APPLICATION OF ARTIFICIAL NEURAL NETWORK (ANN) IN ESTIMATION OF BODY MASS INDEX (BMI) BASED ON THE CONNECTION BETWEEN ENVIRONMENTAL FACTORS AND PHYSICAL ACTIVITY

Seyed Hosein Hoseini *, Ali Soltani**and Meisam Pourahmadi-Nakhli***

* School of Art and Architecture, Shiraz University, Shiraz, Iran
Hoseini860459@gmail.com
* Assistant Professor, School of Art and Architecture, Shiraz University, Shiraz, Iran
(Corresponding author)
soltani@shirazu.ac.ir
*** School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
Meisam.pourahmadi@gmail.com

ABSTRACT

One of the main concerns of people in modern societies is increasing the Body Mass Index (BMI) level. BMI, in fact, can be considered as an indicator of overall health condition. Genetic aspects aside, the BMI level is affected by different factors, such as socio-economic, environmental, and physical activity level. This study investigated the effect of different factors on the BMI level of a sample population of 470 adults of three residential neighbourhoods in Shiraz, Iran. The Pearson correlation test, independent sample T-test and One Way ANOVA were used to extract the variables which significantly influenced the BMI. The statistical analysis showed that despite the apparent association of BMI with physical activity level, it is influenced by several factors such as age, residence record, number of children, distance to bus or taxi stop, indoor or sport exercise. Then, an Artificial Neural Network (ANN) was applied to predict the level of personal BMI. The results of this analysis showed that the generalized estimating ANN model was satisfactory in estimating the BMI based on the introduced pattern.

KEYWORDS

Body Mass Index, Artificial Neural Network, Socio-economics, Built Environment.

1. INTRODUCTION

One of the critical issues for urban planning is how to provide the citizens with an environment in which physical activity is welcomed. In this way, neighbourhood design basing on walking distance criteria is recommended. It is believed that, environmental features such as urban design, land use, buildings, trails, street lights, green and open spaces have potential impacts on likely physical activities by some extent [1]. In addition, socio-economic factors such as age, attitude, gender, job, housing and vehicle ownership affects level of citizens' physical activity. On the other hand, the level of physical activity is related the Body Mass Index (BMI) level as a significant criteria of health measure. BMI is a direct calculation based on weight in kilograms and height in meters (BMI=Weight/Height^2) [2]. Generalization of the environmental impact on

the BMI can generate new insights in health-oriented urban planning. An investigation on a random sample of inhabitants of Ghent, Belgium, on adults aged 18-65 years, revealed that neighbourhood design and recreational environmental factors had reasonably significant correlation with some types of physical activity [3]. A research in Minnesota found that the walking patterns for specific purposes change according to the physical character of the environment [4]. The study by Frank et al. (2005), in Metropolitan Atlanta showed that land use mix, residential density and intersection density were effectively related to daily time spent for moderate physical activity [5]. A survey of 1'200 adults aged between 40 and 60 years, of New South Wales, Australia showed that individual perception of physical environment plays a role in walking pattern and physical activity intensity. The study showed that individuals with negative perception of environment were less involved in physical activities. An attractive built environment provides further convenient access to facilities which in turn, influences on walking as well as overall physical activity [6]. The results of a research in Santiago California showed that the people, who used to walk further, were from neighbourhoods with higher aesthetics, richer land use mix, higher residential density and safer environment. Also, they had higher levels of physical activity and lower BMI [7]. Black and Macinko (2009) reported that environmental factors such as land use diversity and availability of fitness facilities were significantly associated with reduced obesity [8]. Gender does matter. The number of nearby recreational facilities and parks were positively correlated with girls' physical activities; while for boys the retail floor area was positively correlated with the physical activity levels [9].

To sum up, it can be deduced that built environment plays a potential role in decreasing or increasing residents' activities, which in turn influences the BMI level. The existence of relationship between various contributing factors and BMI gives an idea to establish a conceptual model between the explanatory factors and BMI in hope of evaluating the model within new scenarios.

