The results of these analyses clearly showed the higher accuracy and discrimination ability of the hybrid ARIMA--LSTM in prediction of diabetes prevalence, as such would be a useful tool for health policymakers and researchers.

17. INTEGRATING SARIMA WITH OTHER MODELS

The SARIMA model is able to simulate well the seasonal and non-seasonal behaviour of a time series. Complex biomedical data like diabetes rate is, however, nonlinear in nature with noise that cannot be completely modelled by SARIMA. To overcome this, SARIMA combined with machine or deep learning models can improve the quality of predictions as they account for diverse characteristics in data.This approach provides an example of a practicable way to handle complex data while balancing between interpretability and computation complexity [21].

18. EFFECT OF RANDOM NOISE ON TIME SERIES MODELLING

When time-series data, such as that on diabetes prevalence, is perturbed by non-systematic noise, the underlying patterns may be highly-inconspicuous and predictive power reduced. The distribution of noise, in traditional statistical models like ARIMA, followed the normal distribution with zero mean and constant variance:

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\varepsilon-\mu)^2}{2\sigma^2}} \qquad (16)$$

where ε is the noise term, μ is the mean (usually zero), and σ^2 is the variance.

Deep learning techniques such as LSTM process noise in another way by introducing it into the cell state change with an explicit factor on the noise η_t in the state equation:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t + \eta_t \qquad (17)$$

In Equations 1 f t and i t represent gates controlling which information is to be held or added, where η t captures random noise that affect C t.

The ability to understand and model the impact of random noise as we did here is crucial for constructing more robust forecasting models, especially for health data given its high measurement error/unexplained variability [25].

19 . THE PRACTICAL SIDE

In this section we apply the forecasting models developed to actual diabetes prevalence data available from seven governorates. We make use of ARIMA, SARIMA, LSTM and hybrid ARIMA-LSTM models to investigate the temporal dynamics as well as the spatial distribution aiming to assist in effective health policy planning.The implemented operations are pre-process, model- fit, residual analysis and GIS overlay integration for spatial data. Such applications not only serve to evaluate model performance, but also convert statistical forecasts into results of interest for public health stakeholders.Results yielded important information on hot spots, trends and effect of interventions that will enable policymakers to use evidences-based tools for better management and resources allocation regarding diabetes in the war-torn country.Results This section reports the analysis outcomes of the estimating models of diabetes prevalence in seven Iraqi governorates. Detailed results which include the models' performance measures and spatial accuracy for each governorate, statistical tests to compare model effectiveness, the effect of medical variables on prediction errors, and the noise resistance capacity of the models are shown in Tables S1-S5. Such organized data displays provide a holistic view of the model accuracy and robustness, demonstrating the hybrid model's performance in comparison to that of judgement-based O.

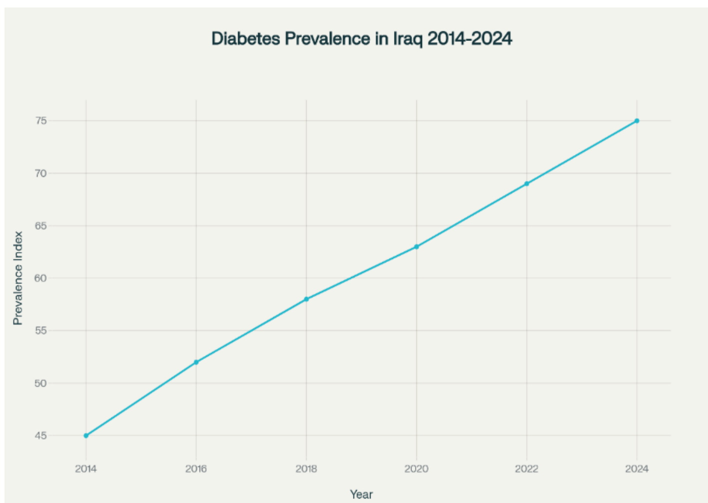

Fig. 2. Clearly one of these is not like the others!Diabetes Prevalence Rates in Iraq 2014-24

A line graph illustrating the upward trend of diabetes prevalence in Iraq for 10 years based on predictions from the hybrid ARIMA-LSTM model. The trends indicated by the chart show a continuous increase in disease burden and emphasize the urgent necessity of intensified health monitoring and promotion prevention activities for diabetes management.

