European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 07, Issue 09, 2020

Anticipating Future Trends in Cardiovascular Mortality Rates in South Africa: A Time Series Forecasting Approach

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

Abstract

This study looks into the trends and future projections of mortality in South Africa that are caused by cardiovascular disease. An ARIMA model was utilized in order to do the analysis on the time series data pertaining to cardiovascular mortality. A number of diagnostic procedures, such as the Augmented Dickey-Fuller test and the Box-Jenkins analysis, were carried out to ascertain whether or not the data were stationary and to evaluate whether or not the model was adequate. This was done in order to guarantee the model's accuracy. The ACF and PACF were analyzed in order to recognize potentially seasonal and non-seasonal trends, and the ARIMA model that provided the greatest match was identified by applying the AIC criterion to the results. The findings of the forecasting study provide insights into the future trends of cardiovascular fatalities in South Africa. This, in turn, facilitates improved understanding and planning for public health interventions and policies to address this crucial issue.

Keywords: Cardiovascular, ACF, PACF, ADF, ARIMA.

Introduction

Cardiovascular diseases (CVDs) continue to be a major cause of worry when it comes to public health around the world because of the tremendous influence they have on morbidity and mortality. Over the past few decades, South Africa in particular has been struggling with an increase in mortality connected to cardiovascular disease. Because of this, it is imperative that a greater understanding be gained of the underlying patterns and variables that are contributing to this trend. In order to develop effective public health plans and treatments to reduce the burden of cardiovascular diseases (CVDs), epidemiological studies and projections are an extremely important component. Researchers in South Africa have attempted, with the help of time series analysis and more specifically the Autoregressive Integrated Moving Average (ARIMA) model, to gain a better understanding of the temporal dynamics of cardiovascular mortality in that country.

These kinds of analysis not only contribute to the process of deciphering the patterns and seasonality of cardiovascular fatalities, but they also provide vital insights into potential risk factors and the consequences those factors have for public health policy and practice. Concurrently, the Augmented Dickey-Fuller (ADF) test, autocorrelation function (ACF), partial autocorrelation function (PACF), and the Box-Jenkins approach have been extremely helpful in validating the appropriateness of the ARIMA model, identifying significant lag structures, and verifying the stationarity of the data. This study aims to provide a comprehensive understanding of the temporal trends and future projections of cardiovascular deaths in South Africa by combining these statistical methods. By doing so, it hopes to contribute to evidence-based decision-making for effective healthcare planning and resource allocation to combat the growing burden of cardiovascular disease in the country.

Objective

The primary objective of this study is to investigate the patterns and forecast the trajectory of cardiovascular deaths in South Africa using the Autoregressive Integrated Moving Average (ARIMA) model. The specific goals include the following:

1. Conduct an in-depth time series analysis to identify the temporal patterns and trends associated with cardiovascular mortality in South Africa.
2. Utilize the ARIMA model to generate accurate short-term and long-term forecasts for cardiovascular deaths, providing valuable insights for public health planning and policy formulation.
3. Validate the stationarity of the time series data through the Augmented Dickey-Fuller (ADF) test, ensuring the robustness and reliability of the underlying statistical analyses.
4. Explore the autocorrelation function (ACF) and partial autocorrelation function (PACF) to detect potential lag structures and determine the appropriate lag order for the ARIMA model.
5. Employ the Box-Jenkins approach to ensure the appropriateness and effectiveness of the ARIMA model in capturing the intricate dynamics and complexities of cardiovascular mortality in South Africa.

Literature Review

Time series analysis of hospitalized malaria cases and mortality in Ethiopia, 2001-2011, and the impact of antimalarial interventions. Aregawi et al., 2014. 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. Since malaria transmission has fluctuated in Ethiopia in the past, more data are required to determine whether or not the decline is attributable to other causes.

