International Journal of Innovative Mathematics, Statistics & Energy Policies
11(1):1-9, Jan.-Mar., 2023

 www.seahipaj.org **ISSN: 2467-852X**

Assessment of the Nature and Pattern of Chronic Disease in Low Income Nations in West Africa, a Machine Learning Approach Case Study Nigeria

***Sunusi Shettima**

Department of Mathematics,
Umar Suleiman College of Education Gashua, Yobe State, Nigeria.
***Corresponding author**:**sunusishettima@gmail.com**

ABSTRACT
This study aim at the nature and pattern of chronic disease in low income nations in West Africa machine learning approach. Secondary data were extracted from the hospital records of patients who hospitalized from the age of 20years to 90years of age with chronic disease (Kidney disease). The objective of the study was to create predictive models using regression modelling and to employ clustering algorithms to form different clusters using states within regions. Therefore, to use distance, similarities, and association measures in order to check the relationship between some groups of individuals in the diseases under study. MINITAP version 16 was used to analyzed the data and from the result obtained it was observed that, the odd ratio for age is 1.0127, which indicates that, males are 1.0127 times more likely to be affected by the disease than the females. Also it was observed that, the odd ratio for glucose is 0.8759, which indicates that females are 0.8759 less likely to be affected by the disease than the males.
Keywords: Assessment, Chronic, disease, CKD, association, relationship, West-Africa, low, income, nation, patient, Nigeria

INTRODUCTION

Healthcare industries are engendering and hoarding a tremendous amount of data, which can be used to forecast and scrutinize the healthcare ratio of the whole nation. By using the data mining (DM) techniques, it is possible to extract confidential and helpful information from the datasets known as Knowledge Discovery in Database (KDD) and Computer-based information system (CBIS) [1, 2].

The healthcare domain is known for its ontological complexity, variety of medical data standards, and variable data quality [3, 4, 5, and 6]. With patient data privacy issues, making a functional and practically usable medical knowledge discovery has been of ongoing importance over recent decades. Current clinical practices also transform diagnosis and treatment methods and the understanding of health and illness concepts, moving from disease-oriented problem solving to a patient-centric approach, where computer-aided knowledge discovery methods play an essential role [7].

Although data mining methods and tools have already been applied in various domains for more than 50 years, their applications in healthcare [8] have started to classify and collect medical publications where knowledge discovery and data mining techniques were applied or researched from 1966 to date.

Since the twentieth century, many countries have chosen e-Health as a prioritized national program, which in essence proposes to benefit from the standardized aggregation of patients' clinical information and healthcare services rendered by providing instant access to this information for healthcare professionals as well as to patients themselves [9, 10]

Rendering to the strategic plans of EU member states, the USA, and many other nations from all continents, many investments are allocated to enable the global computerization of healthcare data. Taking a linear progression would propose that in some years coming to all new medical encounters will be thoroughly digitalized, at least in the developed countries. For the first time in history, the research community will get a complete set of a person's medical history from birth until the deceased date. This likely scenario forecasts the tremendous potential for machine learning, particularly for data mining applications in healthcare.

Diabetes is a chronic health condition that is becoming a global epidemic. Traditional tribal societies adopt a modern lifestyle in developing countries while developing chronic health conditions typically associated with developed nations [11]. The direct and indirect disease burden exceeds the financial and human resources of the healthcare system in sub-Sahara Africa (SSA) [12]. Currently, hypertension, diabetes, and coronary artery disease are the leading chronic health conditions observed in sub-Saharan Africa [13].

Infectious diseases such as human immunodeficiency virus (HIV), tuberculosis (TB), and malaria are the leading cause of death in sub-Sahara Africa. However, with international attention to these conditions, treatment options are improving, and the mortality rates are decreasing [13]; Joint United Nations Programme on HIV/AIDS WHO, 2006). Treatment of infectious diseases has led to increased life expectancy and an increased prevalence of non-communicable diseases [14].

A double disease burden has increased the combination of communicable and non-communicable diseases [14, 15]. According to Unwin (1999), non-infectious diseases in developing countries will soon outpace infectious diseases. The magnitude of these predictions was echoed by others [13, 16, and 17], suggesting chronic health conditions are becoming a significant concern. Currently, mortality from infectious diseases accounts for 69% of the overall mortality in SSA.

