Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks: Modeling and Analysis

Omar Adwan[1], Hossam Faris[1], Khalid Jaradat[1], Osama Harfoushi[1], Nazeeh Ghatasheh [2]

[1.] King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan
[2.] Department of Business Information Technology, The University of Jordan, Aqaba, Jordan
hossam.faris@ju.edu.jo

Abstract: Churn represents the problem of losing a customer to another business competitor which leads to serious profit loss. Therefore, many companies investigate different techniques that can predict churn rates and help in designing effective plans for customer retention. In this study we investigate the application of Multilayer Perceptron (MLP) neural networks with back-propagation learning for churn prediction in a telecommunication company. Different MLP topologies with different settings are used to build churn classification models. Moreover, two MLP based approaches are used and compared in order to rank the most influencing factors on churn rates. For the purpose of this research, real data of customers in a major Jordanian telecom company were provided.

[Omar Adwan, Hossam Faris, Khalid Jaradat, Osama Harfoushi and Nazeeh Ghatasheh. **Predicting Customer Churn in Telecom Industry using MLP Neural Networks: Modeling and Analysis.** *Life Sci J* 2014;11(3):75-81]. (ISSN:1097-8135). http://www.lifesciencesite.com. 11. doi::10.7537/marslsj110314.11

Keywords: Churn Prediction; Telecommunication; Modeling; Analysis; Neural Network; Multilayer Perceptron

1. Introduction

Most of telecommunication companies consider the customer as the most important asset for them. For that reason, nowadays, a challenging problem that encounters telecommunication companies is when the customer leaves the company to another service provider for a reason or another [1]. In most cases, this churn can happen in rates which seriously affect the profitability of the companies since it is easy for the customers to switch companies.

In market, where the competition between the telecommunication companies grows rapidly, companies have shifted their focus from acquiring new customers to retaining their existing ones [1–3]. Basically, churn is one of these significant problems and companies started to seek new Business Intelligence (BI) applications that predict churn customers. When the company is aware of the percentage of customers who leave for another company in a given time period, it would be easier to come up with a detailed analysis of the causes for the churn rate and understand the behavior of customers that unsubscribe and move to other business competitor. This helps in planning effective customer retention strategies for that company [4].

Among many approaches developed in the literature for predicting customer churn, supervised Machine Learning (ML) techniques are the most widely investigated [5–9]. Supervised ML concerns the developing of models which can learn from labeled data. ML includes wide range of algorithms such as Decision trees, k-nearest neighbors, Linear regression, Naive Bayes, Neural Networks, Support Vector Machines (SVM), Genetic Programming and many others.

For example, in [5] authors conducted a comparative analysis of linear regression and two machine learning techniques; neural networks and decision trees for predicting customer churn based on variables related to customer complaints. Their dataset was balanced with 50% ratio of churners and non-churners. Authors showed that neural networks have good capability for predicting churners. In another work, authors in [10] investigated the use of Genetic Algorithms based neural network models to help in predicting churn in cellular wireless services. Their models show better accuracy compared to statistical z-score model.

In general, most of authors in the literature focused on using ML techniques for churn prediction but few of them focused on identifying which variables are more important in churn prediction using these techniques. Identifying important factors can greatly support customer relationship management in telecommunication companies to plan effective customer retention strategies. For that reason, in this research, we use a multilayer perceptron neural network not only for predicting customer churn but also to give an insight on the relative importance of each input variable regarding customer churn. We investigate two different approaches based on neural networks for identifying important variables. The first is based on error change and the second is based on weights contribution in the network. For the purpose of this research, real data were provided by a major Jordanian mobile operator.

2. Multilayer Perceptron Neural Networks (MLP)

An artificial neural network is an information-processing model which consists of a pool of simple processing units called neurons [21-23]. Neurons communicate by sending signals to each other over a large number of weighted connections. Multilayer Perceptron (MLP) is the most widely used type of neural networks nowadays. MLP neural networks have wide range of applications for prediction and classification problems in industrial and business domains [11,24,25]. MLP can be distinguished by a number of performance characteristics which can be summarized in the following two points:

A. MLP architecture

MLP consists of multiple layers (commonly three layers or more) where each layer is fully connected with the next one in a feed forward fashion. The first and the last layers represent the inputs and outputs of the system respectively. Connections between nodes are represented as weights. More complex architectures have more number of layers. An example of simple ANN architecture is shown in Figure 1.

