SPATIAL DATA MINING MODEL FOR LANDFILL SITES SUITABILITY MAPPING BASED ON NEURAL NETWORKS AND MULTIVARIATE ANALYSIS

SOHAIB K. M. ABUJAYYAB

UNIVERSITI SAINS MALAYSIA

2017

SPATIAL DATA MINING MODEL FOR LANDFILL SITES SUITABILITY MAPPING BASED ON NEURAL NETWORKS AND MULTIVARIATE ANALYSIS

by

SOHAIB K. M. ABUJAYYAB

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

June 2017

DEDICATION

I dedicate to my research…

To my kind-hearted mother **Halima K. Abujayyab** and father **Khaled M. M. Abujayyab** who had dreamt to witness these moments, for their unlimited love, sacrifices, supports, protections, inspires, and prayers.

To my life partner, my beloved wife **Madleen T.M Abujayyab** for here priceless support and patienc.

To my Master supervisor and my academic life guidance **Dr. Raed Salha**.

To my family.

To my motivater and supporter, my uncle **Talaat Abujayyab**

To my motivater and supporter **Ahmed Alnnaqla**

To my friends (**Mohammed Abu al-Lail, Yahya AbuHasira and Bilal Abdel Dayem**)

To all,

For their endless support, courage, attention and inspiration in completing this project.

ACKNOWLEDGEMENTS

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

{وَإِذْ تَأَذَّنَ رَبُّكُمْ لَئِن شَكَرْتُمْ لَأَزِيدَنَّكُمْ ۖ وَلَئِن كَفَرْتُمْ إِنَّ عَذَابِي لَشَدِيدٌ}

(Chapter Name: Ibrahim, Verse No: 7)

And (remember) when your Lord proclaimed: "If you give thanks (by accepting Faith and worshipping none but Allah), I will give you more (of My Blessings), but if you are thankless (i.e. disbelievers), verily! My Punishment is indeed severe."

First and foremost, all praises are due to "Allah" the almighty, who gave me the opportunity to accomplish this research and made me overcome all circumstances, Alhamdulillah…

I extend my sincere thanks and appreciation to the Universiti Sains Malaysia, which has long been a beacon of science and scientists, and the director of the generations, as well as for the sponsorship.

It is great to pleasure to express my warmest thanks to the many people that supported me during my study. I wish to extend my gratitude, appreciation and sincere thanks to all of them especially:

I would like to express the most appreciation and gratitude to my main supervisor, Associate Professor **Dr. Mohd Sanusi S. Ahamad** for his guidance, helpful suggestions, constructive criticism, and valuable advices. I thoroughly enjoyed and benefited the high level of attention, continuous support and guidance provided.

Special thanks to Professor **Ahmad Shukri Yahya**, my co-supervisor, for his support, guidance and advices to complete the project.

Last but not least, acknowledgement to all staffs and beloved friends at the School of Civil Engineering USM, for their direct and indirect support in accomplishing my research project.

Thank you.

TABLE OF CONTENTS

CHAPTER FOUR: RESULTS AND DISCUSSION

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

REFERENCES

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MSW	Municipal solid waste
MCDA	Multi-criteria decision analysis
MCE	Multi-criteria evaluation
WLC	Weighted linear combination
OWA	Ordered weighted average
AHP	Analytical hierarchical process
MVA	Multivariate analysis
NN	Neural networks
DA	Discriminant analysis
DT	Decision tree
CHAID	Chi-squared automatic interaction detection
QUSET	Quick unbiased efficient statistical tree
CSE	Consistency subset evaluation
SDM	Spatial data mining
CFNN	Cascade forward neural network
LRN	Layer-recurrent network
NSP	National strategic plan
JICA	Japan international cooperation agency
GIS	Geographic information systems
SMLS	Suitability mapping of landfill sites
FFNN	Feedforward neural network
MaCGDI	Malaysian Centre for Geospatial Data Infrastructure
JUPEM	Department of Surveying and Mapping Malaysia

JMG	Minerals & geoscience department malaysia
NIMBY	Not in my backyard
DLSIC	Determine the landfill siting input criteria
CM	Conventional methods
ANP	Analytical network process
FSAW	Fuzzy simple additive weighting
FMCDA	Fuzzy multi-criteria decision analysis
IPM	Ideal point methods
TOPSIS	Technique for order preference by similarity to the ideal solution
DEMATEL	Decision-making trial and evaluation laboratory
VIKOR	Viekriterijumsko kompromisno rangiranje
MRSS	Median ranked sample set
SAW	Simple additive weighting
AMSL	Above mean sea level
WGS84	World geodetic system 1984
MRSO	Malaysia rectified skew orthomorphic
NASA	National Aeronautics and Space Administration
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
TRMM	Tropical rainfall measuring mission
MODIS	Moderate resolution imaging spectroradiometer
MOD16	Vegetation indices
NDVI	Normalized difference vegetation index
EOS	Nasa's earth observing system
USGS	United states geological survey
JAXA	Japan aerospace exploration agency