Statement of the Problem

The application of machine learning in healthcare raises additional challenges which require specific methods, tools, and methodology. Moreover, cross-domain knowledge is of crucial importance to achieving practical results. The rapid progress in the computerization of the healthcare industry gave a vast amount of heterogeneous, both structured and unstructured, data available for research and secondary use. Hundreds of algorithms are implemented to classify, cluster, and find hidden patterns in data. However, domain-specific issues of healthcare are still to be resolved. As discussed by [5, 18], specific problems shall be determined to apply machine learning methods successfully. According to their studies, without resolving depersonalization, multi-relational and media data pre-processing, clinical data heterogeneity, and quality issues, data mining application is sub-optimal or impossible.

According to the WHO, cancer, heart disease, stroke, and other chronic diseases are often public health issues only for high-income countries. In realism, only 20% of chronic disease deaths occur in high-income nations – while 80% occur in low/middle-income nations, where most of the world's population lives [19]. The growing burden of the disease is a huge challenge for healthcare services the world over. Persons with chronic diseases require sustained engagement with the healthcare delivery system throughout their lives. They also need support for skills to manage their disease condition more than needed for an acute health condition [20].

Thus, there is a need for a methodology that combines machine learning and data mining methods to tackle the problems of the medical domain. Such procedures shall address the following issues: methodological application of predictive and explorative data mining methods for computer-aided Algorithms, data pre-processing (information extraction), feature (selection) extraction, and predictive modelling as it relates to West African countries.

Aims and Objectives of the Study

This research aims to apply machine learning processes to determine the prevalence, nature, and pattern of chronic diseases (Kidney disease) in low income nations in West Africa.

The specific objectives are:

- To create predictive models using regression modelling.

- To employ clustering algorithms to form different clusters using States within regions to explore some characteristics among nations.
- To use distance, similarities, and association measures in order to check the relationship between some groups of individuals in the diseases under study.

Literature Review

With the evolution of machines, we have found that some tiring and routine or complex mathematical calculations can be done using calculators; finding specific information in an extensive database can be done using machines fast and efficiently. We use devices for storing data, reminding us of appointments. As the size of the data increased, computer storage has increased. Due to the vast amount of data, humans invented algorithms that produce results once a query is supplied. Although these tools perform very well, they can perform only routine tasks. Automatic classifications and other machine intelligence algorithms cannot be done using standard database languages. It has led to machine intelligence algorithms that can perform tasks supplied by humans and make decisions without human supervision. From the evolution of machine intelligence came data mining. In data mining, algorithms seek out patterns and rules within the data from which sets of rules are derived. Algorithms can automatically classify the data based on similarities (rules and practices) obtained between the training on the testing data set [1].
Today, machine learning has grown so vast that it can use in many applications. Examples include predicting costs of corporate expense claims, risk management, financial analysis, insurance, in-process control in manufacturing, healthcare, and other fields.
Let us reflect on an example in health care. The number of persons feeling sick and getting admitted to clinics and hospitals increases proportionally. The growing number of patients indirectly increases the amount of data required to be stored. If a small number of patients visit a doctor during a given redundant, the doctor will work efficiently and provide proper care of the patient. Now contemplate when many patients are coming to meet this doctor in the same period. We will find that the quality of care of the doctor will decrease. If the doctor has another colleague at his side, he can, at times, ask him for a second judgment before making decisions about the patient.
The idea of having a colleague next door at all times is not a feasible solution. Using computers to provide a second opinion to the doctor can be a viable solution. The computers will search for patterns within the database and give the doctor a quick view of the patient's diagnosis.

METHODOLOGY

In this study, secondary data were extracted from the hospital records of patients who were hospitalized from the age of 20 years to 90years of age with chronic disease (Kidney disease) admitted within the period of one week in Federal Medical Centre Nguru Yobe State.
In this retrospective study, we extracted and reviewed data from the records of hospitalised patients aged 18 years and older, with chronic illness admitted at OAUTHC Ile-Ife between 2010 and 2014. The sampling included only patients 60 years and above admitted for more than 48 hours. A checklist was used to capture data including age, sex, diagnoses, and indications for admission, intervention, and outcome at discharge

DATA ANALYSIS AND RESULT DISCUSSION

Fitted Model:

$$Li = \log it = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$$

$$Li = \log it = -0.41 + 0.0126A - 0.1325G - 0.0226D - 0.1169 + 0.0189W + 1.403O - 0.2922D$$

GOODNESS OF THE MODEL:

GOODNESS OF FIT TEST	DF	CHI-SQUARE	P-VALUE
DEVIANCE	190	222.85	0.051
PEARSON	190	226.01	0.038
HOSMER-LEMESHOW	8	20.51	0.008

From the above results obtained it was observed that the p-values of all the three tests were less than chi-square calculated. Thus, the fitted model fits the data and is very model for prediction.

