

## Research Article

# Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity

Sasan Golnaraghi, Zahra Zangenehmadar , Osama Moselhi, and Sabah Alkass

*Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, Canada*

Correspondence should be addressed to Zahra Zangenehmadar; [z\\_zange@encs.concordia.ca](mailto:z_zange@encs.concordia.ca)

Received 2 August 2018; Revised 5 October 2018; Accepted 19 November 2018; Published 2 January 2019

Guest Editor: Ali R. Vosoughi

Copyright © 2019 Sasan Golnaraghi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Productivity is described as the quantitative measure between the number of resources used and the output produced, generally referred to man-hours required to produce the final product in comparison to planned man-hours. Productivity is a key element in determining the success and failure of any construction project. Construction as a labour-driven industry is a major contributor to the gross domestic product of an economy and variations in labour productivity have a significant impact on the economy. Attaining a holistic view of labour productivity is not an easy task because productivity is a function of manageable and unmanageable factors. Compound irregularity is a significant issue in modeling construction labour productivity. Artificial Neural Network (ANN) techniques that use supervised learning algorithms have proved to be more useful than statistical regression techniques considering factors like modeling ease and prediction accuracy. In this study, the expected productivity considering environmental and operational variables was modeled. Various ANN techniques were used including General Regression Neural Network (GRNN), Backpropagation Neural Network (BNN), Radial Base Function Neural Network (RBFNN), and Adaptive Neuro-Fuzzy Inference System (ANFIS) to compare their respective results in order to choose the best method for estimating expected productivity. Results show that BNN outperforms other techniques for modeling construction labour productivity.

## 1. Introduction

Artificial Intelligence (AI) has been a powerful tool in the construction industry over the past decade. Several AI modeling techniques have been employed in the construction industry such as expert systems (ES) and Artificial Neural Network (ANN). Modeling labour productivity is challenging as it requires the quantification of substantial factors that affect labour productivity and consideration of influential factor interdependencies. Productivity is a delicate aspect of any construction project. Unquestionably, arriving at a definition of construction productivity can cause confusion because of the various different ways at defining it. Strictly speaking, productivity is a component of cost and is not a method for estimating the cost of resources; rather, it is a quantitative assessment of the correlation among the number of resources used and the amount of output made [1]. Productivity in construction is considered

as a measure of output achieved by a combination of inputs. Considering this perspective, two concepts for measuring productivity described in the literature are total factor productivity and circulating capital [2]. Total factor productivity is the most common construction productivity measurement technique where the output is measured against all inputs. Partial factor productivity is referred to as single-factor productivity, in which the output is measured against a single input or selected inputs. Partial factor productivity is a cost-effective model and very advantageous for developing strategy and assessing the state of the economy; however, it is not beneficial for contractors [2]. Circulating capital is any kind of capital that will be depleted during the course of a project, such as material and operating expenses, whereas fixed capital refers to any kind of capital that is not exhausted during the course of a project.

Productivity modeling has been a topic of interest for many researchers and the various models being developed

today can be classified into two major groups: statistical and AI. Regression analysis is the most common statistical method for modeling labour productivity. The main advantage to regression analysis is that a productivity model can be developed to reach anticipated clarification or forecasting levels with as few predictor variables as possible. However, for regression methods, the degree of relationship (linear and nonlinear) needs to be selected prior to model development. In AI modeling, ANN models are the most common for developing labour construction productivity. Unlike regression methods, the degree of relationship is not a concern in ANN modeling. Those studies that have applied various ANN methods to predict different types of productivity are discussed in the next section.

