VFAST Transactions on Mathematics
http://vfast.org/journals/index.php/VTM@ 2024 ISSN(e): 2309-0022, ISSN(p): 2411-6343 Volume 12, Number 1, January-June, 2024 *pp: 415-423*

Comparative Analysis of Time Series Forecasting using ARIMA, and GRNNs Models: A Case Study of Death Rate of Diabetic Mellitus in Canada

Muhammad Shahbaz Khan[1*,2], **Mir Ghulam Hyder Talpur**[2†], **Muhammad Aslam**[2‡]

[1]Department of Mathematics and Statistics, Karachi Institute of Economics and Technology, Karachi, Pakistan; [2]Department of Mathematics, Institute of Business Management, Karachi, Pakistan

Keywords: *Time series analysis, soft computing, GRNNs, ARIMA, MAD, RMSE, and MAPE.* ***Subject Classification***: *Modeling, forecasting, and soft computing.*

Journal Info:
Submitted:
June 05, 2024
Accepted:
June 28, 2024
Published:
June 30, 2024

Abstract This research aims to compare ARIMA and GRNN models alone. For this comparison the death rate for diabetes mellitus time series data of Canada is used. Autoregressive Integrated Moving Average (ARIMA), and Generalized Regression Neural Networks (GRNN) models were applied for time series prediction of the death rate for diabetes mellitus—trained data for two models from 2000 to 2015. Test data was used to compare the precision through data from 2016 to 2021. The ARIMA model was applied using the auto-command through R package which provided the least BIC and AIC values. The mean absolute deviation (MAD), root mean squared error (RMSE), and mean absolute percentage error (MAPE) were employed to measure the forecasting efficiency of the models. The ARIMA model had the highest prediction accuracy as compared to the GRNN model. ARIMA predicts the death rate for diabetes mellitus more accurately and robustly compared to the GRNNs model.

***Correspondence Author Email Address:**
shahbaz.phd.5002@gmail.com
DOI: 10.21015/vtm.v12i1.1894

1 Introduction

Diabetes is a disease called a silent killer depicted by elevated stages of blood glucose (or blood sugar). These problems initiate severe damage to the human body's organs. Eyesight, kidney failure, cardiovascular diseases, and nerves [12]. Type-2 (T2) diabetes is common in adults. It starts when the body doesn't make enough insulin quantity. Over the past 4 decades, the incidence of T2 diabetes has increased significantly. A researcher used the Grammatical Evolution algorithm model to reduce forecasted high-risk glucose errors for Type 1 diabetic patients [5].

The study of time series data is important for making immediate and valuable decisions on time. For making valuable decisions, forecasting and modeling of time series data has its importance analysis. In each field of life such as gold prices, weather, life stock, stock exchange prices, sales, household consumption, and especially medical science forecast future values [11]. A study applied CTRNNs deep learning model incorporating continuous evolution of the hidden states between observations [6]. ARIMA models are Parametric time series models for forecasting and modeling. Parametric models must satisfy essential assumptions such as linearity and stationery of data sets. These models have been usually working for both modeling and forecasting purposes with some new techniques of wavelets. One of the Non-parametric methods of the time series data, the Generalized Regression Neural Networks (GRNNs) model is used to forecast and model for short time series data sets. GRNNs do not need to satisfy the essential fundamentals of assumptions like the ARIMA model [10].

This paper is structured as part II will deal with the literature review on parametric and non-parametric time series models. Part III will present the procedure of modeling and forecasting death rate time series data. In part IV, we applied the suggested ARIMA and GRNNs models. A comparison of suggested models will help us to predict the future death rate of diabetic patients. Assessment of highly accurate model using precision measures through MAD, RMSE, and MAPE. The paper is concluded in section V.

2 Literature Review

Facing stress, unhealthy food, and excessive work increases the rate of diabetic patients. The death rate of these patients is also high in those places where medical officers are not taking quick action to treat the problem. This study incorporates two techniques data mining and meta-heuristic to foretell the early readmission probability of diabetic patients contained by 30 days of discharge [17].

This paper helps in the assessment of the modern IoMT-based techniques for the continuous observation of DM1 management and enabling an intense representation of diabetic patients. The union of wearable technologies with machine learning approaches yields strong models for short-term blood glucose forecasts [14]. A SEIR model forecasting the COVID-19 epidemic in Kenya using a parametric model called an Autoregressive Integrated moving average (ARIMA). The mean time changes between waves 1 to 4 were monitored to be about 130 days [8].

