South East Asian J. of Math. & Math. Sci. **ISSN : 0972-7752**
Vol.12, No.2 2016, pp. 147-160

WORLD DEVELOPMENT INDICATORS ANALYTICS FOR SOUTH ASIAN ASSOCIATION FOR REGIONAL COOPERATION COUNTRIES

Kavita Pabreja

Associate Professor and Head of Department of Computer Science,
Maharaja Surajmal Institute, New Delhi, India
E-mail: kavita_pabreja@rediffmail.com

Dedicated to Prof. A.M. Mathai on his 80^{th} ***birth anniversary***

Abstract: Data is an asset and offers tremendous opportunities for enabling innovation by observing previously unobserved patterns. Gross Domestic Product growth is considered as one of the most important parameters to understand the economical position of a nation. The economic health of a country depends upon many factors viz. consumption, business investment, government expenditure and net exports. In this study, an attempt has been made to perform analytics on education, health, environment and economic parameters of South Asian Association for Regional Cooperation countries. The datasets under study have been downloaded from The World Bank website and countries under study are Bangladesh, Bhutan, India, Pakistan and Srilanka. Latest analytics tools like R programming language and matlab have been used to give insight into the important world development indicators and finally the Gross Domestic Product of SAARC nations have been forecasted based on education, health and carbon dioxide emissions with a convincing mean square error of 0.037, using artificial neural networks.

Key words and Phrases: Data analytics, health, education, economy, artificial neural networks, world development indicators.

1. Introduction

There is a huge repository of world development indicators available on the World Bank website for over 200 countries worldwide. The South Asian Association for Regional Cooperation (SAARC) is regional intergovernmental organization and geopolitical union in South Asia. Its member states include Afghanistan, Bangladesh, Bhutan, India, Nepal, the Maldives, Pakistan and Sri Lanka. SAARC

compromises 3% of the world's area, 21% of the world's population and 9.12% of the global economy [1], as of 2015. This study focuses on five of the SAARC countries viz. Bangladesh, Bhutan, India, Pakistan and Srilanka. The study throws insights into the most important development indicators viz. health, education and economy of the nations.

The economic development of any nation is closely associated with the increase in the utilization and burning of fossil fuels, coal, oil, and natural gas by factories and electric power plants, motor vehicles, and family units. The consequential carbon dioxide (CO_2) emissions have turned into the largest source of greenhouse gases that do not allow the infrared radiation from the earth to leave the atmosphere and create the risk of global warming [2]. It has been suggested that every country must put in efforts to reduce the CO_2 emissions for the sake of its citizens.

These emissions affect the agriculture directly that in turn brings adverse effects on the economy of a country. As concluded by Smith [3], a small increase in global average temperature (up to 2° C, measured against 1990 levels) would result in net negative market sector impacts in many developing countries and net positive market sector impacts in many developed countries. The author also mentioned that climate change would increase income inequalities between and within countries.

Health of people of a country plays a significant role in its growth and development. The importance of Adult Survival Rate on growth rates for poor countries has been discussed by [4]. The authors discussed that other measures of health such as disease prevalence rates and cognitive functioning are imperative for maintaining a balanced supply of skilled labour which is an important component in growth of economy.

The contribution of health to economic growth has been examined in [5] also. The authors have concluded that good health has a positive, sizable, and numerically significant impact on rates of economic growth.

Education is one of the most dominant mechanism for decreasing poverty and inequality and lays a basis for continual economic growth. The effect of education on GDP has been studied by [6]. The authors argued that the skills available in the labour force and the value of those skills are important determinants of economic performance of any country. There is requirement of workers with higher levels of education in order to handle complex services and production systems [7].