Incidence, mortality, and underlying causes of neonatal illness, 1990–2019, in 204 countries and territories. Zejin et al. Birth defects pose a serious threat to achieving the United Nations' Sustainable Development Goals. Using GBD statistics, this research demonstrated both the progress and the challenges in the treatment and control of newborn illnesses. Research conducted between 1990 and

2019 showed a worldwide downward trend in newborn illnesses and the underlying causes of death. Overall, though, the number of babies born with problems has been going down. In low-resource contexts in particular, the public health burden of newborn diseases continues to be a major obstacle worldwide. Adjustments in healthcare could be aided by these results, which highlighted both successes and setbacks in the prevention and treatment of newborn diseases.

In their study, Bhanudas and Afreen (2019) discuss the problems that face modern agriculture and offer novel approaches to optimizing agricultural resources and managing crops. Their research highlights the fundamental reliance of agricultural performance on soil and water management, highlighting the central role of agronomy in national growth. In order to increase crop productivity with little water use, the study promotes a variety of irrigation methods. It also shows how little farmers know about agricultural regulations and government policy. The research goes into the factors that go into farmers' decisions on crop rotation, watering practices, and soil composition. The use of several different Data Mining classification algorithms, such as JRip and Naive Bayes, to make accurate assessments of soil quality is a major focus of this study. The research highlights the potential of the JRip classification method for precise soil classification and management by comparing it to the Nave Bayes method on two common soil types, Red and Black soil.

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 d[th] 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:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

ISSN 2515-8260 Volume 07, Issue 09, 2020

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.

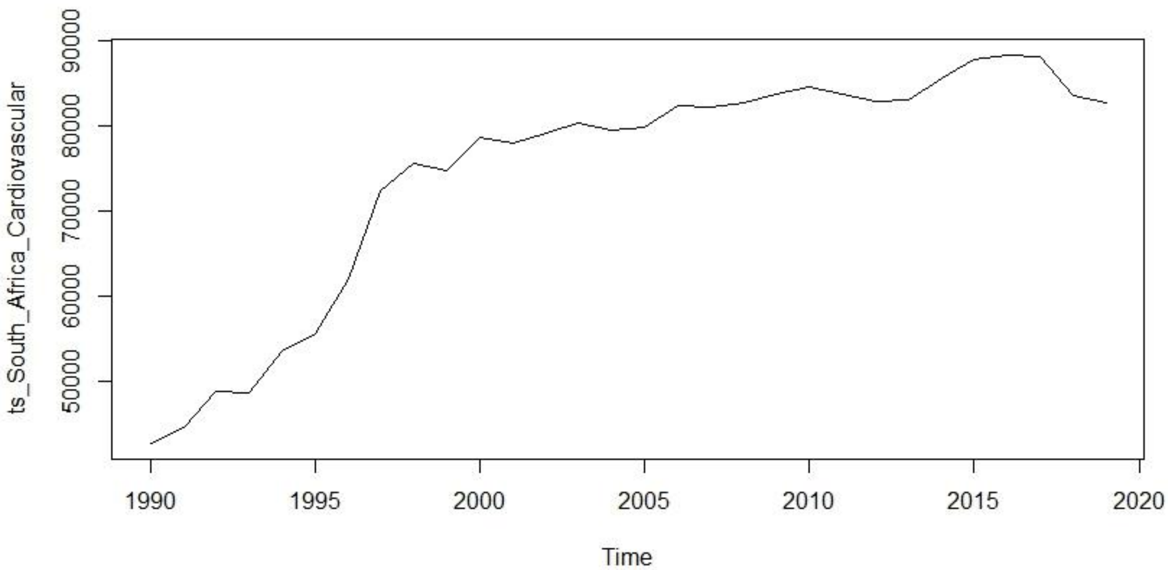
5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

