

Identification of critical risk factors on stomach disorders by means of a regression analysis.

Shahriar Shamiluulu^{1,2,6*}, Umidahan Djakbarova^{1,5}, Rahat Omorov³, Ersan Sander⁴, M. Fatih Abasiyanik¹

¹Department of Genetics and Bioengineering, Fatih University, Turkey

²Department of Computer Engineering, Fatih University, Turkey

³Gastroenterology Department, National Academy of Medical Sciences, Kyrgyzstan

⁴Gastroenterology Department, Endoscopy Unit, Samatya Hospital, Turkey

⁵Medical School, International Ataturk Alatau University, Kyrgyzstan

⁶Faculty of Engineering and Natural Sciences, Suleyman Demirel University, Kazakhstan

Abstract

Currently, the healthcare system is facing many challenges, however computer-aided disease diagnosis might play an essential role in enhancing the quality of medical services and relief these challenges. The aim of this work consists of two parts, where in the first part, we identified the feasible machine learning methods for stomach disease diagnosis, that can be incorporated into Remote Diagnosis and Support Medical System developed by our research team. For the study, a medical dataset with over 1000 instances and 24 attributes for five stomach disorders was used. During the testing process, the implemented machine learning algorithms achieved an accuracy of 98% ($p < 0.001$) for the logistic regression model, and R^2 of 0.88 ($p < 0.001$) for multivariate linear regression model respectively. We have concluded that both machine learning methods are sufficient for stomach disease diagnosis and were integrated into the mentioned system. In the second part, we performed a statistical data analysis for prevalence and critical factor analysis. The risk factors such as painkillers consumption, stress and dental problems found to have a high correlation with stomach disorders and symptoms like nausea and abdominal pain might be an important precursor of a particular stomach disease.

Keywords: Computer-aided disease diagnosis, Stomach disorders, Regression Analysis, Machine Learning.

Accepted on May 11, 2016

Introduction

Presently over the world the healthcare system is facing many challenges due to factors such as understaffed and overwhelmed hospitals, lack of medical equipment in remote areas and increased number of refugees [1]. Medicine is one of the areas that might benefit from the use of computer applications since the advent of Artificial Intelligence (AI) and Machine Learning (ML) to tackle the problems [2,3]. Computer-aided disease diagnosis might play an essential role in enhancing a quality of medical services all over the world [4]. Gastrointestinal diseases are considered to be one of the most common disorders which affect more than 46% of human population, where over 60% of the population is affected by the stomach disorders [5]. Generally, it is more convenient to find a medical data and obtain ethical reports related to digestive disorders, conduct research. Due to the aforementioned reasons, we have selected five stomach diseases: stomach flu, gastroesophageal reflux (heartburn), gastritis, peptic ulcer, and stomach cancer [6]. Currently, there is a continuous on-going research in the field of medical diagnosis and treatment. A tremendous work has been done by

applying the Artificial Neural Network (ANN), Regression models or Support Vector Machines (SVM) algorithms in order to diagnose diseases like diabetes, heart attack, cancer, and kidney diseases [7-9].

Presently, we have developed a web-based remote diagnosis and support system based on machine learning algorithms called Remote Diagnosis and Support Medical System (RDSMED) [10]. These algorithms are based on generalized regression models. This research work consists of two parts. The first part describes the identification of robust and feasible linear as well as non-linear prediction and classification model for mentioned disorders, which were integrated into the RDSMED system. The models have several characteristics like fast model building, decent classification and prediction accuracy, reliability, and easy implementation on clinical data. Based on the simulation results, two ML methods were selected for this study; namely the Logistic regression (Logit) and Multivariate linear regression model (MVRM). As a result, the optimal models have been proposed. In the second part, the prevalence analysis study was conducted, which revealed critical risk factors of stomach diseases. In the literature, there

are many studies which applied complex ML algorithms on medical data and achieved high performance for gastrointestinal diseases.

