Prediction Model for H1N1 Disease

A project submitted to Dean of Research and Postgraduate Studies Office in full

Fulfillment of the requirement for the degree

Master of Science (Intelligent System)

Universiti Utara Malaysia

By

Amy Ling Mei Yin

I

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certifies that)

AMY LING MEI YIN
(805813)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project of the following title)

PREDICTION MODEL FOR H1N1 DISEASE

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that this project is in acceptable form and content, and that a satisfactory knowledge of the field is covered by the project).

Nama Penyelia
(Name of Supervisor) : **MISS ANIZA MOHAMED DIN**

Tandatangan
(Signature) : ANIZA MOHAMED DIN Tarikh (Date) : 24/2/2011
Lecturer
College Of Arts And Sciences

Nama Penilai
(Name of Evaluator) : **MR. TUAN ZALIZAM TUAN MUDA**
Universiti Utara Malaysia

Tandatangan
(Signature) : Tarikh (Date) : 24/2/2011

TUAN ZALIZAM BIN TUAN MUDA
Lecturer
Information Technology Building
College Of Arts & Sciences
Universiti Utara Malaysia

PERMISSION TO USE

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of Postgraduate and Research. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to

Dean of Research and Postgraduate Studies

College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman

Malaysia

II

ABSTRAK

Kajian ini mengunakan data H1N1 daripada Hong Kong yang di kumpulkan daripada pesakit dari klinik (sektor persendirian dan swasta) di seluruh Hong Kong dengan influenza yang sama. Objektif kajian ini adalah untuk menbina model ramalan untuk penyakit H1N1 dengan mengunakan Multilayer Perceptron. Eperiment ini mengunakan WEKA machine learning sebagai perkakas untuk mencipta nilai parameter untuk data tersebut. General Methodology of Design Research (GMDR) and Knowledge Discovery in Databases (KDD) telah digunakan sebagai pengukur rujukan dalam kajian ini. Model ramalan untuk H1N1 mengunakan MLP telah dihasilkan dan MLP menunjukkan keputusan prestasi yang baik dengan nilai ketepatan untuk penyakit H1N1 adalah 88.57%.

Kata kunci: *H1N1, Multilayer Perceptron, Nilai ketepatan*

ABSRACT

This research has used the H1N1 disease based on the data collected from outpatient clinics (private and public sectors) across Hong Kong with influenza like illness. The objective of this project is to develop a prediction model of H1N1 disease using Multilayer Perceptron. The experiment using WEKA machine learning tool produced the best parameter's values for the datasets. The General Methodology of Design Research (GMDR) and Knowledge Discovery in Databases (KDD) has been used throughout the study as a guideline. Prediction model for H1N1 disease using MLP has been generated and MLP has performs the good result where the value of accuracy for the H1N1 disease is 88.57%.

Keywords: *H1N1 disease, Multilayer Perceptron, Accuracy's values*

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

List of table

LIST OF FIGURE

CHAPTER 1

INTRODUCTION

1.1 The Context of the Study

In the spring of 2009, a newly identified flu virus called influenza A (or H1N1) spread rapidly among people (Mabrouk & Marzouk, 2010). Based on the information from the Centers for Disease Control and Prevention (CDC), within a week, the virus spread worldwide to 30 countries by animal-to-human and human-to-human. According to the latest World Health Organization (WHO) statistics, there are more than 18,000 people died because of this virus since it was identified on April 2009. H1N1 virus has spread to enough countries to be considered as a global pandemic. Influenza epidemics can seriously affect the health of all ages particularly children younger than 2 years old and adult age 65 or older. People especially with certain medical conditions such as liver, lung, chronic heart, kidney, blood or metabolic diseases or weakened immune systems are at higher risk of being contacted with this disease.

Patients of H1N1 disease suffer because this disease is still unknown. Consequently, the determination of H1N1 or common flu would require the current model such as Multilayer Perceptron (MLP) .Our project intents to focus on the MLP model and how this model can be used to predict H1N1.

The contents of the thesis is for internal user only

References

Ali, A., Khan, U., Tufail, A. & Minkoo Kim. (2010). Analyzing potential of SVM based classifiers for intelligent and less invasive breast cancer prognosis. *Computer Engineering and Applications (ICCEA)*.

Alenezi, J.K., Awny, M.M., & Fahmy, M.M.M.(2009). Effectiveness of artificial neural networks in forecasting failure risk for pre-medical students. *Computer Engineering & Systems, 2009. ICCES 2009. International Conference*.

Altiparmak, F., Ferhatosmanoglu, H., Erdal, S., & Trost, D.C. (2006). Information mining over heterogeneous and high-dimensional time-series data in clinical trials databases. *Information Technology in Biomedicine*.

Alty, S.R., Millasseau, S.C., Chowienczyc, P.J., & Jakobsson, A. (2003). Cardiovascular disease prediction using support vector machines. *Micro-NanoMechatronics and Human Science, 2003 IEEE International Symposium*.