In the past, different Soft Computing techniques have been applied by many researchers for forecasting GDP of a nation. Neural networks have been applied for forecasting of macro-economical variables and a comparison of different linear and non-linear models has been done [8]. The authors have found that multivariate linear models are better. Artificial Neural Networks (ANN) and Autoregressive

Integrated Moving Average (ARIMA) have been used for anticipating the future of Egypts cereal and it was concluded that ANN performed better than ARIMA [9]. GDP of Britain has been forecasted by authors using ANN and they have compared two different training algorithms [10]. The forecasting ability of ANN to anticipate financial output increase based on monetary and financial variables has been found to be better than linear models, as discussed in [11]. The GDP of Malaysia has been forecasted based on various economical indicators [12]. The authors have also compared econometric approaches with ANN and have demonstrated that ANN has better performance. A combination of ANN and ARIMA has been also been experimented and in comparison with ANN or ARIMA, this hybrid model has produced more convincing results [13].

Most of these studies are based on either statistical relationship between different indicators or application of soft computing techniques using economical indicators as input variables. This paper differs from all of these studies as it is based not only on economical indicators but also on Education and Health statistics. Also this piece of research is not just for a single nation but for five countries of the SAARC segments. In this paper, we have analysed the most recent and precise global development datasets accessible at The World Banks official website and using strong correlations as the basis of selection of a small subset of features, we have forecasted the GDP of SAARC countries. Also, a few statistics on health, education and carbon dioxide emissions have been found using R programming language that throws light on minimum and maximum of these parameters over the years 1998-2013.

The paper is organized as follows:- section 2 explains Data and Methodology including data collection, countries under study, indicators used for the analysis and feature reduction. Section 3 explains the concept of Artificial Neural Network used for forecasting GDP, Results and discussions are elaborated in section 4 followed by conclusions as section 5.

2. Data and Methodology

Datasets utilized for the study have been downloaded from The World Bank website. Various steps followed viz. selection of countries, different indicators responsible for economic growth and development, pre-processing of datasets and finally reduction of datasets, have been explained in following sub-sections.

2.1 Data Collection: There is huge data available for download for 249 countries, for 1343 attributes for 55years for the time period 1960 till date [6]. The countries as per their development status have been categorized by The World Bank in four groups: High Income, Upper Middle Income, Lower Middle Income

and Low Income. The countries can also be categorized based on geographical region and here the datasets corresponding to SAARC nations has been captured.

2.2 Countries under study: Under SAARC countries; the data for five countries Bangladesh, Bhutan, India, Pakistan and Srilanka has been collected for the study.

2.3 Indicators under study: In this paper, data related to important determinants of GDP has been selected. The records corresponding to attributes describing the World Development Indicators; Education statistics; Health, environment and economy related statistics have been collected for the time period 1990-2013 [2] and have been explained below.

World Development Indicators

World Development Indicators (WDI) is the key World Bank repository of development indicators, compiled from officially recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates. The World Bank EdStats Query holds approximately 2,500 internationally comparable education indicators for contact purpose, evolution, conclusion, literacy, educators, population, and money spent by countries. The indicators elaborate the education cycle from pre-primary to tertiary education. The inquiry placed on huge repository of datasets also holds learning result data from international learning assessments (PISA, TIMSS, etc.), equity data from domestic surveys, and protuberance data till year 2050. Data about Key health, nutrition and population statistics gathered from a variety of international sources has been utilized for the study. The indicators considered under the study are as follows:

CO_2 emissions (metric tons per capita), GDP per capita (current US $), GDP per capita growth (annual %),Enrolment in primary education, both sexes (number),Enrolment in secondary education, both sexes (number), Government expenditure on education as % of GDP (%),Percentage of students in secondary vocational education who are female (%), Pupil-teacher ratio in primary education (headcount basis) ,Pupil-teacher ratio in secondary education (headcount basis), Health expenditure per capita (current US$), Health expenditure per capita, PPP (constant 2011 international$), Health expenditure, private (% of GDP), Health expenditure, public (% of GDP), Health expenditure, total (% of GDP), Out-of-pocket health expenditure (% of private expenditure on health), Out-of-pocket health expenditure (% of total expenditure on health), GDP growth (annual %).

2.4 Data Pre-processing and feature reduction: Data for the analysis is approximately 90% complete. The missing values have been filled up with estimated values using linear regression model.