Researchers summarized a decade experience of using computer aided applications in this medical field and analysed the paradoxical situation in the world, which is inflated by an increasing amount of clinical, laboratory and diagnostic imaging information [11]. They showed that using ANNs is a good choice to tackle diagnosis issues and explained why ANN should be used. Çokmak et al. (2007) performed the comparative analysis between two methods; the LS-SVM and BP-ANN, and created a decision support system that classifies the Doppler signals of the heart valve based on extracted features of the patients [8]. As a result, they trained these models that could classify over 94% of unknown cases. In another work, authors used SVM method to classify sound data from Doppler ultrasound of patients with atherosclerosis. Researchers trained a model that achieved 100% accuracy with the 10-fold method. They showed the usability of suggested method for classification of atherosclerosis patients [12]. An exciting study was conducted by Annibale et al. (2007) where they aimed to assess the performance of ANNs in identification of patients with chronic atrophic gastritis, by using only clinical and biochemical attributes. As a result, they were able to develop a model with 98.4% accuracy, using only eight factors [13].

All those above mentioned methods are very complex and difficult to maintain. There are several studies that used moderate methods like Logit and MVRM for prediction and classification purposes. For instance, Kalhori et al. (2010) applied the logistic regression method to predict the probability of fail outcome in Tuberculosis treatment course that might be used to determine the level of patients' supervision and support [14]. They proved that the developed model based on Logit have achieved 95% prediction accuracy based on optimal sensitivity and specificity. In another study, researchers talk about dichotomous diagnostic test and advantages of using logistic regression models in terms of sensitivity, specificity, and likelihood ratios (LRs). The exerted model with an accuracy of 90% allows calculating the LRs of diagnostic test results which are conditional on several covariates. The intended logistic regression approach provides an efficient way to determine the performance of tests at the level of the individual patient risk profile. This method can also be used to examine the effect of patient's characteristics on diagnostic test features [15]. In our previous work, we have used multivariate regression analysis method to develop a prediction model for a dataset obtained from the experimental tracheal system for mucus clearance in pulmonary airways. On the obtained data, by applying several genuine transformations, we developed a model that is able to predict clearance up to 86% accuracy, based on MVRM method [16]. To our knowledge, many research studies were performed by using standalone tools like MATLAB® or Weka®, where only a few of them have been used in real life applications. Many of the studies use complex ML algorithms, however the results in this paper proves that,

less complex ML algorithms are sufficient to use if configured properly.

Theory

Regression models

In the area of data analysis, either linear or logistic regression models are used to predict the future outcome based on historical data. Both methods are considered to be white-box methods and they are clear, usable and easy to implement.

Multivariate linear regression model

Multivariate regression analysis (MVRM) focuses on the relationship between independent and dependent variables. It includes techniques for modelling and analysing the correlations among the attributes. The general formula for linear regression model is provided below (Equation 1). The proper constant coefficients are obtained when model's mean root square error is minimal. The X_i is independent variables, where they can represent several characteristics in a model. The Y_i is dependent variable which relies on x_i values, β_i coefficients and ϵ_i error [17]. The performance of the multivariate linear regression model is measured by the R^2 coefficient of determination. In this work, the MVRM model is used for prediction of stomach disorder type of patients.

$$Y_i = \sum_{i=1}^k \beta_i x_i + \epsilon_i \quad (1)$$

Logistic regression model

The logistic regression (Logit) has wide range of implications in medical research field. The Logit model is used for the classification of the attributes, which might help to classify the outcome. The distinctive feature of the Logit model is that the outcome variable is dichotomous [17]. Generally, patient's data is being used to develop a right model by identifying the important attributes in the dataset, which are important in prediction process. The outcome is not bounded to a linear form. As a result, the created model can be used to classify a newly provided data via placing them in a model for the probability P [9], calculation of a given outcome Y_i . The equation for the dichotomous Logit is given below where the Y_i comes from previous Equation (1).

$$\ln\left(\frac{p}{1-p}\right) = Y_i \quad (2) \text{ and } P(Y_i) = \frac{1}{1 + e^{-Y_i}} \quad (3)$$

$$O = f\left(\sum_{j=1}^n w_j x_j\right) \quad (4)$$

$$\varphi = f(\text{net}) = \begin{cases} 1 & \text{if } w^T x \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The P [9] is the probability of a patient's condition which is provided in Equation (3), for predicted of any digestive system disorder(s) according to threshold value φ shown in Equation (5). The β_0 is an intercept and $\beta_1 \dots \beta_k$ are the model's

parameters, which are the constants for each attribute described in Table 2.

