Investigation of Influential Factors towards Predicting Death Rate in Bangladesh

Md. Nazmus Salehin, Mst. Irin Sultana

Dept. of Computer Science and Engineering, Bangladesh Army University of Engineering & Technology, Natore, Bangladesh

***Abstract*: the concept of death rate is very momentous for development planning in a country. Bangladesh has largely extended its demographic transition from Death Rate (DR). Death Rate observes the economic growth and the standard of living in a country and sometimes controls the country classification defined by the World Economic Situation and Prospects (WESP). This paper looks into numerous social and economic factors responsible for determining the DR in Bangladesh and builds a prediction model that discloses many unexplored and useful observations. The World Bank data repository, based on the World Development Indicator (WDI) - 2015, has been inquired for the investigation and model building process. Primarily, a multiple stepwise linear regression-based method has been applied to build various models, and their performances are measured using MAE, RMSE, and RAE under the shed of k-fold cross-validation. Experimental results show that the final identified model forecasts the DR convincingly.**

***Keywords:* Death Rate, Linear Regression, Risk Factors, Prediction Model, K- cross Validation.**

I. INTRODUCTION

Bangladesh officially named the People's Republic of Bangladesh, is a country situated in South Asia. It is the eighth-most crowed country in the world, with a population almost 163 million people, in an area of 148,560 square kilometers (57,360 sq mi), making it one of the most deeply populated countries in the world [1]. In Bangladesh, About 45% of rural household lie below the poverty line [2]. Despite its widespread poverty and occasionally extreme climate events, the country has made remarkable progress in the last four decades in providing vaccines to children and mothers; reducing vitamin A deficiency as well as mortality of infants, children and mothers; increasing life expectancy; and reducing gender and economic inequalities in health outcomes [3].

Crude death rate introduces the number of deaths occurring during the year, per 1,000 population calculated at midyear. Subtracting the crude death rate from the crude birth rate provides the rate of natural increase, which is equal to the rate of population change in the absence of migration [4]. However, it is clear to us that mortality has already decreased than the past as well as achieved remarkable reductions in maternal mortality due to take several programs to reduce mortality. Several government and non-government organization works continuously about creating awareness among the women, children and also the overall people in Bangladesh. People are now conscious about their health, nutrition and diseases thus reduces mortality rate [5, 6]. Mortality rates and ratios are important indicators reflecting the health situation of the population of a country. The indicator is significantly affected by age distribution and most countries will eventually show an emergence in the overall death rate, in spite of continued degeneration in mortality at all ages, as declining fertility results in an aging population. Crude death rates were higher in the primitive stage but in recent years it is declined due to lack of several programs. Now people know that how to lead a better way of life, how to control all complex situations and also how to implement the policies to reduce death rates [7].

It can be shown from above literatures that there has been a lot of research including death rate decline, but there is not one single, most significant factor that can help people to realize about the argument of death. It has been the findings of a remarkable change in social and economic conditions as well as in values. Much has been written on this topic and depending on the disciplinary background, economic, biological or social dimensions are emphasized. These studies have brought many useful insights into the factors associated with declining death rate.

II. RELATED WORK

Death rates and ratios are important indicators mirror the health situation of the population of a country. Crude Death Rate is the total number of deaths to residents in a specified geographic area such as country, state, county, etc. divided by the total population for the same geographic area multiplied by 100, 000 [3].

The higher mortality rate in the age groups 15-19 and 20-24 for the women may be due to early marriage and maternal mortality old age complication is the principal cause of death as well as asthma and respiratory problem, blood pressure, heart disease are also considerable factors for mortality rate. In case of men, asthma respiratory disease is the principal cause of death [7]. Neurological diseases contributed to 33.0% mortality followed by cardiac 14.3% and renal diseases 13.4% [8]. . The main reason of anesthesia related deaths are attributed to mal-management cardiovascular problems, respiratory events related to anesthesia and drug administration [9]. Crude death rate was a little higher in the government service area in Bangladesh, mainly due to the higher death rates in the neonatal and post-neonatal periods. Mortality was higher for males than females in the neonatal period [10].