In MLP, each hidden layer node actually consists of two parts; the first one contains the *summation function* which calculates the sum of each input value multiplied by the corresponding weight. Summation function can be represented by the following equation:

$$S = \sum_{i=0}^{n} w_i x_i \qquad (1)$$

Where n is the number of inputs to the neuron.

The second part of each neuron is the *Activation Function*. Activation Function is applied by each neuron to its net input (sum of weight input signals) to determine its output signal. This function is usually nonlinear. Sigmoid function such as the S-shaped curve is one of the most commonly used activation functions.

The use of the Sigmoid function has many advantages in the modeling of complex systems based neural. Sigmoid function transfers the results of the summation function to more tangible results using a simple function like in equation 2. In [12], Fausett states that this advantage is because "the relationship between the value of the function at a point and the value of the derivative at that point reduces the computational burden during training". An example of sigmoid function which is used in this work is given by Equation 2.

$$\emptyset(S) = \frac{1}{1 + e^{-S}} \qquad (2)$$

B. The back-propagation learning algorithm

ANN uses a learning algorithm in order to minimize some error criteria. In our experiment we will depend on the back propagation algorithm as a learning algorithm. The back propagation algorithm is an abbreviation for "backwards propagation of errors" created by Rumelhart and McClelland, in 1986, it is one of the most common methods of training artificial neural networks, from the selected inputs.

The back propagation algorithm consists of two phases [26,27]: first one called *propagation phase* where the activations are forwarded from the input to the output layer by adding all the weighted inputs to neurons and then computing the activation output using the sigmoid function. In the second step of the *propagation phase*, the model calculates the error which is the difference between the actual and the predicted value. The error between the desires value and the actual result is propagated backward from the output layer in order to modify the weights[28,29,30].

The second phase is the *weight update* phase. In this phase, the steepest gradient descent method is used to adjust the ANN weights such that a minimization criterion is reached [31].

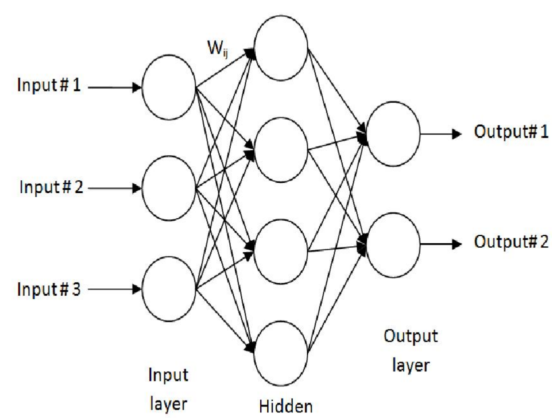

Figure 1. Feed forward MLP architecture.

3. Dataset

For the purpose of this research, real data were provided by Umniah which is a major cellular telecommunication network in Jordan. It is the third GSM cellular phone provider set up in Jordan. Umniah were founded in 2005 and by May 2011, the company had more than 2 million mobile and 25,000 broadband subscribers.

The data set contains 11 attributes of randomly selected 5000 customers for a time interval of three months. The last attribute indicates whether the customer churned (left the company) or not. The total number of churners is 381 (7.6% of total customers).

The attributes along with their description are listed in Table1.

Table 1. Umniah dataset attributes.

Attribute name	Description
3G	The subscriber is provided with 3G service (Yes, No)
Total Consumption	Total monthly fees (calling +SMS) in (JD)
Calling fees	Total monthly calling fees (JD)
Local SMS fees	Monthly local SMS fees(JD)
Int'l SMS fees	Monthly fees for international SMS (JD)
Int'l calling fees	Monthly fees for international calling (JD)
Local SMS count	Number of monthly local SMS
Int'l SMS count	Number of monthly international SMS
Int'l MOU	Total of international outgoing calls in minutes
Total MOU	Total minutes of use for all outgoing calls
On net MOU	Minutes of use for on-net-outgoing calls
Churn	Churning customer status (Yes, No)

4. Experiments and results

A. Evaluation methods

In order to assess the developed ANN models, we use the confusion matrix shown in Table 2 which is the primary source for evaluating classification models. Based on this confusion matrix, the following three different criteria are used for the evaluation:

Table 2. Confusion matrix.

