Review of Machine Learning Techniques in Anemia Recognition

Jameela Ali Akrimi[1], Abdul Rahim Ahmad[2], Loay E. George[3]

[1]College of Graduate Studies
[2]College of Information Technology
University Tenaga National, Malaysia
[3]Baghdad University, Iraq

pt20346@ uniten.edu.my
abdrahim @uniten.edu.my
loayedwar57@yahoo.com

Abstract: ***Machine learning procedure offers a major platform in cases where a model lacks and the amount of data is enormous in explaining the relation and the generation of the data that is set. A research on trends and application of machine learning such as algorithms, techniques, and methods present practical functions for problem solving and application of techniques in settling and automatic data extraction. Anemia is one of the common diseases affecting individuals worldwide. In this study, due to lack of a distinct models, an artificial neural network (ANN) and support vector machines (SVM) have been put in place to establish a non linear function that is continuous and expresses the interdependency of the data collected and erythrocytes levels. This study will identify the use of artificial neural networks, support vector machines and statistical models and methods in the recognition of iron deficiency that leads to anemic conditions.***

Keywords: Artificial neural networks, support vector machines, statistical models

1. Introduction

Machine learning procedure offers a major platform in cases where a model lacks and the amount of data is enormous in explaining the relation and the generation of the data that is set. A research on trends and application of machine learning such as algorithms, techniques, and methods present practical functions for problem solving and application of techniques in settling and automatic data extraction. Anemia is one of the common diseases affecting individual's worldwide. In this study, due to lack of a distinct models, an artificial neural network (ANN) and support vector machines (SVM) have been put in place to establish a non linear function that is continuous and expresses the interdependency of the data collected and erythrocytes and leucocytes levels, networks of neurons are built, taken through cross validation with the use of Excel 4.32 software for neuron-solution hence, forming the artificial neural network (Suzuki, 2011). Consequently, machine learning has emerged as one of the best and most fruitful methods of research in the present world, both in terms of proposing new techniques with effective theoretical algorithms, and also in applying such methods in real life situations such as anemic recognition. The best feature about machine learning is that it combines knowledge form diverse fields such as pattern recognition which involves; artificial neural networks, reinforcement learning, support vector machine, decision tress, and data mining, which entails modeling and time series prediction, statistics involving methods like Bayesian, Montecarlo and bootstrapping, or signal processing using Marvoc Method (Bornn & Tabet, 2010). This paper will look at the artificial neural networks together with the support vector machine method, and how these methods cooperate with the use of statistical methods (Adams, 2010). The performance of neural networks and other learning methods has been established as an effective in the prediction, determination and classification of clinical outcomes involving blood complications. Consequently, studies have shown that it is more accurate as compared to linear regression. The only way to determine the levels of these blood components is looking at aspects such as GIB, blood losses, dialysis efficiency, vitamin B12 deficiency, iron status, folic acid deficiency, pro-inflammatory cytokine activities, aluminum toxicity, and any previous treatments with angiotensin. Such parameters are essential in providing data for the evaluation of the methods.

With the use of ICD-9 codes of GIB, all the available variables that are needed to test and develop various learning methods are identified. This data; demographic data, signs and symptoms, laboratory data, endoscopic diagnosis, co morbidities and outcomes were collected and utilized to create a relative analysis of these models. The methods set out a high ability to generalize symptoms that have come into a machine in the form of data use the polynomial input method of transformation. All methods gave specific but correct outcomes. All the machine learning methods provided distinct though correct results, some more correct than others. The methods studied are artificial neural networks (ANN) and support vector machines (SVM) and statistical models. Although most all the methods used above are very effective in determining, management, and evaluation of acute GIB patients, not all provide the best accuracy. These methods provide identification of the GIB source, allowing health care optimization of resources. These models are mainly considered in monitoring the functional status and individual sense of well being which are generally referred to as the measurement of the quality of life (Kopple & Massry, 2004).