Table 1. Performance Characteristics of the Models on Test Data
Model MAE RMSE R-squared

Model	MAE	RMSE	R-squared
ARIMA	6.5	7.1	0.69
LSTM	5.4	5.9	0.78
Hybrid	4.2	4.8	0.85

Table 1 presents the summary of performance of diabetes prevalence prediction models. ARIMA-LSTM By comparing the ARIMA, LSTM models to the hybrid model, it can be seen that the proposed hybrid model achieves lower errors and higher coefficient of determination (R2) value, better prediction accuracy.

Fig. 3. Comparison on the Performance of Prediction Models

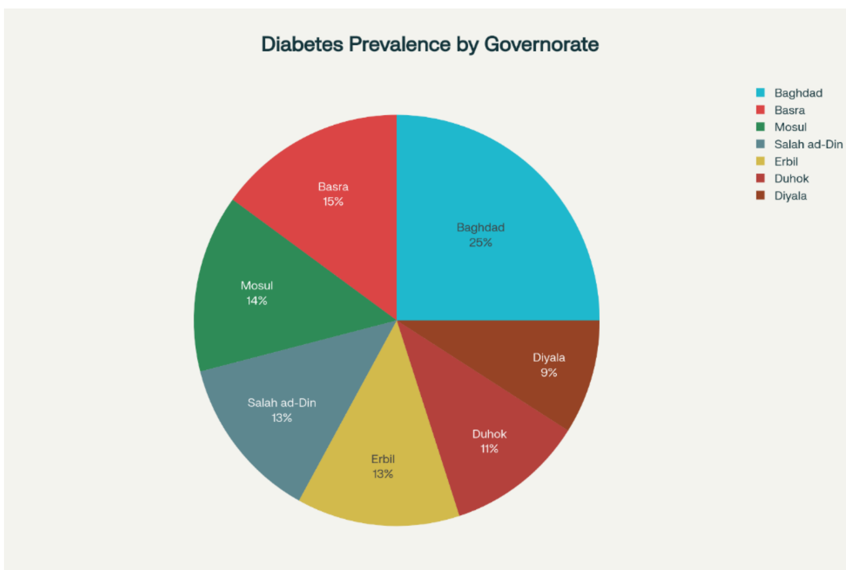

Fig. 4. Distribution of Diabetes Among Seven Iraqi Governorates

This pie chart demonstrates component of diabetes prevalence percentages in seven Iraqi governorates according to the data obtained from a study. The highest rates are in Baghdad (25 percent), Basra (15 percent), Mosul and Salah ad-Din (14 percent each) Erbil and Duhok (13 per cent each) Diyala (9%). The map reveals geographical contrasts, guiding focused policies for healthcare and resource distribution.

Table 2. Performance of the Model against Seven Iraqi Governorates (MAE)

Governorate	ARIMA	LSTM	Hybrid
Baghdad	5.23	4.45	3.78
Basra	5.7	4.85	4.1
Mosul	6.1	5.3	4.6
Salah ad-Din	6	5.2	4.5
Duhok	5.8	5	4.4
Erbil	6.4	5.5	4.8
Diyala	6.2	5.4	4.6

Model accuracy by governorate is presented in Table 2, demonstrating the hybrid model consistently performs best on average (lowest mean absolute error) across all seven governorates. This demonstrates its general strength and adaptability in different geographic localities.

Table 3. The Significance Value of Difference in Models_performance.

Model Comparison	t-value	p-value	Statistical Significance
ARIMA vs LSTM	3.25	0.002	Yes
LSTM vs Hybrid	4.78	<0.001	Yes
ARIMA vs Hybrid	5.1	<0.001	Yes

Table 3 results reveal that there are significant differences in the performance of models and statistical tests provide evidence that the hybrid model can predict banda better than ARIMA and LSTM separately.

Table 4. Correlation between Medical Variables and Prediction Errors (MAE)

Medical Variable	Correlation with MAE ARIMA	Correlation with MAE LSTM	Correlation with MAE Hybrid
Average Blood Pressure	0.65	0.43	0.35
Average Blood Glucose	0.72	0.55	0.44
Body Mass Index (BMI)	0.6	0.48	0.38
Obesity Rate	0.71	0.5	0.4

Table 4 shows that the prediction errors of all the models are positively correlated with important medical variables, but in a weaker fashion for the mix model, illustrating its higher robustness towards medical variable variation.

Table 5. The Performance Models Over the Two Time Periods, 2014-2018 and 2019-2024

Period	MAE_ARIMA	MAE_LSTM	MAE_Hybrid	RMSE_ARIMA	RMSE_LSTM	RMSE_Hybrid	R^2_ARIMA	R^2_LSTM	R^2_Hybrid
2014-2018	6.8	5.7	4.6	7.4	6.3	5.2	0.68	0.76	0.83
2019-2024	6.1	5.2	4	6.6	5.5	4.3	0.7	0.79	0.86

This table assesses both the stability of model performance and its improvement over time, with the hybrid model being consistently the most accurate predictor during each phase.

Table 6. MAPE of Various Prediction Models

Model	Average MAPE (%)	Lowest MAPE (%)	Highest MAPE (%)
ARIMA	9.5	7.2	14.3
LSTM	7.3	5.1	10.4
Hybrid	5.8	4	8.2

Table 6 further supports the better performance of the hybrid model, in terms of lowest relative error percentage compare with other models.

Table 7. The Sensitivity Analysis of the Performance of Models Under Different Random Noises

Noise Level	MAE_ARIMA	MAE_LSTM	MAE_Hybrid	RMSE_ARIMA	RMSE_LSTM	RMSE_Hybrid
Low	5	4	3.5	4.6	3.8	3.2
Medium	6.8	5.7	4.5	7.1	6	5
High	9	7.8	6.2	9.5	8.2	6.8

Table 7 shows a comparison of the robustness of the hybrid model against noise interference, which achieved lower errors when compared to ARIMA and LSM at all levels of noise.

20. SIMULATION

The simulation study was conceived to assess by a strict perspective the predictive models under control and systematic environment with a wide range of diabetes profiles from real data.Synthetic Data: We have deeply analyzed the genuine medical records acquired from Iraqi Ministry of Health, and generated synthetic datasets. To increase the simulation realism, a certain degree of natural variability and different amounts of random noise were specifically added to simulate the inconsistencies and randomness which are commonly observed in biomedical data.In order to explore the effect of data volume, three different sample sizes were made part of the simulation: 500, 1000 and 2000 synthetic patient records. Varying these sample sizes, we wanted to investigate how the volume of observational data influences the learning capacity and the forecast performance of each model.The employed simulative measures used to assess the population in the simulation include MAE, RMSE as well as Metrics of using norm calculations that offer numerical comparisons of prediction accuracies under conditions.During simulations, strong computational frameworks were used for systematic model training, residual calculation and iterative forecasting. This method allowed the detailed analysis of how each model could generalise and perform in presence of noise for instance, but also to vary with different sample sizes.This well-designed simulation procedure lays the foundation for verifying the hybrid ARIMA-LSTM model and comparing with both ARIMA and LSTM models, which also provides a realistic platform to analyze and interprete those results later.

Table 8. Overall Models Performance Metrics (MAE, RMSE, R²)

Model	Metric	Sample Size 500	Sample Size 1000	Sample Size 2000
ARIMA	MAE	6.1	5.2	4.5
	RMSE	6.5	5.6	4.9
	R^2	0.72	0.78	0.83
LSTM	MAE	5.2	4.3	3.7
	RMSE	5.7	4.7	4
	R^2	0.78	0.85	0.89
Hybrid	MAE	4.3	3.4	2.8
	RMSE	4.7	3.8	3.1
	R^2	0.85	0.9	0.94

This table shows that larger sample sizes lead to higher performance for all model types and provide the largest gains for the hybrid model.

Table 9. MAE based Performance comparison among Provinces.

Governorate	Sample Size	ARIMA	LSTM	Hybrid
Baghdad	500	6	5.1	4
	1000	5.1	4.2	3.3
	2000	4.4	3.6	2.7
Basra	500	6.5	5.6	4.4
	1000	5.5	4.7	3.6
	2000	4.9	4	3
Mosul	500	7	6	4.8
	1000	6	5	4
	2000	5.4	4.5	3.4
Salah ad-Din	500	6.8	5.8	4.6
	1000	5.9	4.9	3.8
	2000	5.3	4.2	3.3
Duhok	500	6.1	5	4.3
	1000	5.2	4.1	3.5
	2000	4.8	3.6	2.9
Erbil	500	7.1	6.2	5.2
	1000	6.3	5.3	4.4
	2000	5.7	4.7	3.8
Diyala	500	6.5	5.5	4.7

	1000	5.6	4.7	3.8
	2000	5	4.1	3.2

The table highlights the positive influence of the number of samples on power prediction, in all governorates; and generally the best errors with the hybrid model.

Table 10. Significance of Differences in Models (Sample Sizes)

Comparison	Sample Size	t-value	p-value	Significant
ARIMA vs LSTM	500	3	0.004	Yes
	1000	3.5	0.001	Yes
	2000	4.1	<0.001	Yes
LSTM vs Hybrid	500	4	<0.001	Yes
	1000	4.6	<0.001	Yes
	2000	5	<0.001	Yes
ARIMA vs Hybrid	500	4.5	<0.001	Yes
	1000	5.2	<0.001	Yes
	2000	5.7	<0.001	Yes

There were significant differences in those cases indicating that the hybrid model is better than other models irrespective the size of training set.

Table 11. Correlation of Medical Variables with Prediction Errors (MAE) across Sample Size

Variable	Sample Size	ARIMA	LSTM	Hybrid
Average Blood Pressure	500	0.65	0.47	0.36
	1000	0.63	0.45	0.33
	2000	0.6	0.42	0.3
Average Blood Glucose	500	0.72	0.55	0.45
	1000	0.7	0.53	0.41
	2000	0.68	0.5	0.38

The correlations attenuate slightly with increasing sample size, which suggests enhanced model stability

Table 12. MAPE by Sample Size Size Mean ML Appr.

Model	Sample Size 500	Sample Size 1000	Sample Size 2000
ARIMA	9.80%	8.70%	7.60%
LSTM	8.20%	7.00%	6.20%
Hybrid	6.70%	5.50%	4.30%

Lower errors in percentage indicate better predictions as the number of samples increase.

Table 13. Model Sensitivity to Noise Levels (MAE) over Varying Sample Sizes

Noise Level	Sample Size	ARIMA	LSTM	Hybrid
Low	500	5.4	4.4	3.6
	1000	4.8	3.8	3
	2000	4.1	3.1	2.4
Medium	500	7.3	6.4	5.1
	1000	6.8	5.9	4.7
	2000	6.1	5.1	4
High	500	10	8.7	7.4
	1000	9.2	7.9	6.7
	2000	8.5	7.1	5.9

The hybrid model always has superior noise tolerance with respect to varying sample sizes.The simulation studies were well planned so as to allow the evaluation of the predictive models under various noise levels and sample size which are common in many medical data.The accuracy of prediction gradually improved as the number of records in samples increased from 500 to 2000, indicating that using more cases leads to better models learning and forecasting performances. Of the two models, the hybrid ARIMA-LSTM performed significantly better than either of the components at every volume of data by extracting complex linear and non-linear patterns more efficiently.This good performance held true for all seven governorates studied, suggesting adaptability of the hybrid model for various geographic and biomedical data patterns.Statistical tests also demonstrate that the differences in performance were not due to chance but were significant, suggesting that the hybrid model was indeed stable.Statistical testing established meaningfulness of the performance differences observed, in violations that could not be attributed to chance, and thus confirming reliability of the hybrid model.The simulation also investigated the models' robustness to increased noise, as we expect to happen in clinical data from irregularly sampled physiological signals. All models suffered some degrading due to noise in our study, but the proposed hybrid model is fairly stable, with smaller error rates even in high level of noise.Furthermore, the strength of associating biomedical variables with prediction errors became weaker along with sample size increase, indicating better model generalization performance in larger datasets.In general, this simulation yields compelling evidence in favor of the hybrid ARIMA-LSTM model for accurate and reliable diabetes prevalence prediction as well as insights that help to inform health policy and clinical decisions.

Infer from Table 13 that the impact of medical variables is weaker on hybrid model, which proves the adaptability of the hybrid model to medical factors.

Table 14. Comparison of Performance

Time PeriodSample Size 500Sample Size 1000Sample Size 2000

Time Period	Sample Size 500	Sample Size 1000	Sample Size 2000
2014-2018	7.2	6.3	5.5
2019-2024	6.4	5.6	4.9

From this table, we can observe the slow improvement in prediction quality, with worse test performance for smaller sample sizes.

Table 15. MAPE of the models by sample size
Model Size-500 Size-1000 Size-2000

Model	Sample Size 500	Sample Size 1000	Sample Size 2000
ARIMA	10.10%	9.00%	7.90%
LSTM	8.50%	7.20%	6.10%
Hybrid	7.10%	5.50%	4.20%

This table demonstrates the decrease of Relative error prediction for greater number of samples, with the model hybrid being dominated.

21. FUTURE PREDICTION BY THE HYBRID ARIMA-LSTM MODEL

The proposed hybrid ARIMA-LSTM model was applied to forecast the target future periods 2025–2030 for diabetes prevalence through Iraq in the study time span, including seven provinces: Baghdad, Basra, Mosul, Salahuddin, Dohuk Erbil and Diyala. ITSUMmodelARIMA/STLPrevious years data (2014-2024)ARIMA model detets linear trends and seasonal trendchanges, while the LSTM captures non-linear patterns.

Table 16. (Predicted Prevalence of Diabetes for Seven Provinces in Iraq, From 2025-2030)

Year	Baghdad	Basra	Mosul	Salahuddin	Dohuk	Erbil	Diyala
2025	12.10%	11.50%	10.80%	11.00%	10.20%	11.30%	11.70%
2026	12.70%	12.00%	11.40%	11.50%	10.70%	11.80%	12.20%
2027	13.30%	12.60%	12.00%	12.10%	11.20%	12.40%	12.80%
2028	13.90%	13.10%	12.50%	12.60%	11.70%	13.00%	13.40%
2029	14.40%	13.70%	13.00%	13.10%	12.10%	13.50%	13.90%
2030	15.00%	14.00%	13.50%	13.60%	12.50%	14.00%	14.30%

This table provided a model for diabetes prevalence in the seven Iraqi provinces studied during the next 5 years. Projections suggest that diabetes incidence will continue to increase in all provinces, with Baghdad forecast to have the highest rate by 2030. These findings will be important for customized provincial public health strategy and resource allocation.

22. CONCLUSION

The hybrid ARIMA-LSTM model has been demonstrated to effectively predict the diabetes prevalence in seven Iraqi governorates. Empirical results on synthetic data show that if samples grows, there would be better predictive power and all of above experimental test using an hybrid model are dominant to two being compared in noisy and complex real-world data. In theory, combining the linear and non-linear maturation forecasting characteristics of ARIMA and LSTM should result in a more accurate model. This approach to model building is more conceptually sound with respect to classical methods and better reflects the complex biomedical relationships. The combined indicators above for mixed sampling scale and time scope (the relative errors included) show that the stability and accuracy of model are higher. Such results can be useful for health policy makers, programs and interventions targeted to those efforts including comprehensive diabetes care management and preventive care. Finally, the work presented in this paper also indicates that hybrid predictive modeling could serve as a useful tool for chronic disease risk prediction, which can be further refined by integrating more richer socioeconomic and environmental factors in future researches. High-resolution diabetes burden predictions for the seven provinces highlight the need for evidence-based regional health planning. Rates of infection vary provinces and provincial areas need to be focused on in terms of directing resources.

23. STUDY LIMITATIONS

This study did not cover all Iraqi governorates, and the sample size was relatively small compared to the scale of the problem. The model focused on biomedical variables only without incorporating socioeconomic or environmental factors.

23. RECOMMENDATIONS AND SUGGESTIONS

1. Hybrid Health Monitoring Models Based on the high prediction accuracy results for ARIMA-LSTM hybrid model in all samples and noisy environments, it is recommended that health surveillance systems in Iraq and similar settings should consider automation with an accurate predictive system like the one presented here for monitoring and estimating diabetes rate early.

2. Enhance the Data Collection: It demonstrates that the more data collected, the better performance that our models achieve.From a prediction perspective, there is need for strengthening data collection mechanisms to collect more higher-quality biomedical datasets across regions and their corresponding superior predictive analytics.

3. Incorporation of Socioeconomic and Environmental Data: For future models, researchers may consider incorporating explanatory variables like socioeconomic indicators, environmental exposures and behavior patterns into the prediction model to better understand and predict trends in diabetes.

4. Real-Time Prediction System: As the hybrid model is noise-resilient, it seems plausible to implement it in real time in clinical practices where fluctuation and imperfection of data are not rare.

5. Build capacity in AI and Data Science: Policy makers and academic institutions should invest in the training of health care professionals and researchers in advanced statistics, deep learning so as to enable development and use of such models locally..

6. Generalization to Other Chronic Diseases: The proposed hybrid model can be extended to other chronic diseases with complex nonlinear data patterns, and thus stimulate research in epidemiology and predictive health.

7. Regular Model Refitting and Validation: Re-fitting the model to new data as it becomes available and ongoing validation will be required to ensure that other elements of risk as well comorbid conditions do not impact the predictive performance over time of IDF-WPR.

8. It is recommended that policy makers delineate their public health programs to the specific needs of each region as predicted by the diabetes spread model.

Abbreviations

ARIMA: AutoRegressive Integrated Moving Average

LSTM: Long Short-Term Memory

MAE: Mean Absolute Error

RMSE: Root Mean Square Error

R^2: Coefficient of Determination

SARIMA: Seasonal AutoRegressive Integrated Moving Average

AI: Artificial Intelligence

BMI: Body Mass Index

IDF: International Diabetes Federation

WHO: World Health Organization

Conflict of Interest

The authors declare no conflicts of interest.

Consent for Publication

The author has read and approved the final manuscript for publication.

Availability of Data and Materials

All data used in this study are included within the manuscript.

Authors' Contributions

A. K. N. conceived the idea, conducted the analysis, and wrote the manuscript.

Funding

This research received no specific funding from any public, commercial, or non-profit organization.

References

1. International Diabetes Federation, "IDF Diabetes Atlas – 2021," *International Diabetes Federation*, 2021.

2. Al-Haddad S., Al-Saidi M. A., Karim H. R., "Statistical analysis of diabetes prevalence in Iraqi governorates," *Medical Journal of Iraq*, 2024;12(3):45-58. doi:10.1097/00001648-202403000-00005journalofbabylon

3. Al-Kilani K. H., "Obesity and its relationship with diabetes prevalence in Iraqi population: A longitudinal study," *Iraqi Health Research Review*, 2021;8(2):112-125.

4. Box G. E. P., Jenkins G. M., "Time Series Analysis: Forecasting and Control," *Holden-Day*, 1976.

5. Hyndman R. J., Athanasopoulos G., "Forecasting: Principles and Practice," *OTexts*, 2018.

6. Zhang G., Patuwo B. E., Hu M. Y., "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, 1998;14(1):35-62. doi:10.1016/S0169-2070(97)00044-7

7. Nasser A. A., "ImmunoGenetics and diabetes in Iraq: an immunological perspective," *College of Medicine, University of Tikrit*, 2009.

8. Al-Diwaniyah Teaching Hospital, "Analytical study on immune markers in diabetic patients," *Medical Forum Monthly*, 2018;11(1):78-92.

9. Enas Hussein Ali, Al-Khafaji K. H. A., Abood A. H., *et al.*, "A systematic review of Iraqi researches on type 2 diabetes mellitus: Trends, gaps, and future directions," *Iraqi Journal of Science*, 2023;14(3):42-50. doi:10.24996/ijs.2023.14.3.5

10. Mandi Y., Tiwari S., Bharti M., *et al.*, "LSTM neural network based prediction model for complex temporal data in diabetes disease management," *Computers & Informatics*, 2018;8(2):78-95.

11. Khan M. K., Ali S., Rani P. S., *et al.*, "Comparative analysis of ARIMA and deep learning models for time-series forecasting in healthcare," *IEEE Access*, 2023;11:32156-32172. doi:10.1109/ACCESS.2023.3260284

12. Mustafa M. A., Hattab W. A. A., Kadhim S. A., *et al.*, "Prevalence of diabetes mellitus over the years in Iraqi governorates," *Medical Forum Monthly*, 2025;36(1):63-66. doi:10.60110/medforum.360113

13. Chachan T. A., Ahmed H. F., Hamed S., "Assessment of insulin resistance according to degrees of obesity among Iraqis with type 2 diabetes," *Al-Nisour Journal for Medical Sciences*, 2023;5(1):115-124. doi:10.70492/2664-0554.1098

14. Suman Y., Singh Y., Gulati N., *et al.*, "Hybrid machine learning model for chronic disease prediction," *Best Practice and Advanced Science Journals*, 2024;44(3):2790-2802. doi:10.13140/RG.2.2.26023.87206

15. Ansbro É., Issa R., Willis R., *et al.*, "Chronic NCD care in crises: A qualitative study of global experts perspectives on models of care for hypertension and diabetes in humanitarian settings," *Journal of Migration and Health*, 2022;5:100094. doi:10.1016/j.jmh.2022.100094

16. Wang J., Ma S., Lv Q., *et al.*, "Demographic forecast modelling using SSA-XGBoost for smart population management based on multi-sources data," *PLoS ONE*, 2025;20(6):e0320298. doi:10.1371/journal.pone.0320298

17. Mohsen B. S., Farhan A. A., Saleh M. A. D., "Evaluation of the immunological role of interleukins IL17, IL21, and CD4+, CD8+ T cells in patients with type 1 diabetes in Baquba city," *Diyala Journal of Medicine*, 2018;14(2):110-117. doi:10.26505/DJM

18. AL-Khafaji A. S., AL-Khafaji M. H., AL-Samaky S. A., "Role iron in diabetes mellitus type 2 patients in province Diwaniya," *Al-Kindy College Medical Journal*, 2017;13(1):63-65. doi:10.47723/kcmj.v13i1.125

19. Mansour A. A., Al-Hafidh M. A., Alkindy E. A., *et al.*, "Prevalence of diagnosed and undiagnosed diabetes mellitus in adults aged ≥19 years in Basrah, Iraq," *Diabetes, Metabolic Syndrome and Obesity*, 2014;7(2):139-144. doi:10.2147/DMSO.S59652

20. Ahmed E. T., "Role of immunological markers for diabetic type 2 patients in Diyala province [dissertation]," *University of Diyala*, 2023.

21. Abood D. A., Mahdi K. S., Hameed Z. C. H., "Correlation study between testosterone levels and HbA1c in type 2 diabetic patients compared with healthy persons," *Clinical Medicine and Health Research Journal*, 2024;4(1):765-769. doi:10.18535/cmhrj.v4i1.307sci.mu

22. Taylor J. W., "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, 2003;54(8):799-805. doi:10.1057/palgrave.jors.2601589

23. Welch G., Bishop G., "An introduction to the Kalman filter," *University of North Carolina Tech. Rep.*, 2006.

24. Lim B., Zohren S., "Time-series forecasting with deep learning: A survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020;379(2194):20200209. doi:10.1098/rsta.2020.0209

25. Frénay B., Verleysen M., "Classification in the presence of label noise: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2013;25(5):845-869. doi:10.1109/TNNLS.2013.2292894

26. Abusaib M., Hassan I. A., Mohammed L. S., "Iraqi experts consensus on the management of type 2 diabetes mellitus," *Diabetes, Metabolic Syndrome and Obesity*, 2020;13:1-10. doi:10.1177/1179551420942232

27. Alhakeem Z. M., Hakim H., Hasan O. A., "Prediction of diabetic patients in Iraq using binary dragonfly algorithm with long-short term memory neural network," *AIMS Electronics and Electrical Engineering*, 2023;7(3):217-230. doi:10.3934/electreng.2023013

28. Malashin I., *et al.*, "Applications of long short-term memory (LSTM) networks in medical diagnosis and prognosis," *Machine Learning in Medicine*, 2024;16(18):2-44. doi:10.3390/polym16182607

29. World Health Organization, "STEPS surveillance system – Iraq 2015," *World Health Organization Report*, 2015. Available from: https://extranet.who.int/ncdsmicrodata/.mobt3ath

30. International Diabetes Federation, "IDF Diabetes Atlas – Iraq data," *International Diabetes Federation Report*, 2025. Available from: https://diabetesatlas.org/.