		Predicated	
		Non churn	churn
Truth	positive	tp	fn
	negative	fp	tn

1. Accuracy - measures the rate of the correctly classified instances of both classes.

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn}$$

2. Hit rate - measures the rate of predicted churn in actual churn and actual non-churn

$$hit\ rate = \frac{tn}{fn + tn}$$

3. Churn rate - measures the rate of predicted churn in actual churn.

$$Churn\ rate = \frac{tn}{fp + tn}$$

B. Cross validation

To give a better indication of how well the developed ANN classification model will perform when new data are presented to the model, *k*-cross validation is applied with *k*=5. Umniah data set described in the previous section is split into 5 random subsets of equal sizes. The first four subsets are used for training and the last one is used for testing. Therefore, all criteria described earlier are found for the testing subset. Then the same process is repeated for different four subsets and another subset is used for testing. Consequently, this process is repeated five times. Finally, the average value for the five different tests is calculated.

Table 3. Fixed parameters.

Parameters	Values
Learning rate	0.3
Momentum	0.2
Hidden layers	1

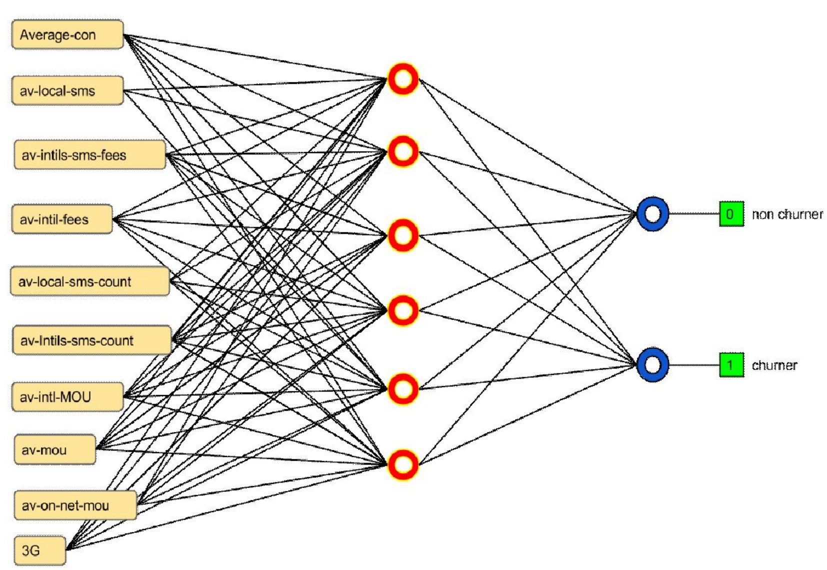

Figure 2. Multilayer Perceptron Neural Network (MLP).

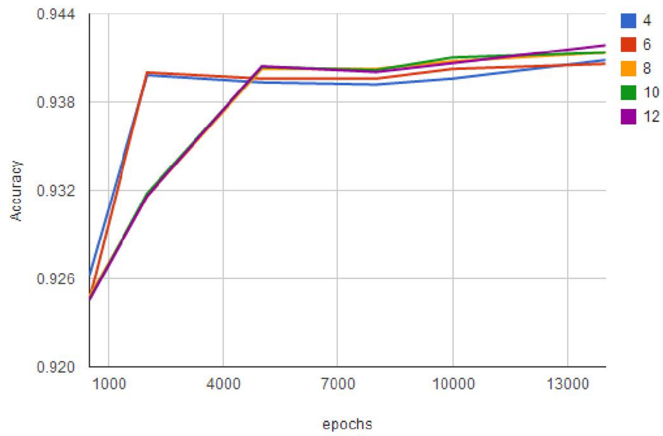

Figure 3. Accuracy rate for MLP with different number of neurons in the hidden layer.