The performance of the proposed system was measured with several performance parameters, like accuracy metrics, a significance of the model and simulation outcomes; where the accuracy metric is the number of truly diagnosed conditions.

Procedures

Collection of medical data and demographics

In this retrospective analysis study, the medical data related to five gastrointestinal disorders were considered. The medical data were collected from years 1999 to 2014 in the process of routine endoscopic practice for over 1000 subjects, in two hospitals i.e. Samatya and Frunze located in Turkey and Kyrgyzstan respectively. Patients’ data were selected for those who were confirmed for having the mentioned stomach disorders. The disorders prevalence in the dataset is 65%, which is close to WHO statistics. In the present study, the patient’s data had 24 independent attributes of different types described in Table 2. Prior applying any ML algorithm, the dataset had undergone several normalization and standardization changes. In this study, patients’ data was used with their permission. Patients’ personal information was kept strictly confidential and anonymous.

Disease groups

The defined disease groups and metrics are shown in Table 1. For the Logit model, there are two final disease conditions, where (0) represents patients without disorder and (1) represents patients with the any stomach disorder. For the MVRM, the groups are defined according to stomach disorder severity metrics.

Table 1. Stomach disorders conditions metrics.

Stomach Disorders	Logit condition metrics	MVRM condition metrics
Flu	[1/0]	0.2
Reflux	[1/0]	0.4
Gastritis	[1/0]	0.6
Ulcer	[1/0]	0.8
Cancer	[1/0]	1

Analysis and feature extraction

Several data pre-processing operations were performed before applying any classification and prediction methods on the data. First of all the missing value interpretation, and outlier deduction by using mode method was applied, where some records were rejected due to a tremendous negative effect on ultimate performance [17]. All attributes in the dataset were quantified after transformations, as most of the methods work with numerical data. The principle component analysis (PCA)

with verimax rotation used on dataset for the extraction and dimensions reduction of underlying construct [18]. The PCA operation on the 24 attributes yielded to five components explaining the 65.2% of the total variance, with Eigen-value threshold of 1.5. Each component was labelled from Comp.1 to Comp.5 and used only for Logit model. These configurations were ideal in order to solve issues related with over fitting and multicollinearity [19].

Table 2. Medical dataset with attributes.

Attributes	Value ranges	Comp. groups	Comp. weights	Corr. and Sig.
History				
Dental probs.	[0.01/1.00]	Comp. 1	0.706	R=0.524 (P<0.001)
Sleep disorders	[0.01/1.00]	Comp. 1	0.706	R=0.544 (P<0.001)
Constipation	[0.01/1.00]	Comp. 1	0.554	R=0.653 (P<0.001)
Age	[0 to 100]	Comp. 1	0.519	R=0.497 (P<0.001)
NSAIDs	[0.01/0.5/1.00]	Comp. 3	0.82	R=0.580 (P<0.001)
Appetite	[0.01/1.00]	Comp. 3	0.819	R=0.579 (P<0.001)
Stress	[0.01/1.00]	Comp. 4	0.645	R=0.470 (P<0.001)
Gender	[0/1]	Comp. 4	0.557	R=0.095 (P=0.002)
Breakfast	[0.01/1.00]	Comp. 4	-0.529	R=-0.373 (P<0.001)
Smoking	[0.01/1.00]	Comp. 5	0.868	R=0.271 (P<0.001)
Alcohol	[0.01/1.00]	Comp. 5	0.86	R=0.048 (P=0.123)
Symptoms				
Swelling	[0.01/1.00]	Comp. 1	0.74	R=0.360 (P<0.001)
Burning	[0.01/1.00]	Comp. 1	0.738	R=0.675 (P<0.001)
Souring	[0.01/1.00]	Comp. 1	0.67	R=0.692 (P<0.001)
Abdom. pain	[0.01/1.00]	Comp. 2	0.841	R=0.501 (P<0.001)
Nausea	[0.01/1.00]	Comp. 2	0.829	R=0.509 (P<0.001)
Weakness	[0.01/1.00]	Comp. 2	0.815	R=0.284 (P<0.001)
Vomiting	[0.01/1.00]	Comp. 2	0.667	R=0.476 (P<0.001)
Diarrhea	[0.01/1.00]	Comp. 2	0.585	R=-0.18 (P<0.001)
Weightloss	[0.01/1.00]	Comp. 3	0.427	R=0.606 (P<0.001)
Lab tests				
Leukocytes (mcL)	[4.0/10.0]	Comp. 2	0.528	R=0.394 (P<0.001)
Hemoglobin (g/dl)	[9.0/17.0]	Comp. 3	0.48	R=0.112 (P<0.001)
Stool blood test	[0.00/1.00]	Comp. 3	0.536	R=0.346 (P<0.001)
CLO test	[0.00/1.00]	Comp. 4	0.778	R=0.450 (P<0.001)