2.5. Research Methods

In order to explore the impact of various factors on BMI, data were collected by questionnaire survey on a sample of adults aged between 15-65 years old lived in the three neighbourhoods of Metropolitan Shiraz, south of Iran. Questions covered three major areas: built environment features, socio-economic and weakly activity pattern. These three themes were assumed to be the major contributing factors for BMI. The questionnaire was based on the International Physical Activity Questionnaire (IPAQ) with very minor changes. The socio-economic variables comprised the age, height, total income of family per month, educational status, marital status, number of children, residence record, floor area, yard area, type and the number of vehicles available for household and accessibility to the internet. The factors related to built environment included the distance to different local facilities as well as resident's perception on their neighbourhoods in terms of neighbourhood safety, beauty, sufficiency of recreational facilities and their potential in breeding the physical activity. The questions about physical activity were in three sub-categories: first, walking and job commuting, intense and moderate indoor activities, and third professional exercise for sport and leisure.

2. Results and Analysis

A preliminary description of the collected data on some characteristics is initially provided. About 48.5 percent were women, 11.9 percent smoking and 30.1 percent on diet. About, 44.5 percent passed an academic education, 47 percent just finished high school, 34.7 percent had bachelor degree and 9.8 percent had master degree or over. Also, 4.3 percent were from one-family households, 33 percent were dependent on their parents, 10.9 percent were a member of families having no kids and 51.9 percent were from families had one child or more. From the

family income aspect, 56.2 percent had monthly income less than $1000, 38.3 percent were with monthly income about $1000 to $2000 and 5.5 percent enjoyed monthly income higher than $2000. Regarding accommodation 56.4 percent lived in town house, 27 percent lived in low-rise apartments and 16.6 percent of respondents lived in high-rise apartments. Also, about 65.1 percent of respondents had one car, 18.1 percent had two cars and 2.6 percent had three cars or more.

The bi-variate Pearson Correlation test was used to examine the correlation between the BMI level with other factors such as distance to bus stop, distance to taxi stop, physical activity level and so on; independent sample t-test was used to compare two groups: men and women, smoking and non-smoking participant, people with or without sport facilities at home and whom have not. Also, to compare three groups or more in terms of educational level and lifestyle, One Way Analysis of Variance (ANOVA) was applied.

The results are presented in Table 1. The first column indicates the Pearson correlation coefficient between the paired contributing factors and the second column includes the significance level of each variable with BMI. The higher the correlation coefficient, the more significant the relationship is.

Table1. The result of Pearson Correlation test

	Pearson Correlation	Sig. (2-tailed)	Mean	Std. Deviation
Age	0.43	0.00	37.46	±14.34
Residence record (year)	0.08	0.09	8.35	±7.37
Number of children	0.34	0.00	1.39	±1.54
Distance to bus stop (min.)	0.11	0.02	6.77	±5.02
Distance to taxi stop(min.)	0.09	0.04	6.67	±4.87
Indoor severe activity(min.)	-0.08	0.09	65.27	±234.86
Sport moderate activity(min.)	-0.13	0.01	33.54	±129.76
Sport severe activity(min.)	-0.08	0.08	59.18	±177.58

(=0.05, n= 470)

The factors with significant correlation whit BMI included age (r=0.43 , p<0.000), residence record (r=0.08, p<0.09), number of children (r=0.34, p<0.00), distance to bus stop (r=0.11, p<0.02), distance to taxi stop (r=0.09, p<0.04), indoor severe activity (r= -0.08, p< 0.09), sport moderate activity (r= -0.13, p<0.01) and sport severe activity (r=-0.08, p< 0.08). Interestingly, all types of physical activity showed negative relationship with BMI.

An independent sample t-test was used to examine the difference between binary groups in terms of gender, smoking, diet status and having sport facilities at home or not. The results revealed that BMI was significantly different between men and women (t=-3.53, p<0.00), have and not have diet (t=-2.39, p<0.02) and have and not have sport facilities at home (t=2.7, p<0.03). However no significant difference between smoking and non smoking adults was found.