Bigus, J.P. (1996). Data Mining with neural networks: Solving Business Problems- from application Development to Decision Support, McGraw Hill, New Work.

Boyd, T., Savel, T., Kesarinath, G., Lee, B., & Stinn, J. (2010). The use of public health grid technology in the united states centers for disease control and prevention H1N1 pandemic response. *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference*.

CDC. Outbreak of swine-origin influenza A (H1N1) virus infection-Mexico, March-April 2009. MMWR Morb Mortal Wkly Rep, vol. 58, pp. 463–466, May 2009.

Camps-Valls, G., Porta-Oltra, B., Soria-Olivas, E., Martin-Guerrero, J.D., Serrano-Lopez, A.J., Perez-Ruixo, J.J., & Jimenez-Torres, N.V.(2003) Prediction of cyclosporine dosage in patients after kidney transplantation using neural networks.

Campos, M.M., Stengard, P.J., & Milenova, B.L. (2005) Data-centric automated data mining *Machine Learning and Applications, 2005. Proceedings Fourth International Conference on Publication Year: 2005*

Cao, J.R. & Liu, Y.J. (2010). Analysis on spatial heterogeneity of H1N1 flu based on GIS spatial analysis technology. *2010 International Conference on Multimedia Technology (ICMT)*.

Chen, C.Y., Bau, D.T., Tsai, M. H., Hsu, Y.M., Ho, T.Y., Huang, H.J., Chang, Y.H., Tsai F.J., Tsai, C.H., & Chen, C.Y. (2009). Drug design for the influenza A virus subtype H1N1. *Biomedical Engineering and Informatics, 2009. BMEI '09. 2nd International Conference*.

Cowling, B. J., Chan, K. H., Fang, V. J., (2010). Comparative epidemiology of pandemic and seasonal influenza A in households. *New England Journal of Medicine, 2010; 362:2175-84.*

Dancea, O., Gordan, M., Dragan, M., Stoian, I., & Nedevschi, S.(2008) Postoperatory risk classification of prostate cancer patients using support vector machines. *Automation, Quality and Testing, Robotics, 2008. AQTR 2008. IEEE International Conference*.

Dancey, D., Bandar, Z.A., & McLean, D. (2010). Rule extraction from neural networks for medical domains Neural Networks (IJCNN), *The 2010 International Joint Conference*

DARPA (1988), Neural Network Study, AFCEA International Press, New York.

Delshadpour, S. (2003) Improved MLP neural network as chromosome classifier. *Biomedical Engineering, 2003. IEEE EMBS Asian-Pacific Conference.*

Fayyad. U., Piatetsky-Shapiro, G., & Smyth P., From data mining to knowledge discovery in database, AI Magazine (1996) pp. 37-54

Garcia-Orellana, C.J., Gallardo-Caballero, R., Macias-Macias, M., & Gonzalez-Velasco, H. (2007). SVM and neural networks comparison in mammographic CAD.

Glotsos. D., Spyridonos. P., Petalas P., Cavouras. D., Zolota. V., Dadioti. P., Lekka. I., & Nikiforidis. G. (2003). A hierarchical decision tree classification scheme for brain tumors aztrocytoma grading using support vector machines. *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (2003).*

Hafner. M., Gangl. Wrba. R., Kastinger. Ch., Uhl. A., Thonhauser. K., Schmidt. H.-P., & Vecsei. A. (2007). Comparison of K-NN, SVM and NN in pit pattern classification of zoom-endoscopic colon images using co-occurrence histograms.

Haykin, S. (2009). Neural Network and Learning Machines. (3rd ed.). Prentice Hall.

Jaffar, M.A., Ahmed, B., Hussain, A., Naveed, N., Jabeen, F., & Mirza, A.M. (2009) Multi domain features based classification of mammogram images using SVM and MLP. *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference.*

Jajoo, R., Mital, D., Haque, S., & Srinivasan, S. (2002) Prediction of hepatitis C using artificial neural network. *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference.*

Joshi, S., Shenoy, D., Vibhudendra Simha, G.G., Rrashmi, P.L., Venugopal, K.R., & Patnaik, L.M. (2010) . *Machine Learning and Computing (ICMLC), 2010 Second International Conference.*

Koay, J., Herry, C., & Frize, M. (2004) Analysis of breast thermography with an artificial neural network. *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE.*

Kholghi, M., Hassanzadeh, H., & Keyvanpour, M. (2010) Classification and evaluation of data mining techniques for data stream requirements. *Computer Communication Control and Automation (3CA), 2010 International Symposium*

Koç, S., Yilmaz, G., & Kabak, Y. (2010). The clinical guidelines usage towards the diagnosis and treatment of H1N1. *Health Informatics and Bioinformatics (HIBIT), 2010 5th International Symposium.*

Li. T., Li. Q., Zhu. S. H., & Ogihara. M., (2002). A Survey on Wavelet Applications in Data Mining. SIGKDD Explor. Newsl.