2. Methodology

In this study of accuracy and effectiveness of SVM, ANN and statistical models in the diagnosis of iron deficiency, the optimum conditions for a stable hemoglobin level has to be maintained in the range of between 11 to 12 g/dl as being the recommended level, and the concentration of the hemoglobin set above 12 g/dl. However, considering the possibility of thrombotic activities, it should not go above 14 g/dl (Quaglini Barahona, & Andreassen, 2001). Accurate identification of certain myocardial infarction affiliated with iron deficiency in mature patients has remained elusive for a long time. However, the ANN has been able to overcome the huddle in medical diagnosis with its powerful statistical paradigm in recognition of complex behavioral patterns and ability to maintain high levels of accuracy. Moreover, it has come in handy in instances when some data that is required for some function to take place is missing (Anagnostou, Remzi, Lykourinas, & Djavan, 2003).

This conclusion was reached after a survey on 2000 adult patients in an emergency department complaining of chest pains. Forty variables relating to the histories of the patients, ECG results, data determination through chemical analysis and physical examination were applied in the train and testing the network functionality. The ANNs accuracy and maintaining of this accuracy when data required is unavailable suggests the importance of artificial neural networks in being a potential aid in diagnosis of anemic conditions during patient evaluation (Baxt, Shofer, Sites & Hollander, 2002). An example of the range that can be used is input layers sum up to17 units, 15 units in the hidden layer, and 8 units for the output layer. The highest performed results were obtained when the hidden layer units were 15, 0.7 learning rate, and 0.1 momentum. Consequently, it emerges that there is 71.56 percent testing and 72.78 percent correctness. This shows that the potential of multilayered perception in the recognition and predicting of anemic cases and levels can be used by medical staff and hematologists (Suzuki, 2011).

The exclusion criteria used in the support vector machine include erythropoietin or therapy of androgen, presence of hepatic and inflammatory diseases, and blood transfusion in recent months. The major classification was either suffering or not suffering from iron deficiency (Quaglini Barahona, & Andreassen, 2001). This is in accordance with the response of administering iron therapy, leading us to a dichotomous response identified by NR for No response meaning no iron deficiency and R for Response to iron deficiency. The variables used included serum ferritin, red cells (GR), hemoglobin (Hb), mean corpuscular volume (MCV), iron binding capacity, serum iron, and hematocrit. The problem with this classification was faced with the theory of the SVMs involving linear approach, instead of ensuring that the errors are minimized in the training data.

The frequency and length of occurrence of iron deficiency cases are closely linked with morbidity in patients that are undergoing MD. Such rates are common outcomes in the statistical measurement and analysis of anemic conditions in patients. These modalities may influence inflammatory and nutritional processes in patients that are undergoing MD. Consequently, the degrees of refractoriness of anemic conditions in patients that have undergone dialysis are other outcomes established by statistical models. The end-stage renal disease or ESRD anemia considered as a multi-factorial disorder that is managing well through the recombination of iron therapy and erythropoietin (Zhang, & Rutgers, 2008). Through statistical analysis, it is easy to establish and classify the iron and EPO requirements that will maintain the optimum hemoglobin concentration from 11 to 12 g/L.

3. Results

The purpose of applying ANN in hematology is to initiate the processes that human experts have developed to achieve reliable diagnosis that are separable from pattern analysis. Owing to the limits of this study, we can conclusively suggest that support vector machines and the artificial neural network through the use of statistical models have to be embraced as innovative ways in the current computer technology world as an effective means of approaching clinical problems such as anemia recognition (Bornn & Tabet, 2010). However, the greatest limitations could be the complexity of the involved calculations when sampling large sizes, and in implementing relationships that are non linear. Consequently, this has been addressed through bring statistical models on board (Quaglini Barahona, & Andreassen, 2001). More over, identification of the causes of various forms of anemia such as iron deficiency, which could be due iron deficiency, bone marrow problems or vitamin deficiency complications. The next step after making considerations that have been put forward in this discussion is to establish a multi-center study that will enlarge samples through which we will carry out an evaluation of the possibility of modifying the artificial neural networks, and support vector machines to suite the study at hand (Suzuki, 2011).

