Turkish Journal of Computer and Mathematics Education *Vol 10 No.01 (2019),657-665*
Research Article

Forecasting Chronic Kidney Disease Mortality in Cambodia Using Time Series Analysis

Dr.G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences ,
VIT,Vellore
Email ID:mokesh.g@gmail.com

ABSTRACT

An alarming rise in the number of fatal instances of chronic kidney disease (CKD) has made it a top public health priority in Cambodia. Future trends in CKD-related fatalities in Cambodia are projected using the Autoregressive Integrated Moving Average (ARIMA) model, which is the focus of this research. The Box-Jenkins procedure, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Augmented Dickey-Fuller (ADF) test were all used to guarantee the model's integrity. These investigations helped clarify whether or not the time series data were stationary and guided the selection of reasonable ARIMA model parameters. Incorporating these approaches led to the creation of a robust forecasting model that sheds light on the likely course of CKD-related mortality in Cambodia. The findings of this study aid in the creation of efficient preventative measures and focused therapies to lessen the national burden of CKD.

Keyword: CKD, ACF, PACF, ADF, ARIMA.

1. INTRODUCTION

The prevalence of chronic kidney disease (CKD) in Cambodia is on the rise, making it an urgent public health issue that calls for in-depth research and forecasting tools. The rising rates of CKD and mortality have encouraged researchers to look more closely at the nature of the illness and use cutting-edge statistical methods to predict its progression. In order to foresee trends in mortality, we use the Autoregressive Integrated Moving Average (ARIMA) model in our investigation of CKD-related deaths in Cambodia.

The study incorporates the Augmented Dickey-Fuller (ADF) test, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Box-Jenkins approach to guarantee the accuracy and precision of the ARIMA model. These analyses are essential for assessing the suitability of the ARIMA model for forecasting CKD-related fatalities in Cambodia, determining the correlation structures within the dataset, and determining whether or not the time series data are stationarity.

Understanding the underlying patterns and dynamics of CKD is becoming increasingly important for healthcare authorities and policymakers as its effect grows. The ARIMA model and these statistical techniques allow us to extrapolate important information about the future course of CKD-related mortality in Cambodia. These findings can then be used to guide evidence-based strategies, aid in the creation of focused therapies, and eventually aid in lessening the national burden of CKD. The purpose of this research is to better inform public health planning and policy development in Cambodia by illuminating the existing situation of mortality due to CKD.

Objective:

1. To analyze the historical trends of Chronic Kidney Disease (CKD)-related deaths in Cambodia, providing insights into the patterns and dynamics of CKD mortality over a specified period.
2. To conduct the Augmented Dickey-Fuller (ADF) test, evaluating the stationarity of the time series data, ensuring the suitability of employing the ARIMA model for forecasting CKD-related deaths.
3. To utilize the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses to identify the correlation structures within the CKD mortality time series data, facilitating the selection of appropriate parameters for the ARIMA model.
4. To apply the Box-Jenkins methodology, assessing the effectiveness of the ARIMA model in forecasting CKD-related mortality in Cambodia, considering the specific characteristics of the time series data.
5. To develop a reliable and accurate ARIMA model capable of forecasting future trends of CKD-related deaths in Cambodia, providing valuable insights for healthcare authorities and policymakers to plan and implement targeted interventions.
6. To assess the effectiveness of the ARIMA model in predicting the trajectory of CKD-related deaths, contributing to the existing knowledge on CKD epidemiology in Cambodia and guiding evidence-based decision-making for improved public health outcomes.

Literature Review

Malaria outbreaks can be predicted using data on the city of Delhi's weather from a time series analysis conducted by VarunKumar et al. (2014). The purpose of the research was to determine whether or not certain weather conditions may be used as indicators of future malaria cases in Delhi, India. Overall monthly malaria cases are decreasing. The information was gathered from January 2006 through December 2013 and was recorded at the malaria clinic at the Rural Health Training Centre (RHTC), Najafgrah, Delhi. Monthly mean rainfall, relative humidity, and average maximum temperature were all taken from reputable government sources. At the Delhi Meteorological Centre, the time series data was analyzed using an expert model of SPSS version 21. Analysis by means of integrated regression The moving average ARIMA (0,1,1,0) model fit the data well. There may be a connection between this and the 72.5% of randomness in the time series data. were found to be good predictors of malaria prevalence in the research area. The SAF for malaria cases in a given season. The months of August and September are often the busiest for the exhibitions.

