TWO STAGES HYBRID MODEL OF FUZZY LINEAR REGRESSION WITH SUPPORT VECTOR MACHINES FOR COLORECTAL CANCER

MUHAMMAD AMMAR BIN SHAFI

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy in Science

Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia

JANUARY 2020

DEDICATION

This thesis is dedicated exclusively to my lovely parents that:

Haji Shafi Bin Haji Shaid

Hajah Hasidah Binti Haji Razali

Blessing both of you always accompanies me now and hereafter.

My beautiful wife:

Nur Azia Hazida Binti Mohamad Azmi

My cutety daughters:

Nur Husna Humaira Binti Muhammad Ammar

Nur Aina Raniya Binti Muhammad Ammar

My siblings:

Muhammad Ismail Bin Haji Shafi

Nur Syafiqah Binti Haji Shafi

Muhammad Firdaus Bin Haji Shafi

And my sister and brother in-law

Nik Suhaily Binti Ismail

Muhammad Kamil Bin Ramli

Always support and also to the loved ones who are always there for me.

ACKNOWLEDGEMENT

"In the name of Allah SWT, the Merciful and Beneficent"

Praise is to Allah S.W.T for the blessing and giving me the strength and health to complete this thesis. First and foremost, I would like to thanks to my kindness supervisor, Assoc. Prof. Dr. Mohd Saifullah Bin Rusiman for his invaluable advice, comments and constant guidance during the progress of my thesis. Thank you to Assoc. Prof. Dr. Kavikumar S/O Jacob, Assoc. Prof. Dr. Nor Shamsidah Binti Amir Hamzah and Ts. Dr. Shuhaida Binti Ismail for the guidance as co-supervisor. The cooperation given by the general hospital in Kuala Lumpur is also highly appreciated. Sincere appreciations to my beloved family, especially my parents Hj. Shafi bin Hj. Shaid and Hajah. Hasidah Binti Haji Razali who are given full supports and encouragement in the process to finish this thesis, my thank you will be given to my beautiful wife Nur Azia Hazida Binti Mohamad Azmi that give encouragement and help me during thesis process and all my siblings Muhammad Ismail, Nur Syafiqah and Muhammad Firdaus, my sister in-law Nik Suhaily and my brother in law Muhammad Kamil who are never give up giving me an advised and supported. Last but not least, special thanks to all beloved friends who are shared their experiences, time and commitment in finishing this thesis. No such valuable words than "Thank You" to appreciate your eagerness and kindness. More than words could express on paper or could be spoken in words. Hopefully this research thesis would be guidance in future. Thank You.

MUHAMMAD AMMAR BIN HJ.SHAFI

ABSTRACT

Fuzzy linear regression analysis has become popular among researchers and standard model in analyzing data in vagueness phenomena. However, the factor and symptoms to predict tumor size of colorectal cancer still ambiguous and not clear. The problem in using a linear regression will arise when uncertain data and not precise data were presented. Since the fuzzy set theory's concept can deal with data not to a precise point value (uncertainty data), fuzzy linear regression was applied. In this study, two new models for hybrid model namely the multiple linear regression clustering with support vector machine model (MLRCSVM) and fuzzy linear regression with symmetric parameter with support vector machine (FLRWSPCSVM) were proposed to analyze colorectal cancer data. Other than that, the parameter, error and explanation of the five procedures to both new models were included. These models applying five statistical models such as multiple linear regression, fuzzy linear regression, fuzzy linear regression with symmetric parameter, fuzzy linear regression with asymmetric parameter and support vector machine model. At first, the proposed models were applied to the 1000 simulated data. Furthermore, secondary data of 180 colorectal cancer patients who received treatment in general hospital with twenty five independent variables with different combination of variable types were considered to find the best models to predict the tumor size of CRC. The main objective of this study is to determine the best model to predicting the tumor size of CRC and to identify the factors and symptoms that contribute to the size of CRC. The comparisons among all the models were carried out to find the best model by using statistical measurements of mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The results showed that the FLRWSPCSVM was found to be the best model, having the lowest MSE, RMSE, MAE and MAPE value by 100.605, 10.030, 7.556 and 14.769. Hence, the size of colorectal cancer could be predicted by managing twenty five independent variables.

ABSTRAK

Analisis regresi linear kabur telah menjadi popular di kalangan penyelidik dan menjadi model yang biasa digunakan di dalam fenomena kabur. Walau bagaimanapun, factor-faktor dan symptom-simptom untuk meramal saiz tumor untuk kanser usus masih samar-samar dan tidak jelas. Masalah dalam menggunakan regresi linear akan timbul apabila data yang tidak pasti dan data yang tidak tepat digunakan. Oleh kerana konsep teori set kabur dapat menangani data bukan pada nilai titik yang tepat (ketidakpastian data), regresi linear kabur telah digunakan. Dalam kajian ini, dua model hybrid baru iaitu kluster regresi linear berganda dengan model mesin vektor sokongan (MLRCSVM) dan regresi linear kabur dengan parameter simetri dengan mesin vektor sokongan (FLRWSPCSVM) dicadangkan untuk menganalisis data kanser usus. Selain itu, parameter, ralat dan penjelasan lima prosedur untuk kedua-dua model baru dimasukkan. Lima model yang sedia ada di dalam statistic digunakan seperti regresi linear berganda, regresi linear kabur, regresi linear kabur dengan parameter simetri, regresi linear fuzzy dengan parameter tidak simetrik dan model mesin vektor sokongan. Pada mulanya, 1000 data digunakan untuk simulasi. Tambahan pula, data sekunder dari 180 pesakit kanser usus yang mendapat rawatan di hospital umum dengan dua puluh lima pembolehubah tidak bersandar dengan pelbagai jenis kombinasi pemboleh ubah telah digunakan untuk mencari model yang terbaik untuk menjangkakan saiz tumor. Objektif utama kajian ini adalah untuk menentukan model terbaik untuk meramalkan saiz tumor dan mengenal pasti faktor dan gejala yang menyumbang kepada saiz tumor. Perbandingan antara semua model telah dijalankan untuk mencari model yang terbaik dengan menggunakan ukuran statistic iaitu ralat kuasa dua min (MSE), ralat punca kuasa dua min (RMSE), ralat mutlak min (MAE) dan ralat peratusan mutlak min (MAPE). Hasilnya menunjukkan bahawa FLRWSPCSVM didapati model terbaik, mempunyai nilai MSE, RMSE, MAE dan MAPE terendah sebanyak 100.605, 10.030, 7.556 dan 14.769. Oleh itu, saiz tumor boleh dijangkakan oleh dua puluh lima pembolehubah tak bersandar.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