Association between dependent variable and independent variables: Wald analysis

SOURCE	DF	CHI-SQUARE	P-VALUE
REGRESSION	1	35.17	0.000
AGE	1	O.80	0.372
GLU	1	2.15	0.143
DBP	1	1.69	0.193
BMI	1	2.09	0.148
WGT	1	0.70	0.402
OCP	1	29.91	0.000
DIT	1	0.72	0.397

From the above result obtained, it was observed that, all the values of chi-square are greater than all their corresponding values of probability. Thus, the coefficients of the fitted model are significant.

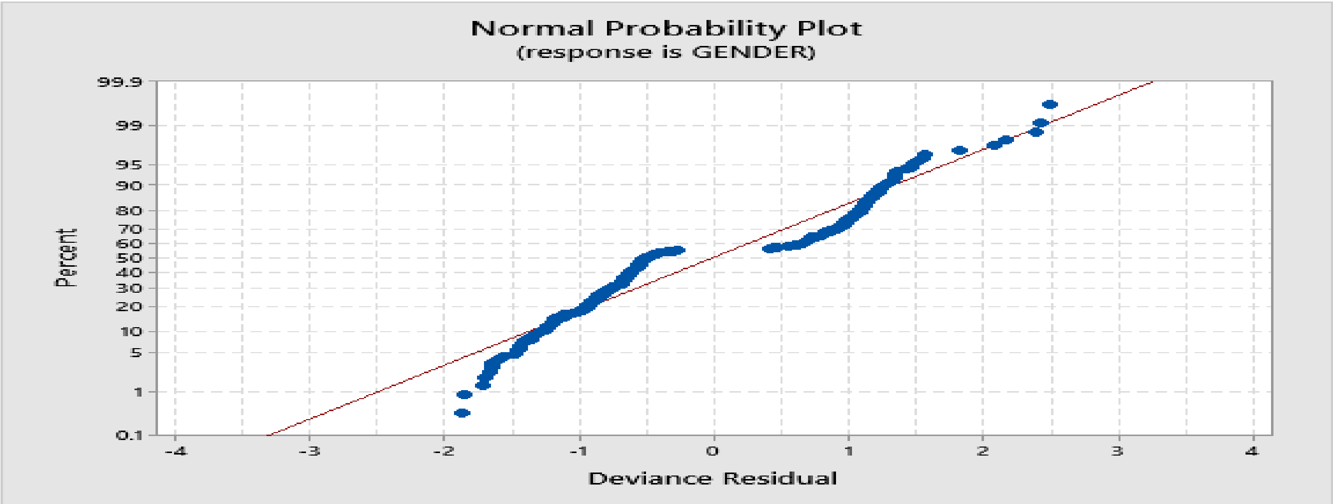

Figure 1: Checking the Normality of The Residuals Of The Fitted Model:

From the above result obtain from the normal probability plot, it was observed that the residuals of the fitted model are approximately normal, since there is no many outliers and the residuals are approximately straight line.

Odd Ratios:

	ODD RATIO	95% C.I
AGE	1.0127	(0.9851, 1.0410)
GLU	0.8759	(0.7336, 1.0458)
DBP	0.9776	(0.9449, 1.0115)
BMI	0.8897	(0.7592,1.0425)
WGT	1.0191	(0.9750, 1.0652)
OCP	4.0657	(2.4594, 6.7212)
DIT	0.7469	(0.3802, 1.4672)

From the above result obtained it was observed that, the odd ratio for age is 1.0127, which indicates that, males are 1.0127 times more likely to be affected by the disease than the females. Also it was observed that, the odd ratio for glucose is 0.8759, which indicates that females are 0.8759 less likely to be affected by the disease than the males. However the odd ratio for diastolic blood pressure is 0.9776, which indicates that females are 0.9776 less likely to be affected by the disease than the males. However the odd ratio for BMI is 0.8897, which indicates that females are 0.8897 less likely to be affected by the disease than the males. However the odd ratio for weight is 1.0191, which indicates that males are 1.0191 times more likely to be affected by the disease. Also the odd ratio for occupation is 4.0657 which also indicate that males that are not occupant are 4.0657 more likely to be affected. However the odd ratio for diet is 0.7469 which shows that females with less die are 0.7469 times more likely to be affected by the diseases.