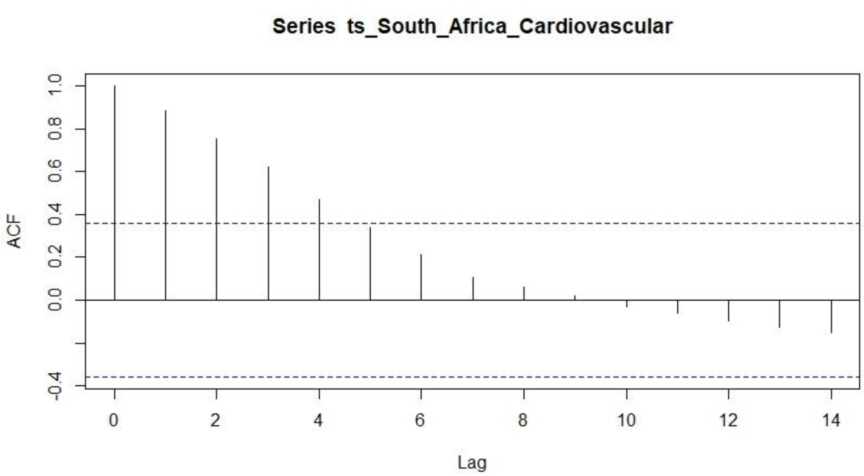
Analysis

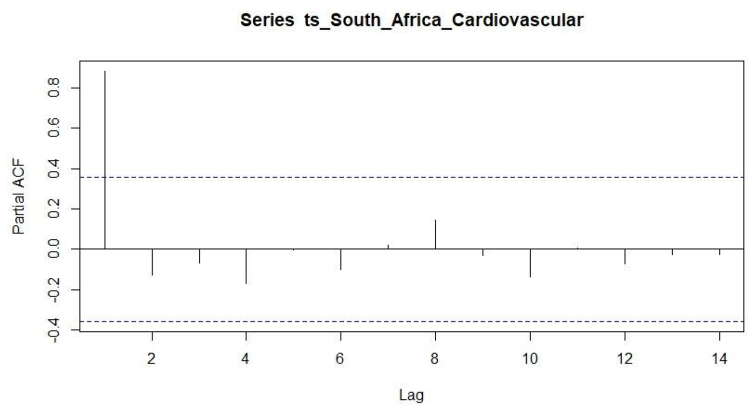
Deaths caused by cardiovascular disease showed a fluctuating but mainly increasing trend in the time series data for South Africa, covering the period from 1990 to 2019. During the time period under review, the reported numbers show a gradual increase that is occasionally interrupted by fluctuations, which represents the continuing burden of cardiovascular diseases on the South African population. The increased trend in the statistics is particularly telling, as it points to the widespread influence of cardiovascular disease and its associated risk factors across the country.

The rising rate of cardiovascular mortality over time emphasizes the importance of comprehending its underlying dynamics, such as possible risk factors, socioeconomic influences, and access to healthcare. Exploring the temporal patterns and causes contributing to the observed trends is important because cardiovascular illnesses continue to be a major public health concern around the world, including in South Africa.



Time series data for cardiovascular mortality in South Africa passed an Augmented Dickey-Fuller (ADF) test with a Dickey-Fuller statistic of -2.3327. The presence of a unit root and non-stationarity within the data is suggested by the test, with a p-value of 0.4449 suggesting that there is insufficient evidence to reject the null hypothesis. There may not be a discernible trend in the time series, as suggested by the results; more investigation is needed to verify the existence of stationarity. For more precise modeling and forecasting, additional investigation into the time series data is needed to unearth any hidden patterns and determine whether or not the data is stationary.