For the MVRM model, a QR decomposition optimization algorithm and the Logit model, a Newton-Raphson optimization algorithm used with 24 attributes. Both optimization algorithms are based on least-squares method

[20]. For selected methods, both gender attributes were considered together. The finalized input parameters after applying feature selection methods are shown in Table 2.

Applying algorithms

After the pre-processing operations on a dataset, further classification and prediction operations were performed using R[®] (R v3.2.3, Core Team, Armonk, Vienna, Austria). Both methods are implemented in the in stats package. The dataset divided into two parts for training and testing models.

Results and Discussions

In this study, we applied two supervised ML methods i.e., Logit and MVRM models. Both methods are considered to be a white-box approach model which can be easily implemented. Running time of such models is notably faster but accuracy is lower than other machine learning algorithm like ANNs or SVMs [21]. After successful training of both algorithms, the ability increased for diagnosing the unknown cases and making classifications, predictions and performance metrics calculations that are explained further.

Logit Model

The Logit model used sigmoid activation formula provided in Equation (3). The purpose of the model was to predict the severity of the case according to provided patients' symptomatic data. In other words, we needed a model that is capable of accurately predicting whether a patient has any kind of stomach disorder with provided clinical data. The Table 3 shows results of model testing simulations.

Table 3. Classification performance results of the logit model.

Model Parameters	Performance Output
Training subjects no.	730
Testing subjects no	311
Accuracy	0.98
Sensitivity	0.98
Specificity	0.99
Duration in min	0.001
R ² (McFadden)	0.91
Model Sig.	P<0.001

The total dataset contained 1041 records. The training and testing simulations were performed by 70% and 30% strategy, respectively. Learning rate was set to 0.25 for training as well as testing operation. This value gave a maximum accuracy of 0.98. The model R² (McFadden) is 0.91 and significance metric has been calculated as p<0.001. The threshold parameter was set to 0.5, changing it affects the sensitivity and specificity of the model. Thus, increasing it, affects the

sensitivity in positive and specificity in the negative direction [22]. The properly pre-processed data and developed model results in better prediction performance. It can be clearly observed in the results presented in Table 3. Rudzki et al. (1996) applied FF-ANNs for classification of gastrointestinal diseases which contained 7 features with average disease prevalence and achieved 97% accuracy [23]. Das et al. (2007) applied BP-ANNs for acute haemorrhage with 26 features and achieved 95% classification accuracy [24]. In our case, we achieved classification accuracy of 98% with 5 features and 65% disease prevalence. There are several reasons for notably well performance of the Logit. The first reason is the disorders prevalence is over 60%, which makes it easier for the algorithm to learn. The second reason is there is a clear cut off between healthy and unhealthy patients in the dataset. The last reason is the structure of the dataset and pre-processing techniques are significant, because only the attributes related to stomach disorders were selected for modelling. Thus it is clear with proper data pre-processing and feature selection techniques, the classification performance of the Logit model can be substantially enhanced.