α	- Center for fuzzy parameter
β_i	- Coefficient in multiple linear regression
$\mathbf{B_i}$	- Vector of constants
e	- Index regression number from 1 to n
ε	- Random error of parameter
$\boldsymbol{\varepsilon}$	- Vector of independent normal random variables
$\Im$	- Infinite dimensional feature space
f	- Index regression number from 1 to n
H	- Height of fuzzy triangular
ith	- Sample of *i*
$\mathbf{J}$	- An nxn matrix
n	- Number or the respondents less than 30
N	- Number of observations 30 and above
P	- Amount of predictor variables
s	-symmetric measurement
S	- Standard deviation
$\mu_{\mathbf{A}}(a)$	- Membership function of element *a*in set A
X_i	- Parameter of independent variables
$\boldsymbol{X_i}$	- Matrix parameters
$\bar{x}$	- Mean or average
$\mathbf{X}^{\mathrm{T}}$	- Transpose for matrix **X**
Y	- Dependent variables/ observations
y_i	- Observation of data
$\hat{y}$	- Output of fuzzy model
$\hat{y}_i$	-Predicted data
y^*_i	- Estimated value for y
$\boldsymbol{Y_i}$	- Vector of responses

$\sum$	- Covariance matrix
A	- Fuzzy set of A
AHP	- Analytical hierarchy process
ANOVA	- Analysis of variance
B	- Level of existence corresponds to the level membership
BMI	- Body mass index
CI	- Condition index
CI	- Confident interval
CRC	- Colorectal cancer
c_i	- Width of fuzzy parameter
C_{ii}	- ith diagonal element of matrix $(\mathbf{X}^T\mathbf{X})^{-1}$
DM	- Data mining
EFLRBM	- Extended fuzzy linear regression under benchmarking model
EORTC	- European Organisation for Research and Treatment of Cancer
FCM	- Fuzzy c-mean model
FCRM	- Fuzzy c-regression model
FIT	- Fecal immunochemical test
FLR	- Fuzzy linear regression
FLRC	- Fuzzy linear regression clustering
FLRWSP	- Fuzzy linear regression with symmetric parameter
FLRWSPCSVM	- Fuzzy linear regression with symmetric parameter clustering with support vector machine model
FOBT	- Fecal occult blood test
$f(x, \mathbf{A})$	- Fuzzy function of set **A**
$f(\mathbf{X}, \mathbf{A})$	- Fuzzy model
gFOBT	- Guaiac based fecal occult blood test
ICU	- Intensive care unit
icd10	- Place where CRC existed by patient
IDA	- Intelligent data analysis
IRFC	- Iterative relative fuzzy connectedness
MAE	- Mean absolute error

MANOVA	- Multivariate analysis of variance
MAPE	- Mean absolute percentage error
MF	- Membership function
MLR	- Multiple linear regression
MLRC	- Multiple liner regression clustering
MLRCSVM	- Multiple linear regression clustering with support vector machine model
Mm	- Milimetre
MOHM	- Ministry of Health Malaysia
MOS	- Malaysian Oncology Society
MRA	- Magnetic resonance angiography
MRI	- Magnetic resonance imaging
MSE	- Mean sqaure error
MSR	- Mean square regression
NCR	- National Cancer Registry
Q-Q plot	- Plot of quartile
R	- Right reference function
R^2	- Coefficient of determination
RMSE	- Root mean sqaure error
RS2337107	- Gene of polymorphism on risk colorectal cancer
Sig.	- Significant value
SMAD7	- Functional candidate gene for colorectal cancer
SPSS	- Software package for statistical analysis
SSE	- Sum of error
SSR	- Sum of regression
SST	- Sum of total
TNM	- Tumor, nodes and metastases
VIF	- Variance inflation factor
WHI	- Waist circumference to height index
WHO	- World Health Organization

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter discussed the background of regression analysis, fuzzy logic, support vector machine model and background of colorectal cancer (CRC) of the research. In addition, the problem statement, research objectives, the scope of the study, research contribution and thesis organization are also given and lastly, summary of each chapter in this study is stated.

1.2 Background of research methods

Carl F. Gauss called as father of regression analysis was the first to made contributions to physics, mathematics and astronomy in 1777-1855 and the term ‘’regression’’ was first used in 1877 by Francis Galton. The regression analysis is a technique of studying the dependence of one variable (dependent variable) on one or more variables (independent variables) with a view to estimate or predict the average value of the dependent variables in terms of the known or fixed values of the independent variables (John, 2012).

The objective of regression analysis is primarily used to estimate the relationship between variables, determine the effects of all other independent variables and predict the value of dependent variable toward independent variables. Regression analysis is the most often applied technique and tools of statistical analysis and modeling such as in business analysis and medicine analysis. This is because regression analysis is easy to use and can applies to many situations in real life. The statistical equation is derived obtained from the analysis which explains the relationship of dependent and independent variables. It provides much explanatory power, especially due to its multivariate nature. It is available in computer packages and can be easily interpreted. Plus, it also extensively used in applied sciences, economic, engineering, computer, social sciences and other fields (Agresti, 1996).

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which one among the independent variables are related to the dependent variable and to explore the forms of these relationships. Hence, regression analysis can be used to infer causal relationships between the independent and dependent variables (Kutner *et al*., 2004). However, this can lead to illusions or false relationships, therefore caution about the data that applied is advisable.