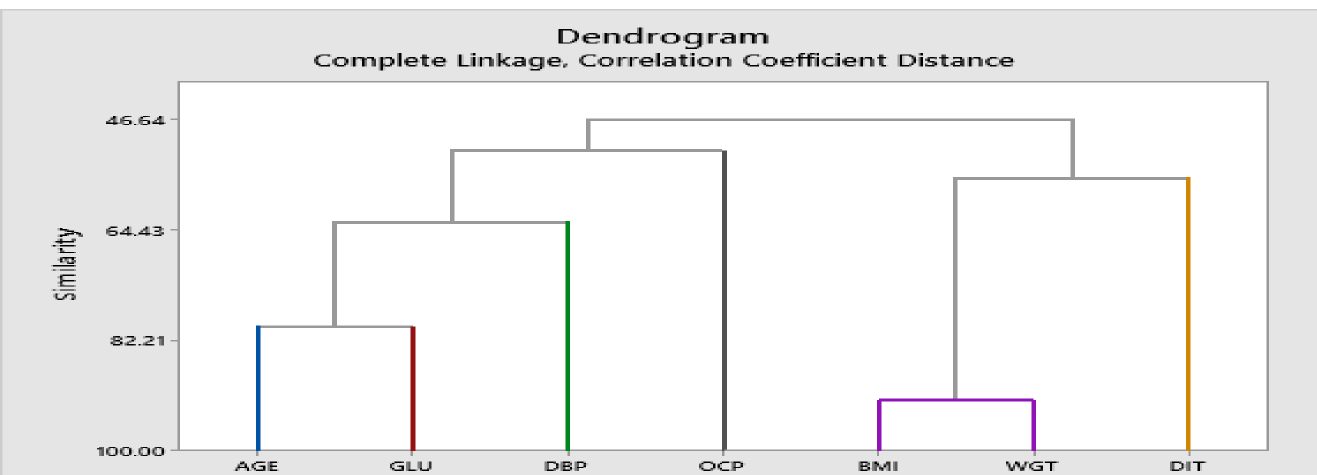

Figure 2: Distance or Similarity measure

From the above result obtained it was observes that the body mass index and weight are the most similar, since the height of the link that joins them together is the smallest. Next followed by age and glucose then glucose and diastolic blood pressure, then weight and height and finally diastolic blood pressure and occupation.

CLUSTERS:

STEP	NO OF CLUSTERS	SIMILARITY LEVEL	DISTANCE LEVEL	CLUSTER JOINED	NEW CLUSTERS	NUMBER OF OBSERVATIONS IN NEW CLUSTERS	
1	6	91.8855	0.16229	4 5	4	2	
2	5	79.9516	0.40097	1 2	1	2	
3	4	63.2017	0.73597	1 3	1	3	
4	3	56.0916	0.87817	4 7	4	3	
5	2	51.7441	0.96512	1 6	1	4	
6	1	46.6417	1.06717	1 4	1	7	

FINAL PARTITION:

CLUSTERS	VARIABLES
Cluster 1	Age
Cluster 2	Glucose
Cluster 3	Diastolic blood pressure
Cluster 4	Body mass index weight
Cluster 5	Occupation
Cluster 6	Dieet

ASSOCIATION MEASURE:
METHOD

CORRELATION	PEARSON
ROW USED	198

CORRELATION:

GLU	AGE	GLU	DBP	BMI	WGT	OCP
D BP	0.599					
BMI	0.264	0.324				
WGT	0.041	0.085	0.118	0.838		
OCP	0.035	0.062	0.075	0.010	0.144	
DIT	-0.067	0.085	-0.040	0.122	0.135	0.065

From the above result obtained it was observed that, the association between DBP and AGE is 0.8, which is strong positive association. Also the association between the BMI and AGE is 0.3, which shows weak positive association between the variables. However the association between the WGT and AGE is 0.04, which shows weak positive association. Moreover the correlation between the OCP and the AGE is 0.0.03, which is weak positive association. Also the correlation between DIT and AGE is -0.1, which is very strong negative association.