Multivariate linear regression model

In this study, we have used the MVRM method for prediction. Totally 24 attributes have been selected based on the Pearson correlation analysis, related to dependent variable (i.e., patient's disorder condition), where the correlation results are provided in Table 1. In order to decrease multicollinearity effect, some of the attributes had been excluded from the experimental study. The residual plots for prediction output (predicted values) with reference points (actual values, refer Table 1) of each stomach disorder are shown in Figure 1. During the study, 20 patients' clinical data were simulated for each disorder group and the outcome was averaged. It was observed that for the more severe disorders i.e. gastritis, ulcer, and cancer, the classification accuracy was close to the reference point. On the other hand, for the less severe disorders, there are high fluctuations and thus overlapping. This phenomenon is due to the fact that, generally the disorders like stomach flu, and reflux, many of the symptoms experienced by patients are the same and classification is done only according to the lab tests. In such cases, the selection boundary from 80% to 95% can be applied and the prediction outcome is evaluated according to disorders. The trade-off for this high selection boundary might identify the less severe disorders accurately but introduce overlapping for more severe ones. In order to get more sense out of the data groups in Figure 1, noisy outcomes have been eliminated. Overall, we can propose that, the multivariate linear model can classify disorders with a good precision. The performance of the model is as follows: R² as 0.88, where a root mean square error is 0.115. The significance metric found as, p<0.001 which implies the trustworthiness of the model. Generally, MVRM models are used for econometrics problems. In this work, our model is tested for prediction of stomach disorders, where final results seem promising. Schneider et al. (2010) summarizes several studies which apply regression models in medical field.

They show that these models preferable for predicting continue outcomes by providing several examples where R^2 is over 0.80 [25]. In another work, researchers applied linear regression model to predict ages of people by analysing text message send by user. In their model, there are over 10 features and as a result they created a model with an $R^2 > 0.55$. It is clearly seen that the R^2 of 0.88 implies that 88% of variation in metrics related to stomach disorders can be explained by independent attributes, which is considered as a good performance result for the model.

Risk factors analysis

During the research study, critical risk factor analysis using descriptive statistics techniques has been carried out. The goal was to identify attributes and significance levels by Chi-square (χ^2) statistics that are vital to stomach disorders. Furthermore, it was found that seasonal periods play an important role and change the severity of the gastrointestinal diseases; this phenomenon is planned to be studied in the future. Initially, the stomach disorders prevalence analysis has been conducted based on gender and disease groups. As shown in Table 4, dataset consists of 45% of males and 55% of females records. In general, 65% of subjects have one or more mentioned stomach disorder. Out of that 65%, 33% are females, and 32% are males. It has been observed during the gender-wise analysis; males are more prone to stomach disorders, with the prevalence of 71%. Disorder-wise studies also reveal a higher disease percentage in males for severe disorders like gastritis, ulcer, and cancer. An explanation to this might be the males experience the more stressful lifestyle, and thus more affected by stomach disorders. According to APA (American Psychological Association), women can handle stressful situations better than men and being less affected emotionally [26]. In the Middle East and Central Asian regions, generally, the males are those who are working to support their family, which makes them prone to experience more stressful conditions [27].

Table 4. Prevalence analysis of stomach disorders.

Gender	Population dataset	General disease %	Gender wise disease %*	Gender wise and disorder wise %**
Females	55%	33%	60%	45%
Males	45%	32%	71%	48%

*P<0.001; **P>0.05 (χ^2 test was performed)

The living conditions of patients' also have an impact on stomach disorders. The patients' demographic and historical data attributes have been studied in order to reveal risk factors related to stomach disorders shown in Figure 2. During the analysis, it was observed that, patients who regularly use painkiller medicines have at least one kind gastroenterological disorder, with the value of χ^2 is 638 and $p < 0.001$. According to several studies conducted by a different research institution boards concluded that long usage of painkillers i.e., non-steroidal anti-inflammatory (NSAIDs), paracetamol and

opioids can cause stomach bleeding and further lead to cancer and significant side-effects [28,29]. Stress, dental problems, smoking, drinking and drug consumption are also highly correlated with stomach disorders, where the value of χ^2 is 422 and $p < 0.001$. In fact, the human body is so sensitive to surrounding environment that the effect of stress on the stomach may go far beyond known outcomes [30-32].

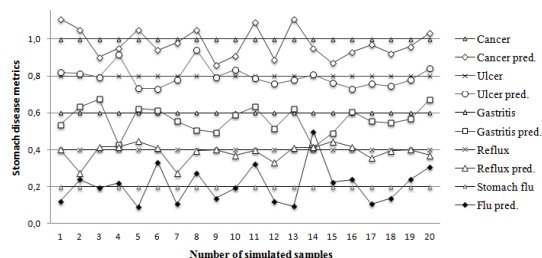


Figure 1. Stomach disorders predictions with reference points.