Many techniques for carrying out regression analysis have been developed. Familiar methods such as linear regression, fuzzy linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions which may be infinite-dimensional.

The performance of regression analysis results depends on the form of the data generating process and how it relates to the regression approach being used. Since the data generating process is generally good or no missing values, the process or results regression analysis on making assumptions will be acceptable. These assumptions are sometimes testable if a sufficient quantity of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally. Though, this may happened in many applications, especially with small effects or questions of causality based on observational data (Schneider *et al*., 2010).

However, regression models are very sensitive to outliers. An outlier is a data point that differs significantly from other observations. The variability in the measurement may indicate experimental error and an outlier can cause serious problem in regression analysis. A researcher found another linear model that is not focus on outliers such as support vector machine model (SVM). Support vector machine is widely applied to classifying something into a group objects. In machine learning, support-vector machine (SVM) is supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. There is a lot of using a support vector machine versus artificial neural network to find the minimum errors and the sigmoid function in both (Mariette *et al*., 2015). Vladimir N. Vapnik and Alexey Ya. Chervonenkis were the persons who develop the original SVM algorithm in 1963 and extend the algorithm to non-linear classifier by applying the kernel trick to maximum-margin hyperplanes in 1992.

Support vector machine can be used in any applications to solve various real world problems in machine learning area such as text and hypertext categorization, classification of images, hand-written character, biological and other sciences. Image classification can be greatly improved by SVM and be able to classify thousand or millions of images rather than use of smartphones and applications like instagram. Moreover, the National Institute of Health also has even developed a SVM protein software library for protein classification into functional family. Support vector machine have been used to classification scenarios with up to 90% compounds classified correctly (Chu *et al*., 2005).

Support vector machine divided into two categories classification such as linear and non-linear. There are two types of linear SVM which are hard margin linear refer to maximum margin in hyperlane and soft margin linear that refer to minimum margin hyperlane in SVM. Beside that, non-linear SVM which are primal, dual and kernel trick (Xiaojin, 2010).

Models of support vector machine and regression models cannot handle the real world data or problems that too complicated and difficulty involves with the level of uncertainty which come from human, measurement devices or environmental conditions. A researcher, Lotfi A. Zadeh is the first person developed the model that can handle the vangunes phenomenon such as fuzzy model.

Lotfi A. Zadeh studied in University of California at Berkeley introduced the paper on fuzzy sets in 1964. Among the contents described in the paper are the idea of grade membership was born, sharp criticism from academic community and waste of government funds. Moreover, on 1965 until 1975 Lotfi A. Zadeh continued to broaden the foundation of fuzzy set theory. The concept fuzzy set theory provides a fuzzy multistage decision making, fuzzy similarity relations, fuzzy restrictions and linguistic hedges. Fuzzy logic can be interpreted in a wider sense as theory of fuzzy sets. As such two objectives, fuzzy logic alleviate difficulties in developing and analyzing complex systems encountered by conventional mathematical tools and observing that human reasoning can utilize concepts and knowledge that do not have well-defined and sharp boundaries. Such for examples are tall of human, the lighter of lamp and else.

Fuzzy logic is a form of many-valued logic which can be any real number or point number between 0 and 1. In contrast with traditional logic theory, binary sets have two valued logic which is the truth value that ranges in degree between 0 and 1 (true and false). The truth value of fuzzy logic may be at range between completely true and completely false. Furthermore, if linguistic variables are used, specific function degree are also managed (Husain *et al*., 2015).

Fuzzy logic has been applied to many fields such as aerospace, automotive, business, chemistry industry, financial and medical. It allows getting the approximate values and numbers as well as incomplete and ambiguous data in all fields of fuzzy data. Fuzzy logic is able to solve incomplete data using controlling and decision making part.

Other than that, Hideo Tanaka was the first person that developed fuzzy linear regression both the research as well as statistic in 1982. In his study, he concerned with the application of fuzzy linear function to a regression analysis in a vague phenomenon. Usually in regression model, deviations between the observed values and the estimated values are supposed to be due to observation errors. It assumed that these deviations of system parameter depend on vagueness of the system structure. The data considered input and output relations whose vagueness the systems structure (Husain *et al*., 2015).

There are significant advantage of fuzzy model in analysis which is can be used without any assumptions. If the error of data is not normally distributed, then the data still can be used. It is difference with another regression analysis in statistic. Fuzzy logic provides a basic mathematical framework for dealing with vagueness.

Fuzzy regression analysis gives a fuzzy functional relationship between the dependent and independent variables in a vagueness environment. Linear regression is recommended as initial analysis before fuzzy regression analysis to make the greater decisions in fuzzy data. The input of the fuzzy data may be crisp or fuzzy. There are two types of fuzzy regression models such as Tanaka's linear programming approach and the fuzzy least-squares approach. Several methods have been presented to estimate fuzzy regression models. The first model is fuzzy regression was proposed by Tanaka et al. in 1982 for linear case by focusing on extension principle (Taheri, 2003).

Making sense of data is an ongoing task for researchers and professionals in almost every practical endeavor. The age of information technology, characterized by a vast array of data, has enormously amplified this quest and made it even more challenging. Data collection has become the reality of our lives at any time and from everywhere. It is reported that understanding the data, revealing underlying phenomena, and visualizing major tendencies are the major undertakings to pursue in intelligent data analysis (IDA), data mining (DM), and system modeling (Pedryoz, 2005).

Fuzzy regression used in complex systems such as in industry, economy, finance, marketing, and ecology function in the real world and it is more imprecision. Such systems require decisions based on human thinking and judgmental and involve human–machine interactions. In such environments, human often not be able to obtain exact numerical data about the system. The nature of information about the complex systems with vagueness is frequently fuzzy. In general, fuzzy regression seems to be intuitively more adequate for real life problems. Therefore, fuzzy regression analysis is more effective for modeling of complex systems. The pioneering work in this field reported that the authors used Zadeh's extension principle, A-level procedure, interval arithmetic, and linear programming techniques to develop a fuzzy linear regression analysis. Minimization of these distances in the fuzzy number space with respect to the unknown parameters of regression models leads to solving systems of equations (Aliev *et al*., 2002).