OSM	Openstreetmap
KML	Keyhole markup language
DEM	Digital elevation model
VIF	Variance inflation factor
CSE	Consistencysubseteval
ROC	Receiver operating characteristic
CM	Confusion matrix
SVM	Support vector machine
ESRI	Environmental systems research institute
ASCII	American Standard Code for Information Interchange
HRCFNN	Hybrid recurrent cascade forward neural network
GPS	Global positioning system
MSE	Mean square error
RMSE	Root mean square error
SCG	Scaled conjugate gradient
BFG	BFGS quasi-Newton
RP	Resilient backpropagation
LM	Levenberg–marquardt
CGB	Conjugate gradient with Powell/Beale restarts
CGF	Fletcher–Powell conjugate gradient
CGP	Polak–Ribiére conjugate gradient
OSS	One-step secant
GDX	Variable learning rate backpropagation
GDM	Gradient descent with momentum backpropagation
GD	Gradient descent backpropagation

GDA	Gradient descent with adaptive learning rate backpropagation
HDF	Hierarchical data format
Automation	Transforming thing to Automatic

MODEL PERLOMBONGAN DATA SPATIAL UNTUK PEMETAAN KESESUAIAN TAPAK PELUPUSAN BERDASARKAN RANGKAIAN NEURAL DAN ANALISIS MULTIVARIAT

ABSTRAK

Keperluan aliran kerja yang tepat untuk pemetaan kesesuaian tapak pelupusan baru adalah penting dalam perancangan pembangunan sistem pengurusan sisa pepejal perbandaran. Kesesuaian pemilihan tapak pelupusan boleh melindungi alam sekitar dan kesihatan awam. Namun demikian, wujud kerumitan dalam proses pemetaan kesesuaian tapak apabila usaha untuk mengintegrasikan maklumat atau keputusan dari bidang kepakaran berbeza yang akhirnya memberi kesan kepada keputusan pemodelan pemilihan tapak pelupusan yang tidak cekap. Terdapat beberapa kaedah Perlombongan Data Spatial (SDM) dan alir kerja Analisis Keputusan Pelbagai Kriteria (MCDA), tetapi aplikasinya dalam pemilihan tapak pelupusan adalah terhad dan menampilkan beberapa kelemahan. Dalam kajian ini, peningkatan model SDM dibangunkan untuk memenuhi empat tujuan: 1) alir kerja baru dalam penghasilan peta-peta kesesuaian berskala regional untuk perancangan tapak pelupusan sisa pepejal menggunakan Rangkaian Neural; 2) metodologi untuk memilih kriteria input yang relevan untuk model tapak pelupusan GIS berdasarkan Analisis Kaedah Multi-Variat untuk prestasi maksimum; 3) rangkaian hibrid yang menggabungkan rangkaian neural berulang lapisan dan rangkaian neural lata hadapan untuk mencapai prestasi tinggi tanpa keperluan pengetahuan manusia; dan 4) mengautomasi kotak alatan perlombongan data ruang berangkaian neural ArcGIS untuk pemetaan kesesuaian tapak pelupusan berskala regional. Kes kajian kesesuaian tapak pelupusan dijalankan di empat negeri bahagian utara Malaysia untuk menunjukkan kesahihan model SDM. Sejumlah 31 kriteria telah di proses awal untuk menetapkan set data input untuk

pemodelan NN. Sejumlah 22 kriteria telah diambil sebagai set data input selepas semakan awal kekolinearan berbilang. Rangkaian dipelajari telah digunakan untuk mendapatkan pemberat kriteria. Struktur optima cadangan rangkaian dipilih menggunakan 600,000 kes terpakai. Enam kaedah MVA digunakan untuk memilih kriteria yang relevan. Rangkaian neural hibrid digunakan sebagai kaedah penilaian dalam pemilihan kaedah optima dan algoritma latihan optima. Penggunaan kotak alat automatik adalah proses jelas dan mudah dibina dari lapan sub-alatan untuk menyedia, melatih dan memproses data. Ketepatan 99.2% telah dicapai untuk set data ujian. Struktur rangkaian terlatih yang akhir digunakan untuk menghasilkan peta indeks kesesuaian. Hasil menunjukkan fungsi latihan LM dengan kaedah pemilihan 'Consistency-Subset-Eval' telah mengenal pasti secara effisien 14 kriteria pada ketepatan prestasi 99.2%. Di samping itu, lima daripada enam kaedah telah memilih tujuh kriteria seiras yang paling relevan. Aliran kerja didapati mampu mengurangkan interferens manusia dalam penjanaan peta-peta boleh percaya. Rangkaian yang dibangunkan dan cadangan aliran kerja menunjukkan keteguhan dan kebolehgunaan NN dalam menjana peta kesesuaian tapak pelupusan dan kebolehlaksanaan pengintegrasian dengan aliran kerja MCDA yang ada. Hasil kajian menunjukkan bahawa kaedah pemilihan dan pemeringkatan kriteria adalah lebih cepat, berekonomi, dan tepat. Ia boleh menjadi satu alternatif kepada kaedah sedia yang memakan masa dalam pemilihan kriteria yang relevan. Akhir sekali, model automatik yang dijanakan sudah tentu boleh menyediakan platform yang efektif kepada pembuat keputusan melaksanakan hasil aliran kerja dan metodologi termasuk rangkaiannya. Kesimpulannya, model SDM dibangunkan adalah disyorkan untuk perancangan jangka panjang pengurusan sisa pepejal dan untuk menghasilkan peta kesesuaian untuk tapak pelupusan baru.