In developed countries, mortality rates raises during uphill cycles in the economy, and reduce during downward cycles.

Overall, every 1% increase in Gross Domestic Product (GDP), the death rate for men 70 to 74 years of age increased. On the other hand, the effect is similar for women. The mortality rate for women aged 70 to 74 increased by 0.18% for every 1% increase in Gross Domestic Product and 0.15% among middle-aged women [11].

One study has been carried out in the State of Alagoas, Brazil that measures the effect of prime sanitation services on the mortality rate. The estimates show that the 10% increase in access to household sanitary sewage is related to a reduction of 5.7 deaths per 1000 born alive [12]. The improved water source, improved sanitation facilities and prevalence of undernourishment have inversed association with maternal mortality [13]. Life expectancy at birth reflects the overall mortality level of a population and it summarizes the mortality pattern that prevails across all age groups [14].

III. DATA ANALYSIS STEPS

The proposed system utilizes to death rates to identify risk factors and generate death rate prediction models in Bangladesh. Our proposed architecture consists of background study, data collection and preparation, analyzing data for building models and model validation. The architecture of our suggested system is depicted in Figure 1. To give you a better idea of the system, we'll go over each stage of the suggested design.

A. Background Study

We have searched Elsevier, IEEE Xplore, Google Scholar, Wiley Online Library, ACM Digital Library, and Springer for a substantial percentage of death rate research articles. Following that, around 50 numbers of research papers are identified based on the title, abstract, and keywords of the papers being reviewed. Finally, we have selected best 21 numbers of papers for the study which are given as reference. Then we focus on the remaining studies to find various factors that have a substantial impact on death rate.

B. Data Collectiona and Preparation

This study makes use of the World Bank's "World Development Indicator, 2015 (WDI- 2015)" Bangladesh dataset, which collects data from officially recognized international sources [4]. It shows the most up-to-date and accurate global development data available, with national, regional, and worldwide estimations. Based on the existence of the relevant data to death rate, we retrieved appropriate data for this study. The data in the collection spans 55 years, from 1960 to 2015. We ready up our data set and explore the relevant factors from the dataset. The dataset is consisted by 1350 indicators. All indicators related to economic, political, socio-cultural, demographic, and health systems with direct and indirect effects on birth rate and we separate 400 indicators by our deep studies. Then we pick out 100 indicators that contain available data from 1990 to 2013. There are several indicators having missing values that have been avoided for this study. It will be considered in future by filling the missing values.

C. Applying ANOVA Test

At this point, we perform an Analysis of Variance (ANOVA) test on each and every indicators with a F statistic measurement to determine the significance of the risk factors for death rate in Bangladesh. The relevant elements that have been found will be used to develop our prediction model. The significance test in this study is based on a p-value of less than 0.05.

D. Building Regression Model

At this point, we plan to employ linear regression, a frequently used approach utilizing all of the important and relevant factors determined by ANOVA testing. As a consequence, a regression model is created to elucidate the relationship between variables, and it is often used to predict the value of a dependent variable (y) using the independent variable (x). A regression equation of the from is shown in equation (1).

$$y = \varepsilon + x_1 \beta_1 + x_2 \beta_2 + \cdots + x_k \beta_k \qquad (1)$$

where y is the dependent variable, ε is an unobservable random variable, $x = [x_1, x_2, ..., x_k]$ are independent variables, k is the number of observations and $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of that independent variables.