1.3 Research background of colorectal cancer

Colorectal cancer (CRC) is the cancer which affecting colon (bowel or large intestine) or rectum. Usually, the colon is about 5 feet long and as a part of insider body human. Plus, CRC can occur in any section of the colon or the rectum. These parts colon and rectum are the main important role to digest food and past waste in human body. Colorectal cancer can be called as colon cancer, colorectal carcinoma and vaginal metastases (Hwei *et al*., 2013). Colorectal cancer is one of the most

common diseases malignancies in the world (Malaysian Oncological Society, 2007). According to World Health Organization (WHO) (2018), cancer is the second leading cause of death globally and accounted for 8.8 million death in 2015.

REFERENCES

Abbas, M. A. and Suhad, M. A. (2017). Modelling the Strength of Lightweight Foamed Concrete using Support Vector Machine (SVM). *Case Studies in Construction Materials*, (6), 8-15.

Abolfazl, K., Hossien, S. G. and Norouz, M. (2018). The Data on the Effective Qualifications of Teachers in Medical Sciences: An Application of Combined Fuzzy AHP and Fuzzy TOPSIS Methods. *Data in Brief*, (21), 2689-2693.

Adriana, L. D., Nuria, S., Mina, L. B., Lazaro, C. (2015). Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology,* (149), 563-566.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.New York : John Wiley & Sons, Inc.*

Ahmed, K. A., Najim, A., Al-haideri, Ali, A. B. (2019). Implementing Artificial Neural Networks and Support Vector Machines to Predict Lost Circulation. *Egyptian Journal of Petroleum*, https://doi.org/10.1016/j.ejpe.2019.06.006.

Akbari, Z., Safari, A.S. and Montazer, M. (2014). Lack of Influence of the SMAD7 Gene rs2337107 Polymorphism on Risk of Colorectal Cancer in an Iranian Population.*Asian Pacific Journal of Cancer Prevention*, (15), 4437-4441.

Albion, D. M. D. (2017). Fuzzy Risk Stratification and Risk Assessment Model for Clinical Monitoring in the ICU. *Computers in Biology and Medicine,* (87), 168-178.

Alejandro, M., Valencia, F., Collado, J., Espinosa J. J. (2015). Robust Energy Management System Based on Interval Fuzzy Models. *IEEE*, (24), 140-157.

Aliev, R. A., Fazlollahi, B. and Vahidov,R. (2002). Genetic algorithms-based fuzzy regression analysis. *Journal of Soft Computing,* (6), 470 – 475.

Alison, R., Buron, A. P., Blanks, R., Pieri, K. (2017). Heterogeneity of Colorectal Cancer Risk by Tumour Characteristics: Large Prospective Study of UK Women. *International Journal of Cancer*, 1082-1090.

American Cancer Society. (2012). *How is colorectal cancer staged?* American Cancer Society.

American Cancer Society. Cancer Facts & Figures 2014 Atlanta.(2014*). American Cancer Society*.

Amin, J. and Mohammad, H. M. (2017). Fuzzy Evidential Network and Its Application as Medical Prognosis and Diagnosis Models. *Journal of Biomedical Informatics*, (72), 96-107.

Amit, S. and Neeru, A. (2018). Fuzzy System Model for Gene Expression. *The Egyptian Journal of Medical Human Genetics*, (19), 301-306.

Anupam, G., Rajat, K. D. (2016). Fuzzy Correlated Association Mining: Selecting Altered Associations among the Genes, and Some Possible Marker Genes Mediating Certain Cancers. *Applied Soft Computing*, (38), 587-605.

Ashanira, M. D., Azlan, M. Z. and Roselina, S. (2011). Overview of Support Vector Machine in Modeling Machining Performances. *Procedia Engineering*, (24), 308-312.

Baig, F., Khan, M. S., Noor, Y., and Imran M.. (2011). Design Model of Fuzzy Logic Medical Diagnosis Control System. *International Journal of Computer Science and Engineering* (IJCSE),(3), No.5, 2093-2108.

Benitez-Pena S., Blanquero, E. Carrizosa, P. Ramirez-Cobo. (2019). Cost-sensitive Feature Selection for Support Vector Machines. *Computer & Operations Research*, (106), 169-178.

Bennet, D.H. and Hardcastle JD.(1996). Early Diagnosis and Screening.*Colorectal Cancer.Ed Williams NS. Churchill Livingstone*, 21-37.

Bezdek, J. C. (1974). Cluster validity with fuzzy set. *Journal Cybernetic*, (3), 58–72.

Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. USA,*Kluwer Academic Publishers.*

Bin, Y., Hu, B. Q. (2016). A Fuzzy Covering-Based Rough Set Model and Its Generalization over Fuzzy Lattice. *Journal Information Sciences*, (367), 463-486.

Bingrong, S. and Byungkyu, B. P. (2017). Route Choice Modeling with Support Vector Machine. *Transportation Research Procedia*, (25), 1806-1814.

Bottomley, A. (2002). The cancer patient and quality of life. *The Oncologist Cancer Research UK. Cancer Stats Key Facts*,(7), 120-5.

Brand, M. R., David, D. J., Henry, T. L., Randali, E. B., Patrice, W., Ramesh, A. and Hmeant, K. R. (2006). Risk of Colon Cancer in Hereditary Non-Polypsis Colorectal Cancer Patients as Predicted by Fuzzy Modeling: Influence of Smoking. *World J Gastroenterol* , 12(28), 4485-4491.

Cappell, S. M. (2005). The pathophysiology, clinical presentation and diagnosis of coloncancer and adenomatous polyps. *The Medical Clinics of North America*, (89), 1-42

Carlos, A. E. and Ruben, M. (2019). Process Monitoring for Quality-A Model Selection Criterion for Support Vector Machine. *Procedia Manufacturing*, (34), 1010-1017.