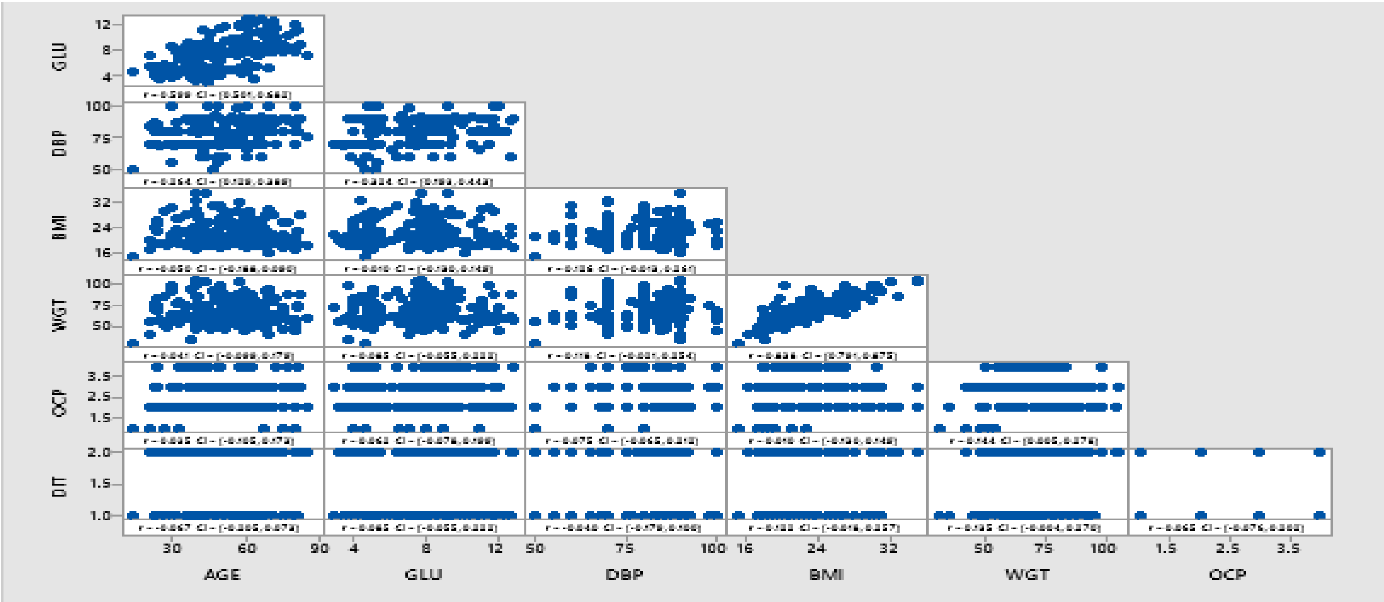

Figure 3: Correlation Matrix (Matrix plot of Age Glu, DBP, BMI, WGT,OCP and DIT) 95% CI for Pearson Correlation

The above result obtained is a correlation matrix which indicates or shows the general direction of the variables association.

SUMMARY OF THE CORRELATION:

VARIABLE	OTHER VARIABLES OF CORRELATION	CORRELATION TYPE
AGE	DIT, OCP,WGT, BMI, DBP,GLU	strong and moderate
GLU	DIT, OCP,WGT, BMI, DBP,	strong and moderate
DBP	DIT, OCP,WGT, BMI,	strong and no correlation
BMI	DIT, OCP,WGT,	strong and negative
WGT	WGT	strong
OCP	OCP	no correlation.
DIT		

In the above table, the age of the patients associated with the diet, occupation, weight body mass index, diastolic blood pressure and glucose are correlated. The glucose of the patients associated with the diet, occupation, weight, body mass index and diastolic blood pressure are also correlated. The diastolic blood pressure of the patients associated with the diet, occupation, weight and body mass index shows no correlation between them. The body mass index associated with the diet, occupation, and weight are correlated.

CONCLUSION

Generally, there are several chronic disease such as Chronic Kidney disease, Heard disease, Diabetes, Hypertension, etc. CKD are among the killer disease that low income nations are seriously battling with.

CKD are the most common disease in West Africa due to the unhealthy food/water that are using. Therefore, the low income family cannot afford the hospital bill and these can result to loss of lives.