One-Way ANOVA was applied to examine the effects of education, lifestyle, housing, family income and vehicle model and accessibility to the internet on BMI level. The analyses indicated that adults with different literacy status (f= 4.443, p<0.004) or different lifestyle (f= 35.54, p<0.000) had different BMI level. On the other hand, other grouping variables such as vehicle class, income level and accessibility to the internet made no significant difference in BMI level.

3.1. Artificial neural network –based BMI estimator

ANN is a novel computational tool constructed with some interconnected principal units called neurons (Fig 1). Pattern recognition is one of the most important applications of the Feed Forward Artificial Neural Networks (FFANN). Two important types of FFANN are Multi Layer Perceptron (MLP) and Radial Basis Function (RBF), selecting the most suitable structure for ANN and thanks to nonlinear nature it can adapt to approximate even such complex patterns – between some impact factors and targets – that conventional models never afford (Pourahmadi-Nakhli and Safavi, 2011)[10]. The MLP-ANN scheme is shown in Fig. 1.

Fig.1. The ANN scheme
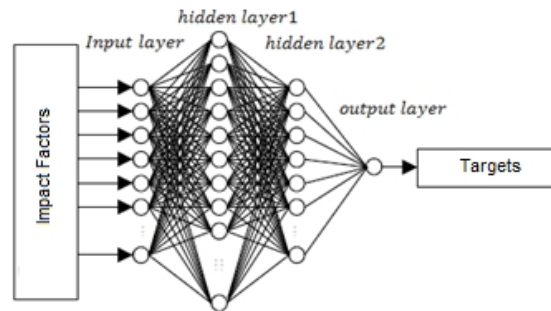
At least two studies can be found discussing the application of ANN in estimating physical activities which used the information collected by accelerometer, a promising tool for recording physical activity patterns in daily life. Staudenmayer et al. (2009) attempted to measure and classifying physical activity level by ANN and under the data collected with a uniaxial accelerometer for low-level activities, locomotion, vigorous sports, and household activities. Their approach led to a true classification about 89 percent of the time [11]. Similarly, Rothney et al. (2007) evaluated the energy consumption of sampled adults collected by biaxial raw (32 Hz) acceleration signals. The results showed a considerable reduction in estimation error comparing with regression models [12]. In this study, by using a state-of-art approach, it is tried to extend the intrinsic correlation of physical activities, socioeconomic and environmental factors, into a pattern to be estimated by an intelligent pattern estimator model.

3.2. Training process using relationship between BMI and impact factors

Artificial neural networks (ANN) aim to create a model considering relationships between some input and output data. The data set must be representative of all scenarios possible with low redundancy as well as with adequate dimensions. Providing the data set under what a model develops is the first step. Although ANN has high potential in approximation the complex patterns, it may fail in predicting the patterns with low correlation between input and output. So, herein, the key was selecting the data set which reflected the goal function as well as possible. Therefore, the significant link between BMI and various environmental, socio-economic and activity-based factors suggests a significant pattern to be learned by ANN model on the basis of statistical correlation between BMI and different influential factors, the input-output set to the ANN model (Table 2).

Table 2. ANN input variables

Input No.	Significant impact factors
1	Age
2	Residence record
3	Number of children
4	Distance to bus stop
5	Distance to taxi stop
6	Indoor severe activity
7	Sport moderate activity
8	Sport severe activity
9	Lifestyle
10	Gender
11	Diet
12	Sport facility in home
13	Education level
14	BMI

One of the major concerns of experts in ANN implementation is selecting the best structure under which model reveals precise estimation with lowest number of layers and neurons [10]. Two MLP structures for two different training algorithms and one RBF network are developed for estimation purpose. The first MLP network comprises 23 neurons for the input layer, 25 neurons for only one hidden layer and one neuron for output layer. Another MLP of four layers (two hidden layers) is selected with 20 neurons for the input layer, 15 and 5 neurons for the hidden layers and one neuron for the output layer. Hidden layer was used as it managed a better prediction with less number of nodes. One Hyperbolic tangent (Tansig) function was used for the hidden layers and a linear transfer function (Purelin) applied for the output layer.