References

[1] Abedini, R., Zanganeh, I., & Mohagheghian, M. (2011). Simulation and Estimation of Vapor-Liquid Equilibrium for Asymmetric Binary Systems (CO2-Alcohols) Using Artificial Neural Network. Journal of Phase Equilibria and Diffusion, 32(2), 105-114

[2] Abdolmaleki, P., Buadu, L. D., Murayama, S., Murakami, J., Hashiguchi, N., Yabuuchi, H., &

[3] Masuda, K. (1997). Neural network analysis of breast cancer from MRI findings. Píloha diplomové práce, 15(5) 283-93.

[4] Adams, D.C. (2010). Parallel evolution of character displacement driven by competitive selection in terrestrial salamanders. BMC Evolutionary Biology, 10(72), 1-10.

[5] Anagnostou, T., Remzi, M., Lykourinas, M., & Djavan, B. (2003). Artificial neural networks for decision-making in urologic oncology. Eur Urol, 43(6), 596-603.

[6] Atkov, Y. O., Gorokhova, G. S., & Sboev, G. A. (2012). Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters. Journal of cardiology, doi:10.1016/j.jjcc.2011.11.005.

[7] Baxt, W. G., Shofer, F. S,, Sites, F. D., & Hollander, J. E. (2002). A neural computational aid to the diagnosis of acute myocardial infarction. Ann Emerg Med, 39(4), 366-73.
[8] Basak, D., Pal, S., & Patranabis, D. C. (2007). Support Vector Regression. Neural Information Processing – Letters and Reviews, 11(10), 203-224.
[9] Benke, S. K., , D. M. (2010) Methods: Genetic Epidemiology. Clinics in Laboratory Medicine, 30(4), 795-814.
[10] Bornn, L., Farrar, C.R., & Park, G. (2010). Damage Detection in Initially Nonlinear Systems. International Journal of Engineering Science, 48, 909-920.
[11] Bornn, L., & Tabet, A. (2010). Comment on "Particle Markov Chain Monte Carlo Methods". Journal of the Royal Statistical Society Series B, 72, 269-342.
[12] Cortes, C., & Vapnik, N. V. (1995). Support-Vector Networks. Journal Machine Learning, 20(3), Doi: 10.1023/A:1022627411411
[13] El-Zammar, D., Yan, M., Huang, C., Fang, D., Petigara, F., Bornn, L., & Ngai, T. (2011) Assessment and Management of Anemia in a Population of Children Living in the Indian Himalayas: A Student-Led Initiative. UBC Medical Journal, 2(2), 12-18.
[14] Hannan, S. A., Bhagile, D. V., Manza, R. R., & Ramteke, J. R. (2010). Development of an Expert System for Diagnosis and appropriate Medical Prescription of Heart Disease Using SVM and RBF. International Journal of Computer science and Information security, 8(5), 122-201.
[15] Hong, D. W., Ji, F. Y., Wang, D., Chen, Z. T., & Zhu, H. Q. (2011). Use of artificial neural network to predict esophageal varices in patients with HBV related cirrhosis. Hepatitis monthly, 11(7), 544-547.
[16] Jorgensen, J. A., & Pirmohamed, M. (2011). Risk modeling strategies for pharmacogenetic studies. Pharmacogenomics, 12(3), 397-410
[17] Kenji Suzuki. (2011). Artificial Neural Networks-Methodological Advances and Biomedical Applications. Introduction to the Artificial Neural Networks. Retrieved from <http://www.ltfe.org/wp-content/uploads/2011/04/Artificial_Neural_Networks_-_Methodological_Advances_and_Biomedical_Applications.pdf>
[18] Kopple, D. J., & Massry, G. S. (2004). Kopple and Massry's nutritional management of renal disease. Philadelphia: Lippincott Williams & Wilkins.
[19] Lisboa, P. J. G. (2004). Neural networks in medical journals: current trends and implications for BioPattern. Proc. 1st European Workshop on Assessment of Diagnostic Performance (EWADP), 7(9), 99-112.