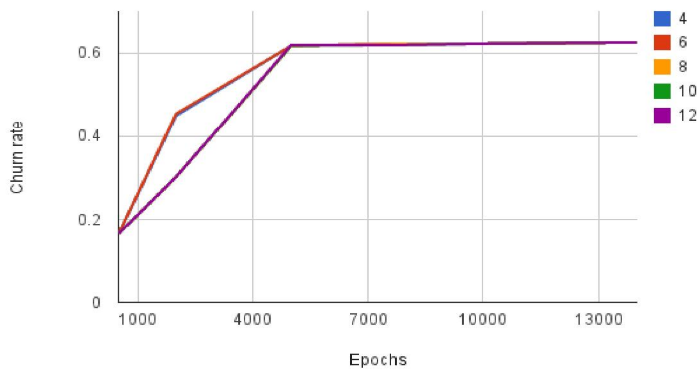

Figure 4. Churn rate for MLP with different number of neurons in the hidden layer.

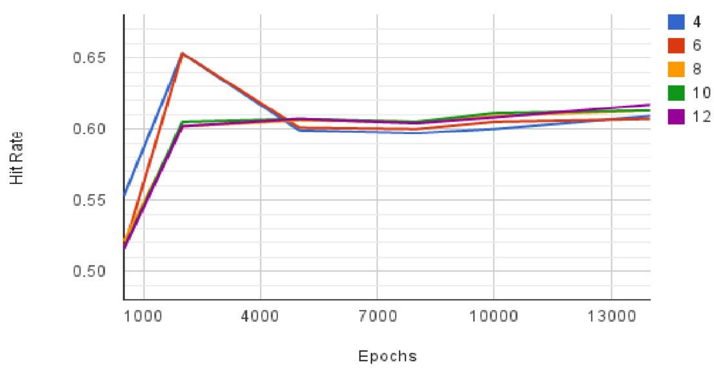

Figure 5. Hit rate for MLP with different number of neurons in the hidden layer.

C. Tuning ANN parameters

MLP is a parametric approach where the initial values of the parameters can affect the quality of the developed model. Such parameters include; setting number of epochs and number of neurons in the hidden layer. In this part of our experiments, we study the effect of changing number of epochs and number of neurons in the hidden layer (i.e; 2, 4, 6, 8, 10 and 12) on the quality of the model while the other parameters are fixed. Fixed parameters are listed in Table 3. Figure 2 shows an example of MLP model with 6 neurons in the hidden layer for the churn prediction problem. Figure 3, 4 and 5 show that the quality of the MLP model according to the accuracy, churn rate and hit rate varies by changing number of epochs each time. In Figure 3, we notice that MLP models with 4 and 6 parameters in the hidden layer converge to a better accuracy faster than the other models. In Figure 4, all models achieve the best churn rate of 62% at 5000 epochs. For our case, the best model obtained has 4 neurons in the hidden layer with 5000 epochs. This model will be used as a base model to perform the *Impact Factor analysis* in the next section.

D. *Impact Factor analysis*

To study the influence of each customer attribute on the churn, we apply two different approaches. The first is *Change on Error (CoE)* approach. CoE ranks input variables according to the change of some quality criteria when each input is deleted from the dataset in the training process. In our case, there are 10 input variables. In order to measure the importance of each variable, the dataset is trained 10 times with different 9 variables excluding one each time. Therefore, attributes that makes the largest change in the Accuracy and Churn rate are considered the most important. We performed CoE on our dataset twice, using the two different criteria each time. First we measured the change on accuracy then we measured the change in churn rate criteria. Results of applying CoE for both cases are shown in Figure 6 and Figure 7, respectively. It can be noticed from the two figures that the Total monthly fees, Total of international outgoing calls and 3G service have the highest impact. The disadvantage of this approach is that noise and redundancy in the training set can lead to degradation in its reliability [13-17].