$$\varphi(s) = \frac{1}{1+e^{-s}} \qquad (1)$$

Training is the step which goes to adjust parameters of priory structure. It is tried to fit parameters of a nonlinear model through the common optimization methods such as maximum likelihood and least square methods constraining inputs meet outputs. The training process was performed by Levenburg-Marguardt (LM) technique and Resilient back Propagation (RP) techniques. Network structures and training technique were selected based on trial and error approach to achieve the best estimation results. The RBF structure consists of 14 neurons for the input layer, one hidden layer of 15 neurons with *Gaussian* radial basis functions and 1 neuron for the output layer.

## 3.3 Evaluation of neural BMI estimator

The most important step of implementation is evaluating the trained ANN within new scenarios never used before to test the accuracy of model on its pivotal estimation task. This stage evaluates the ability the trained neural network to measure generalization under new scenarios that neural fault locator never experienced before. New scenarios are coherently randomly sampled within the total society. Here 30 input vectors which never used in training stage are introduced to ANNs. For a better judgment, five criteria including the mean value, Standard Deviation (SD), the correlation coefficient, the Normalized Root Mean Square of Errors (NRMSE) and the

Maximum Absolute of Errors (MAE) between ANN results and target values are calculated. The NRMSE is defined as:

$$\text{NRMSE} = \sqrt{\frac{\sum_{n}[Y(n) - T(n)]^2}{\sum_{n} T(n)^2}} \tag{2}$$

R_ criterion reveals how well the estimated values are represented by Target values. The nearer R_ criterion to 1, the higher correlation between Y and T exists. Linear fit also represents the accuracy of the estimation so zero crossing line with the slope of 1 shows a perfect estimation(Y=T). Figure 2 (a), (b) and (c) visually reveals the performance of MLP (RP), MLP (LM) and RBF, respectively under the new scenarios. From linear fit and R_ criterion, one can see that RBF whose MAE=8.582 and NRMSE=0.190 yields a less satisfactory results in comparison with the MLP-based networks.
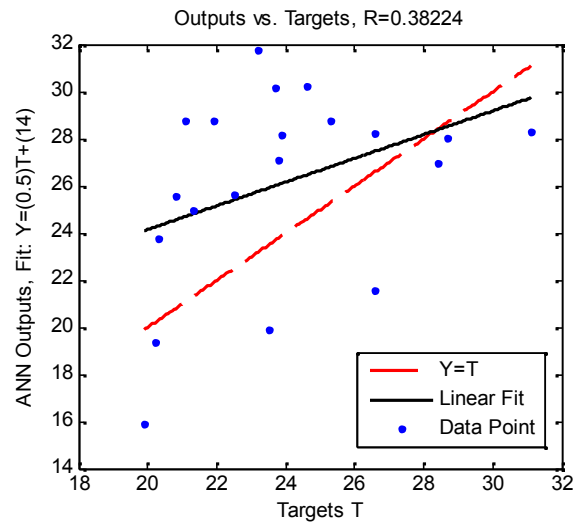
The MLP-based networks.
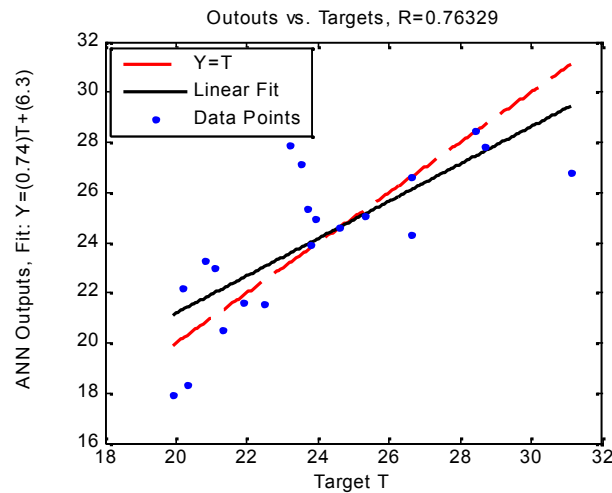
(2 a)
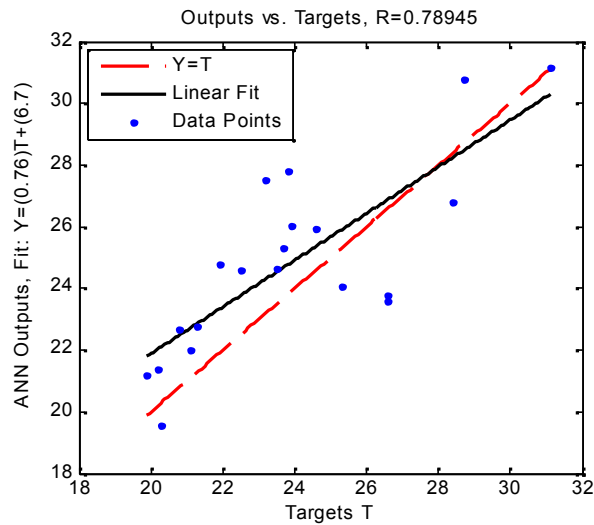
(2 b)

