

STROKE RISK PREDICTION MODEL BASED ON DEMOGRAPHIC AND  
MEDICAL SCREENING DATA

ACTING SUB LT. TEERAPAT KANSADUB 5736274 EGIT/M

M.Sc. (INFORMATION TECHNOLOGY MANAGEMENT)

THESIS ADVISORY COMMITTEE: SOTARAT THAMMABOOSADEE, Ph.D.,  
CHUTIMA JALAYONDEJA, Dr. P.H., SUPAPORN KIATTISIN, Ph.D.

ABSTRACT

Nowadays, strokes are the third leading cause of Thai's mortality in all age groups. The statistical data during 1994 - 2013 found that strokes caused 255,307 mortalities. In this paper, we present the data mining model for stroke prediction to screen people having strokes. Three classification algorithms including Neural Network, Decision Tree, and Naïve Bayes, are used for stroke prediction with different datasets: demographic data, medical screening data, and integrated data. This research was initialized with attributes and data selection, data collection, data resampling, data integration, data grouping, modeling, evaluation, and deployment. The best experimental result is Neural Network applied with integrated data result with 0.84 accuracy, 0.12 false positive rate, 0.25 false negative rate, and 0.9 area under ROC curve (AUC). Furthermore, the factor analysis using the best integrated data based on decision tree found that hemophilia and balance loss are the new, discovered risk factors compared with prior research. Finally, the best model was also used to develop an application for user-friendliness.

KEY WORDS: STROKE/ DATA MINING/ PREDICTION/  
DEMOGRAPHIC DATA

60 pages