The second approach used to identify the relative importance of input variables is based on the final connection weights of the neural network. Different equations were proposed in the literature based on magnitudes of the weights. In this work, we apply the basic one which was introduced by Garson in [18-20]. This approach can be represented by the following equation

$$RI\% = \frac{\sum_{j=1}^{h} \frac{|w_{ji}||w_{kj}|}{\sum_{i=1}^{n}|w_{ji}|}}{\sum_{i=1}^{n}\sum_{i=1}^{h} \frac{|w_{ji}||w_{kj}|}{\sum_{i=1}^{n}|w_{ji}|}}$$

Where n is the number of input nodes, h is the number of neurons in the hidden layer and k is number of neurons in the output layer.

By applying this equation to the matrix of weights of the best MLP model, we get relative frequencies for customers attributes as shown in Figure 8. We can notice that Garson method identified Total monthly fees, Total of international outgoing calls in minutes, Total minutes of use for all outgoing calls and 3G service as the most influencing method on churn.

It is important to point out that there are three attributes identified in common by both methods, CoE and Garson's one. These attributes are Total monthly fees, Total of international outgoing calls and 3G service. Such conclusion can be highly valuable for Customer Relationship Management in telecommunication companies in order to plan more effective strategies for churn handling.

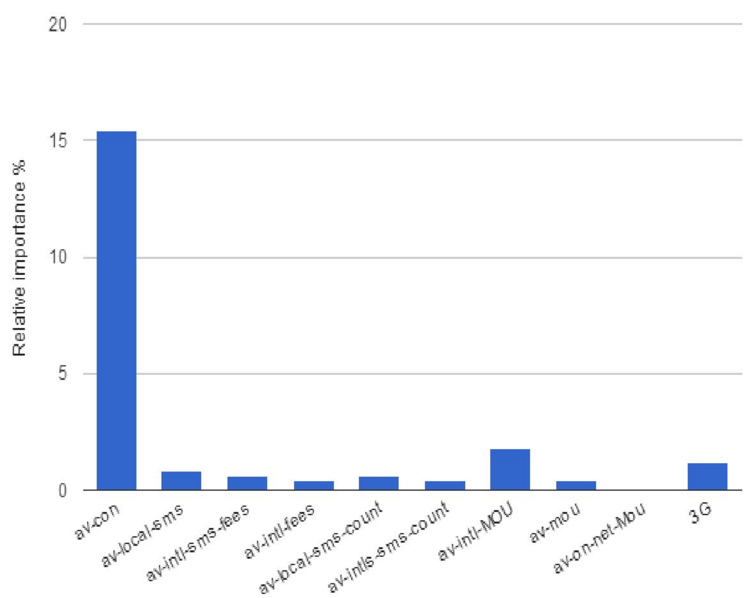

Figure 6. Relative importance of customer attributes using CoE method based on accuracy criteria.

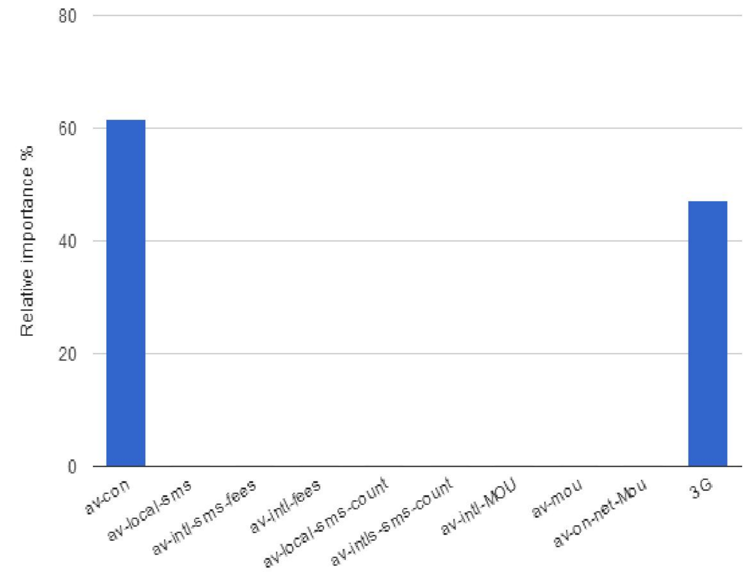

Figure 7. Relative importance of customer attributes using CoE method based on churn rate criteria.