This study used the Box-Jenkins method on the training data set and built many time series models to precisely capture the pattern behavior in the monthly RTA data. SARIMA $(0,1,1)\times(1,0,0)_{12}$ model was obtained to model for predicting RTAs [2]. In this research, a novel model was used Neural Network to forecast basal insulin. The experimental part showed evidence of the high efficiency of the proposed method [4]. A hybrid model based on ARIMA-Long Short-Term Memory (LSTM) established on the random forest lag choice principle has been proposed to predict the average effect. The GARCH model is also used with the residuals determined from the ARIMA model to predict the change behavior of the time series

[13]. ARIMA-GARCH models portray outstanding statistical models on ARIMA models observing the lowest values of AIC and BIC. The LSTM model is also used on all normalized training data series.

Two time series models ARIMA and ANN applied to forecast Karachi stock exchange prices [9]. A model Deep Belief Network is used to foretell the diabetes mellitus. This research is utilized to predict the complications of diabetic mellitus [15].

3 Tools and Suggested Methodology

This study used a leading parametric model (ARIMA) and a Non-Parametric model (GRNN) to analyze the death rate of diabetic patients. Data from 2000 to 2015 were observed for the training set to guide the models, and from 2016 considered as a test set to test the model's foretelling efficiency.

A. Dataset This study used the time series data which was obtained from Statistica website [16]. The death rate of diabetics (per 100,000 population) in Canada from 2000 to 2021.

B. Autoregressive Integrated Moving Modeling:

The ARIMA model is used to analyze time series data. The fundamental assumption is data should be stationary. It is also known as the Box-Jenkins model. The Parametric model is applied to provide efficient and precise forecasting the future observations [11]. The ARIMA model is based on three parameters. The "P" represents the order of the (AR) term, the "q" represents the order of the (MA), and the "d" represents an integer that adjusts the level of differencing. The (AR) model is defined as

$$R(p) : Y_t = \alpha + \sum \omega Y_{t-1} + \epsilon_t. \tag{1}$$

for i = 1, 2, . . . , p. It shows the linear combination of past values.

The (MA) model forms the estimates by utilizing the past forecast errors such that

$$\mathrm{MA}(q) : Y_t = \varepsilon_t + \sum \theta_i * \varepsilon_{t-1}. \tag{2}$$

for i = 1, 2, . . . , q. Which is equal to a linear order of q past forecast errors. We estimate the effective forecasted number at time t (Re(t)) as follows:

$$\mathrm{Re}(t) = \omega_1 y'_{t-1} + \cdots + \omega_1 y'_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_q \varepsilon_{t-q}. \tag{3}$$

The selection of model criteria are Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. The concept is that any model that has the least AIC and BIC values will be considered as the best model [3]. The Augmented Dickey-Fuller (ADF) test [citation] was used to check the condition of stationarity. The residual plot, histogram, and "qq" plot were used to check the fundamental assumption fulfillment.

C. Generalized Regression Neural Networks Model:

The generalized regression neural network (GRNNs) is one of the neural network models. The learning process of GRNNs is a feedforward training type that belongs to the radial basis model. The GRNNs model was designed to be suitable for small data sets. The structure of the GRNN model is shown in Fig. 1 and described in [1].

The algorithm of GRNN is presented in Fig. 2 which describes the procedure of forecasting values. This procedure consists of the Input layer, pattern layer, summation layer, and output layer.

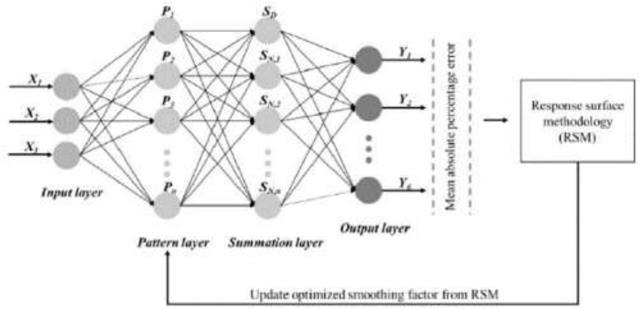

Figure 1. Structure of Generalized Regression Neural Networks (GRNNs)

Input Layer: The input layer is the first layer of the GRNN model. (Data set for testing Networks)
Pattern Layer: The pattern Layer is known as one of the hidden layers of the GRNNs model
Summation Layer: The summation layer is part of the hidden layer of this network
Output Layer: The predictions and results of the GRNNs model were generated in this layer

Figure 2. Procedure of GRNN Model

4 Implementation and Results

The Time series data for diabetic patients from Canada is presented in Table 1 and plotted in Fig. 3(a). The trained data set for modeling is plotted in Fig. 3(b).