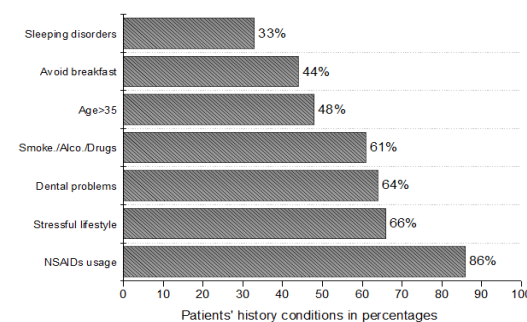


Figure 2. Effect of patients demographic and historical attributes.

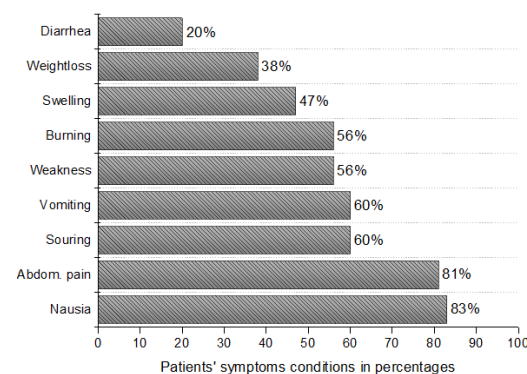


Figure 3. Effect of patients' symptomatic data attributes.

In order to avoid any major digestive system problems, patients should change their life habits that are highly correlated with the mentioned stomach disorders. In addition, proton pump inhibitor drugs should be prescribed by doctors to anyone taking NSAIDs for a long-term period. These drugs help to protect stomach lining by cutting stomach acid production. Without this layer, stomach acid can damage the gut lining that can cause gastroenterological bleeding. The symptomatic attributes described in Table 2, are common to many disorders related to internal medicine. The primary emphasis during the study was to show if anyone has a set of highly correlated symptoms with stomach disorders, he/she must immediately contact a physician. For instance, as shown in Figure 3, if a person has symptoms of nausea and abdominal pain, he/she

should not avoid it but rather start taking immediate actions. The study also revealed that patients with any stomach disorder, 83% had nausea, where the value of χ^2 is 470 and $p < 0.001$ and 81% had abdominal pain with the value of χ^2 is 474 and $p < 0.001$. Several studies also mention about the important correlated symptoms of stomach disease like in our case [32,33].

Conclusions

In the recent years, computer-based disease diagnosis has played an important role in improving the quality of medical services [34]. The aim of research was to identify feasible white-box algorithms for classification and prediction of stomach disorders. The methods aimed to have several characteristics like fast model building, decent classification and prediction accuracy, reliability, and easy implementation. The real medical data were used with over 1000 instances and 24 attributes for five stomach disorders. The Logit model was used to classify the patient's severity condition and the MVRM model to identify the particular disorder. Both are white-box and supervised learning methods. During the benchmarking study, the accuracy of 98% with $p < 0.001$ for the Logit and R^2 of 0.88 with $p < 0.001$ for the MVRM models were achieved respectively. As a result both methods are concluded to be sufficient to be used in order to diagnose stomach disorders and integrated into RDSMED. The critical risk factors analysis revealed that patients who regularly use NSAIDs drugs have at least one kind of stomach disorder. Also, symptoms like nausea and abdominal pain might be an important precursor for a particular stomach disorder.