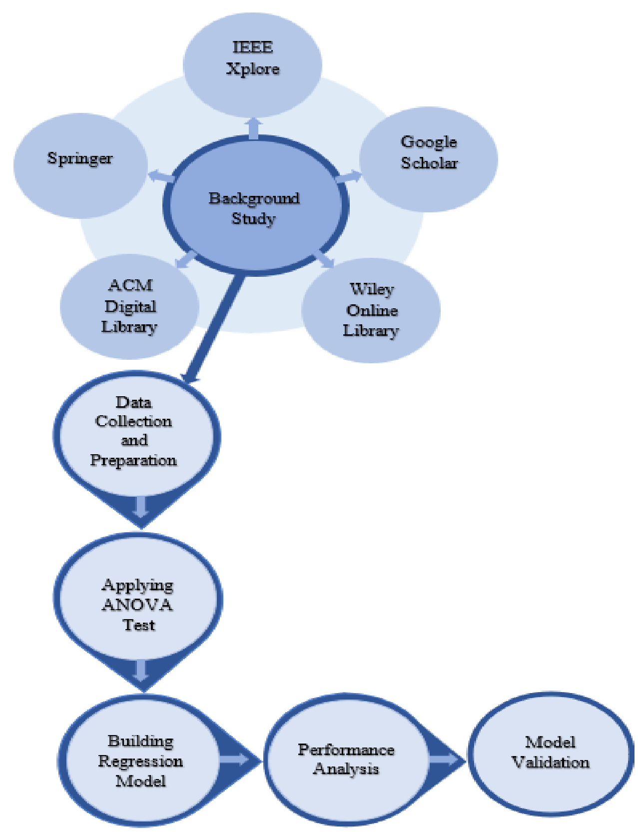

Fig. 1 Proposed Methodology.

A stepwise multiple linear regression analysis with death rate as the dependent variable and the recovered macro-structural components as predictors is planned to evaluate the connections between death rate and the retrieved macro-structural factors [15]. When a direct linear regression with all significant factors does not fit the dependent variables well, this technique is effective. Furthermore, because it is an exploratory study, there is no precise hypothesis concerning the order of the variables in terms of their likely causal connection. As a result, for adding variables in multiple regression models, the stepwise technique is used.

E. Performance Analysis

Performance evaluation is a difficult process. We define a set of commonly used performance assessment criteria to test and evaluate model performance for this purpose. The following is an explanation of the performance measuring process:

1) Model Validation: Using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Relative Absolute Error (RAE) benchmarks, we hope to identify various faults in the investigated dataset and choose the optimal model using the k-fold cross-validation approach [16]. We choose the two top models from the 4-fold cross-validation for comparison. Formulas for MAE, RMSE, and RAE are presented in (2), (3) and (4) respectively.

$$MAE = \sum (| y_a - y_p|)/ n \qquad (2)$$

$$RAE = \sum (|y_a - y_p|)/\sum(|y_a-a|) \qquad (3)$$

$$RMSE = \sqrt{(\sum(| y_a - y_p|)^2/ n)} \qquad (4)$$

Here, the actual value from the data set is represented by y_a and the predicted value created from the model is represented by y_p and $a = (\sum ya/n)$. We also look at the R-squared and Adjusted R-squared values to see how close the data is to the fitted regression model.

IV. IMPLEMENTATION

The proposed system utilizes to death rates to identify risk factors

A. Data Arrangement and Recognizing Significant Factors:

Following by our proposed methodology, firstly we ready up our data set and explore the relevant factors from the dataset. The dataset is consisted by 1350 indicators and we separate 400 indicators by our deep studies. Then we pick out 100 indicators that contain available data from 1990 to 2013. From the dataset for the period of 1990-2013, we perform the ANOVA test to find out significant and relevant factors. The significance tests were applied at three levels of p-values, that is, “p < 0.05”, “p < 0.01” and “p < 0.001”. As a consequence, 50 indicators are found not significant. After the significance testing, we observed 20 explanatory variables that have significance on death rate which is an immensely high number of variables for building an appropriate linear regression model. Then we conduct stepwise regression by considering two or more factors analyzed together to check whether they become significant altogether or not. From Fig.2 to Fig.6, the histograms of these five components are represented.