Center, M.M., Jemal, A. and Ward, E. (2009). International trends in colorectal cancer incidence rates. *Cancer Epidemiology Biomarkers*, (19), 1688-94.

Charles, Y. (2016). Principles of Risk Analysis: Decision Making Under Uncertainty. *1st Edition, Kindle Edition.*

Choi, S. H. and Buckley, J. J. (2008). Fuzzy Regression Models using the Least Sqaure Method Based on The Concept of Distance. *IEEE Trans, Fuzzy Syst*, (17), 1257-1272.

Chu, F., Jin, G. and wang, L. (2005). Cancer Diagnosis and Protein Secondary Structure Prediction using Support Vector Machines. *Springer, Berlin.*

Chung, W. (2012). Using the Fuzzy Linear Regression Method to Benchmark the Energy Efficiency of Commercial Buildings. *Applied Energy*, 45-49.

Chung, W. (2012). Construction of Benchmarking Models Using Fuzzy Linear Regression Techniques. *International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012) IEEE*. 555-559.

Chunqiu, Z., Lei, X., Tian, L., Tingting, L., Huan, Y., Jia, C., Badong, C., Ziyuan, Z. and Le, Z. (2016). Developing A Robust Colorectal Cancer (Ere) Risk Predictive Model with the Big Genetic and Environment Related CRC Data. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM*, 1885-1893.

Deanna, N. S. (2017). Multicollinearity: What Is It, Why Should We Care and How Can It Be Controlled?. *National University*, 1404.

Debasish, B., Srimanta, P. and Dipak, C. P. (2007). Support Vector Regression. *Neural Information Processing-Letters and Reviews*, (11), 203-224.

Devendra, K., Yadav and Akhiles, B. (2018). Segmenting Critical Success Factors of Humanitarian Supply Chains using Fuzzy DEMATEL. *Benchmarking: An International Journal*, (25), 400-425

Dubois, D. J. and Henry. (1980). Fuzzy set and systems: *Theory and applications*. Academic Press Inc.

Dunn, J. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well separated cluster. *J. Cybernetics*, 3(3), 32-57.

Duygu, I. and Haydar, D. (2016). Error Measures for Fuzzy Linear Regression: Monte Carlo Simulation Approach. *Applied Soft Computing*, (46), 104-114.

Eline, A., Avdem, A., Johansen, V. (2018). Laparoscopic Versus Open Resection for Colorectal Liver Metastases. *Analysis of Surgery*, (267), 2.

Elizabeth, S. and Sujathan, L. (2013). Apllication of Fuzzy Membership Matrix in Medical Diagnosis and Decision Making. *Applied Mathematical Sciences*,(7), No. 127, 6297-6307.

El-Melegy, M. T. and Mokhtar, H. M. (2014). Tumor Segmentation in Brain MRI Using A Fuzzy Approach with Class Center Priors. *Journal on Image and Video Processing*. 1-14.

Esther, O. I., Asier, I., Onintza, S., Azucena, A. (2019). Prediction of Irinotecan Toxicity in Metastatic Colorectal Cancer Patients Based on Machine Learning Models with Pharmacokinetic. *Parameters. Journal of Pharmacological Sciences*, (140), 20-25.

European Organisation for Research and Treatment of Cancer (EORTC). (2014). *Colorectal Cancer Facts and Figures In support of the Spectacolor Biobank Project in United Kingdom.*

Fang, T., Xiaoxin, Z., Zhihong, Y., Dongyu, S., Yong, C., Yanhao, H. (2019). A Preventive Transient Stability Control Method Based on Support Vector Machine. *Electric Power System Research*, (170), 286-293.

Fangning, C., Yizeng, C., Jian, Z., Yuanyuan, L. (2016). Optimizing *h* Value for Fuzzy Linear Regression with Assymetric Triangular Fuzzy Coefficient. *Engineering Applications of Artificial Intelligence*, (47), 16-24.

Frieden. (2002). Physical Activity, Dietary Fat and Colorectal Cancer. *PhD Thesis Martina Perše University of Ljubljana, Faculty of Medicine, Institute of Pathology, MEC,Slovenia.*

Gogtay, N. J., Thatte, U. M. (2017). Principles of Correlation Analysis. *Journal of The Association of Physicians of India*, (65), 78-81.

Hasan, P., Ramin, N. N., Amir, H. M., Kamyar, Y. and Zabihalah, C. (2019). Landfill Site Selection using A Hybrid System of AHP-Fuzzy in GIS Environment: A Case Study in Shiraz City, Iran. *MethodsX*, (6), 1454-1466.

Hata. (1998). Medical Image Segmentation by Fuzzy Logic Techniques. *IEEE 1998 Department of Computer Engineering, Himeji Institute of Technology, Japan*, 4098-4103.

Husain, M. S., Wahab, A. F., Gobithasaan, R. U. (2015). Fuzzy Linguistic in Geometric Modeling. *Malaysian Journal of Fundamental and Applied Sciences*, (11), 36-41.

Hwei, J. N., Aly E. H. (2013). Vaginal Metastases from Colorectal Cancer, *International Journal of Surgery*, (11), 1048-1055.

Hyodo, I., Suzuki, H. and Takahashi, K. (2010). Present Status and Perspective of Colorectal Cancer in Asia. Colorectal Cancer Working Group Report in 30th Asia-Pacific Cancer Conference. *Jpn J ClinOncol*, (40), 38-43.

Ibrahim, B. S. K. K., Tokhi, M. O., Huq, M. S. and Gharooni, S. C. (2011). Optimized Fuzzy Control for Natural Trajectory Based Fes-Swinging Motion. *International Journal of Integrated Engineering*, (3), No. 2, 17-23.

Jakub, K. and Marek, K. (2018). On a New Method of Dynamic Integration of Fuzzy Linear Regression Models. *International Conference on Computer Recognition Systems*, 182-190.