Pearsons correlation coefficient has been calculated to find out the statistical relationships between two or more variables of health, climate change, education and other development related indicators so as to perform feature reduction. A correlation coefficient is a numerical measure of the degree to which changes to the value of one variable envisage change to the value of another. In positively correlated variables, the value increases or decreases accordingly. In negatively correlated variables, the value of one variable decreases as the value of the other increases. Out of the 57 attributes, there are many that are strongly correlated and there are certain findings that are quite unexpected and are discussed in Results and Discussions section. Further only following 8 attributes, as given in Table I, have been selected based on correlation coefficient values, for anticipation of GDP corresponding to SAARC countries using Artificial Neural Networks.

Table I List of 8 selected indicators for forecasting GDP (current US $)

Country name	Percentage of students in secondary vocational education who are female (%)
Year	Pupil-teacher ratio in primary education (headcount basis)
CO_2 emissions (metric tons per capita)	Pupil-teacher ratio in secondary education (headcount basis)
Government expenditure on education as % of GDP (%)	Health expenditure, total (% of GDP)

3 About Artificial Neural Network

An ANN is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and processes information using a connectionist approach to computation [14, 15].

In this paper, ANN has been implemented using Matlab 7.0 (matrix laboratory), designed and developed by Math Works Inc. A two layer MLP Back Propagation network has been used for the training and testing of artificial neural network. Tangent-sigmoid, transfer function is used in hidden layer of the back propagation network and pure linear transfer function is used in output layer. The Back propagation learning algorithms viz. trainrp has been used for prediction of percentage of GDP. The input dataset comprises of attributes selected based on correlation coefficient, mentioned in Table I. The output data corresponds to GDP of the

country under consideration. A sample of dataset is shown in Table II. Columns 1 to 8 of this table are used as input and column 9 corresponds to target. Before training, the range of inputs and outputs have been adjusted so that they fall in the range [-1,1]. A pre-defined function premnmx() provided by Matlab has been used for the purpose.

Table II Sample of dataset for forecasting GDP

Country name	Time	CO2 emissions (metric tons per capita)	Government expenditure on education as % of GDP (%)	Percentage of students in secondary vocational education who are female (%)	Pupil-teacher ratio in primary education (head-count basis)	Pupil-teacher ratio in secondary education (head-count basis)	Health expenditure, total (% of GDP)	GDP growth (annual %)
Bangladesh	1998	0.19	2.10	23.42	43.50	36.31	2.25	5.18
Bangladesh	1999	0.20	2.13	23.83	44.00	37.41	2.29	4.67
Bangladesh	2000	0.21	2.13	24.76	44.50	38.43	2.33	5.29
Bhutan	2010	0.68	4.02	37.00	25.90	21.40	5.17	11.73
Bhutan	2011	0.77	4.65	38.00	25.38	20.54	4.73	7.89
Bhutan	2012	0.86	5.00	39.00	23.99	19.89	3.70	5.07
India	2001	1.12	4.00	16.62	40.15	33.44	4.50	4.82
India	2002	1.13	3.80	17.44	40.66	32.32	4.40	3.80
India	2003	1.16	3.55	15.17	41.33	32.32	4.30	7.86
Pakistan	2010	0.96	2.29	41.36	40.45	21.60	3.02	1.61
Pakistan	2011	0.94	2.22	41.64	39.83	21.30	3.01	2.75
Pakistan	2012	0.93	2.14	42.76	41.35	21.04	2.76	3.51
Srilanka	1998	0.42	3.05	35.00	27.10	16.86	3.71	4.70
Srilanka	1999	0.47	2.90	35.80	26.80	16.89	3.65	4.30
Srilanka	2000	0.55	2.81	36.30	26.50	16.92	3.77	6.00

4. Results and Discussions

The Correlation coefficients between all 19 variables under the categories have been calculated and the variables having positive correlation of more than 0.7 have been selected. Since a change in the value of one variable will forecast a change in the same direction in the second variable, we have selected a subset of variables with 7 features to be used for forecasting GDP in case of SAARC countries. There are certain correlations that are implied but there are certain findings which are quite novel, unusual and interesting. Certain graphs for various parameters have been plotted as discussed here.