[20] Magalhães, R. J. S., & Clements, A. C. A. (2011). Mapping the Risk of Anemia in Preschool-Age Children: The Contribution of Malnutrition, Malaria, and Helminthes Infections in West Africa. PLoS Med, 8(6) doi:10.1371/journal.pmed.1000438
[21] Mandal, I., & Sairam, N. (2012). Accurate Prediction of Coronary Artery Disease Using Reliable Diagnosis System. Journal of Medical Systems, 2(3), 93-235 DOI: 10.1007/s10916-012-9828-0
[22] Min, L., Wu, W., Joseph, R., Fulton, D.B., Berg, L., & Andreotti, A.H. (2010). Disrupting the intermolecular self-association of Itk enhances T cell signaling. Journal of Immunology, 184, 4228-4235.
[23] Nasr, B. M., & Chtourou, M. (2010).Training recurrent neural networks using a hybrid algorithm. Neural Computing & Applications, 21(1), 1-203.
[24] Quaglini, S., Barahona, P., & Andreassen, S. (2001). Artificial intelligence in medicine: 8th Conference on Artificial Intelligence in Medicine in Europe, Berlin: Springer.
[25] Ranganath, H., & Gunasekaran, N. (2004). Artificial Neural Network Approach in Estimation of Hemoglobin in Human Blood. International Computer Engineering Conference on New Technologies for the Information Society, ICENCO, Cairo, 341-344.
[26] Shi, H., Lu, Y., & Du, J. (2011). Application of Back Propagation Artificial Neural Network on Genetic Variants in Adiponectin ADIPOQ, Peroxisome Proliferator-Activated Receptor-γ, and Retinoid X Receptor-α Genes and Type 2 Diabetes Risk in a Chinese Han Population. Diabetes technology & therapeutic., doi:10.1089/dia.2011.0071
[27] Shukla, A., Tiwari, R., & Kala, R. (2010). Towards Hybrid and Adaptive Computing. Studies in Computational Intelligence, 307, 31-58.
[28] Tahmasebi, P., & Hezarkhani, A. (2011). Application of a Modular Feedforward Neural Network for Grade Estimation. Natural Resources Research, 20(1), 25-32.
[29] Tong, L. L. D., & Schierz, C. A. (2011). Hybrid genetic algorithm-neural network: feature extraction for unpreprocessed microarray data. Artificial intelligence in medicine, 53(1), 47-56 doi:10.1016/j.artmed.2011.06.008
[30] Wehe, Chang, W.-C., Eulenstein, O., & Aluru, S. (2010). A scalable parallelization of the gene duplication problem. Journal of Parallel and Distributed Computing, 70, 237-244.
[31] Xu, M., Brar, H., Grosic, S., Palmer, R., & Bhattacharyya, M.K. (2010). Excision of an active CACTA-like transposable element from DFR2 led to variegated flowers in soybean. Genetics, 184, 53-63.
[32] Zahir, S., Rejaul, G. C. & Payne, W. (2006). Automated Assessment of Erythrocyte Disorders Using Artificial Neural Network. IEEE International Symposium on Signal Processing and Information Technology, Vancouver, 776-780.
[33] Zhang, X., Wang, Z., Tang, L., Sun, Y., Cao, K., & Gao, Y. (2011). Support vector machine model for diagnosis of lymph node metastasis in gastric cancer with multidetector computed tomography: a preliminary study. BMC cancer. Retrieved from http://www.biomedcentral.com/1471-2407/11/10
[34] Zhang, G. P. (2000). Neural Networks for Classification: A Survey. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, 30(4), 451-462.
[35] Zhang, Q., & Rutgers, the State University of New Jersey - New Brunswick Graduate School. (2008). Maternal anemia and adverse pregnancy outcomes. Michigan: ProQuest.