The Auto-Arima command of package "tseries" of R software found the best fit was ARIMA (0,1,0) [7] which is AR(0), and MA(0) with a 1-time difference and was used to estimate the Re(t) values up to 16 values from 2000 to 2015 as trained data presented in Fig. 3(b). The next six years' forecasted values as test data from 2016 to 2021 are presented in Table 2.

The confidence intervals for each forecasted value using ARIMA model presented in Table 2 with 80% and 95% confidence levels and presented in figure 5(a).

The accuracy measurements MAD, RMSE, and MAPE of the ARIMA model are presented in Table 3. It shows the accuracy measurements MAD = 0.933, RMSE = 1.055, and MAPE = 4.96 % of the ARIMA model.

For implementing the GRNN model, we used the package "tsfgrnn" of R software. For smoothing standard deviation is used in grnn model command [5]. First, we applied the GRNN model over the trained

year	Death Rate of Diabetic Patients	Year	Death Rate of Diabetic Patients
2000	21.8	2011	21
2001	22.8	2012	20.1
2002	25.1	2013	20
2003	25.2	2014	18.9
2004	24.5	2015	19
2005	24.4	2016	18.5
2006	22.2	2017	19
2007	22.5	2018	18.5
2008	22.6	2019	18.6
2009	20.5	2020	20.1
2010	20.3	2021	19.5

Table 1. Death Rate of Diabetic Patients in Canada

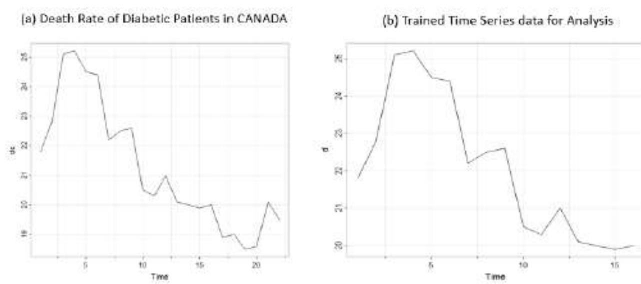

Figure 3. Time Series Plot of Death Rate of Diabetic Patients in CANADA

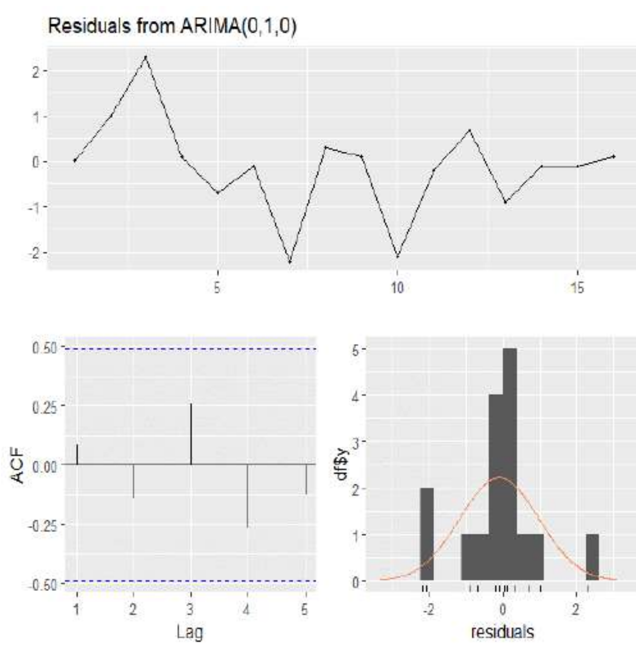

Figure 4. Trained Time Series data for Analysis

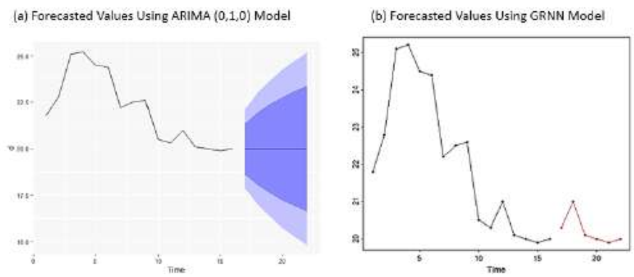

Figure 5. Forecasted Values of Set Data for Model Accuracy

Year	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2016	20	18.615	21.385	17.882	22.118
2017	20	18.041	21.959	17.004	22.996
2018	20	17.601	22.399	16.331	223.669
2019	20	17.23	22.77	15.764	24.236
2020	20	16.903	23.097	15.263	24.737
2021	20	16.607	23.393	14.811	25.189