Jaspin, J. S. C. and Suganthi, C. (2019). Automatic Brain Tumor Segmentation from MRI using Greedy Snake Model and Fuzzy C-Means Optimization. *Journal of King Saud University –Computer and Information Sciences*, 1319-1578.

Jemal, A., Siegel, R. and Ward, E. (2008). Cancer Statistics. C.A. *A Cancer J Clin*, (58), 71-96.

John, L. and Teng, S. L. (2017). Using Information Management systems And Processes to Support Shared Care for Colorectal Cancer Survivors. *IEEE International Symposium on Technology in Society (ISTAS) Proceedings Paul Cunningham and Miriam Cunningham (Eds)*, ISBN: 978-1-5386-0487-8.

Jung, H. K. and Ling, R. (2016). Applying Fuzzy Linear Regression To Understand Metacognitive Judgments In Human-In-The-Loop Simulation Environment. *IEEE Transactions on Human-Machine Systems*, (46), NO. 3, Kaneko, R., Nakazaki, N., Tagawa, T. and Ohishi, C. (2014). A New Index of Abdominal Obesity which Effectively Predicts Risk of Colon Tumor Development in Female Japanese. *Asian Pasific Journal of Cancer Prevention*, (5), 1005-1010.

Kang, S. J., Kim, J. Y., Jeong, I. K. (2019). An Improved Gas Classification Technique Using New Features and Support Vector Machines. *10th International Conference on Soft Computing and Pattern Recognition*, (942), 158-166.

Ker, K. T. and Gerald, C. H. K. (2018). Could Spouses of Colorectal Cancer Patients Possess Higher Risk of Developing Colorectal Cancer?. *International Journal of Colorectal Disease*, 33:353.

Kong, C. K., Roslani, A. C. and Law, C. W. (2010). Impact of Socioeconomic Class on Colorectal Cancer Patient Outcomes in Kuala Lumpur and Kuching, Malaysia.*APJCP*, (11), 969-74.

Konstantina, N., Pendaraki and Konstantinos, P., Tsagarakis. (2016). Linear Regression Versus Fuzzy Linear Regression: Does it Make a Difference in the Evaluation of the Performance of Mutual Fund Managers?, *Artificial Intelligence In Financial Markets*, 311-335.

Kumru, D. A., Ergun, E., Oya, C. (2015). A Hybrid Algorithm Based on Fuzzy Linear Regression Analysis by Quadratic Programming for Time Estimation: An Experimental Study in Manufacturing Industry. *Journal of Manufacturing System*, (36), 182-188.

Kutner, H. M., Nachtsheim, N. J. and Li, W. (2004). Applied Linear Statistical Models.Fifth Edition. *Applied Linear Statistical Models Mc Graw Hill*, 197-209.

Lazim, A., Nurnadiah, Z. (2012). Road Traffic Accidents Models using Threshold Levels of Fuzzy Linear Regression. *International Conference on Statistics in Science, Business and Engineering (ICSSBE).*

Lazim, A. and Nadia. (2012). Matrix Driven Multivariate Fuzzy Linear Regression Model in Car Sales. *Journal of Applied Sciences*, 1-8.

Lavdaniti, M., Barbas, G., Fratzana, A. and Zyga, S. (2012). Evaluation of Depression in Colon Cancer Patients. *Health Science Journal*, (6), Issue 4, 681-692.

Leong W. C., Kelani, R. O., Ahmad Z. (2019). Prediction of Air Pollution Index (API) Using Support Vector Machine (SVM). *Journal of environmental chemical engineering*, 103208.

Li, J. H., Shi, M. M., Lin, X. B. (2019). Character Localization Based on Support Vector Machine. *Computer Vision Conference, CVC 2019; Las Vegas; United States*, (944), 752-763.

Lingling, S., Dongzhi, H. (2019). Multi-join Query in Database Based on Genetic and Ant Colony Algorithm Optimization. *Proceeding ICBDC 2019*, 29-33.

Liu, X., Ma, L., Zhang, S., Mathew, J. (2005). Using fuzzy c-means and fuzzy integrals for machinery fault diagnosis. *In International Conference on Condition Monitoring*, Cambridge, England.1-9.

Madar, T., Jorma, S., Elisabete W. (2019). Workplace Diesel Exhausts and Gasoline Exposure and Risk of Colorectal Cancer in Four Nordic Countries. *Safety and Health at Work*, (10), 141-150.

Malaysian Oncological Society Novartis Corporation (Malaysia). (2007). *The Lancet Oncology*,(8), 773-783.

Man, S.K., Dongsan, K. and Jeong, R. K. (2018). Stage-Dependent Gene Expression Profiling in Colorectal Cancer. *IEEE Transactions on Computational Biology and Bioinformatics.*

Mariette, A., Khanna, R. (2015). Support Vector Machines for Classification. *Journal of Bionic Engineering*, 39-66.

Mehran, H., Bector, C. R., Kamal, S. (2005). A Simple Method for Computation of Fuzzy Linear Regression. *European Journal of Operational Research*, (166), 172-184.

Mehrbakhsh, N., Hossein, A., Leila, S., Othman, I. and Elnaz, A. (2019). A Predictive Method for Hepatitis Disease Diagnosis using Ensembles of Neuro-Fuzzy Technique. *Journal of Infection and Public Health*, (12), 13-20.

Micheal, H. K., Christoper, J. N. and John, N. (2008). Applied Linear Regression Models, *International Edition*.

Milind, S. K., Santosh, B. R., Surya, P. S. (2018). Integrated SEM-FTOPSIS Framework for Modeling and Prioritization of Risk Sources In Medical Device Development Process. *An International Journal*, (25), 178-200

Ministry Of Health, Malaysia. (1995). *Information and Documentation System Unit. Planning & Development Division.*

Mohammad, A., Gandomani, H. S., Yousefi, S. M., Salehiniya, H. (2017). Colorectal Cancer in The World: Incidence, Mortality and Risk Factors. *Biomedical Research and Therapy,* (4), 1656-1675.