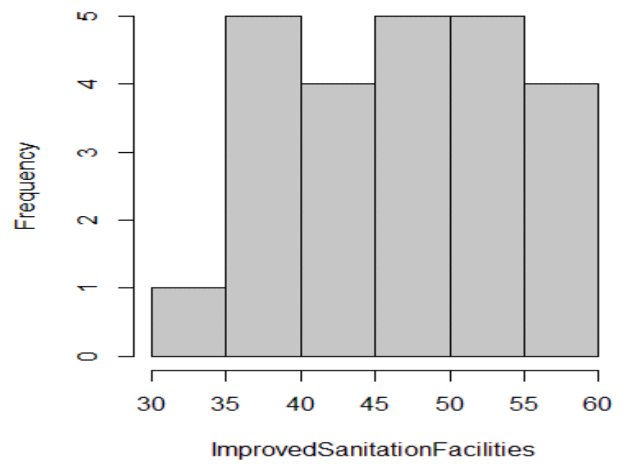

Fig. 2 Histogram of Improved Sanitation Facilities.

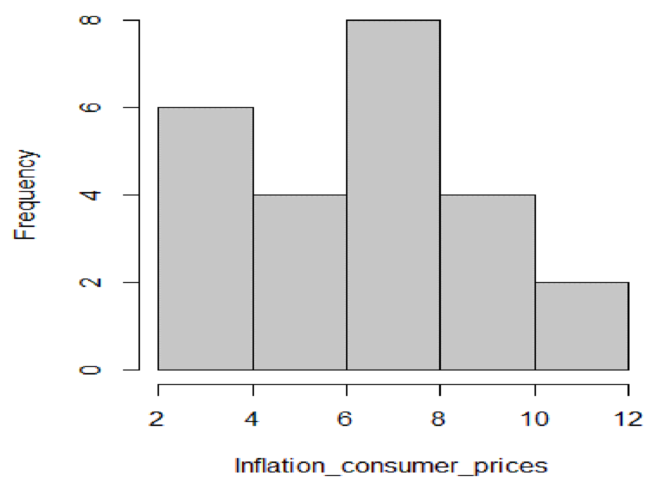

Fig. 3 Histogram of Inflation_consumer_prices.

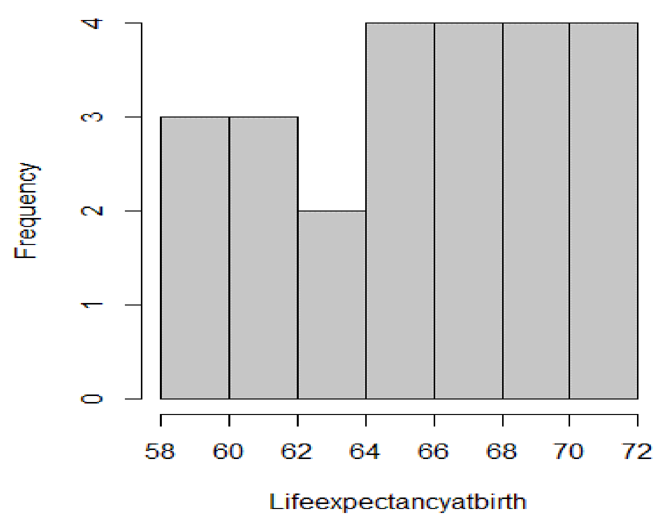

Fig. 4 Histogram of Life expectancy at birth.

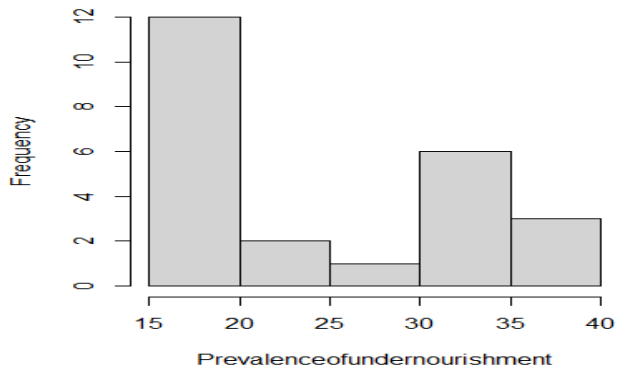

Fig. 5 Histogram of Prevalence of undernourishment.