Lu et al. [3] estimated construction labour productivity using real historical data from local construction companies. They applied a Probability Inference Neural Network (PINN) model and compared it to a feed-forward back-propagation neural network model. AbouRizk et al. [4] developed a two-stage ANN model for predicting labour productivity rates. They stated that understanding input factors and having a sufficient historical database are the most important parts in productivity prediction. Later, [5] introduced a neural network model for defining the impact of a change order on labour productivity and found that the ANN model shows better performance in comparison to other techniques. Ezledin and Sharara [6] established an ANN model for productivity prediction in formwork activity, steel fixing, and concrete-pouring activities. Ok and Sinha [7] applied ANN to estimate the daily productivity of earthmoving equipment. Song and AbouRizk [8] presented a productivity model for steel drafting and fabrication productivity through ANN and discrete-event simulation using actual data. Oral and Oral [9] utilized a Self Organizing Map (SOM) to analyze the relationship between construction crew productivity and different factors. They also predicted productivity in given situations for ready-mixed concrete, formwork, and reinforcement crew. Data were collected randomly from a construction site in Turkey. They concluded that SOM can predict productivity better than regression methods due to its complexity. Muqem et al. [10] predicted production rate values for installation of beam formwork using ANN. Meanwhile, Mohammed et al. [11] predicted the productivity of ceramic wall construction using data from general contractor companies. They applied ANN since analysis required performing complex mapping of environment and management factors to productivity. AL-Zwainy et al. [12] developed a model for estimating construction labour productivity in marble finishing works. They used multilayer perceptron training through a BNN algorithm. Moselhi and Khan [13] ranked labour productivity-influencing parameters in construction using Fuzzy Subtractive Clustering, Neural Network Modeling, and Stepwise Variable Selection. They determined that work type, floor level, and temperature were the parameters with a larger effect on productivity. Heravi and Eslamdoost [14] considered 15 important factors in the motivation of labour, supervision sufficiency, and competency and suggested a model for labour productivity rate estimation. Aswed [15]

applied ANN for estimating the bricklayer (builder) productivity and modeled 13 productivity-influencing factors. El-Gohary et al. [16] proposed a framework to document, control, and predict contractor labour productivity using ANN and hyperbolic tan function. They considered factors at micro and macro/microlevels and applied the models to construction crafts, carpentry, and fixing reinforcing steel bars.

A considerable issue in the construction industry is that a lot of problems such as last-minute bids, design under pressure, and so on are analogy-based in form. Thus, ANN techniques as compared to other conventional practices are more appropriate in modeling construction industry problems that demand analogy-based resolutions [17]. There are four major steps for modeling analogy-based problems using ANN: (i) gathering historical data, (ii) building and configuring relevant network, (iii) initializing weights and biases, and (iv) training and validation step. In this research, four types of ANN were applied for modeling labour productivity. These were Backpropagation Neural Network (BNN), Radial Basis Network (RBF), Generalized Regression Neural Network (GRNN), and Adaptive Neuro-Fuzzy Inference System (ANFIS). A detailed explanation of all the applied modeling techniques will be discussed along with the data collection procedure in the following sections.

## 2. ANN Model Development

Construction projects are highly dynamic with many challenges in the areas of costs, delays and disruptions, impaired productivity, quality issues, safety aspects, materials unavailability, and escalation among others. These challenges are highly complicated in nature and information related to these challenges is vague. Therefore, construction projects are within the purview of ANN in which the given ambiguous information can be effectively interpreted in order to arrive at meaningful conclusions. In other words, because of ANN's capability to draw the relationships between input and output provided via a training dataset, ANN is suitable for nonlinear problems where vague information, subjective judgment, experience, and surrounding conditions are key features, and traditional approaches are insufficient to calculate the complex input-output relationship necessary for predicting construction labour productivity. This paper reviews the application of various ANNs in predicting construction labour productivity for formwork assembly with a limited given dataset. Figure 1 shows the overall flowchart of the research. The first step in the model development process was the choice of inputs from the available data and appropriate model outputs. Then, the data were processed through normalization and for handling missing data. The data were divided into training and testing, and the BNN, RBF, ANFIS, and GRNN AI models were applied to the datasets. The models were calibrated via a performance evaluation and were compared using a determination coefficient ( $R$ -squared), Mean Squared Error (MSE), and Mean Absolute Error (MAE). Ultimately, the best AI model was selected. Development of each model and their results is presented in the following sections.