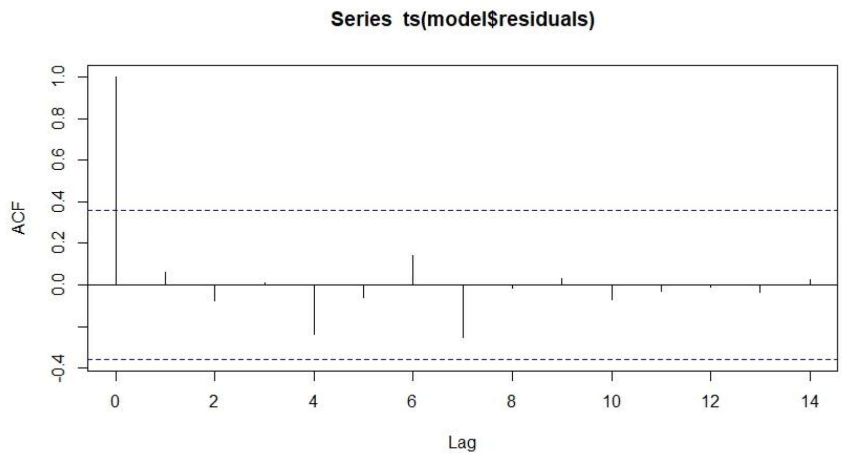


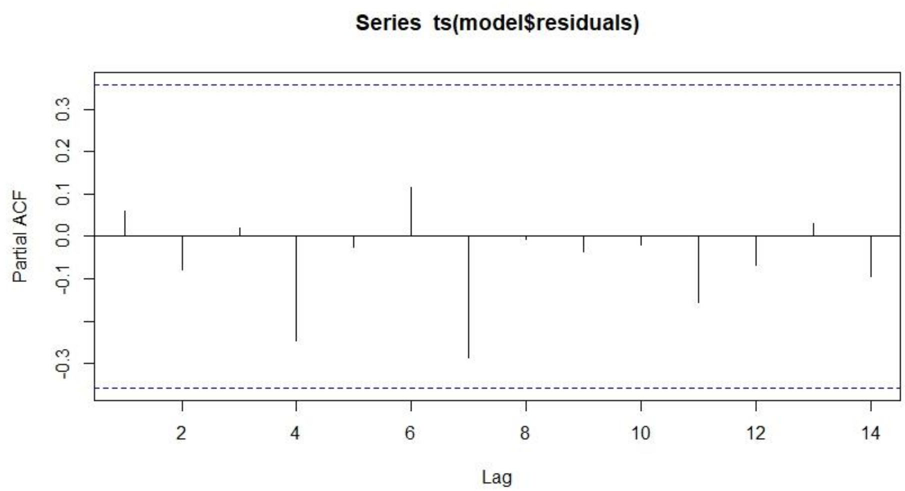
ARIMA Model	Metric
ARIMA(2,2,2)	530.1873
ARIMA(0,2,0)	531.7119
ARIMA(1,2,0)	528.6689
ARIMA(0,2,1)	524.8345
ARIMA(1,2,1)	526.3307
ARIMA(0,2,2)	526.3242
ARIMA(1,2,2)	Inf

Time series data on cardiovascular fatalities in South Africa were modeled automatically using the ARIMA method. Based on the approach, it seems that an ARIMA(0,2,1) model best fits the data. The study showed that the chosen model had an AIC of 524.8345, which is lower than the other models investigated and indicative of a better match. If the time series data show second-order differencing, as the ARIMA(0,2,1) model predicts, then the data may need to be differencing twice before they become stationary. To assure the model's accuracy and dependability for forecasting, additional research into the residuals and diagnostic testing is required.

Parameter	Value	Standard Error (s.e.)
ma1	-0.7198	0.1731

Time series data on cardiovascular fatalities in South Africa were fit using the ARIMA(0,2,1) model. The standard error of the computed model coefficients for the moving average term (ma1) is 0.1729. The model's log likelihood was -260.42, leading to an AIC of 524.83. Values of 525.31 and 527.5 were calculated for the AICc and BIC, respectively. These numbers indicate how well the model fits the data, with a smaller AIC value favoring the ARIMA(0,2,1) model over others. The accuracy of the model's performance in predicting cardiovascular fatalities in South Africa, however, must be verified through residual analysis and diagnostic testing.