B. Building Prediction Models :

From the stepwise regression analysis, we attain numerous amounts of model; among them one model consisting five risk factors: The risk factors are Gross Domestic Product (GDP), Improved sanitation facilities (ISF), Prevalence of undernourishment (PUN), Life expectancy at birth (LEB) and Inflation Consumer Prices (ICP). TABLE I shows a set of coefficients determined from the regression model and the regression model itself is presented in (5).

*Death Rate = ((5.003e + 01) + (GDP * 3.038e-13) + (ISF * 1.220e-01) - (PUN * 2.601e-02) - (LEB * 7.516e-01) + (ICP * 8.048e-03))* (5)

TABLE I.
PREDICTION MODEL PARAMETERS AND SUMMARY

No.	Factors Summary		
	Indicators	Coefficients	p-values
1	GDP	0.0000000000003038	0.00126
2	Improved Sanitation Facilities (ISF)	0.122	0.03795
3	Inflation Consumer Prices (ICP)	0.008048	0.04343
4	Life expectancy at birth (LEB)	-0.7516	0.00000000467
5	Prevalence of undernourishment (PUN)	-0.02601	0.00000000304
6	Intercept	50.03	0.000000000000107
7	R-squared	0.9995	
8	Adjusted R-squared	0.9994	

This multiple linear regression-based model exploits five important and relevant factors to death rate where the ultimate model significance is observed with p-value < 0.05 and adjusted-R2 is equal to 0.9995. We plot histograms of datasets of model variables, observe zero lines of residuals versus fitted graph (Fig. 7) and the normal Q-Q graph depicted in Fig. 8 which prove that the model is quite linear with negligible errors. Therefore, we can accept that the sample mean of the fitted model has a normal distribution and our dataset does not have many outliers.

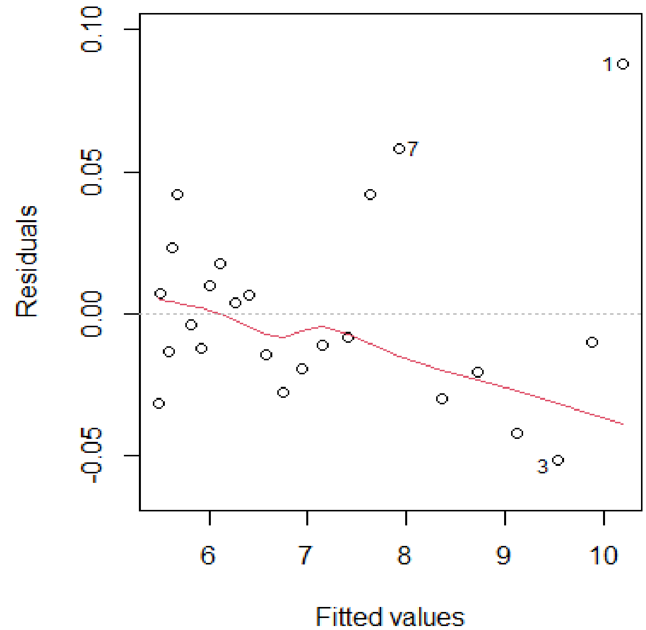

Fig. 6 Residual vs Fitted.

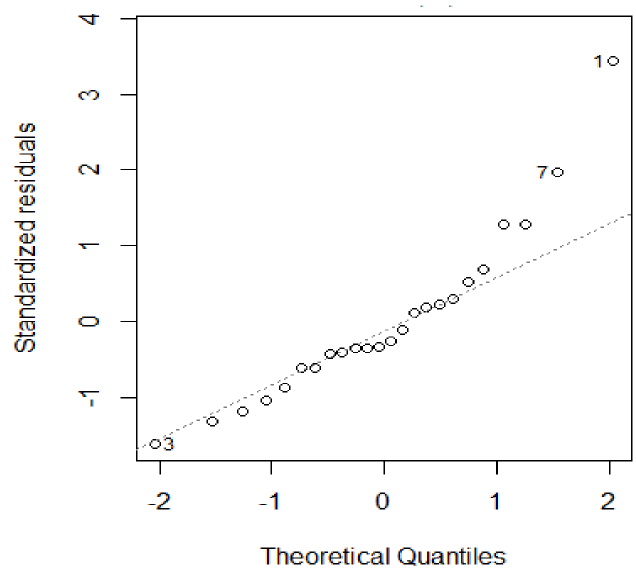

Fig. 7 Normal Q-Q.