SPATIAL DATA MINING MODEL FOR LANDFILL SITES SUITABILITY MAPPING BASED ON NEURAL NETWORKS AND MULTIVARIATE ANALYSIS

ABSTRACT

It is very crucial to have a precise suitability mapping workflow for new landfill sites in the development planning of municipal solid waste management systems. An appropriate siting of landfill sites will protect both environment and public health. However, the complexity in the process of suitability mapping that arises from the attempt to integrate information or decisions from different disciplines has affected the results and leads to inefficient landfill siting model. There are several Spatial Data Mining (SDM) methods and Multi Criteria Decision Analysis (MCDA) workflows that are currently available, but their application in landfill sites selection is limited and reveals a number of drawbacks. In this study, the enhancement of the SDM model was constructed to serve four purposes; (1) new workflow in creating suitability maps at the regional scale for solid waste planning based on neural network (NN); 2) a hybrid network that combines layer-recurrent network and cascade forward neural network to achieve high performance without requiring prior human knowledge; 3) a methodology for selecting the relevant input criteria for landfill GIS model based on multivariate analysis (MVA) methods for maximal performance; and 4) automating an ArcGIS neural network spatial data mining toolbox for mapping the suitability of landfill sites at a regional scale. A case study on landfill site selection in four northern states of Malaysia was conducted to demonstrate the validity of the new SDM model. A total of 31 criteria were pre-processed to establish the input dataset for NN modeling. From these, 22 criteria were adopted as input datasets after prechecking for multicollinearity. The learned network was used to acquire the weights of the criteria. The optimum structure of the proposed network was selected using 600,000 use cases. Six MVA methods were employed to select the relevant criteria. Hybrid neural network was utilized as an evaluation method to select the optimal selection method and optimal training algorithm. The employment of automated toolbox is a straightforward process constructed from eight sub-tools to prepare, train, and processes the data. An accuracy of 99.2% was achieved for the test dataset. The final structure of the trained network was used to produce the suitability index map. The result showed that the LM training function with ‘Consistency-Subset-Eval’ selection method has efficiently identified 14 criteria with a performance accuracy of 99.2%. In addition, five out of the six methods has selected seven identical criteria that were most relevant. The workflow was found to be capable of reducing human interference to generate highly reliable maps. The developed network and the proposed workflow reveal the robust and the applicability of NN in generating landfill suitability maps and the feasibility of integrating them with existing MCDA workflows. The research outcomes show that the methodology of selecting and ranking criteria is quicker, economical, and precise. It can be an alternative to the existing time-consuming methodologies for selecting relevant criteria. Lastly, the automated model generated can certainly and effectively provides platform for decision makers to implement the developed workflow and methodology as well as the network. In conclusion, developed SDM model is recommended for long-term planning of solid waste management and to produce suitability maps for new landfill sites.

CHAPTER ONE
INTRODUCTION

1.1 Background

From a global perspective, the release of municipal solid waste (MSW) in enormous volumes raises serious concerns for the public. The amount of MSW generated is being augmented by rapid urbanisation, increasing public living standards, and prosperous economies; and thereby threatens public health, urban environment, and long-term sustainable development. Landfill sites are among the most hazardous locations that can cause the deterioration of the environment, industrial areas, future land use, tourism industry, and properties (Demesouka et al., 2013). This risk is largely attributed to poor decisions in the suitability mapping of landfill sites (Xu et al., 2013). Then again, the cognizance of landfill suitability mapping workflow has progressively focused on the employment of environmental, engineering, and economic criteria to satisfy common goals such as: (1) reduce threats to public health, (2) reduce the impact on the ecosystem, (3) to increase the level of facilities offered by the site, and (4) cost reduction in the use of facilities.

The process of suitability mapping of new landfill sites is considerably complex (Yesilnacar et al., 2011). The complexity comes from the incorporation of considerable information from different disciplines to many parties either responsible or affected by the results. Such complexity leads to inefficient landfill modelling which is burdened by additional financial costs, considerable time consumption, and obstacles brought about by the need for data collection, geoprocessing, and dealing with experts. This inefficiency is also attributed to low modelling accuracies or uncertain results (Eskandari et al., 2013). Therefore, the most favourable landfill criteria selection, and other pre-requisites are indispensable. This process must be