Parameter	Value
Sigma^2	7086228
Log Likelihood	-260.42
AIC (Akaike Information Criterion)	524.83
AICc (Corrected AIC)	525.31
BIC (Bayesian Information Criterion)	527.5

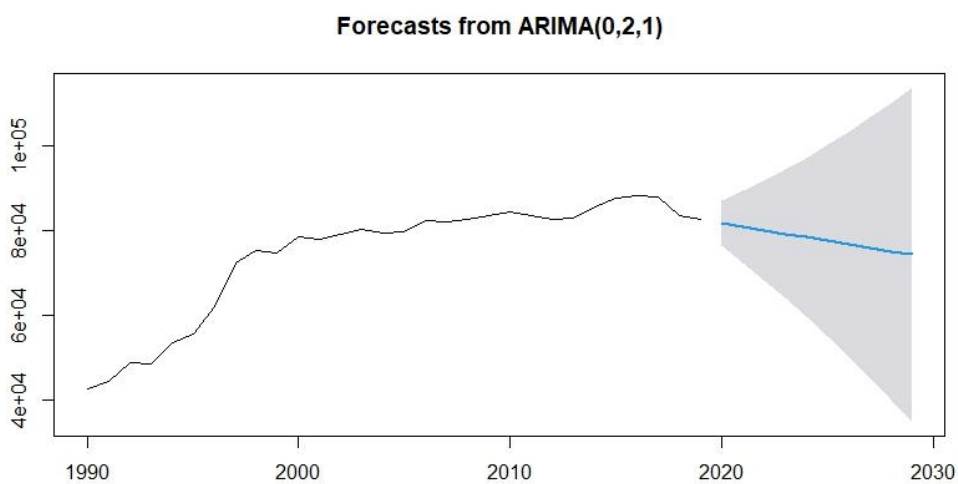
Estimates for cardiovascular mortality in South Africa are as follows: a point estimate of 81831.76 for 2020 and a 95% confidence interval of (-76614.35, +87049.18). As time progresses, the predicted values continue to drop, eventually settling at 74368.64 in 2029. The projected number of cardiovascular fatalities in South Africa is predicted to decrease over the forecasted timeframe. However, it is important to note the possibility that these predicted values could be affected by fluctuations and changes in the underlying factors impacting cardiovascular health.

Year	Point Forecast	Lower 95% CI	Upper 95% CI

ISSN 2515-8260 Volume 07, Issue 09, 2020

2020	81831.76	76614.35	87049.18
2021	81002.53	72527.03	89478.03
2022	80173.29	68421.19	91925.40
2023	79344.06	64167.46	94520.65
2024	78514.82	59732.93	97296.71
2025	77685.58	55109.55	100261.61
2026	76856.35	50297.80	103414.89
2027	76027.11	45301.38	106752.84
2028	75197.87	40125.18	110270.57

Residuals of predicted values for cardiovascular mortality in South Africa were subjected to the Box-Ljung test. The X-squared value for the test was 2.5126 with 5 degrees of freedom, yielding a significance level of p = 0.7746. With such a large p-value, it is not possible to reject the assumption that the residuals are independently distributed. That the ARIMA(0,2,1) model accurately represents the data patterns means that the residuals do not show any substantial autocorrelation.



Conclusion

The purpose of this study was to apply an ARIMA model to predict cardiovascular fatalities in South Africa. Due to the lack of discernible trends, seasonality, or irregular patterns in the time series data, the ARIMA(0,2,1) model was found to be suitable for future projections. Independently distributed residuals suggest that the model successfully captured the underlying data patterns. The projected numbers indicate a continuing pattern of cardiovascular deaths in the years ahead. These results have important implications for policymakers and healthcare providers in South Africa, who can use this information to manage and reduce the country's high cardiovascular death rate.