Liu Y. (2010). Investigation of prediction and establishment of SIR model for H1N1 epidemic disease. *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference.*

Lo, J.Y. (1999). Application of artificial neural networks for diagnosis of breast cancer. *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress.*

Mabrouk, M., S, & Marzouk. S., Y. (2010). A chaotic study on pandemic and classical (H1N1) using EIIP sequence indicators. *2010 2nd International Conference on Computer Technology and Development (ICCTD 2010).*

Nirkhi, S. (2010) Potential use of artificial neural network in data mining. *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on Volume: 2.*

Palaniappan, S., & Awang, R.(2008) Intelligent heart disease prediction system using data mining techniques. *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference.*

Qi, F., Zhu, C. J., & Liu, Y. (2010) Predicting breast cancer recurrence using data mining techniques. *Bioinformatics and Biomedical Technology (ICBBT), 2010 International Conference*

Sadeghzadeh, N., Afshar, A., & Menhaj, M.B. (2008). An MLP neural network for time delay prediction in networked control systems. *Control and Decision Conference, 2008. CCDC 2008.*

Schwarzer, G., Vach, W., & Schumacher, M., (2000). On the misuses of artificial neural network for prognostic and diagnostic classification in oncology. *Statistics in Medicine.*

Seker, H., Odetayo, M., Petrovic, D., Naguib, R.N.G., Bartoli, C., Alasio, L., Lakshmi, M.S., & Sherbet, G.V.(2001) Prognostic comparison of statistical, neural and fuzzy methods of analysis of breast cancer image cytometric data. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*

Sewak. M., Vaidya. P., Chan. C. C., & Duan. Z. D. (2007). SVM approach to breast cancer classification. *Second International Multisymposium on Computer and Computational Sciences 2007.*

Segovia-Vargas M.J., Gil-Fana J.A., Heras-Martinez A., Vilar-Zanon, J.L. and Sanchis-Arellano A. (2003). Using rough sets to predict insolvency of Spanish non life insurance companies.

Smith, G. J., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G. *et al.* (2009). Origins and evolutionary genomics of the 2009 swineorigin H1N1 influenza A epidemic. *Nature, vol. 459, pp. 1122-1125, June 2009.*

Su, M. W., Chen, P.C., Chu, W.C., & Yuan, H.S. (2010). Host specific codon usage pattern of H1N1 influenza A viruses. *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference.*

Tay, S. S. (2009). Predicting Employment Condition of TAR's ICT Graduates Using Backpropagation Neural Network.

Tian, Z. G., & Zuo, M.J.(2010). Health Condition Prediction of Gears Using a Recurrent Neural Network Approach Reliability, IEEE Transactions .

Tourassi, G.D., Floyd, C.E., Jr, & Lo, J.Y. (1999)A constraint satisfaction neural network for medical diagnosis. *Neural Networks, 1999. IJCNN '99. International Joint Conference.*

Tsoukalas, L. H. & Uhrig, R. E. (1997). Fuzzy and Neural Approaches in Engineering. New Work: John Wileys & Sons.

Ultsch, A., Korus, D., & Kleine, T. O. (1995). Integration of neural networks and knowledge-based system in medicine. *Hans-Meerwein-Strabe, Lahnberge, Marburg.*

Yarmand, H., Ivy, J. S., Roberts, S. D., Bengtson, M.W., & Bengtson, N. M. (2010). Cost-effectiveness analysis of vaccination and self-isolation in case of H1N1. *Winter Simulation Conference (WSC), Proceedings of the 2010.*

Walczak, S. (2005) Artificial neural network medical decision support tool: predicting transfusion requirements of ER patients. *Information Technology in Biomedicine, IEEE Transactions.*

Wang, F.P., & He, X. H. (2010) Prediction of HLA-A*0201 restricted cytotoxic T lymphocyte epitopes in Influenza A H1N1 Virus and the similarity analysis of these epitopes with those existing in other influenza viruses. *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference.*

Wang, L. M., Chen, J. X., Pei, Y., Zhao, X., Cui, H. T., & Cui, H. Z. (2010) Feature selection and prediction of sub-health state using SVM-RFE. *Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference.*

Xing, Y. W., Wang, J., Zhao, Z.H., & Yonghong Gao. (2007) Combination data mining methods with new medical data to predicting outcome of coronary heart disease convergence. *Information Technology, 2007. International Conference.*

Zhang, W., Zeng. F., Wu. X., Zhang. X., & Jiang. R. (2009). A comparative study of ensemble learning approaches in the classification of breast cancer metastasis. *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing.*

Zhang, Z., & Fenstermacher, D. (2003) SEP: score for expression profile-a novel method for predicting clinical outcome in breast cancer. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*