Table 2. Forecasted values with Confidence Interval

Years	Actual	Forecasted	MAD	RMSE	MAPE
2016	18.9	20	1.1	1.21	0.058201
2017	19	20	1	1	0.052632
2018	18.5	20	1.5	2.25	0.081081
2019	18.6	20	1.4	1.96	0.0750269
2020	20.1	20	0.1	0.01	0.004975
2021	19.5	20	0.5	0.25	0.025641
			5.6	6.68	0.297799
				1.11333	
			0.933	1.0551	4.96

Table 3. ARIMA Model Accuracy Measurements

data set based on the years 2000 to 2015. Forecast test data values from 2016 to 2021 to measure the accuracy of the GRNN model as we applied it to the ARIMA model. The forecasted values from the trained set data are presented in Fig. 5(b) using the GRNN model.

Table 4 shows the accuracy measurements MAD=1.1816, RMSE=1.3372, and MAPE = 6.26% of the GRNNs model.

In this study, MAD, RMSE, and MAPE clarify the performance of ARIMA and GRNN models. In Figure 8, these metrics explain that ARIMA outperforms GRNN due to low values. The MAD of the GRNN model is higher than the ARIMA model. Similarly, RMSE and MAPE are also higher values for the GRNN model. The ARIMA model which was applied through the auto-arima command outclass the GRNN model with high accuracy. Therefore, the future death rate of diabetic patients from 2022 to 2030 is presented in Fig. 6(a). A comparison of accuracy measurements using ARIMA and GRNNs models is presented in Fig. 6(b).

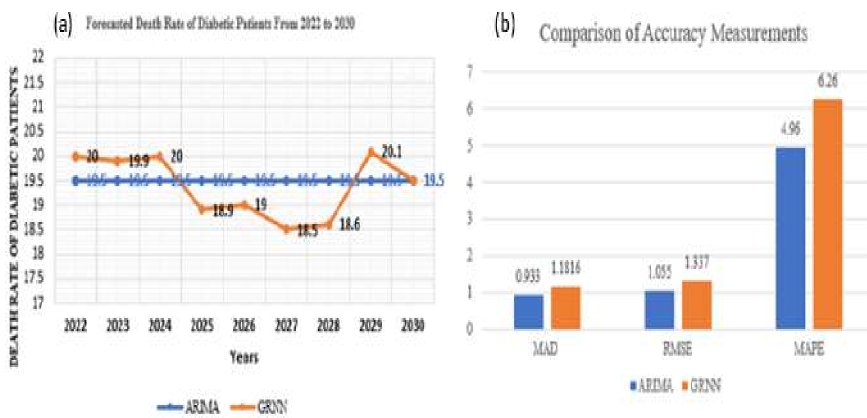

Figure 6. Forecasted Values Using GRNN Model

Years	Actual	Forecasted	MAD	RMSE	MAPE
2016	18.9	20.3	1.4	1.96	0.074074
2017	19	20.99	1.99	3.96	0.104737
2018	18.5	20.1	1.6	2.56	0.086486
2019	18.6	20	1.4	1.96	0.075269
2020	20.1	19.9	0.2	0.04	0.00995
2021	19.5	20	0.5	0.25	0.025641
			7.09	10.73	0.376157
			1.1816	1.78	6.26929
			1.1816	1.33	6.26

Table 4. GRNN Model Accuracy Measurements

5 Conclusion and Limitation

The comparison of ARIMA and GRNN models' accuracy measurements identified that ARIMA and GRNN both have approximately the same accuracy. ARIMA outclass the GRNN model with a slightly better accuracy result. However, the disadvantage of the ARIMA model is not fulfill the fundamental assumptions that are required for parametric time series model implementation. On the other hand, GRNN presents approximately similar accuracy and there is no need for fundamental assumptions of the parametric time series model. Therefore, we can say that GRNN is a better time series model to apply over short data as well as where fundamental assumptions are not fulfilled by parametric models. The limitation of this research paper is two or more time series data will be used for analysis. There are different variants of the GRNN model which would help increase accuracy.

Author Contributions

Muhammad Shahbaz Khan: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. **Mir Ghulam Hyder Talpur**: Visualization, Investigation and Supervision . **Muhammad Aslam**: Validation, Writing- Reviewing and Editing.