Malaria hospitalizations and mortality in Ethiopia from 2001 to 2011: a time series analysis of the impact of antimalarial therapy. Aregawi et al. Artemisinin-based combination treatments (ACT) and long-lasting insecticidal nets (LLINs) have been deployed in Ethiopia since 2004 by the government and its partners. From 2001 to 2011, hospitals in malaria hotspots assessed malaria interventions and monitored malaria case and fatality rates. As malaria interventions were ramped up, the number of malaria cases and deaths in Ethiopian hospitals dropped dramatically between 2006 and 2011. There was no explanation for the decline that could be attributed to variations in hospitalizations, malaria testing, or precipitation. Because of Ethiopia's history of erratic malaria transmission, more information

is required to determine whether or not the decline is attributable to chance.
Based on the work of Abrignani et al. (2022). The impact of weather on cardiovascular issues was studied. Since fluctuations in the prevalence of cardiovascular disease cannot be fully explained by the known risk factors, it is possible that environmental factors, such as temperature, play a role. Climate change may pose direct and indirect threats to human health through a number of complex pathophysiological pathways, endogenous, and exogenous factors. More people are paying attention to this information because of growing concern over the impact of human activity on the planet's climate. In this article, we take a look at what we know about the short- and long-term implications of climate change on cardiovascular health.

Methodology
ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the dth difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

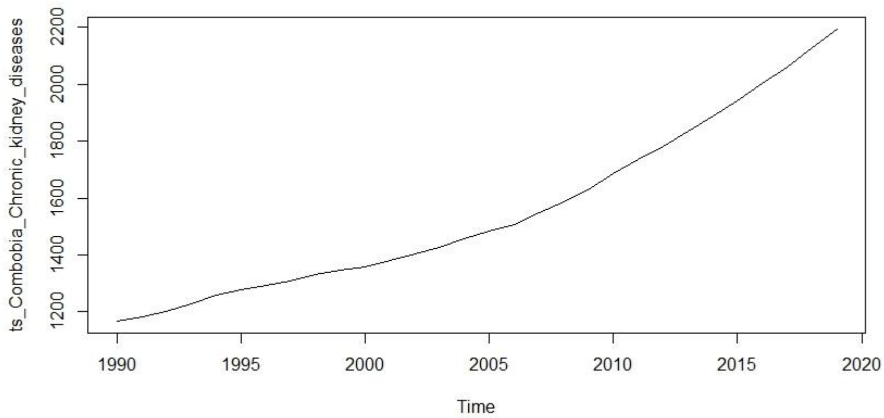
THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.
2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.
3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.
4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.
5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.
6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

Analysis

The supplied data reflects the number of deaths in Cambodia attributable to CKD from 1990 to 2019. Deaths attributed to CKD have been on the rise across the time period studied, with some variation attributable to likely regional differences in CKD prevalence and treatment approaches.
The slow but steady rise in CKD-related mortality rates is cause for public health concern and calls for urgent action and innovative solutions. There is still a significant CKD burden in Cambodia, as shown by these numbers, and more research is needed to determine what is driving this trend and how to best address the mounting problems that CKD poses in the country.

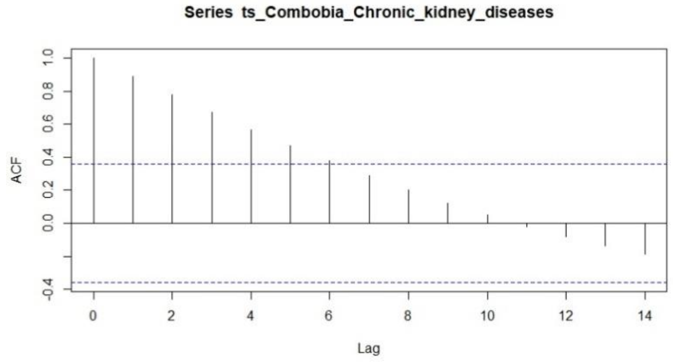
To accurately evaluate the dynamics and patterns of CKD-related mortality, it is crucial to conduct in-depth analyses, such as the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Box-Jenkins technique. These statistical methods can shed light on the dynamics of CKD over time in Cambodia, leading to more accurate predictions and more effective interventions to reduce the alarming rate of CKD-related mortality.