It has been observed that in India if females are given education, it will result in less expenditure on health. This is clearly visible from the bubble size in Fig.1 and Fig. 2, as they are inversely proportional. Also the percentage of female population in secondary vocational education is minimum in India in comparison with other SAARC countries.

It has been found that Bhutan has been spending the most on health of its people among the selected five SAARC nations, as is shown in Fig. 3 and Fig. 4. Even Srilanka has catched up over the years with Bhutan in terms of health expenditure per capita. Bangladesh has been spending the least on health services per capita and health expenditure in totality.

Annual GDP per capita growth was highest in Bhutan in year 2007 and otherwise in general GDP per capita growth on annual basis was least in all countries in 2008, as is evident from Fig. 5.

Government expenditure on education as percentage of GDP has been always highest in Bhutan among the five SAARC countries under study and Government of Srilanka has been spending the least on education over the years 1998 to 2013. This is depicted in Fig. 6.

India has been the most polluting countries among SAARC nations as far as carbon dioxide emissions per capita are concerned as shown in Fig. 7. Summary of these findings is shown in Table III.

Table III Summary of various parameters for SAARC countries during year 1998 - 2013

Parameter	Minimum value/ Country name/ year	Maximum value/ Country name/ year
Health expenditure per capita	8.63467 Bangladesh 1998	120.4155648 Srilanka 2013
Annual GDP per capita growth	-2.289173205 Srilanka 2001	15.40783 Bhutan 2007
Government expenditure on education as percentage of GDP	1.50000 Srilanka 2013	7.07810 Bhutan 2005
Carbon dioxide emissions, metrics tonnes per capita	0.1906021 Bangladesh 1998	1.6628735 India 2011

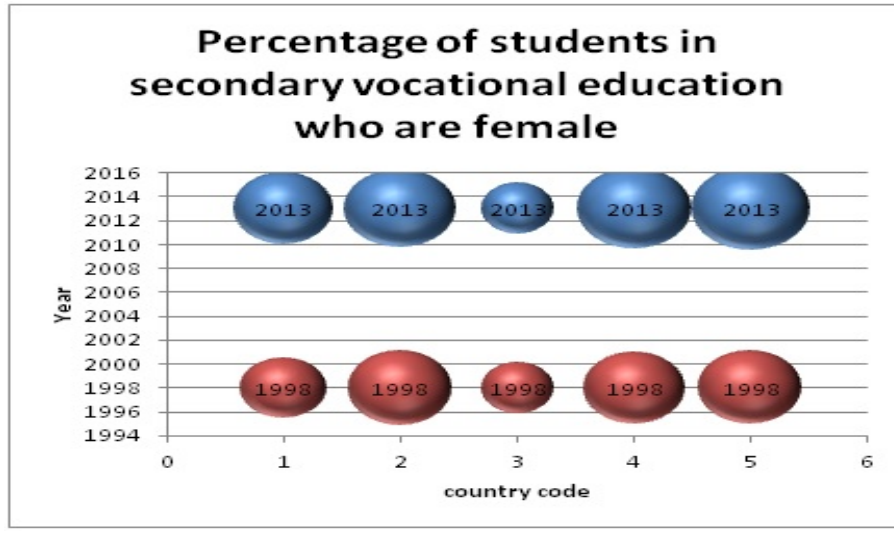

Fig. 1 Bubble chart for SAARC countries depicting percentage of students in secondary education who are females, for years 1998 and 2013(country code 1=Bangladesh, 2=Bhutan, 3=India, 4=Pakistan, 5=Srilanka)