Author Contributions

Muhammad Shahbaz Khan: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. **Mir Ghulam Hyder Talpur**: Visualization, Investigation and Supervision . **Muhammad Aslam**: Validation, Writing- Reviewing and Editing.

Compliance with Ethical Standards

It is declared that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in this study.

Funding Information

No funding.

Author Information

ORCID:
Muhammad Shahbaz Khan: 0009-0009-1828-0462
Talpur Mir Ghulam Hyder: 0000-0002-5273-6963

References

[1] Aengchuan, P., Wiangkham, A., Klinkaew, N., Theinnoi, K. and Sukjit, E. [2022], 'Prediction of the influence of castor oil–ethanol–diesel blends on single-cylinder diesel engine characteristics using generalized regression neural networks (grnns)', *Energy Reports* **8**, 38–47.

[2] Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E. and Nortey, E. N. [2023], 'Time series based road traffic accidents forecasting via sarima and facebook prophet model with potential changepoints', *Heliyon* **9**(12).

[3] Bai, Z., Choi, K. P., Fujikoshi, Y. and Hu, J. [2022], 'Asymptotics of aic, bic and c p model selection rules in high-dimensional regression', *Bernoulli* **28**(4), 2375–2403.

[4] Deja, R. and Froelich, W. [2022], 'Forecasting basal insulin for the clinical therapy of juvenile diabetes at onset', *Procedia Computer Science* **207**, 138–144.

[5] Della Cioppa, A., De Falco, I., Koutny, T., Scafuri, U., Ubl, M. and Tarantino, E. [2023], 'Reducing high-risk glucose forecasting errors by evolving interpretable models for type 1 diabetes', *Applied Soft Computing* **134**, 110012.

[6] Fitzgerald, O., Perez-Concha, O., Gallego-Luxan, B., Metke-Jimenez, A., Rudd, L. and Jorm, L. [2023], 'Continuous time recurrent neural networks: overview and benchmarking at forecasting blood glucose in the intensive care unit', *Journal of Biomedical Informatics* p. 104498.

[7] Kumar, S. K., Reddy, R. and Patil, V. [2022], 'Prediction of covid-19 very critically ill (equivalent to deaths of the patients) in the united states of america using arima model in r programming.', *International Journal of Early Childhood Special Education* **14**(6).

[8] Lukman, A. F., Benedicta, A., Awotunde, J. B., Okon, C. E., Oludoun, O., Oluwakemi, A., Ayinde, O. E., Alabi, O. O. and Adeniyi, A. E. [2022], 'Robust statistical modeling of covid-19 prevalence in african epicentres", *Modeling, Control and Drug Development for COVID-19 Outbreak Prevention* pp. 315–358.

[9] Marri, A. A., Ghulam, M. and Talpur, H. [2022], 'Evaluation of stochastic and ann model for karachi stock exchange prices pre-diction', *Int. Trans. J. Eng. Manag. Appl. Sci. Technol* **13**, 1–11.

[10] Martínez, F., Charte, F., Frías, M. P. and Martínez-Rodríguez, A. M. [2022], 'Strategies for time series forecasting with generalized regression neural networks', *Neurocomputing* **491**, 509–521.

[11] Mills, T. [2019], *Applied time series analysis: A practical guide to modeling and forecasting*, Academic Press.

[12] Organization, . W. H. [2023], 'Diabetes'.
URL: *Diabetes (who.int)*

[13] Ray, S., Lama, A., Mishra, P., Biswas, T., Das, S. S. and Gurung, B. [2023], 'An arima-lstm model for predicting volatile agricultural price series with random forest technique', *Applied Soft Computing* **149**, 110939.

[14] Rodríguez-Rodríguez, I., Campo-Valera, M. and Rodríguez, J.-V. [2023], 'Forecasting glycaemia for type 1 diabetes mellitus patients by means of iomt devices', *Internet of Things* **24**, 100945.

[15] Shahin, O. R., Alshammari, H. H., Alzahrani, A. A., Alkhiri, H. and Taloba, A. I. [2023], 'A robust deep neural network framework for the detection of diabetes', *Alexandria Engineering Journal* **74**, 715–724.

[16] Statista [2023], 'Death rate for diabetes mellitus in canada from 2000 to 2022'.
URL: *Diabetes death rate Canada 2000-2022 | Statista*

[17] Zeinalnezhad, M. and Shishehchi, S. [2024], 'An integrated data mining algorithms and meta-heuristic technique to predict the readmission risk of diabetic patients', *Healthcare Analytics* **5**, 100292.