Moses, E. E., Philip, I. E. and Tenderwealth, C. J. (2019). Fuzzy-Multidimensional Deep Learning for Efficient Prediction of Patient Response to Antiretroviral Therapy. *Heliyon*, (5), e02080.

Mumtaz, H. S., Gaetano, G., Andrea, L. (2017). Haralick's Texture Analysis Applied To Colorectal T2-Weighted MRI: A Preliminary Study of Significance for Cancer Evolution. *Proceedings of the lASTED International Conference Biomedical Engineering,* 16-19.

Nadeer, S., Aljunid S. A., Salim, M. S., Badlishah, R. B., Kamaruddin, R., Abd M. M. R. (2013). Fuzzy Inference System: Short Review and Design. *International Review of Automatic Control*, (6), 441-449.

Nasrabadi, E., Hashemi, M. and Ghatee, M. (2007). An LP-based Approach to Outliers Detection in Fuzzy Regression Analaysis. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst*, (15), 441-457.

Natrah, M. S., SharifaEzat, W. P., Syed, M. A., Mohd, R. A. M. and Saperi, S. (2012). Quality of Life in Malaysian Colorectal Cancer Patients: A Preliminary Result. *APJCP*, (13), 1-6.

National Cancer Registry, Ministry of Health Malaysia. (2006). *Malaysian Cancer Statistics – Data and Figure Peninsular Malaysia*.

National Cancer Registry, Ministry of Health Malaysia. (2010). *Malaysian Cancer Statistics – Data and Figure Peninsular Malaysia.*

Ni, Y. (2005). Fuzzy Correlation and Regression Analysis. *PhD Thesis of University of Oklahoma Graduate College,* UMI number:3163014.

Nishihara, R., Wu, K. and Lochhead, P. (2013). Long Term Colorectal Cancer Incidence and Mortality After Lower Endoscopy. *The New England Journal of Medicine*, (369), No. 32, 1095-1105.

Nwoye, E., Khor, L. C., Dlay, S. S. and Woo, W. L. (2008). *Spectral and Statistical Features in Fuzzy Neural Expert Machine for Colorectal Adenomas and Adenocarcinoma Classification.*

Obi, J. C. and Imianvan, A. A. (2012). Fuzzy Neural Approach for Colon Cancer Prediction. *Scienta Africana*, (11), No.1, 65-76.

Palanikumar, G., Shnmugan, S., Chithambaram, V. and Selvaraju, P. (2019). Evaluation of Fuzzy Inference in Box Type Solar Cooking Food Image of Thermal Effect. *Environmental & Sustainability Indicators*, INDIC 100002.

Pamela, A., Colangelo, T., Polgaro, G. (2017). Friend or Foe?: The Tumour Microenvironment Dilemma in Colorectal Cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer,* (1867), 1-18.

Paul, J., Arnaud, S., Patricia, D., Helene, B. and Catherine, E. (2019). A Twelve-Year Study of the Prevalence, Risk Factors and Characteristics of Interval Colorectal Cancers after Negative Colonoscopy. *Clinics and Research in Hepatology and Gastroenterology,* 1-9.

Pedryoz, W. (2005). Knowledge-based Clustering: From Data to Information Granules. *John Wiley & Sons*.

Pereira, M., Graca, A., Paula, F. and Frank, D. F. (2011). Anxiety, Depression, Traumatic Stress and Quality of Life in Colorectal Cancer after Different Treatments: A Study With Portuguese Patients and Their Partners. *European Journal of Oncology Nursing,* 1-6.

Phillips, L., Karon and Smith, L. M. (2013). Correlates of Initiating Colorectal Cancer Screening Beginning at Age 50. *J Community Health*,(38), 23-30.

Ping-Feng, P., Chih-Sheng, L. (2005). A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting. *The International Journal of Management Science*, (33), 497-505.

Prabhpreet, K., Gurvinder, S. and Parminder, K. (2019). An Intelligent Validation System for Diagnostic and Prognosis of Ultrasound Fetal Growth Analysis using Neuro-Fuzzy Based on Genetic Algorithm. *Egyptian Informatics Journal*, (20), 55-87.

Qiuling, H., Jinxin, Z., Liming, L., Yiju, Wa., Ling, J. (2019). Discriminative Information-Based Nonparallel Support Vector Machine. *Signal Processing*, (162), 169-179.

Qureshi, M. A., Mahendra, R. and Jayaram, M. (2012). Screening for Colorectal Cancer in Malaysia Consensus/ Clinical Practice Guidelines. *Academy of Medicine, Malaysia*, 2-12.

Rajeswari, K. and Vaithiyanathan, V. (2011). Fuzzy Based Modeling for Diabetic Diagnosis Decision Support using Artificial Neural Network. *International Journal of Computer Science and Network Society,* (11), No. 4, 126-130.

Rachel, A. M., Salazar, R., Paul, R. (2014). Molecular Subtyping of Colorectal Cancer to Identify a Mesenchymal Tumor Type that Might Benefit from TGF-Beta Pathway Inhibition. *Journal of Clinical Oncology*, (32), 456-456.

Rodriguez, G. G., Blanco, A., Colubi, A. and Libiano, M. A. (2009). Estimation of a Simple Linear Regression Model for Fuzzy Random Variables. *Fuzzy Set Syst*, (160), 357-370.

Roslani, A. C., Taufiq, A. and Kulenthran, A. (2012). Screening for Colorectal Neoplasias with Fecal Occult Blood Tests: False-positive Impact of Non-Dietary Restriction. *Asian Pacific Journal of Cancer Prevention*, (13), 237-241.

Rousseeuw, Peter, J., Annick, M. L. (1987). Robust regression and Outlier Detection. *John Wiley & Sons.*

Rusiman, S., Robiah, A., Efendi, N. and Kavikumar, J.(2012). Adjustment of an Intensive Care Unit (ICU) Data in Fuzzy C-Regression Models. *Journal of Science and Technology*, 4 (2), 99-108.

Second Paper of the National Cancer Registry. (2003*). Case Studies on Decision for Cervical Cancer Screening among Working Women.*

Sekeroglu, B., Hasan, S. S., Abdullah, S. M. (2019). Comparison of Machine Learning Algorithms for Classification Problems. *Computer Vision Conference, CVC 2019; Las Vegas; United States*, (944), 491-499.

Shchneider, J., Bitterlich, N. and Schulze. (2005). Improved Sensitivity In The Diagnosis Of Gastro-Intestinal Tumours By Fuzzy Logic-Based Tumour Marker Profiles Including The Tumour M2-PK. *Anticancer Research*,(25), 1507-1516.

Sohrab, A., Mina, R., Mehdi, D., Ali, N. (2010). A Goal Programming Model for Computation of Fuzzy Linear Regression with Least Error. *IEEE's Publication*, 4244-6349.

Stefani, D. E., Hugo, D. P. and Alvaro, L. R. (2011). Dietary Patterns and Risk of Colorectal Cancer: a Factor Analysis in Uruguay. *Asian Pacific J Cancer Prev*, (12), 753-759.

Taheri, S. M. (2003). Trends in Fuzzy Statistics.*Austrian Journal Of Statistics*,(32), Number 3, 239-257.

Taheri, S. M. and Kelkinnama, M. (2012). Fuzzy Linear Regression Based on Least Absolute Deviation, *Iran. J. Fuzzy Syst*, (9), 121-140.

Tanaka, H. (1987). Fuzzy Data Analysis by Possibilities Linear Models. *Fuzzy Set Syst*, (24), 363-375.

Tanaka, H., Uejima, S. and Asai, K. (1982). Linear Regression Analysis with Fuzzy Model. *IEEE Transactions On Systems, Man and Cybernetics*, SMC-12, 903-907.

Thomas, R., Gerrit, A. M., David, J. H., Godfrey, G. (2019). The Landscape of Genomic Copy Number Alterations in Colorectal Cancer and Their Consequences on Gene Expression Levels and Disease Outcome. *Molecular Aspects of Medicine*, 149-153.

Udupa, J. K., Odhner, D. and Zhao, L. (2014). Body-Wide Hierarchical Fuzzy Modeling, Recognition and Delineation of Anatomy in Medical Images. *Medical Image Analysis*, (18), 752-771.

Vanagunas. (2011). Role of Endoscopy in the Staging and Management of Colorectal Cancer. *Gastrointestinal Endoscopy*,(78), No. 1, 8-12.

Varela, A., Lina, J. and Katherine, D. (2010). Understanding Factors Related to Colorectal Cancer (CRC) Screening among Urban Hipanics: Use of Focus Groups Methodology. *J Cane Edu*, (25), 70-5.

Vladimir, K. (2019). A Model of Molecular Vector Machine of Proteins. *BioSystems*, (180), 7-18.

Wassawa, W., Andrew, W., Annabella, H. B. and Johnes, O. (2019). Cervical Cancer Classification from Pap-Smears using an Enhanced Fuzzy Cmeans Algorithm. *Informatics in Medicine Unlocked*, (14), 23-33.

Wendy and Radzi. (2008). Editorial Cell Therapy Centre, Universiti Kebangsaan Malaysia Medical Centre.*Med J Malaysia*, (63), No 4, 57-58.

Wenyi, Z., Qilei, F., Junhong, L. (2017). Fuzzy Least Absolute Linear Regression. *Applied Soft Computing*, (52), 1009-1019.

World Health Organization. (2018).*Publications of the World Health Organization are available on the WHO web site (www.who.int).*

World Health Organization Data. (2010).*Publications of the World Health Organization are available on the WHO web site (www.who.int).*

Wu, S., Liu, C., Wang, Z. (2019). Regression with Support Vector Machines and VGG Neural Networks. *4th International Conference on Advanced Machine Learning Technologies and Applications, AMLTA 2019*, (921), 301-311.

Xiaojin, Z. (2010). Support Vector Machines. *Advance natural language processing*, 1-5.

Xu, Y., Wen, J. and Liu, L.(2009). Comparison of AdaBoost and Logistic Regression for Detecting Colorectal Cancer Patients with Synchronous Liver Metastasis.*IEEE Advancing Technology and Human Huminity*, 4244-4764.

Yazhou, H., Yuhan, O., Xue, L., Farhat, V. N. (2019). Performance of Prediction Models on Survival Outcomes of Colorectal Cancer with Surgical Resection: A Systematic Review and Meta-Analysis. *Surgical Oncology*, (29), 196-202.

Yilmaz, A. and Ayan, K. (2013). Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of The Model. *Turkish Journal of Engineering & Computer Science*, (21), 897-912.

Yuanyuan, L., Yizeng, C., Jian, Z., Shuya, Z. (2015). Fuzzy Linear Regression Models For QFD Using Optimized *h* Values. *Engineering Applications of Artificial Intelligence*, (39), 45-54.

Yusoff, H. M., Norwati, D., Norhayati, M. N. and Amry, A. R. (2012). Participants and Barriers to Colorectal Cancer Screening in Malaysia. *Asian Pacific J Cancer Prev*, (13), 3983-3987.

Zadeh, L. A. (1965). Fuzzy sets. *Inform Control*, (8), 338-358.

Zhang, W. (2004). Hospital Quality and Patient Trust in Care for Colorectal Cancer. *Certificate Degree of Doctor Philosophy Harvard University, Cambridge, Massachusetts, United States.*

Zhe, H. (2016). Assessing the Population Representativeness of Colorectal Cancer Treatment Clinical Trials. *IEEE*, 2970-2973.

Zeqian, X., Tongling, L. V., Liming, L., Zhiqiang, Z., Junyan, T. (2019). A Regression-Type Support Vector Machine for K-Class Problem. *Neurocomputing*, (340), 1-7.

Zolfaghari, Z. S., Mohebbi, M. and Najariyan, M. (2014). Application of Fuzzy Linear Regression Method for Sensory Evaluation of Fried Donut. *Applied Soft Computing*, (22), 417-423.