V. RESULT AND DISCUSSION

By our acute investigation, we finalized a list of influential risk factors responsible for finding the root and Rate of death in Bangladesh by analyzing the dataset from 1990 to 2013. The stepwise linear regression model in TABLE I exhibits five factors (this five factors are called independent variables for our study) and the consequence of R-squared statistics expose that the independent factors of the model are good predictors for Death Rate (it is considered as outcome or dependent variable for our study) and also reveal that there is a high correlation among outcome variable and independent variables

(Adjusted R squared = .9995). For validating the model, we apply k-fold cross-validation to check whether any subset has any bias effect on the final model. Our target to build regression models using the training sets and validate them with the main or base model found in (5). The K-fold validation process also provides knowledge of any outliers among observations. In this analysis of k-fold cross-validation, we use k = 4 that regenerates four regression models using training sets and test model parameters using testing sets. Each model yields a similar outline as of TABLE I. In Table II, fold 1 illustrate the training set of 18 data in the span of 1996-2013 and the testing set contains 6 data from 1990- 1995. Other folds display the same pattern.

TABLE II.
SAMPLE DIVISION OF K-FOLD CROSS VALIDATION

Cases	K-fold Distribution Summary	
	Training Set (Years)	Testing Set (Years)
1	1996-2013	1990-1995
2	1990-1995, 2002-2013	1996-2001
3	1990-2001, 2008-2013	2002-2007
4	1990-2007	2008-2013

According to the performance measurement criteria in our proposed system, we obtain the error measurements of MAE, RMSE, and RAE for 4 models and compared them with the base model in TABLE III. It is noted that the model 2c from these 4-fold cross-validation models are quite near to the base model with minimal errors which is interesting to explore further. To investigate this model further we attempt to apply the coefficients from these models to (5) where the whole dataset of 24 data is taken into account. We name them as 'Best'. From the analytical outcome, we see that the overall performance of the best model has been improved which concludes that the overall dataset has a better impact on the regression models than the cross-validation datasets. However, the improvement in the accuracy of this model has not suppressed the performance of the base model. Therefore, the base model remains the superior model.

TABLE III.
PERFORMANCE COMPARISON OF BASE MODEL WITH MODELS FROM 4-FOLD CROSS VALIDATION

No.	Model Summary			
	Name of Models	Mean Absolute Error (MAE)	Relative Absolute Error (RAE)	Root Mean Square Error (RMSE)
1	Base	0.024488227	0.019740389	0.032033684
2	4 Models from 4-Fold Cross Validation	0.159307166	0.271007371	0.215055245
		4.898752057	12.92546717	5.584424286
		0.011458761	0.0586626	0.012464251
		0.07418106	0.698722698	0.094428601
3	Best	0.026849691	0.021644007	0.03267528

We also analyze box plots of the base model errors: RAE and RMSE are displayed in Fig. 9 and Fig. 10 respectively to illustrate the performance graphically. It can be seen that the mean values of the errors are well under the standard error values of 0.05.

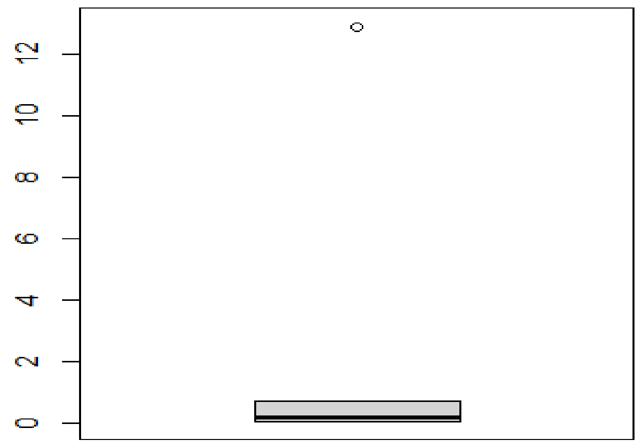

Fig. 8 Boxplot of RAE.