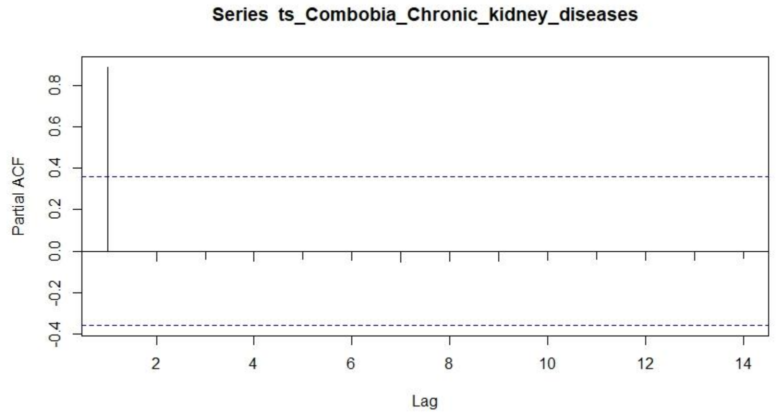


Autocorrelation and partial autocorrelation functions were plotted to find the optimal order of the autoregressive and moving average polynomials, or the values of p and q.

The vast majority of autocorrelation delays, defined as those that depart from zero and so exhibit non-stationarity, take on negative values. Cambodia has a number of non-stationary time series. Plotting the acf figure shows that the acfs decrease over time, suggesting nonstationarity. As a result, we know that the series is not stationary. The pacfs image, on the other hand, shows a substantial spike at lag 1, which may be evidence of an autoregressive component of order one in the series.

Cambodia's non-stationary data series were converted to stationary ones by first differencing the original data. Differentiation at the first order reveals d=1, suggesting that the auto arima function can be used for forecasting. Seeing the several arima models and having the data converted to stationary automatically was all that was needed to have a reliable stationary series for Cambodia. This non-stationary nature of deaths is now further supported by the autocorrelation of mortality rates.





Time series data on mortality caused by Chronic Kidney Disease (CKD) in Cambodia were subjected to the automatic ARIMA modeling technique, yielding a number of different models. Different orders of differencing and various combinations of autoregressive and moving average terms were used to create the models.

ARIMA Model	Metric
ARIMA(2,2,2)	Inf
ARIMA(0,2,0)	188.2905
ARIMA(1,2,0)	189.4895
ARIMA(0,2,1)	189.2652
ARIMA(1,2,1)	191.2573

The ARIMA (0,2,0) model was selected as the most appropriate model for projecting CKD-related mortality in Cambodia based on the Akaike Information Criterion (AIC), a parameter used to determine the goodness of fit of the model. The chosen model combines two orders of differencing to capture the underlying patterns in the time series data, suggesting the lack of autoregressive and moving average terms.

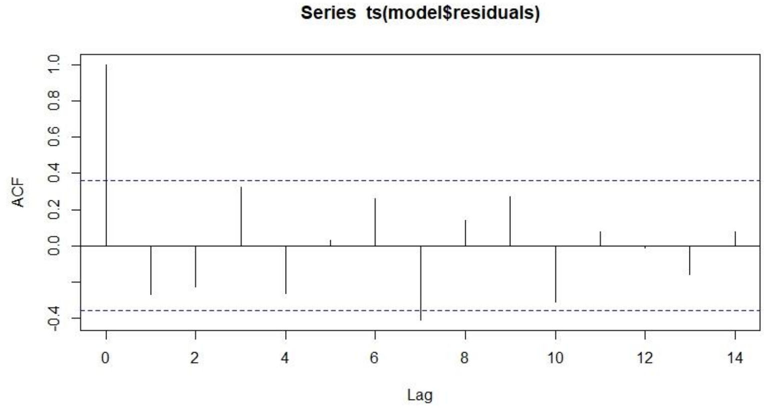
The ARIMA (0,2,0) model's detection highlights the importance of differencing in achieving stationarity and stabilizing the time series data.