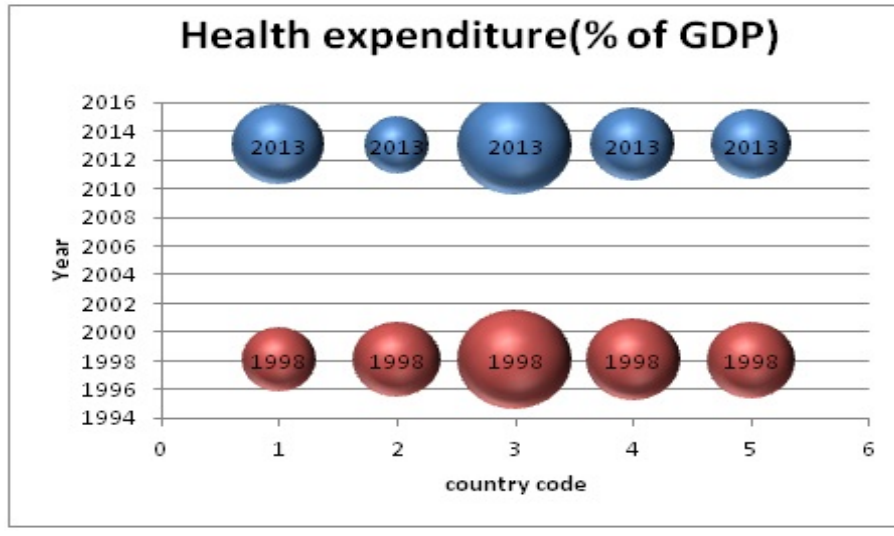

Fig. 2 Bubble chart for SAARC countries depicting Health expenditure as percentage of GDP, by SAARC countries, for years 1998 and 2013(country code 1=Bangladesh, 2=Bhutan, 3=India, 4=Pakistan, 5=Srilanka)

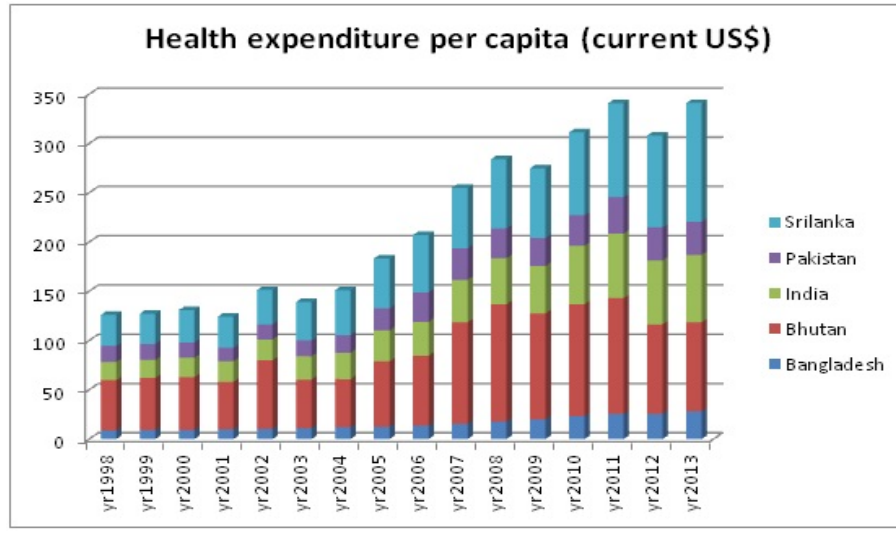

Fig. 3 Histogram depicting year-wise Health Expenditure per capita by SAARC countries

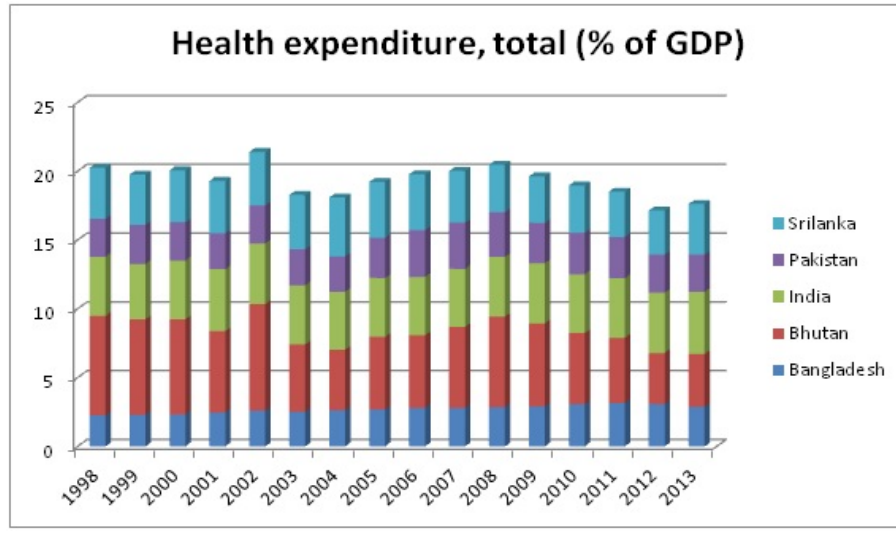

Fig. 4 Histogram depicting Total health expenditure as percentage of GDP, by SAARC countries

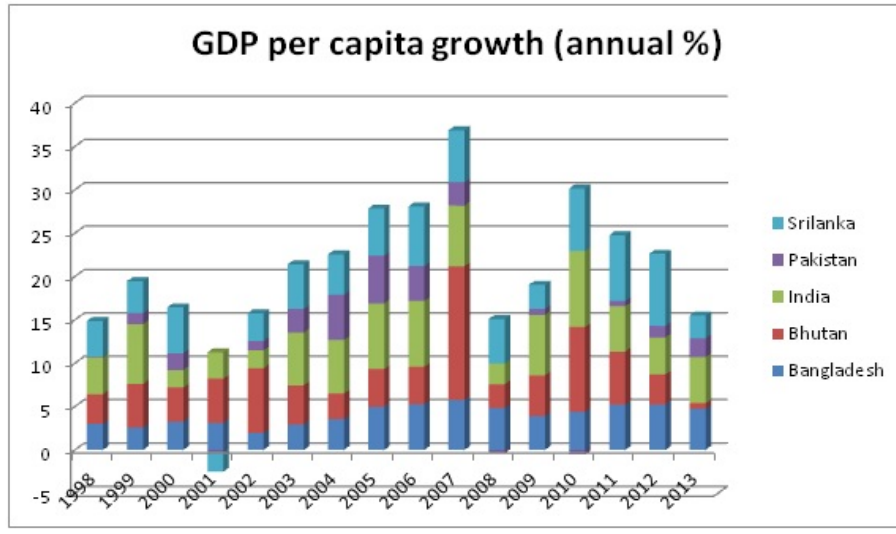

Fig. 5 Histogram depicting year-wise GDP per capita growth(annual percentage) by SAARC countries

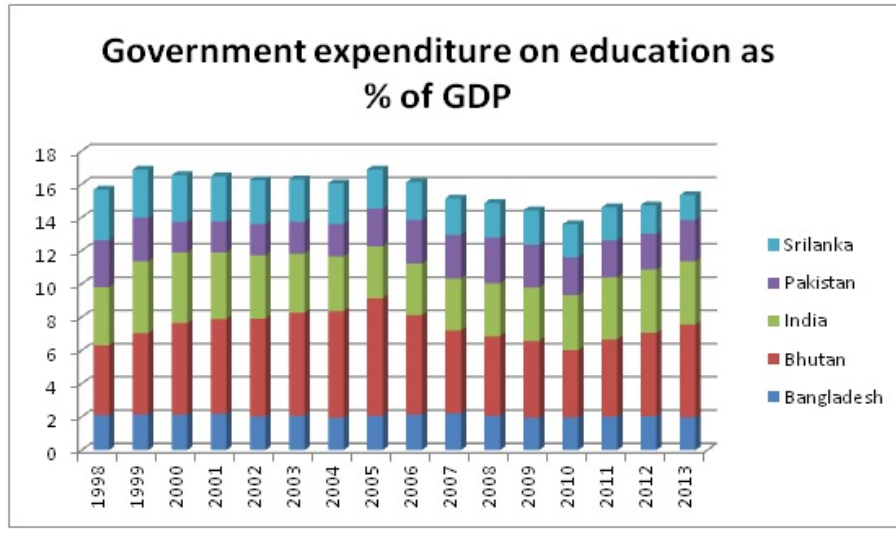

Fig. 6 Histogram depicting year-wise Government expenditure on education as percentage of GDP, by SAARC countries

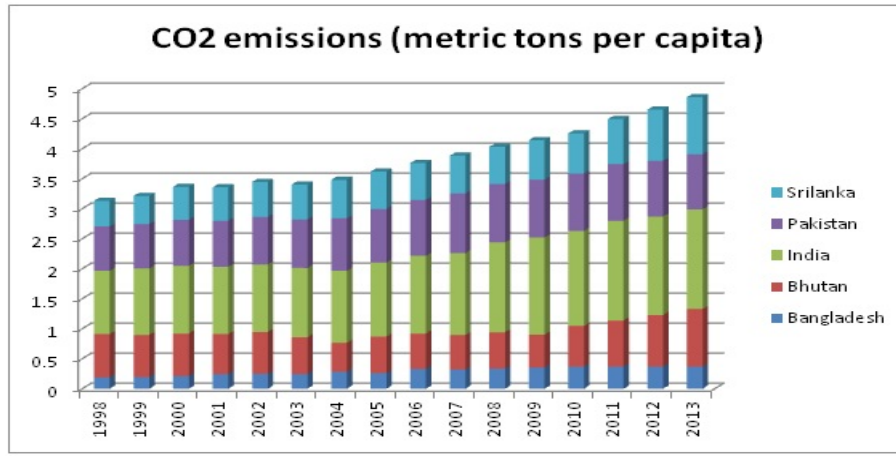

Fig. 7 Histogram depicting year-wise Carbon dioxide emissions (metric tons per capita), by SAARC countries

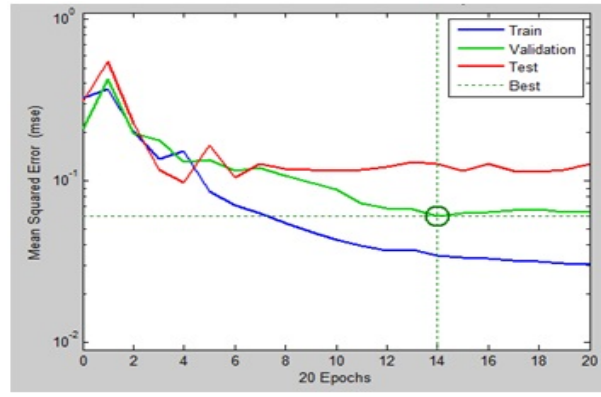

Fig.8. Result of training, testing and validating ANN for SAARC countries data using learning function trainrp

Hence after selecting the important features based on correlation coefficient, we have trained the ANN for data of SAARC countries, so as to forecast GDP. The performance measurement in terms of Mean Square Error between actual and forecasted GDP stands at 0.037 at 14 epochs, which is quite convincing, shown in Fig. 8. The sample data corresponding to actual Annual GDP Growth and forecasted Annual GDP Growth by ANN, is shown in Table IV.

Hence, ANN with trainrp back propagation learning algorithm, transfer function namely tangent-sigmoid used in hidden layer of the back propagation network and pure linear transfer function in output layer is able to predict the GDP correctly.

Table IV Table showing the comparison of Actual Annual GDP Growth and the Annual GDP Growth forecasted by Artificial Neural Network

Actual Annual GDP Growth(%)	Forecasted Annual GDP Growth(%)
5.9085022	6.013789759
5.567119371	5.571802274
5.936902755	6.184415821
3.135415426	3.660132744
3.670653998	3.224429973
4.035717573	4.300540498
5.092274715	5.940269078
5.930745063	6.241748045
6.064963977	5.950088145
6.66147781	7.058636206

5. Conclusion

Important correlations between various World Development Indicators have been found and used for data subset selection so as to forecast GDP growth. Also important relation between health expenditure and female education is found. The countries that are performing best and worst in terms of Health expenditure per capita, Annual GDP per capita growth, Government expenditure on education as percentage of GDP and carbon dioxide emissions, metrics tonnes per capita; have been mentioned in Results and discussion section.

It is also concluded that ANN has been proved to be providing convincing results and is very suitable for anticipating Gross Domestic Product growth. Using the input parameters describing Education, Health, Climate change indicators, the ANN has been trained to predict the GDP growth for five of the SAARC countries viz. Bangladesh, Bhutan, India, Pakistan and Srilanka. This study has clearly brought out that application of Soft Computing techniques can help in providing advance information for forecast of GDP growth.

This paper differs from all of the earlier studies explained in Introduction section, as it is based not only on economical indicators but also on Education statistics and Health, Nutrition, population and pollution statistics. Also the anticipation of GDP is not done just for a single country but for five of the SAARC economies. In this paper, selection of a small subset of features has been done on the basis of strong correlations between more than 15 indicators.

In future the study can also be extended for western countries or based on economic development status of the countries.

Acknowledgements

The author would like to express deepest sense of gratitude to Dr. Rattan K. Datta, Former Advisor, Department of Science & Technology, Government of India and currently Director, Mohyal Educational Research Institute of Technology, for his encouragement, guidance and mentoring. Without his support, it would not have been possible to take up research in this challenging field.

References

[1] About Asian Association for Regional Cooperation. Available from: https://en.wikipedia.org/wiki/South Asian Association for Regional Cooperation.

[2] Economic Development and the Risk of Global Climate Change, Available from:- www.worldbank.org/depweb/beyond/beyondbw/begbw 14.pdf.

[3] Smith, J.B., Vulnerability to Climate Change and Reasons for Concern: A Synthesis. In: Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change [J.J. McCarthy et al. Eds.]. Cambridge University Press, Cambridge, UK, and New York, N.Y., U.S.A., (2001).

[4] Bhargava, A., Jamison, D.,T., Lau, l., Murray, C.,J.,L. : Modeling The Effects Of Health On Economic Growth, GPE Discussion Paper Series: No. 33 Evidence and Information for Policy World Health Organization.

[5] David, E., Bloom, D.,E., Canning, D., Sevilla, J., The Effect of Health on Economic Growth: Theory and Evidence, NBER Working Paper No. 8587, Issued in November 2001.

[6] From: Education at a Glance 2012 Highlights Access the complete publication at: http://dx.doi.org/10.1787/eag highlights-2012-en

[7] Hanushek, E.A., Jamison, D.T., Jamison, E.,A., Woessmann, L., Education and Economic Growth, Education and Economic growth, Education Next, spring 2008, Vol.8, No.2.

[8] Swanson,N.,R.,White,H., A model selection approach to real time macroeconomic forecasting using linear models and artificial neural networks, Review of Economics and Statistics, Vol.79, pp. 540-550, (1997).

[9] Kohzadi, N., Milton, S., B., Kaastra, I., Kermanshahi, B., S., Scuse, D., Neural Networks for forecasting : An introduction, Canadian journal of Agricultural Economics, Vol. 43, pp.463-474, (1995).

[10] Li, Y., Macroeconomics modeling on UK GDP growth by neural computing, technical report, CSC-95009, 1995.

[11] Tkacz, G., Hu, S., Forecasting GDP growth using artificial neural networks, Working paper 1999-3/ Bank of Canada, pp.1-24, (1999).

[12] Junoh, M.,Z.,H.,M., Predicting GDP growth in Malaysia using knowledge based economy indicators: a comparison between neural networks and econometric approach, Sunway college journal, vol. 1, pp. 39-50, (2004).

[13] Zhang, G., P., Time series forecasting using a hybrid ARIMA and neural network model, neurocomputing, Vol. 50, pp. 159-175, (2003).

[14] Sivanandam S.N., Sumathi S., Deepa S.N., Introduction to Neural Networks using Matlab, Tata McGraw Hill Education Private Ltd., 2009.

[15] Kosko B., Neural Networks and Fuzzy Systems, Prentice Hall of India Ltd., 2005.