(2 c)

Figure 2. Idealized estimated Y values versus target T values. (a) MLP (RP), (b) MLP(LM), (c) RBF

Overall performance of the neural three estimators at the test stage is shown in Table 3. The estimation error of three neural estimators at test stage (Fig. 3) is calculated by:

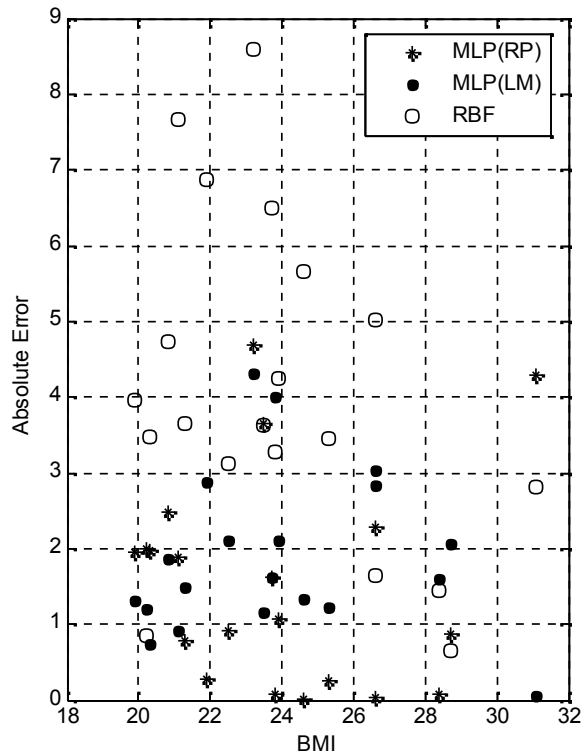Absolute Error = |real BMI − neural BMI estimator output |

Fig 3. Absolute Error of 30 estimation case for three ANN models

It can be concluded from the result, that the proposed method leads to estimation with satisfactory result of Table 3. Nevertheless, there is no perfect fit in estimation of BMI; because, this index greatly is influenced by genetic traits. Furthermore, some external factors including environmental, socio-economic, and physical activity-based factors have considerable influences on BMI. The estimated results of MLP-based ANN prove that the significant influence of input factors can be reasonably generalized to new scenarios with similar introduced patterns.

Table 3. MAE, NRMSE criteria for train and the test result

Net. Type	Net. Size	NRMSE	MAE	Mean	Std	R
MLP (RP)	23×18×1	0.086	4.688	0.198	2.124	0.763
MLP (LM)	23×25×1	0.089	4.310	0.948	1.991	0.789
RBF	23×24×1	0.190	8.582	2.226	4.110	0.382

3. CONCLUSIONS

In this study the impacts of various environmental and non-environmental factors on BMI as one of important health indicators was explored. The statistical examinations on the data collected by questionnaire survey revealed that there was a reasonable relationship between the environmental/non-environmental impact external factors and the target (BMI). This relationship formed the idea of generalizing an estimating model and evaluation of the reliability to the generalized model in estimating new scenarios. For this purpose, an (ANN)-based BMI estimator was trained to establish a model reflecting the link between BMI and its influential factors. Although BMI is eventually calculated by body characteristics, the neural BMI estimator anticipated satisfactory results in BMI estimation basing on indirect input factors associated with the environment and personal physical activities.