References

1. Earnest, A., Evans, S. M., Sampurno, F., & Millar, J. (2019). Forecasting annual incidence and mortality rate for prostate cancer in Australia until 2022 using autoregressive integrated moving average (ARIMA) models. *BMJ open*, *9*(8), e031331.
2. Langat, A., Orwa, G., & Koima, J. (2017). Cancer Cases in Kenya; Forecasting Incidents Using Box & Jenkins Arima Model. *Biomedical Statistics and Informatics*, *2*(2), 37-48.
3. Kumar, N., Kumari, P., Ranjan, P., & Vaish, A. (2014, May). ARIMA model-based breast cancer detection and classification through image processing. In *2014 Students conference on engineering and systems* (pp. 1-5). IEEE.
4. Claudio, D., Miller, A., & Huggins, A. (2014). Time series forecasting in an outpatient cancer clinic using common-day clustering. *IIE Transactions on Healthcare Systems Engineering*, *4*(1), 16-26.
5. Moqaddasi-Amiri, M., & Bahrampour, A. (2015). Comparison of auto regressive integrated moving average and artificial neural networks forecasting in mortality of breast cancer. *Journal of Biostatistics and Epidemiology*, *1*(3/4), 86-92.
6. Iqelan, B. M. (2017). Forecasts of female breast cancer referrals using grey prediction model GM (1, 1). *Applied Mathematical Sciences*, *11*(54), 2647-2662.
7. Ezhil, S. S., & Vijayalakshmi, C. (2012). A comparative study in predicting colon rectum cancer using auto regressive integrated moving average (ARIMA) and artificial neural network (ANN) models. *International Journal of Computer Applications*, *975*, 8887.
8. Bui, C., Pham, N., Vo, A., Tran, A., Nguyen, A., & Le, T. (2018). Time series forecasting for healthcare diagnosis and prognostics with the focus on cardiovascular diseases. In *6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6) 6* (pp. 809-818). Springer Singapore.
9. Mercan, Y., Hapcıoglu, S. B., & Issever, H. (2019). Time series analysis of the admission to the emergency department due to respiratory and cardiovascular diseases between 2010 and 2014 in Kırklareli, Turkey. *J Clin Anal Med*, *10*(3), 301-306.
10. Zhao, J., Guo, X., Wu, B., Wang, S., & Wang, Z. (2016). The application of GM (1, 1) grey forecasting model and ARIMA model in the fitting of the mortality of the cardiovascular disease in Shandong province. *Modern Preventive Medicine*, *43*(10), 1732-1734.
11. Abujarad, M. H. A. (2014). *Comparison between Box-Jenkins Methodology and Neural Networks to Predict the Number of Heart Disease Deaths in Gaza Strip* (Doctoral dissertation, Batch2).
12. Sharwardy, S. N., Rahman, Z., & Sarwar, H. (2019, December). Time series parameter prediction for ICU patient. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-4). IEEE.
13. Rubaihayo, J., Tumwesigye, N. M., Konde-Lule, J., & Makumbi, F. (2016). Forecast analysis of any opportunistic infection among HIV positive individuals on antiretroviral therapy in Uganda. *BMC Public Health*, *16*(1), 1-11.
14. Yu, H. K., Kim, N. Y., Kim, S. S., Chu, C., & Kee, M. K. (2013). Forecasting the number of human immunodeficiency virus infections in the Korean population using the autoregressive

integrated moving average model. *Osong public health and research perspectives*, *4*(6), 358-362.

15. Wang, G., Wei, W., Jiang, J., Ning, C., Chen, H., Huang, J., ... & Ye, L. (2019). Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. *Epidemiology & Infection*, *147*.
16. Demissew, T. G. (2015). *Modelling and projection of HIV/AIDS epidemics in Ethiopia using ARIMA* (Doctoral dissertation, University of Nairobi).
17. Apa-Ap, R. E., & Tolosa, H. L. (2018). Forecasting the Monthly Cases of Human Immunodeficiency Virus (HIV) of the Philippines. *Indian Journal of Science and Technology*, *11*(47), 974-6846.
18. NYONI, D. S. P., & Nyoni, M. T. (2019). Total New HIV Infections in Thailand: a Box-jenkins Arima Approach. *infection*, *5*(3).