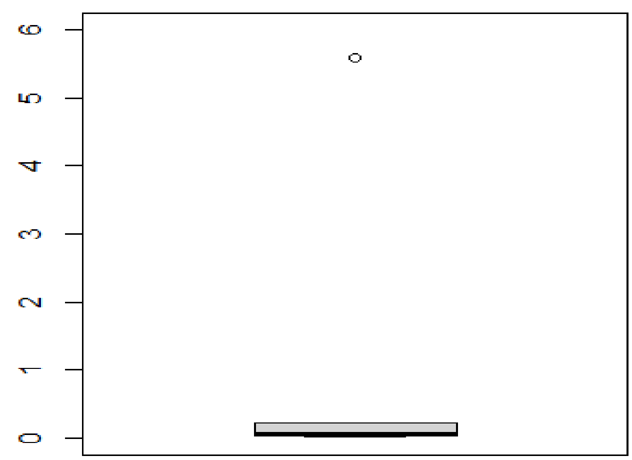

Fig. 9 Boxplot of RMSE.

In summary, the base model built with a stepwise linear regression model yields by far a better and significant model using five factors as model parameters. The risk factors are GDP, Improved sanitation facilities (% of population with access), Prevalence of undernourishment (% of population), Life expectancy at birth, total (years) and Inflation, consumer prices (annual %).

It has been shown in TABLE I that a strong negative and significant effect of mortality rate on per-capita GDP growth when socioeconomic development in a country are getting better, mortality rate is decreasing; accordingly [18].There are strong co relation between health and income. Female mortality is more sensitive than male mortality to negative economic shocks. However, economic shocks in the developing world generally lead to more infant deaths, especially of girls [19].

On the other hand, unimproved water and sanitation significantly increased the risk of post-neonatal and child mortality. Unimproved water cause diarrheal diseases, which has result in high mortality [20].

Again, a strong and consistent relation was observed between nutritional status and death. Undernourishment is the underlying cause for most deaths associated with severe infections [21]. Finally, our studies also show that Life expectancy at birth, total (years) and Inflation, consumer prices (annual %) have significant effect on death rate. In conclusion, the identified factors are found useful to uncover some potential facts relevant to Death Rate in Bangladesh. The final model can potentially be used in determining the Death Rate in upcoming years where the model parameters have their usual values.

VI. CONCLUSION

Bangladesh is one of the developing countries in the world in recent times as well as Death Rate is also momentous indicator that needs close attention. This study is an attempt to identify such influential factors and has measured the effect of those factors for measuring Death Rate in Bangladesh. It was planned to identify the factors which significantly influence the Death Rate in Bangladesh and build a prediction model that can determine the Death Rate in upcoming years. We proposed multiple stepwise regression analysis that, among 20 significant factors, revealed only 5 factors that are highly correlated to Death Rate. Our standard cross-validation process and the performance measurement criteria confirm that the built model has been the best model. In WDI-2015 dataset, for the period of 1990-2013, there were several missing values that have been avoided for Death Rate. It will be considered in future by filling the missing values. The analysis would be additionally examined by filling the missing values from 1960-2015 for the factors identified. Overall, the proposed model has a valuable precision on the informational index accessible.

Finally, we have identified five risk factors that are related with Date rate in Bangladesh as well as we have build a predictive model using the identifying risk factors that will help the policy makers to predict the future Date rate in Bangladesh.

REFERENCES

[1] Wikipedia, "Bangladesh", https://en.wikipedia.org/wiki/Bangladesh, 26 September 2021.

[2] M. A. Kabir, M. F. Begum, M. A. Hossain, "Rural Poverty in Bangladesh and its Associated Factors: Evidence from Micro Level Study", in The Jahangirnagar Economic Review, Vol. 17 No. 1 June 2006.

[3] P. Das, R. Horton, "Bangladesh: Innovating For Health" in The Lancet, Volume 382, Issue 9906, P1681-1682, November 23, 2013.

[4] World Development Indicator, World Health Organozation, https://data.worldbank.org/country/bang ladesh, 29 December 2015.

[5] Bangladesh Country Report, General Economics Division (GED), Bangladesh Planning Commission, Government of the People's Republic of Bangladesh, 2013

[6] Bangladesh Demographic And Health Survey "National Institute Of Population Research And Training Ministry Of Health And Family Welfare", Dhaka, Bangladesh, 2014.

[7] M. P. Moon, M. Kamruzzaman, "Mortality Experience Trends in Bangladesh: Present and Past Facts", in Humanity & Social Sciences Journal 13 (2), December 2018

[8] A. Tasie and T. Getinet, "Assessing Pediatric Index Of Mortality 2 Score And Factors Affecting Outcome Of Patients Admitted To Pediatric Intensive Care Unit In St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia, 2016 To 2019", in 32nd World Pediatrics Conference, Neonat Pediatr Med 2019, Volume: 05

[9] H. S. M. Abdalaziz1 , D. E. M. Essayed and E. F. M. Khalid, "Estimation of Anaesthesia Related Deaths in Elective Surgical Operations in Khartoum, Khartoum North and Omdurman Teaching Hospitals from October 2011 to October 2012", in Community Med Health Educ, 2014.

[10] N. Alam, H. R. Chowdhury, A. Ahmed, M. Rahman and P. K. Streatfield, "Distribution Of Cause Of Death In Rural Bangladesh During 20032010: Evidence From Two Rural Areas Within Matlab Health And Demographic Surveillance Site", in Global Health Action, 2014.

[11] R. Morin, "A Downside To An Up Economy? Mortality Rates Increase In Better Times", in Pew Research Center, October 16, 2013

[12] A. Cavalcanti, A. Teixeira and K. Pontes, "Regression Model to Evaluate the Impact of Basic Sanitation Services in Households and Schools on Child Mortality in the Municipalities of the State of Alagoas, Brazil", Sustainability, Received: 26 June 2019; Accepted: 30 July 2019; Published: 1 August 2019.

[13] M. I. Sultana, M. L. S. Lovely, M. M. Hasan, "Building Prediction Models for Maternal Mortality Rate in Bangladesh", in Proceedings of the 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 26-28 September, 2019, Dhaka, Bangladesh.

[14] C. J. Murray, B. D. Ferguson, A. D. Lopez, M. Guillot, J. A. Salomon, O. Ahmad, "Modified Logit Life Table System: Principles, Empirical Validation And Application", in Evidence and Information for Policy (EIP), World Health Organization, 2003.

[15] D. C. Montgomery, E. A. Peck, and G. G. Vining, "Introduction to linear regression analysis", in John Wiley & Sons , vol. 821, April 2012.

[16] J. Brownlee, "A Gentle Introduction to k-fold Cross-Validation", in Statistical Methods, May 2018.

[17] Garner and R. Stephen, "Weka: The waikato environment for knowledge analysis", in Proceedings of the New Zealand computer science research students conference, pp. 57-64, April 1995.

[18] G. Miladinov "Socioeconomic Development And Life Expectancy Relationship: Evidence From The EU Accession Candidate Countries" in Springer Open, 10 January 2020.

[19] S. Baird, J. Friedman, and N. Schady, "Aggregate Income Shocks And Infant Mortality In The Developing World", in The Review of Economics and Statistics, August 2011, 93(3): 847–856.

[20] O. K. Ezeh, K. E. Agho, M. J. Dibley, J. Hall, and A. N. Page , "The Impact of Water and Sanitation on Childhood Mortality in Nigeria: Evidence from Demographic and Health Surveys, 2003–2013", in International Journal of Environment Research and Public Health, September 2014.

[21] L. E. Caulfield, M. Onis, M. Blössner, R. E. Black, "Undernutrition As An Underlying Cause Of Child Deaths Associated With Diarrhea, Pneumonia, Malaria, And Measles", in The American Journal of Clinical Nutrition, Volume 80, Issue 1, July 2004, Pages 193–198.