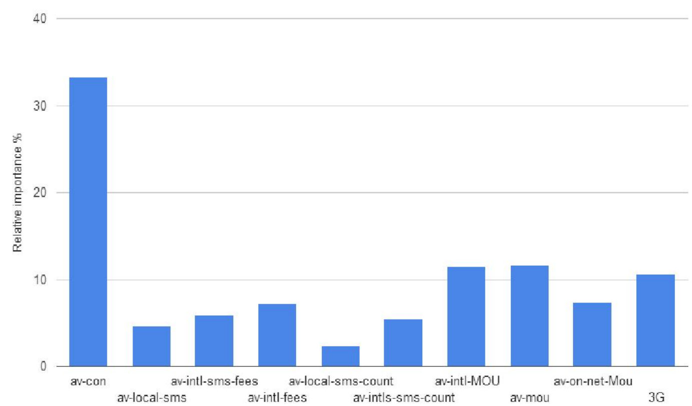

Figure 8. Relative importance of customer attributes in churn rate using weights method.

5. Conclusion

In this work we investigated the application of multilayer perceptron neural networks with back propagation learning for churn prediction in a Jordanian telecommunication company. Unlike most of previous studies, we used the developed MLP model for identifying of most influencing factors in customers churn. Two methods were examined and compared; the typical change on error method and the ANN weights based method.

Acknowledgements

Authors would like to thank the Jordanian mobile telecommunication operator Umniah for providing the required technical assistance and the data for this research work.

References

1. Y. Seog Kim, H. Lee, and J. Johnson, "Churn management optimization with controllable marketing variables and associated management costs," Expert Systems with Applications, 2012.
2. A. Keramati and S. Ardabili, "Churn analysis for an Iranian mobile operator," Telecommunications Policy, vol. 35, no. 4, pp. 344–356,2011.
3. T. Saba, A. Rehman, M. Elarbi-Boudihir (2014). Methods and strategies on off-line cursive touched characters segmentation: a directional review, Artificial Intelligence Review, vol. 42 (4), 1047-1066 doi.10.1007/s10462-011-9271-5.
4. T. Saba, A. Rehman, G. Sulong (2010) Improved Offline Connected Script Recognition Based on Hybrid Strategy, International Journal of Engineering Science and Technology, vol. 2 (6), pp. 1603-1611.
5. H.S. Kim and C.H. Yoon, "Determinants of subscriber churn and customer loyalty in the Korean