References

- Feldstein P. Healthcare economics. Cengage learning. 2011.
- Norvig P, Russell S. Artificial Intelligence: A Modern Approach, Prentice Hall. 2002.
- Dunham MH. Data mining: Introductory and advanced topics, Pearson Education, 2006.
- Ramesh A. "Artificial intelligence in medicine", Annals of The Royal College of Surgeons of England. 2004; 86; 334.
- WHO, World health statistics 2010, World Health Organization. 2010.
- Yamada T. Textbook of Gastroenterology, Wiley-Blackwell. 2009.
- Alkım E. "A fast and adaptive automated disease diagnosis method with an innovative neural network model". Neural Networks. 2012; 33: 88-96.
- Çomak, E. "A decision support system based on support vector machines for diagnosis of the heart valve diseases". Computers in Biology and Medicine. 2007; 37: 21-27.
- Chaudhry B. "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care". Annals of internal medicine. 2006; 144: 742-752.
- Shamiliulu S, Abasiyanik MF. Web-based Remote Diagnosis and Support Medical System (Beta 3 ed). 2013.
- Pace F, Savarino V. "The use of artificial neural network in gastroenterology: the experience of the first 10 years". Eu J Gastroenterol Hepatol. 2007; 19: 1043-1045.
- Tufan K. "Non-invasive diagnosis of atherosclerosis by using empirical mode decom-position, singular spectral analysis, and support vector machines". Biomedical Research. 2013; 24; 303-313.
- Annibale B, Lahner E. "Assessing the severity of atrophic gastritis", Eu J Gastroenterol Hepatol. 2007; 19: 1059-1063.
- Kalhari SRN. "A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion". IN J Apld Math. 2010; 40: 102-107.
- Janssens ACJ. "A new logistic regression approach for the evaluation of diagnostic test results", Medical decision making. 2005; 25: 168-177.
- Shamiliulu S. "Estimation of Mucus Clearance in Pulmonary Airways by Means of a Regression Analysis", Mathematical and Computational Applications. 2014; 19: 144-151.
- Montgomery DC. Engineering Statistics, John Wiley & Sons. 2009.
- Öz HR. "Application of Artificial Neural Networks Method in Mucus Clearance in Pulmonary Airways", IN J Ntrl Engnrng Sciences. 2009; 3: 14-16.
- Delen D. "A comparative analysis of machine learning systems for measuring the impact of knowledge management practices", Decision Support Systems. 2013; 54: 1150-1160.
- Snyman J. Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms, Springer Science & Business Media. 2005.
- Park SS. A comparative study of logit and artificial neural networks in predicting bankruptcy in the hospitality industry. ProQuest. 2008.
- Berner ES. "Testing system accuracy". Clinical Decision Support Systems. 1999; 61-74.
- Rudzki K. "Focal liver disease: neural network-aided diagnosis based on clinical and laboratory data". Gastroentérologie clinique et biologique. 1996; 21: 98-102.
- Das A, Wong RC. "Prediction of outcome in acute lower gastrointestinal hemorrhage: role of artificial neural network". Europ J Gstroenterol Hepa. 2007; 19: 1064-1069.
- Schneider A. "Linear Regression Analysis", Dtsch Ä rztebl Int. 2010; 107: 776-782.
- Johnson LG. "Managing stress among adult women students in community colleges", Community College Journal of Research and Practice. 2000; 24: 289-300.
- World bank, Unemployment statistics by gender in Middle East, World Bank. 2014.
- Crichton B, Green M. "GP and patient perspectives on treatment with non-steroidal anti-inflammatory drugs for the treatment of pain in osteoarthritis", Current Medical Research and Opinion. 2002; 18: 92-96.

29. Russell R. "Non-steroidal anti-inflammatory drugs and gastrointestinal damage—problems and solutions". Postgraduate medical journal. 2001; 77: 82-88.
30. Das D, Banerjee RK. "Effect of stress on the antioxidant enzymes and gastric ulceration". Molecular and cellular biochemistry. 1993; 125: 115-125.
31. Pritchard S. "Effect of experimental stress on the small bowel and colon in healthy humans". Neurogastroenterology and Motility. 2015; 27: 542-549.
32. Bayyurt N. "Canonical correlation analysis of factors involved in the occurrence of peptic ulcers". Digestive diseases and sciences. 2007; 52: 140-146.
33. Salih BA. "H pylori infection and other risk factors associated with peptic ulcers in Turkish patients: a retrospective study". World J Gastroenterol. 2007; 13: 3245-3248.
34. Graber ML, Mathew A. "Performance of a web-based clinical diagnosis support system for internists". J Gen Intern Med. 2008; 23: 37-40.

***Correspondence to:**

Shahriar Shamiluulu
Department of Genetics and Bioengineering
Fatih University
Turkey