ACKNOWLEDGEMENT

This research work was sponsored by Tertiary Education Trust Fund (Tetfund) Abuja Nigeria led by Umar Suleiman College of Education Gashua, Yobe state. I want to acknowledge the contribution of Dr. Musa Uba of Department of Statistics Kano Sate University of Science and Technology Wudil.

REFERENCES

1. Neesha jothi, NurAini, Wahidah, Data mining in healthcare, Procedia computer science, vol. 72, pp. 306-313, 2015.
2. G. E. Vlahos, Ferratt, and Knoepfle, The use of computer-based information systems by German managers to support decision-making, inf. Manage, Vol, 41, no. 6, pp. 763-779, 2004.
3. Bodenreider, O., 2008. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearbook of medical informatics,* p. 67.
4. Chen, H., Fuller, S. S., Friedman, C. & Hersh, W., 2006. *Medical informatics: knowledge management and data mining in biomedicine.* Springer.
5. Cios, K. J. & Moore, W. G., 2002. The uniqueness of medical data mining. *Artificial intelligence in medicine*, 26(1), pp. 1–24.
6. Esfandiari, N., Babavalian, M. R., Moghadam, A.-M. E. & Tabar, V. K., 2014. Knowledge discovery in medicine: Current issue and future trend. *Expert Systems with Applications,* 41(9), pp. 4434–4463.
7. Rudnick, A., 2004. An introductory course in philosophy of medicine. *Medical humanities,* 30(1), pp. 54–56.
8. Wilson, A., Thabane, L. & Holbrook, A., 2004. Application of data mining techniques in pharmacovigilance. *British journal of clinical pharmacology,* 57(2), 127–134.
9. Castro, D., 2009. Explaining international IT application leadership: Health IT. *Available at SSRN 1477486.*
10. Stroetmann, K. A., Artmann, J., Stroetmann, V. N. & Whitehouse, D., 2011. European countries on their journey towards national eHealth infrastructures. *Final European progress report*, pp. 1–47.
11. Assah, F. K., Ekelund, U., Brage, S., Mbanya, J. C., & Wareham, N. J. (2011). Urbanization, Physical Activity, and Metabolic Health in Sub-Saharan Africa. *Diabetes Care, 34*(2), 491-496. doi: 10.2337/dc10-0990
12. Kirigia, J., Sambo, H., Sambo, L., & Barry, S. (2009). Economic burden of diabetes mellitus in the WHO African region. *BMC International Health and Human Rights, 9*(1), 6.
13. Dalal, S., Beunza, J. J., Volmink, J., Adebamowo, C., Bajunirwe, F., Njelekela, M., . . . Holmes, M. D. (2011). Non-communicable diseases in sub-Saharan Africa: what we know now. *International Journal of Epidemiology, 40*(4), 885-901. doi: 10.1093/ije/dyr050
14. Levitt, N. S., Steyn, K., Dave, J., & Bradshaw, D. (2011). Chronic non-communicable diseases and HIV-AIDS on a collision course: relevance for health care delivery, particularly in low-resource settings—insights from South Africa. *The American Journal of Clinical Nutrition, 94*(6), 1690S-1696S. doi: 10.3945/ajcn.111.019075.
15. De-Graft Aikins, A., Unwin, N., Agyemang, C., Allotey, P., Campbell, C., & Arhinful, D. (2010). Tackling Africa's chronic disease burden: from the local to the global. *Globalization and Health, 6*(1), 5.
16. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D., T., & Murray, C. J. L. (2006). *Global Burden of Disease and Risk Factors*. Washington DC: The World Bank Oxford Press.
17. Habib, S. H., & Saha, S. (2010). The burden of non-communicable disease: Global Review. *Diabetes & Metabolic Syndrome: Clinical Research& Reviews, 4*(1), 41-47.
18. Špečkauskienė, V. & Lukoševičius, A., 2009. Methodology of adaptation of data mining methods for medical decision support: A case study. *Electronics and Electrical Engineering,* 2(90), pp. 25–28.

Stacey, M. & McGregor, C., 2007. Temporal abstraction in intelligent clinical

19. World Health Organization, Chronic diseases in low and middle-income Countries 2005; (wwwwho. int/CHP)
20. Nolte EE, McKee M, eds. caring for people with chronic conditions: a health system perspective. European Observatory on Health Systems and Policies series, 2008: XXI, 259. ISBN: 9789289042949.