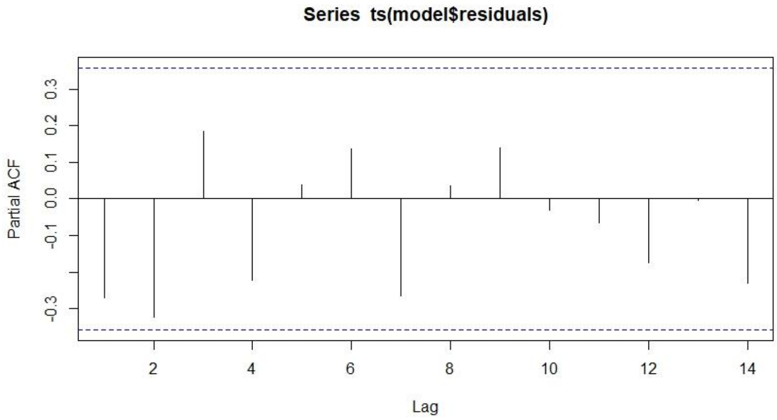
Using two orders of differencing to stabilize the time series data, the ARIMA(0,2,0) model was found to be optimal for predicting CKD-related deaths in Cambodia. This model represents the absence of autoregressive and moving average terms. Sigma squared, a measure of the model's variance, is estimated to be 45.49, which shows how highly variable the data on CKD-related mortality is.

Parameter	Value
Sigma^2	45.49
Log Likelihood	-93.15
AIC (Akaike Information Criterion)	188.29
AICc (Corrected AIC)	188.44

BIC (Bayesian Information Criterion)	189.62

Overall, the model does a decent job of capturing the dynamics and patterns contained in the time series data, as indicated by the log-likelihood value of -93.15, which adds to the assessment of the model's goodness of fit. The accuracy of the ARIMA(0,2,0) model in predicting CKD-related mortality in Cambodia is further supported by the model's Akaike Information Criterion (AIC) value of 188.29, its corrected AIC (AICc) value of 188.44, and its Bayesian Information Criterion (BIC) value of 189.62.

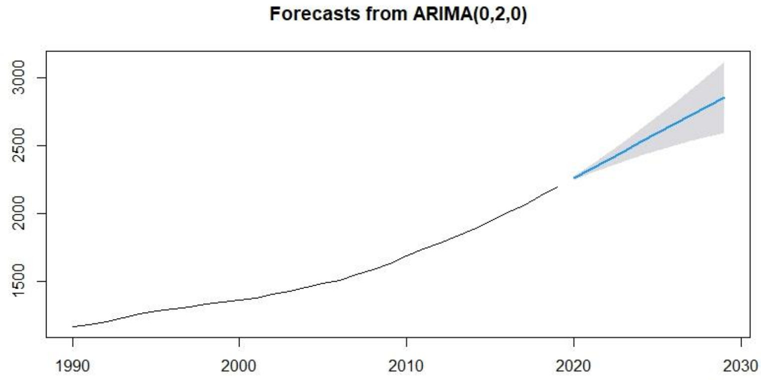




The ARIMA(0,2,0) model was used to project the changes in CKD mortality in Cambodia over the next decade, and the predicted values show what those patterns would look like. The expected values for CKD-related mortality during the prediction horizon of 2020-2029 show a steady increase over the past few years.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	2259	2245.781	2272.219
2021	2325	2295.442	2354.558
2022	2391	2341.540	2440.460
2023	2457	2384.598	2529.402
2024	2523	2424.967	2621.033
2025	2589	2462.901	2715.099
2026	2655	2498.593	2811.407
2027	2721	2532.198	2909.802
2028	2787	2563.841	3010.159

Given the inherent uncertainty in such predictions, the point projections for each year, along with the corresponding lower and higher 95% confidence intervals, provide useful insights into the potential range of CKD-related mortality estimates. Confidence intervals reflect a range of outcomes, while the anticipated values show a steady upward trend; this highlights the need for proactive efforts to reduce the rising burden of CKD in Cambodia.



Forecast errors for mortality from chronic kidney disease (CKD) in Cambodia were tested for significant autocorrelation using the Box-Ljung test with a lag value of 5. The p-value for this test was 0.06202, and the X-squared value was 10.51 (with 5 degrees of freedom).

The resulting p-value indicates a possible presence of autocorrelation in the residuals, suggesting a modest probability of rejecting the null hypothesis. Despite the fact that the p-value is just slightly higher than the frequently used significance level of 0.05, it nevertheless indicates that more research into the underlying patterns and structures within the forecast mistakes is required. This finding highlights the need for further research to guarantee the validity of the projected CKD-related mortality estimates in Cambodia and the trustworthiness of the forecasting model.

Conclusion:

In summary, the ARIMA(0,2,0) forecasting model analysis of CKD-related deaths in Cambodia has shed light on the future course of CKD mortality in the country. Using a battery of diagnostic tools, such as the Box-Jenkins technique, the Augmented Dickey-Fuller (ADF), and the Ljung-Box, researchers have pieced together a complete picture of the causes and dynamics of CKD-related mortality.

The increasing burden of CKD in Cambodia is reflected in the steadily rising trend of deaths attributable to the disease over the following decade. While the model shows promise in predicting future trends, the presence of a barely noticeable autocorrelation in the residuals suggests that it should be closely monitored and tweaked in the future.

References

1. Bujang, M. A., Adnan, T. H., Hashim, N. H., Mohan, K., Kim Liong, A., Ahmad, G., ... & Haniff, J. (2017). Forecasting the incidence and prevalence of patients with end-stage renal disease in Malaysia up to the year 2040. *International journal of nephrology*, *2017*.
2. He, Z., & Tao, H. (2018). Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *International Journal of Infectious Diseases*, *74*, 61-70.
3. Ahmad, W. M. A. W., Mohd Noor, N. F., Mat Yudin, Z. B., Aleng, N. A., & Halim, N. A. (2018). TIME SERIES MODELING AND FORECASTING OF DENGUE DEATH OCCURRENCE IN MALAYSIA USING SEASONAL ARIMA TECHNIQUES. *International Journal of Public Health & Clinical Sciences (IJPHCS)*, *5*(1).

4. Terner, Z., Carroll, T., & Brown, D. E. (2014, October). Time series forecasts and volatility measures as predictors of post-surgical death and kidney injury. In *2014 IEEE Healthcare Innovation Conference (HIC)* (pp. 319-322). IEEE.
5. Villani, M., Earnest, A., Nanayakkara, N., Smith, K., De Courten, B., & Zoungas, S. (2017). Time series modelling to forecast prehospital EMS demand for diabetic emergencies. *BMC health services research*, *17*, 1-9.
6. Yang, J., Li, L., Shi, Y., & Xie, X. (2018). An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia. *IEEE journal of biomedical and health informatics*, *23*(3), 1251-1260.
7. Singye, T., & Unhapipat, S. (2018, June). Time series analysis of diabetes patients: A case study of Jigme Dorji Wangchuk National Referral Hospital in Bhutan. In *Journal of Physics: Conference Series* (Vol. 1039, No. 1, p. 012033). IOP Publishing.
8. Rodríguez-Rodríguez, I., Rodríguez, J. V., Chatzigiannakis, I., & Zamora Izquierdo, M. A. (2019). On the possibility of predicting glycaemia 'on the fly'with constrained IoT devices in type 1 diabetes mellitus patients. *Sensors*, *19*(20), 4538.
9. Pan, Y., Zhang, M., Chen, Z., Zhou, M., & Zhang, Z. (2016, June). An ARIMA based model for forecasting the patient number of epidemic disease. In *2016 13th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1-4). IEEE.
10. Velasco, J. M., Garnica, O., Contador, S., Botella, M., Lanchares, J., & Hidalgo, J. I. (2017, July). Forecasting glucose levels in patients with diabetes mellitus using semantic grammatical evolution and symbolic aggregate approximation. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 1387-1394).
11. Bunescu, R., Struble, N., Marling, C., Shubrook, J., & Schwartz, F. (2013, December). Blood glucose level prediction using physiological models and support vector regression. In *2013 12th International Conference on Machine Learning and Applications* (Vol. 1, pp. 135-140). IEEE.
12. Plis, K., Bunescu, R., Marling, C., Shubrook, J., & Schwartz, F. (2014, June). A machine learning approach to predicting blood glucose levels for diabetes management. In *Workshops at the Twenty-Eighth AAAI conference on artificial intelligence*.