mobile telephony market," Telecommunications Policy, vol. 28, no. 9, pp. 751–765, 2004.
6. J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," Computers and Operations Research, vol. 34, no. 10, pp. 2902 – 2917, 2007.
7. J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn prediction using complaints data," in Proceedings Of World Academy Of Science, Engineering and Technology, 2006.
8. T. Saba, A. Rehman and G. Sulong (2011) Improved statistical features for cursive character recognition, International Journal of Innovative Computing, Information and Control (IJICIC) vol.7(9), pp. 5211-5224.
9. C.F. Tsai and Y.H. Lu, "Customer churn prediction by hybrid neural networks," Expert Systems with Applications, vol. 36, no. 10,pp. 12547–12553, 2009.
10. R. Obiedat, M. Alkasassbeh, H. Faris, and O. Harfoushi, "Customer churn prediction using a hybrid genetic programming approach," Scientific Research and Essays, vol. 8, no. 27, pp. 1289–1295, 2013.
11. T. Saba, A. Rehman, A. G. Sulong, (2010). An intelligent approach to image denoising, Journal of Theoretical and Applied Information Technology, vol. 17 (2), 32-36.
12. S.Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," Expert Systems with Applications, vol. 31(3), pp. 515–524, 2006.
13. P. C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services," Expert Syst. Appl., vol. 36, pp. 6714–6720, Apr. 2009.
14. T. Saba, G. Sulong, A. Rehman(2010) Non-linear segmentation of touched roman characters based on genetic algorithm, International Journal on Computer Science and Engineering, vol. 2 (6), pp.2167-2172.
15. L. V. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Prentice Hall, 1994.
16. A. SUNG, "Ranking importance of input parameters of neural networks," Expert Systems with Applications, vol. 15, no. 34, pp. 405 – 411, 1998.
17. AR Khan, Z Mohammad (2008). A simple segmentation approach for unconstrained cursive handwritten words in conjunction with the neural network, International Journal of Image Processing (IJIP) vol. 2 (3), pp. 29-35.
18. A. Rehman, S. Alqahtani, A. Altameem, T Saba (2013). Virtual machine security challenges: case studies, International Journal of Machine Learning and Cybernetics, vol. 5(5), pp.729–742, doi. 10.1007/s13042-013-0166-4.
19. F. Kurniawan, A.R. Khan, D Mohamad (2009). Contour vs non-contour based word segmentation from handwritten text lines: an experimental analysis, International Journal of Digital Content Technology and its Applications, vol. 3(2) , pp. 127-131.
20. K. Meethongjan, M. Dzulkifli, A. Rehman, A. Altameem, T Saba.(2013) An intelligent fused approach for face recognition Journal of Intelligent Systems, vol. 22 (2), pp. 197-212, doi: 10.1515/jisys-2013-0010.
21. F. Kurniawan, MSM Rahim, D.Daman, A. Rehman, M Dzulkifli, S M. Shamsudi (2011) Region-based touched character segmentation in handwritten words, International Journal of Innovative Computing, Information and Control, Vol. 7(6), pp. 3107-3120.
22. T Saba, A Rehman, G Sulong (2010) Cursive script segmentation with neural confidence Int. J. Innovative Comput. Inform. Control (IJICIC) vol. 7 (8), 4955-4964.
23. Rehman, A. and Saba, T. (2014). Evaluation of artificial intelligent techniques to secure information in enterprises, Artificial Intelligence Review, vol. 42(4), pp. 1029-1044, DOI. 10.1007/s10462-012-9372-9.
24. A Rehman, D Mohammad, G Sulong, T Saba (2009). Simple and effective techniques for core-region detection and slant correction in offline script recognition,Proceedings of IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 15-20.
25. M. Harouni, MSM Rahim, M. Al-Rodhaan, T. Saba, A. Rehman, A. Al-Dhelaan (2014) Online Persian/Arabic script classification without contextual information, The Imaging Science Journal, vol. 62(8), pp. 437-448, doi. 10.1179/1743131X14Y.0000000083.
26. A Nodehi, G Sulong, M Al-Rodhaan, A Al-Dhelaan, A Rehman, T Saba (2014) Intelligent fuzzy approach for fast fractal image compression, EURASIP Journal on Advances in Signal Processing, doi. 10.1186/1687-6180-2014-112.
27. K. Meethongjan, M. Dzulkifli, A. Rehman, A. Altameem, T. Saba (2013) An intelligent fused approach for face recognition, Journal of Intelligent Systems, vol. 22 (2), pp. 197-212, doi. 10.1515/jisys-2013-0010.
28. K. Neamah, D. Mohamad, D. T. Saba, T. A. Rehman (2014). Discriminative features mining for offline handwritten signature verification, 3D Research vol. 5(3), doi. 10.1007/s13319-013-0002-3.
29. S.Joudaki, D. Mohamad, T. Saba, A. Rehman, M. Al-Rodhaan, A. Al-Dhelaan (2014) Vision-Based Sign Language Classification: A Directional Review, IETE Technical Review, Vol.31 (5), 383-391, doi. 10.1080/02564602.2014.961576.
30. T. Saba S. Al-Zahrani S, A. Rehman (2012) Expert system for offline clinical guidelines and treatment Life Sci Journal, vol.9(4), pp. 2639-2658.
31. H. Haron, A. Rehman, DIS. Adi, S.P. Lim and T. Saba (2012). Parameterization method on B-Spline curve. mathematical problems in engineering, vol. 2012, doi:10.1155/2012/640472.

1/19/2014