This study used some statistical tests to discover the intrinsic relationship, yet more advanced techniques such as factor analysis and principal component analysis (PCA) method can help improving the analysis to achieve more accurate estimation.

REFERENCES

[1]   Rutt, CD. and Karen, JC., "Examining the Relationships Among Built Environment, Physical Activity, and Body Mass Index in El Paso, TX", Preventive Medicine, Vol. 40: 831-841, 2005.
[2]   Rolls, BJ. "The Supersizing of America: Portion Size and the Obesity Epidemic," Nutrition Today, 38, No. 2: 42–53, 2003.
[3]   Bourdeaudhuij, DI., Sallis, FJ., and Saelens, EB., "Environmental Correlates of Physical Activity in a Sample of Belgian Adults", American Journal of Health Promotion, Vol. 18: 83-92, 2003.
[4]   Forsyth, A., Hearst, M., Oakes, JM., and Schmitz, HK., "Design and Destinations: Factors Influencing Walking and Total Physical Activity", Urban Studies Journal, Vol. 18: 1973–1996, 2008.
[5]   Frank, DL., Schmid, LT., Sallis, FJ., Chapman, J., and Saelens, EB., "Linking Objectively Measured Physical Activity with Objectively Measured Urban Form Findings from SMARTRAQ". American Journal of Preventive Medicine, Vol. 28: 117–125, 2005.
[6]   Carnegie, MA., Bauman, A., Marshall, AL., Mohsin, M., Westley-Wise, V., and Booth, ML." Perceptions of the Physical Environment, Stage of Change for Physical Activity, and Walking Among Australian Adults", Research Quarterly for Exercise and Sport, 146-155, 2002
[7]   Saelens, BE., Sallis, JF., Black, JB., and Chen, D., "Neighbourhood-based Differences in Physical Activity:An Environment Scale Evaluation", American Journal of Public Health, Vol.93: 1552-1558, 2003.
[8]   Black, JB., and Macinko, J., "The Changing Distribution and Determinants of Obesity in the Neighbourhoods of New York City, 2003–2007. American Journal of Epidemiology, Vol. 171: 765-775, 2009.

[9]   Norman GJ, Nutter SK, Ryan S, Sallis JF, Calfas KJ, and Patrick K. "Community Design and Access to Recreational Facilities as Correlates of Adolescent Physical Activity and Body-Mass Index", Journal of Physical Activity and Health, Vol. 3: 118-128, 2006.

[10] Pourahmadi-Nakhli, M., and Safavi, AA.," Path Characteristic Frequency Based Fault Locating in Radial Distribution Systems Using Wavelets and Neural Networks", IEEE Transactions on Power Delivery, 772-781, 2011.

[11] Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., and Freedson, P., "An Artificial Neural Network to Estimate Physical Activity Energy Expenditure and Identify Physical Activity Type from an Accelerometer", Jornal of Applied Physiology, Vol. 107: 1300–1307, 2009.

[12] Rothney, MP., Neumann, M., Béziat, A. and Chen, KY., "An Artificial Neural Network Model of Energy Expenditure Using Nonintegrated Acceleration Signal", Journal of applied physiology, Vol. 103: 1419–1427, 2007.

Authors

Master of Urban and Regional Planning from the Faculty of Art and Architecture, Shiraz Unive rsity, Shiraz, Iran

Assistant Professor of Urban Planning, Department of Urbanism, Faculty of Art and Architecture, Shiraz University, Shiraz, Iran.

Master Student of Power and Control Engineering, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran.