IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (1): 814-825

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521
University of Kirkuk Journal For Administrative and Economic Science

Hussein Hemin Mohammad & Hasan Samyia Khalid. A Comparative Time Series Analysis Using ARIMA Models for Forecasting Types Leukemia Disease at Nanakali and Hiwa Hospitals in Kurdistan region of Iraq. ***University of Kirkuk Journal For Administrative and Economic Science*** **(2026) 16 (1):814-825.**

A Comparative Time Series Analysis Using ARIMA Models for Forecasting Types Leukemia Disease at Nanakali and Hiwa Hospitals in Kurdistan region of Iraq

1Hemin Mohammad Hussein [1], Samyia Khalid Hasan [2]

[1] *PhD candidate, Department of Statistics and Informatics-College of Administration and Economics/Salahaddin University-Erbil, Erbil, Iraq*

[2] *Department of Statistics and Informatics-College of Administration and Economics/Salahaddin University-Erbil, Erbil, Iraq*

hemin.hussein@su.edu.krd [1], saiya.hasan@su.edu.krd [2]

Abstract: Leukemia is a group of malignant blood disorders characterized by the uncontrolled proliferation of abnormal white blood cells, leading to impaired bone marrow function and reduced immunity. The main objective of this study is to apply ARIMA time series models to the four major types of leukemia Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML) using monthly patient data collected from Nanakali and Hiwa Hospitals. The study aims to evaluate, compare, and select the most suitable ARIMA models for forecasting leukemia trends in both hospitals. A total of 17 years of monthly observations were analyzed, representing one of the most comprehensive leukemia datasets available in the Kurdistan Region. Model performance was assessed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to identify the most suitable ARIMA structure for each leukemia type. The study applied ARIMA time series models to forecast four major leukemia types ALL, AML, CLL, and CMLusing monthly data from Nanakali and Hiwa Hospitals (2007–2023). Results revealed statistically significant differences in mean observed and forecasted values between the hospitals for most leukemia types. The best-fitting ARIMA models varied by type and hospital, providing accurate predictions for hospital planning and resource allocation.

Keywords: Leukemia, Time series, Akaike Information Criterion, Bayesian Information Criterion

تحليل مقارن للسلاسل الزمنية باستخدام نماذج ARIMA للتنبؤ بأنواع مرض اللوكيميا في مستشفيي نانكلي وهيوا في إقليم كردستان العراق

م.م. هيمن محمد حسين [١]، أ.م. سامية خالد حسن [٢]

IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (1): 814-825

[1] مرشح لنيل شهادة الدكتوراه، قسم الإحصاء والمعلوماتية-كلية الإدارة والاقتصاد/جامعة صلاح الدين-أربيل، أربيل، العراق
[2] قسم الإحصاء والمعلوماتية-كلية الإدارة والاقتصاد/جامعة صلاح الدين-أربيل، العراق

hemin.hussein@su.edu.krd [1], saiya.hasan@su.edu.krd [2]

المستخلص: اللوكيميا هي مجموعة من اضطرابات الدم الخبيثة التي تتميز بتكاثر غير منضبط للخلايا البيضاء غير الطبيعية، مما يؤدي إلى ضعف وظيفة نخاع العظم وانخفاض المناعة. الهدف الرئيسي من هذه الدراسة هو تطبيق نماذج السلاسل الزمنية ARIMA على الأنواع الأربعة الرئيسية من اللوكيميا، وهي اللوكيميا الليمفاوية الحادة (ALL) واللوكيميا النقوية الحادة (AML) واللوكيميا الليمفاوية المزمنة (CLL) واللوكيميا النقوية المزمنة (CML) باستخدام بيانات المرضى الشهرية التي تم جمعها من مستشفيي نانكلي وهيوا. تهدف الدراسة إلى تقييم ومقارنة واختيار نماذج ARIMA الأكثر ملاءمة للتنبؤ باتجاهات سرطان الدم في كلا المستشفيين. تم تحليل ما مجموعه ١٧ عامًا من الملاحظات الشهرية، والتي تمثل واحدة من أكثر مجموعات بيانات سرطان الدم شمولاً المتوفرة في إقليم كردستان. تم تقييم أداء النموذج باستخدام معيار معلومات Akaike (AIC) ومعيار معلومات Bayesian (BIC) لتحديد هيكل ARIMA الأكثر ملاءمة لكل نوع من أنواع سرطان الدم. طبقت الدراسة نماذج السلاسل الزمنية ARIMA للتنبؤ بأربعة أنواع رئيسية من سرطان الدم، وهي ALL وAML وCLL وCML، باستخدام البيانات الشهرية من مستشفيي نانكلي وهيوا (٢٠٠٧-٢٠٢٣). كشفت النتائج عن اختلافات ذات دلالة إحصائية في متوسط القيم الملاحظة والمتوقعة بين المستشفيين لمعظم أنواع سرطان الدم. اختلفت نماذج ARIMA الأكثر ملاءمة حسب النوع والمستشفى، مما وفر تنبؤات دقيقة لتخطيط المستشفيات وتخصيص الموارد.

الكلمات المفتاحية: اللوكيميا، السلاسل الزمنية، معيار معلومات أكايكي، معيار المعلومات البيزين.

Corresponding Author: E-mail: hemin.hussein@su.edu.krd

Introduction

Leukemia refers to a group of cancers affecting blood-forming tissues, primarily the bone marrow and lymphatic system. It disrupts the normal production of white blood cells, leading to immunodeficiency, anemia, and bleeding disorders (Bispo et al., 2019; Whiteley et al., 2021). The disease can progress rapidly or slowly depending on its type, age group, and biological characteristics. Leukemia remains a major global health burden, particularly in developing countries where early detection and treatment resources are limited. Understanding its epidemiological trends is essential for improving clinical planning and public health strategies.

The Four main types of Leukemia, which was classified into four major types based on cell lineage and progression. Acute Lymphoblastic Leukemia (ALL) is characterized by the rapid accumulation of immature lymphoid cells and is the most common cancer in children (Pui et al., 2021). Acute Myeloid Leukemia (AML) involves the uncontrolled growth of myeloid precursors, leading to bone marrow failure and rapid progression if untreated. Chronic Lymphocytic Leukemia (CLL) affects mature lymphocytes and progresses slowly, primarily impacting older adults. In contrast, Chronic Myeloid Leukemia (CML) is associated with the Philadelphia chromosome abnormality and typically progresses in three phases, from chronic to blast crisis (Jabbour & Kantarjian, 2020). These four types vary significantly in incidence, clinical features, survival rates, and treatment responses, making comparative epidemiological studies crucial for improving disease management and forecasting future healthcare needs.

Time series refers to a sequence of observations collected at regular intervals over time (Omer, et al, 2023). It is widely used to analyze temporal patterns, detect trends, identify seasonal fluctuations, and understand past behavior for improved decision-making (Ahmed, et al, 2023). In healthcare, time series analysis helps track disease incidence, monitor outbreaks, and evaluate long-term public health interventions. By examining how values evolve chronologically, researchers can identify systematic movements that may not be visible through cross-sectional data. Understanding these patterns is essential for effective forecasting, planning, and policy development, particularly for chronic diseases such as leukemia that show long-term variability in hospital admissions and diagnosis rates. (Chatfield, 2019)

Time series forecasting involves predicting future values based on historical patterns and mathematical modeling. It is a fundamental tool in healthcare analytics, enabling hospitals to anticipate patient loads, estimate resource needs, and plan treatment capacity (Hyndman &

Athanasopoulos, 2018). Forecasting methods capture key components such as trend, seasonality, and random fluctuations to project likely future outcomes. In disease surveillance, accurate forecasting supports early warning systems, reduces uncertainty, and enhances the preparedness of medical institutions. Forecasting leukemia cases is particularly important for oncology centers, where treatment resources such as chemotherapy units, laboratory diagnostic capacity, and blood product supplies must be carefully planned. Time series forecasting provides a scientific foundation for improving operational efficiency and health service delivery

The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used time series forecasting techniques due to its flexibility and strong predictive performance. ARIMA combines three key components autoregression (AR), differencing (I), and moving average (MA) to capture temporal dependencies and stabilize non-stationary data (Box et al., 2016). It is particularly effective for forecasting medical and epidemiological data where trends and irregular patterns frequently occur. ARIMA models can be optimized using diagnostic criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which help identify the best-fitting structure. In healthcare forecasting, ARIMA provides reliable short-term predictions and supports strategic planning, making it suitable for modeling leukemia incidence trends.

1st: Objective of the Research

The main objective of this study is to apply ARIMA time series models to the four major types of leukemia Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML) using monthly patient data collected from Nanakali and Hiwa Hospitals. The study aims to compare the epidemiological patterns and monthly incidence trends of these leukemia types between the two hospitals from 2007 to 2023, and to apply ARIMA models for forecasting in order to examine differences in model behavior and predictive performance across leukemia categories. In addition, the research seeks to identify the most suitable ARIMA model for each leukemia type through AIC and BIC evaluation, producing accurate and data-driven forecasts that can enhance hospital planning, early preparedness, and oncology resource allocation

2nd: Materials and Methods

This study employed a quantitative time series approach using monthly leukemia case data from Nanakali and Hiwa Hospitals for the period 2007–2023. The dataset included four leukemia types: ALL, AML, CLL, and CML. Preprocessing steps included data cleaning, handling missing values, and converting raw data into monthly time series format. Stationarity was assessed using the Augmented Dickey–Fuller (ADF) test. ARIMA models were applied to each leukemia type for both hospitals. Candidate models were evaluated through ACF/PACF diagnostics, residual analysis, and comparison of AIC and BIC values. The best-fitting ARIMA model for each leukemia type was selected and used to generate forecasts. All analyses were conducted using R software with the forecast and series packages

3rd: Forecasting Model

Leukemia is a malignant disorder of blood-forming tissues characterized by uncontrolled proliferation of abnormal white blood cells, impairing immunity, bone marrow function, and overall hematologic stability. A time series is a sequence of observations recorded at regular intervals, used to analyze temporal patterns, trends, and fluctuations, thereby supporting statistical modeling of dynamic processes over time. Forecasting leukemia types involves examining historical incidence data to predict future trends in ALL, AML, CLL, and CML, enabling improved clinical preparedness and optimized oncology resource planning. The AutoRegressive Integrated Moving Average (ARIMA) model is a statistical forecasting technique that integrates autoregressive components, differencing, and moving averages to effectively capture temporal structures and generate reliable short-term predictions. This study introduces the concepts of the ARIMA model, providing more detailed information as follows.

- **Auto Regressive Integrated Moving Average ARIMA Model**

he ARIMA (Auto Regressive Integrated Moving Average) model is one of the most widely used statistical methods for time series forecasting. It combines three fundamental components: The Auto Regressive (AR) process, the Moving Average (MA) process, and differencing to handle non-stationarity in the data (Box & Jenkins, 1976, Muangkhoua,2019). These components have been extensively utilized in diverse applications, serving as foundational tools in univariate time series modeling (Fattah, et al ,2018)

- **Auto Regressive (AR) Process:**

The AR process models the current value of the series as a linear combination of its previous values and a stochastic error term. Mathematically (Kadri, et al, 2023), an AR process of order p is represented as:

$$X_t = c + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \emptyset_3 X_{t-3} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t \quad t = 1,2, \ldots T \qquad (1)$$

where X_t is the current value, c is a constant term, $\emptyset_i$ are the autoregressive coefficients, and $\varepsilon_t$is a white noise error term. This model captures the linear dependence between an observation and its lagged values. (Guha & Bandyopadhyay, 2016).

- **Moving Average (MA) process:**

The MA process models the current value of the series as a linear combination of past error terms. An MA process of order q is expressed as

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \cdots - \theta_q \varepsilon_{t-q} \qquad (2)$$

where θ_i are the moving average coefficients, representing the influence of previous shocks on the current observation (Zhang & Meng,2023)

- **Auto Regressive Moving Average (ARMA) process:**

By combining AR and MA processes, the ARMA model captures both lagged observations and lagged errors:

$$X_t = c + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_3 \varepsilon_{t-3} - \cdots - \theta_q \varepsilon_{t-q} \qquad (3)$$

ARMA models assume the time series is stationary. However, many real-world datasets, including fertility rates or airport passenger flows, exhibit trends or non-stationarity (Gorgess,2017)

- **Integrated component (I): ARIMA (p, d, q) model**

The ARIMA model extends ARMA to non-stationary series by introducing differencing. For a series X_t that is non-stationary, first-order differencing can be applied: (T Kahwachi & Khalid, 2023)

$$\Delta X_t = X_t - X_{t-1} \qquad (4)$$

The general ARIMA (p, d, q) model is then expressed as

$$\Delta^d X_t = c + \emptyset_1 \Delta^d X_{t-1} + \cdots + \emptyset_p \Delta^d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \qquad (5)$$

where d denotes the order of differencing

If $p = q = 0$ and $d = 1$, the model reduces to a **random walk** . ARIMA (p, d, q) which implies the next value is equal to the previous value plus a random error, reflecting a simple non-stationary process (Jenkins & P, 1976). The ARIMA model is highly flexible, allowing it to model a wide range of linear time series behaviors (Hussein,2024). AR captures the memory effect of past observations, MA captures the influence of past shocks or errors, and the integration component ensures stationarity (Kharista, et al, 2015).

4th: Testing for stationary using the Dickey-Fuller Test

The Dickey-Fuller test assesses stationary in time series data, identifying whether a unit root is present (Dickey & Fuller, 1979; Hamilton, 1994; Enders, 2014). Stationary is essential for reliable forecasting with models like ARIMA. The test's formula is:

$$\Delta Y_t = \propto + \beta Y_{t-1} + \delta t + \epsilon_t \qquad (6)$$

Where : $\Delta Y_t = Y_t - Y_{t-1}$ and $\propto$: The intercept or constant term, representing a fixed value that affects allobservations in the time series. Including α\alphaα allows the test to account for a non-zero mean in the series. β:: The coefficient of Y_{t-1}, representing the effect of the previous time period's value on the current change. δt represents a deterministic trend component ϵ_t is white noise.
The following presumptions guide the test's execution:
H0 :The unit root is present and is non-stationary. vs H1: The unit root is not existing and is stationary.

5th: Measuring Forecast Accuracy

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are widely used to evaluate and compare the performance of forecasting models. Both criteria consider the goodness of fit of the model while penalizing model complexity to avoid overfitting (Xin, 2022; Zhang & Meng,2023). The smaller the AIC or BIC value, the better the model balances fit and simplicity.

$$.AIC = n \times ln(\frac{SSE}{n}) + 2k \qquad (7)$$

$$BIC = n \times ln(\frac{SSE}{n}) + k \times \ln(n) \qquad (8)$$

Where: n is number of observations (time periods), k is number of estimated parameters in the model, $SSE = \sum_{t=1}^{n}(x_i - \hat{x}_i)^2 =$ is sum of squared error. x_i is actual value at time t and $\hat{x}_i$ is forecasted value at time t

6th: Results

In this study, monthly time series data for the four major leukemia types ALL, AML, CLL, and CML were collected from Nanakali and Hiwa Hospitals for the period 2007 to 2023 to analyze temporal patterns and generate forecasts. The ARIMA modeling framework discussed earlier was applied to each leukemia type in both hospitals to evaluate incidence trends and identify the most suitable forecasting structure. All analyses were performed using RStudio, where model parameters were estimated and diagnostic checks were conducted. Figures representing the observed data and model behavior are presented in the results section. The original monthly incidence data for the four major leukemia types Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML) collected from 2007 to 2023 were analyzed to identify temporal patterns. The data can be seen in Figure 1, which shows the distribution of ALL, AML, CLL, and CML cases at Nanakali Hospital and Figure 2 presents the corresponding data for Hiwa Hospital, showing the distribution of ALL, AML, CLL, and CML cases. To determine the best forecasting models for leukemia incidence from 2024 to 2030, candidate ARIMA models were compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the lowest AIC and BIC was selected as the optimal forecasting model for each leukemia type. Forecast results were further examined for consistency and reliability, enabling a comparative assessment of leukemia trends between the two hospitals. Overall, the forecasting outputs provide valuable insights for oncology planning and early preparedness in the Kurdistan Region.

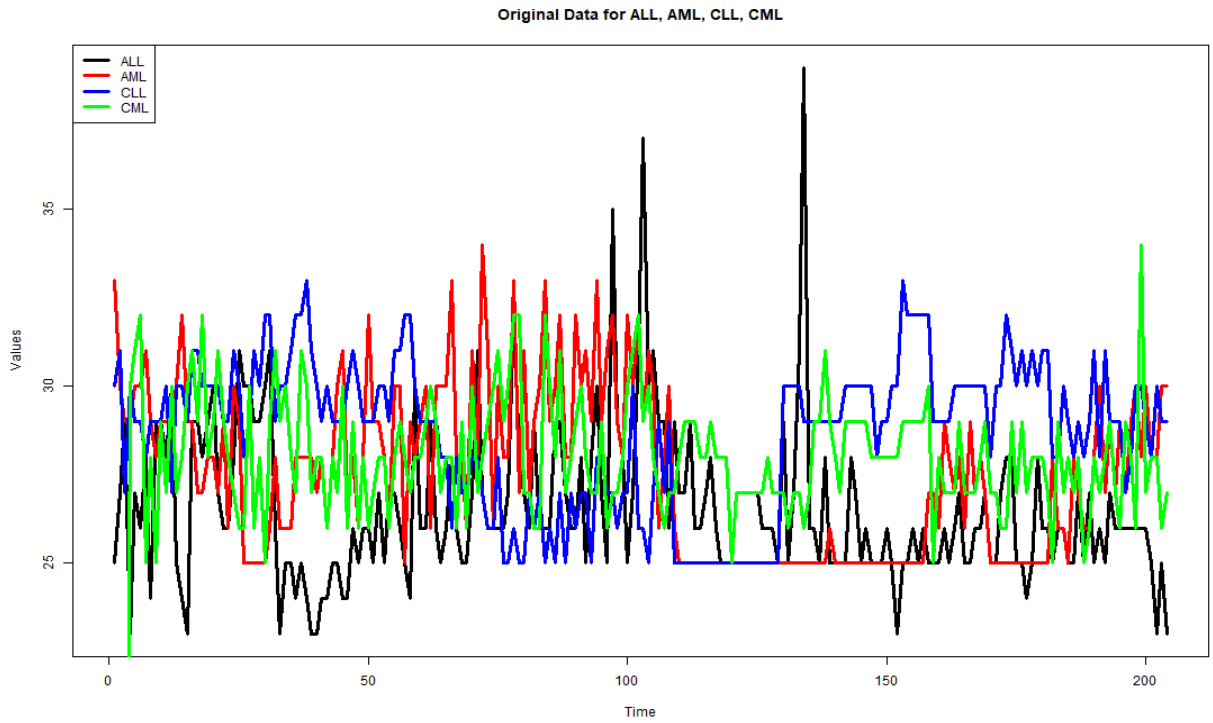

Fig. (1): Types of leukemia in Nanakali Hospital

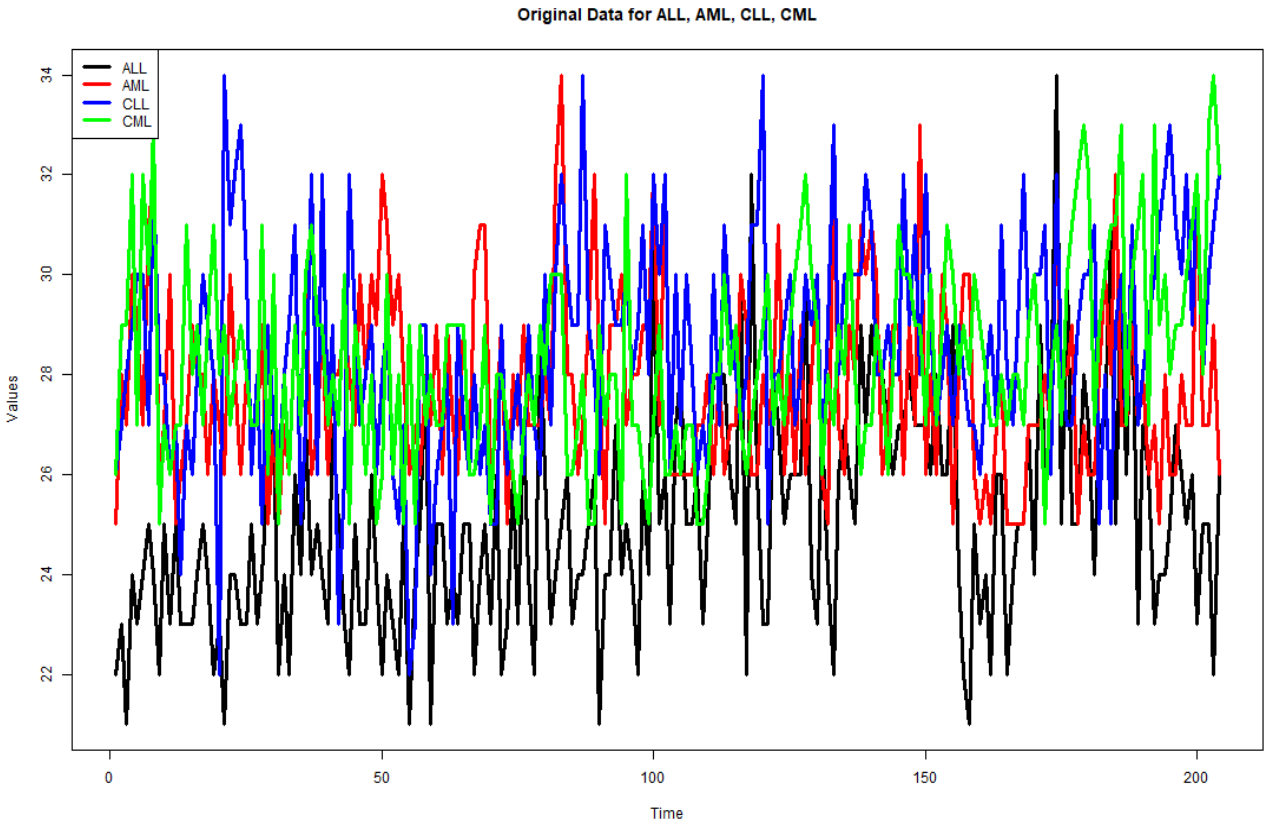

Fig. (2): Types of leukemia in Hiwa Hospital

Testing for stationary using the Dickey-Fuller Test:

The following presumptions guide the test's execution
H0: The unit root is present and is non-stationary's H1: The unit root is not existing and is stationary.

Table (1): Augmented Dickey–Fuller (ADF) Test for Original Series

Hospital	Series	ADF Original	P Original	Diff Used	ADF Stationary	P Stationary
Nanakali	ALL	-3.65716	0.02926	0	-3.65716	0.02926
	AML	-1.49705	0.7868	1	-8.86657	0.01
	CLL	-2.12074	0.52549	1	-8.01183	0.01
	CML	-5.43087	0.01	0	-5.43087	0.01
Hiwa	ALL	-3.8241	0.01898	0	-3.8241	0.01898
	AML	-5.35375	0.01	0	-5.35375	0.01
	CLL	-4.55925	0.01	0	-4.55925	0.01
	CML	-4.01732	0.01	0	-4.01732	0.01

The Augmented Dickey–Fuller (ADF) test was conducted to assess the stationarity of the leukemia time series at Nanakali and Hiwa Hospitals (Table 1). For Nanakali Hospital, the ALL and CML series were stationary in their original form, with ADF statistics of -3.65716 and -5.43087 and corresponding p-values of 0.02926 and 0.01, indicating that no differencing was required (d = 0). In contrast, the AML and CLL series were non-stationary initially, with p-values of 0.7868 and 0.52549 exceeding 0.05. After first differencing, both series became stationary (ADF = -8.86657, p = 0.01 for AML; ADF = -8.01183, p = 0.01 for CLL), suggesting a differencing order of d = 1. At Hiwa Hospital, all four leukemia series (ALL, AML, CLL, and CML) were stationary in their original form, with ADF statistics ranging from -3.8241 to -5.35375 and p-values below 0.05, indicating that differencing was unnecessary (d = 0)

Table (2): compare the real value types of the Leukemia between Nanakali and Hiwa Hospitals

Leukemia	Hospital	N	Mean	S.D	t-test	p-value
ALL	Nanakali	204	26.5686	2.25912	6.845	0.000
	Hiwa	204	25.0343	2.26861		
AML	Nanakali	204	27.5196	2.33724	-1.284	0.200
	Hiwa	204	27.7892	1.88059		
CLL	Nanakali	204	28.6176	2.09171	0.430	0.668
	Hiwa	204	28.5245	2.28389		
CML	Nanakali	204	27.6569	3.74428	-2.132	0.034
	Hiwa	204	28.2892	1.98012		

The comparative analysis of leukemia types between Nanakali and Hiwa Hospitals, as shown in Table 2, reveals notable differences in the mean values for each type. For Acute Lymphoblastic Leukemia (ALL), the mean at Nanakali Hospital (26.57) is higher than at Hiwa Hospital (25.03), and this difference is statistically significant (t = 6.845, p < 0.05), indicating a meaningful variation between the two hospitals for this leukemia type. In contrast, for Acute Myeloid Leukemia (AML), Nanakali's mean (27.52) is slightly lower than Hiwa's (27.79), but this difference is not statistically significant (t = -1.284, p = 0.200), suggesting no substantial difference between the hospitals. Similarly, for Chronic Lymphocytic Leukemia (CLL), the means are very close (Nanakali: 28.62, Hiwa: 28.52) with a non-significant t-test result (t = 0.430, p = 0.668), indicating similar levels between the two settings. However, for Chronic Myeloid Leukemia (CML), Nanakali's mean (27.66) is slightly lower than Hiwa's (28.29), and this difference reaches statistical significance (t = -2.132, p = 0.034), suggesting a modest but significant difference between the hospitals in CML cases.

Estimated Best Model forecasted

As in the previous process, the data on types of leukemia were analyzed using ARIMA models. Prior to modeling, it was ensured that the data met all necessary assumptions and were suitable for time series forecasting. Table 3 presents the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for several ARIMA models tested to forecast each type of leukemia in both Nanakali and Hiwa Hospitals. The main objective of this analysis is to identify the ARIMA model that provides the most accurate and reliable forecasts for each leukemia type, as indicated by the lowest AIC and BIC values.

Table (3): Best Models to forecasting types of Leukemia between Nanakali and Hiwa

Hospitals	Nanakali			Hiwa		
Leukemia	Best Model	Best AIC	Best BIC	Best Model	Best AIC	Best BIC
ALL	ARIMA(1,1,1)	877.972	891.225	ARIMA(2,0,2)	866.302	889.529
AML	ARIMA(1,1,1)	779.64	792.893	ARIMA(1,1,1)	843.187	856.44
CLL	ARIMA(1,1,1)	661.588	674.84	ARIMA(1,0,1)	901.895	918.486
CML	ARIMA(2,0,2)	997.609	1,020.84	ARIMA(2,1,2)	818.462	838.341

The results in Table 3 indicate the optimal ARIMA models for forecasting different types of leukemia at Nanakali and Hiwa Hospitals based on the lowest AIC and BIC values. For Nanakali Hospital, ALL, AML, and CLL are best modeled using ARIMA (1,1,1), suggesting that these series exhibit similar patterns of trend and autocorrelation that are effectively captured by a first-order autoregressive and moving average component with one level of differencing. CML at Nanakali, however, is best represented by ARIMA (2,0,2), indicating a more complex structure with higher-order autoregressive and moving average components without differencing. At Hiwa Hospital, the optimal models differ across leukemia types: ALL is best forecasted by ARIMA (2,0,2), AML by ARIMA (1,1,1), CLL by ARIMA (1,0,1), and CML by ARIMA(2,1,2).

Estimated Model forecasting

The forecasting results presented in Table 4 illustrate the predicted values of different leukemia types at Nanakali and Hiwa Hospitals based on the best-selected ARIMA models. For Nanakali, ALL, AML, and CLL were modeled using ARIMA (1,1,1), while CML used ARIMA (2,0,2). At Hiwa Hospital, ALL followed ARIMA (2,0,2), AML ARIMA (1,1,1), CLL ARIMA (1,0,1), and CML ARIMA (2,1,2). The forecasted values across 50 steps show gradual and consistent trends for each leukemia type, indicating that the models capture the underlying temporal patterns effectively. These predictions provide valuable insights for hospital planning, resource allocation, and proactive clinical management of leukemia cases. The results are visually represented in Fig. 3, showing the forecasting values of leukemia types in Nanakali, and Fig. 4, showing the forecasting values in Hiwa Hospital.

Table (4): forecasting value of types of Leukemia between Nanakali and Hiwa depends on the Best Models

Hospitals	Nanakali				Hiwa			
Step	ALL	AML	CLL	CML	ALL	AML	CLL	CML
1	24.5277	29.1726	29.0505	28.038	25.2394	27.4149	30.1216	30.6269
2	24.9682	29.233	29.0585	28.4932	26.0629	27.5385	30.06	30.7984
3	25.091	29.2208	29.0581	28.6776	26.9405	27.5474	30.009	30.3804
4	25.1211	29.2146	29.056	28.5586	27.4574	27.5461	29.967	30.6868
5	25.124	29.208	29.0535	28.3361	27.4718	27.5439	29.9327	30.417
6	25.119	29.2013	29.051	28.214	27.1114	27.5417	29.905	30.6668
7	25.1116	29.1947	29.0484	28.2497	26.6458	27.5394	29.8829	30.4665

8	25.1036	29.188	29.0459	28.3654	26.3254	27.5371	29.8656	30.661
9	25.0954	29.1814	29.0433	28.4527	26.2696	27.5348	29.8524	30.5106
10	25.0872	29.1747	29.0408	28.4628	26.4444	27.5325	29.8426	30.6619
11	25.0789	29.1681	29.0382	28.4206	26.7173	27.5302	29.8359	30.5494
12	25.0706	29.1614	29.0357	28.3802	26.9437	27.528	29.8316	30.6676
13	25.0624	29.1547	29.0331	28.375	27.0372	27.5257	29.8296	30.584
14	25.0541	29.1481	29.0306	28.4008	26.9951	27.5234	29.8294	30.6769
15	25.0458	29.1414	29.0281	28.4322	26.8785	27.5211	29.8308	30.6155
16	25.0375	29.1348	29.0255	28.4491	26.7676	27.5188	29.8335	30.689
17	25.0292	29.1281	29.023	28.4488	26.7198	27.5165	29.8374	30.6445
18	25.021	29.1215	29.0204	28.4424	26.7483	27.5143	29.8423	30.7032
19	25.0127	29.1148	29.0179	28.4414	26.8278	27.512	29.848	30.6717
20	25.0044	29.1082	29.0153	28.4491	26.9156	27.5097	29.8544	30.7191
21	24.9961	29.1015	29.0128	28.4614	26.9758	27.5074	29.8614	30.6974
22	24.9878	29.0949	29.0102	28.4723	26.9946	27.5051	29.869	30.7361
23	24.9796	29.0882	29.0077	28.479	26.9808	27.5028	29.8769	30.7221
24	24.9713	29.0816	29.0051	28.4827	26.9564	27.5006	29.8853	30.7542
25	24.963	29.0749	29.0026	28.4864	26.9424	27.4983	29.894	30.7459
26	24.9547	29.0683	29	28.492	26.9499	27.496	29.903	30.773
27	24.9464	29.0616	28.9975	28.4994	26.9771	27.4937	29.9122	30.7691
28	24.9382	29.055	28.9949	28.5071	27.0135	27.4914	29.9216	30.7923
29	24.9299	29.0483	28.9924	28.514	27.047	27.4892	29.9312	30.7918
30	24.9216	29.0417	28.9898	28.52	27.0699	27.4869	29.9409	30.812
31	24.9133	29.035	28.9873	28.5256	27.0814	27.4846	29.9508	30.8142
32	24.905	29.0284	28.9848	28.5315	27.0861	27.4823	29.9607	30.8321
33	24.8968	29.0217	28.9822	28.5379	27.0908	27.48	29.9708	30.8362
34	24.8885	29.0151	28.9797	28.5446	27.1004	27.4777	29.9809	30.8524
35	24.8802	29.0084	28.9771	28.5512	27.1163	27.4755	29.9911	30.8581
36	24.8719	29.0018	28.9746	28.5575	27.1367	27.4732	30.0014	30.8729
37	24.8637	28.9951	28.972	28.5637	27.1581	27.4709	30.0117	30.8797
38	24.8554	28.9885	28.9695	28.5699	27.1775	27.4686	30.022	30.8936
39	24.8471	28.9818	28.9669	28.5761	27.1934	27.4663	30.0324	30.9013
40	24.8388	28.9752	28.9644	28.5825	27.2066	27.464	30.0428	30.9144
41	24.8305	28.9685	28.9618	28.5889	27.2186	27.4618	30.0533	30.9227
42	24.8223	28.9619	28.9593	28.5953	27.2315	27.4595	30.0637	30.9352
43	24.814	28.9552	28.9567	28.6016	27.246	27.4572	30.0742	30.9441
44	24.8057	28.9486	28.9542	28.6079	27.2622	27.4549	30.0847	30.9561
45	24.7974	28.9419	28.9516	28.6142	27.2793	27.4526	30.0952	30.9654
46	24.7891	28.9353	28.9491	28.6205	27.2962	27.4503	30.1057	30.9771
47	24.7809	28.9286	28.9466	28.6268	27.3123	27.4481	30.1162	30.9867
48	24.7726	28.922	28.944	28.6332	27.3274	27.4458	30.1268	30.9981
49	24.7643	28.9153	28.9415	28.6395	27.3419	27.4435	30.1373	31.0079
50	24.756	28.9087	28.9389	28.6458	27.3564	27.4412	30.1479	31.0191

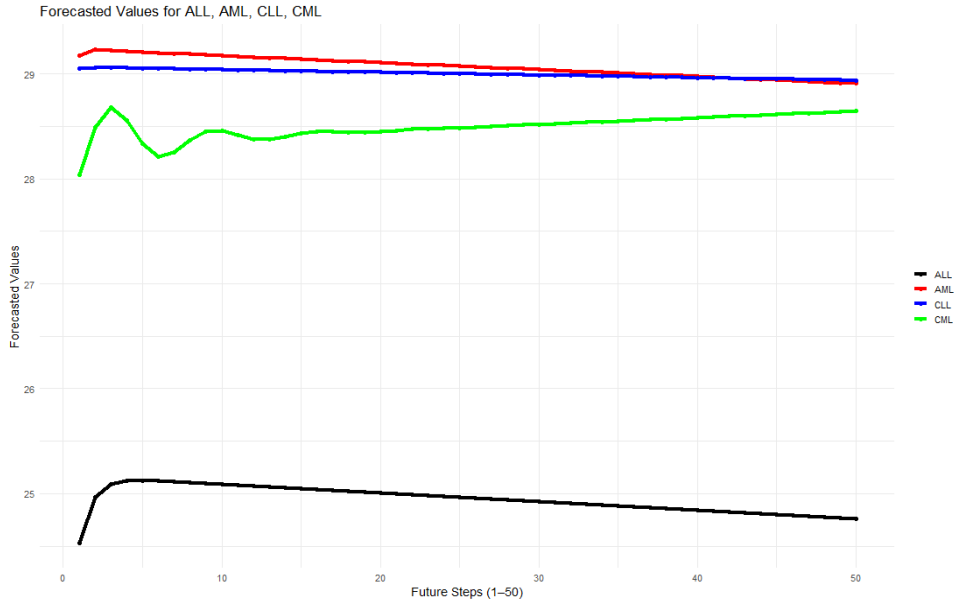

Fig. (3): forecasting value of types of Leukemia in Nanakali

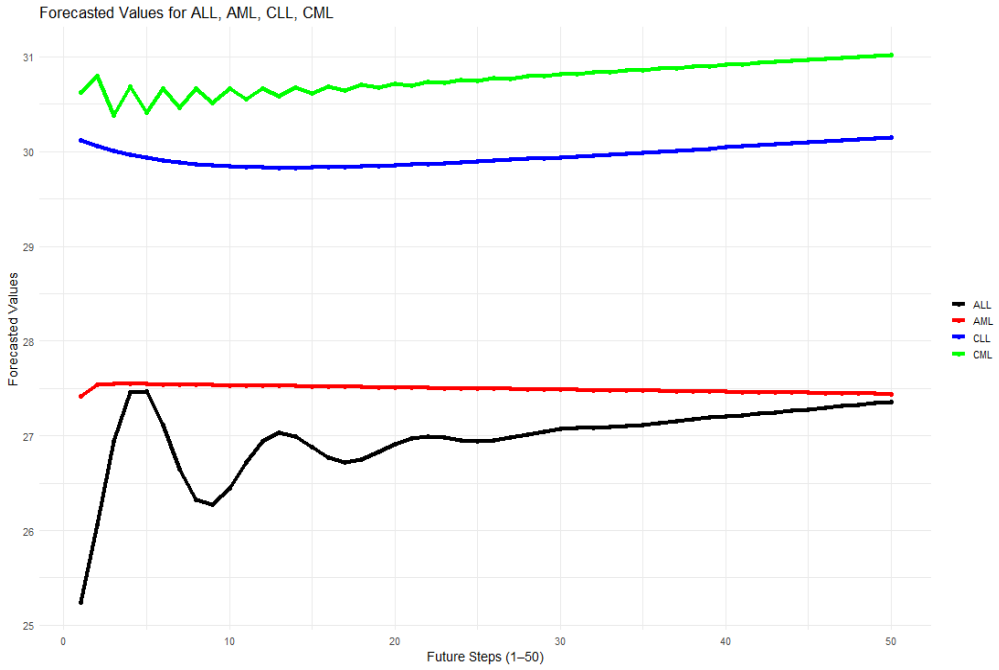

Fig. (4): forecasting value of types of Leukemia in Hiwa

Table (5): compare of the forecasting value of types of Leukemia between Nanakali and Hiwa depends on the Best Models

Leukemia	Hospital	Mean	S.D	t-test	p-value
ALL	Nanakali	24.9409	0.12616	-35.59	0.000
	Hiwa	26.9729	0.38349		
AML	Nanakali	29.0705	0.09537	110.149	0.000
	Hiwa	27.4941	0.03385		
CLL	Nanakali	29.0010	0.03661	-62.583	0.000
	Hiwa	29.9555	0.10145		
CML	Nanakali	28.4943	0.11803	-82.202	0.000
	Hiwa	30.7672	0.15587		

Table 5 presents a statistical comparison of the forecasted values of leukemia types between Nanakali and Hiwa Hospitals based on the best-selected ARIMA models. For ALL, the mean forecast at Nanakali (24.94) is significantly lower than at Hiwa (26.97), with a highly significant t-test result ($t = -35.59$, $p < 0.05$), indicating a clear difference in predicted trends. Similarly, AML shows a higher mean at Nanakali (29.07) compared to Hiwa (27.49), also statistically significant ($t = 110.15$, $p < 0.05$). For CLL, Nanakali's forecasted mean (29.00) is significantly lower than Hiwa's (29.96, $t = -62.58$, $p < 0.05$), and for CML, Nanakali (28.49) is again lower than Hiwa (30.77, $t = -82.20$, $p < 0.05$). Overall, these results indicate that the forecasted trends of all leukemia types differ significantly between the two hospitals, reflecting distinct temporal patterns and potential differences in hospital-specific case dynamics.

The comparative analysis of both observed and forecasted leukemia values at Nanakali and Hiwa Hospitals reveals important insights into the temporal dynamics and inter-hospital differences in leukemia cases. In the original data Table 2, ALL and CML exhibit statistically significant differences between hospitals, while AML and CLL show no significant variation, indicating that historical incidence was relatively similar for some types but differed for others. In contrast, the forecasted values based on ARIMA models Table 5 show significant differences for all leukemia types, suggesting that predictive trends amplify disparities between hospitals. Notably, ALL and CLL are projected lower at Nanakali compared to Hiwa, whereas AML shows an inverse pattern, and CML is markedly higher at Hiwa.

Conclusion and Recommendation

This study applied ARIMA time series models to analyze and forecast the monthly incidence of four major leukemia types Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML) at Nanakali and Hiwa Hospitals in the Kurdistan Region of Iraq, using data covering the period from 2007 to 2023. The comparative analysis successfully revealed differences in both observed patterns and forecasted trends between the two hospitals, providing valuable insights into the epidemiological dynamics of leukemia in the region.

- The results from the historical (real) data showed that ALL and CML exhibited statistically significant differences between Nanakali and Hiwa Hospitals, while AML and CLL did not. This suggests that certain leukemia types may be influenced by hospital-specific factors such as patient referral patterns, diagnostic capacity, or demographic variations in catchment areas.
- The ARIMA modeling process identified distinct optimal models for each leukemia type and hospital, demonstrating that the temporal structures of leukemia incidence vary across settings. For Nanakali Hospital, ARIMA (1,1,1) best fit the patterns of ALL, AML, and CLL, while ARIMA (2,0,2) was optimal for CML. In contrast, Hiwa Hospital displayed more model diversity, with best fits of ARIMA (2,0,2) for ALL, ARIMA (1,1,1) for AML, ARIMA (1,0,1) for CLL, and ARIMA (2,1,2) for CML. These differences highlight the importance of hospital-specific modeling rather than applying a single forecasting model to all data sources.
- Forecasting results further indicated significant differences between hospitals for all leukemia types. Nanakali showed lower projected means for ALL, CLL, and CML compared to Hiwa, whereas AML was forecasted higher at Nanakali. These findings imply that future workload and resource requirements for leukemia care may differ substantially between hospitals.
- Overall, this study confirms the usefulness of ARIMA models for understanding leukemia trends and generating reliable forecasts that support health planning and decision-making. The significant differences in both observed and predicted leukemia patterns between Nanakali and Hiwa Hospitals emphasize the need for tailored strategies in cancer management, resource allocation, and future epidemiological surveillance. Continued monitoring using advanced time series models is recommended to track evolving leukemia trends and enhance healthcare preparedness in the Kurdistan Region of Iraq.

References

1. Ahmed, D. H., Mohamad, S. H., & Karim, R. H. R. (2023). Using Single Exponential Smoothing Model and Grey Model to Forecast Corn Production in Iraq during the period (2022-2030). University of Kirkuk Journal For Administrative and Economic Science, 13(3).
2. Bispo, J. A. B., Pinheiro, P. S., & Kobetz, E. K. (2019). Epidemiology and Etiology of Leukemia and Lymphoma. Cold Spring Harbor Perspectives in Medicine, 10, a034819. https://doi.org/10.1101/cshperspect.a034819
3. Box, G., & Jenkins, G. M. (1976). Analysis: Forecasting and Control. San francisco.
4. Chatfield, C. (2019). The analysis of time series: An introduction (8th ed.). Chapman and Hall/CRC.
5. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431. https://doi.org/10.2307/2286348
6. Enders, W. (2014). Applied econometric time series (4th ed.). Wiley.
7. Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 1847979018808673.
8. Gorgess, H. M. (2017). Time Series Forecasting by Using Box-Jenkins Models. Ibn AL-Haitham Journal For Pure and Applied Science, 26(1), 337-345.
9. Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. Journal of Advanced Management Science, 4(2).
10. Hamilton, J. D. (1994). Time series analysis. Princeton University Press.
11. Hussein, M. M. F., Saeed, A. A., & Mohamad, S. H. (2023). Comparison Markov Chain and Neural network models for forecasting population growth data in Iraq. University of Kirkuk Journal For Administrative and Economic Science, 13(4).
12. Hussein, O. M. M. (2024). A Comparison of ARIMA and GVM (1, 1) models for Forecasting Rice Production in Iraq. Journal of Kurdistani for Strategic Studies, (11).
13. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts.
14. Jabbour, E., & Kantarjian, H. (2020). Chronic myeloid leukemia: 2020 update on diagnosis, therapy, and monitoring. American Journal of Hematology, 95(6), 691–709.
15. Kadri, R., Kahoui, H., & Sahed, A. (2023). Electricity Consumption Forecasting in Algeria: A Comparison of ARIMA and GM (1, 1) Models. Globalization and Business, 8(16), 15-24.
16. Kharista, A., Permanasari, A. E., & Hidayah, I. (2015, May). The performance of GM (1, 1) and ARIMA for forecasting of foreign tourists visit to Indonesia. In 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 33-38). IEEE.
17. Muangkhoua, S. (2019). Time series forecasting by using Box-Jenkins method. Vajira Medical Journal: Journal of Urban Medicine, 63(Supplement), S185-S192.
18. Omer, A., Faraj, S. M., & Mohamad, S. H. (2023). An application of two classification methods: hierarchical clustering and factor analysis to the plays PUBG. IRAQI JOURNAL OF STATISTICAL SCIENCES, 20(20), 25-42.
19. Pui, C. H., Robison, L. L., & Look, A. T. (2021). Acute lymphoblastic leukaemia. The Lancet, 398(10299), 2146–2162. https://doi.org/10.1016/S0140-6736(21)01817-6
20. T Kahwachi, W., & Khalid Hasan, S. (2023). Forecasting Electricity Generation in Kurdistan Region Using BOX-Jenkins Model. Eurasian Journal of Science and Engineering, 9(1).
21. Whiteley, A. E., Price, T. T., Cantelli, G., & Sipkins, D. A. (2021). Leukaemia: a model metastatic disease. Nature Reviews Cancer, 21(7), 461–475. https://doi.org/10.1038/s41568-021-00355-z
22. Xin, Y. (2022, December). Research on the Gold Price Forecasting Based on Machine Learning Models. In 2022 2nd International Conference on Business Administration and Data Science (BADS 2022) (pp. 1339-1346). Atlantis Press.
23. Zhang, Y., & Meng, G. (2023, March). Simulation of an adaptive model based on AIC and BIC ARIMA predictions. In Journal of Physics: Conference Series (Vol. 2449, No. 1, p. 012027). IOP Publishing.