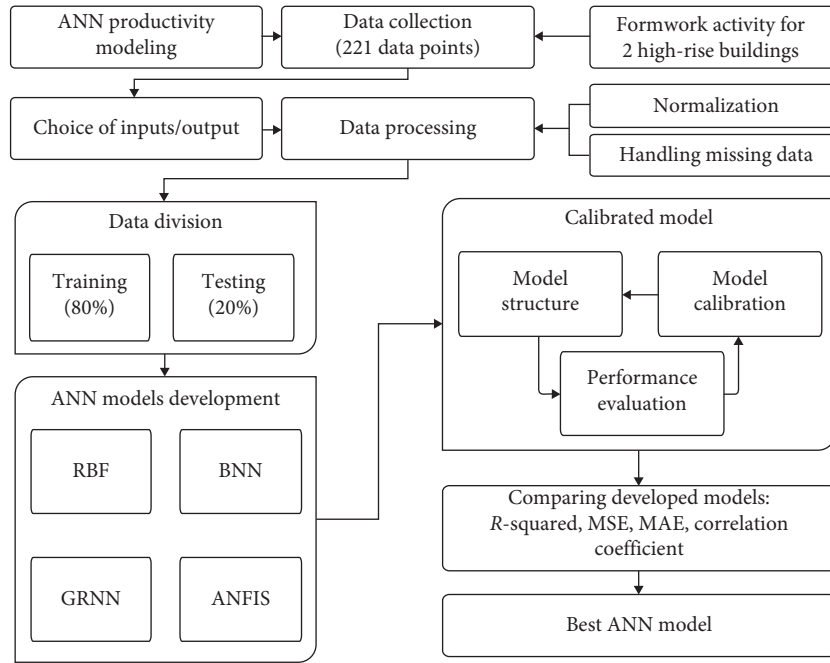


FIGURE 1: Overall flowchart of the study.

**2.1. Data Collection.** The dataset used for modeling labour productivity in this research was gathered by Khan [1] and collected from field observations and data collection from two high-rise buildings located in downtown Montreal over a period of eighteen months. The buildings were 17 and 16 floors. The first building is a concrete, mainly flat slab structure with Roller Compacted Concrete (RCC) construction and several typical levels with a surface of 68,000 m<sup>2</sup>. The project was constructed over three years. The second building also has a similar structure system and is a flat slab building. Two hundred and twenty-one data points were collected from both projects for formwork activity. The collected data was classified into three groups of weather, crew, and project. Data related to temperature, humidity, wind speed, and precipitation were classified into weather data while gang size and labour percentage as crew. Floor level, work type, and method were the parameters used in the project category. These variables were selected because they cause variations in productivity on a daily basis [1]. Data of nine factors classified into three major categories were available, as shown in Table 1, for performing the task.

These variables were chosen since they can cause differences in productivity on a daily basis or in the short-term. Short-term influence means that factors change value every day and do not have a cumulated or ripple effect impact on other activities. Thus, it is worthwhile to consider the abovementioned labour productivity factors for modeling labour productivity. Table 2 shows the descriptive statistics of the collected data which can provide a summary of the dataset.

The Pearson correlation coefficient is used to examine the strength and direction of a linear relationship between two variables in a database. A correlation coefficient ranges between  $-1$  and  $+1$ . A larger absolute coefficient value results in a stronger relationship between variables. In the case of a

TABLE 1: Labour productivity factors.

Weather	Crew	Project
Temperature (T)	Gang size (GS)	Work type (WT)
Humidity (H)	Labour percentage (LP)	Floor level (FL)
Wind speed (WS)		Work method (WM)
Precipitation (P)		

Temperature: based on an average of eight working hours a day (°C). Humidity: average humidity of eight working hours a day in percentage form. Precipitation: includes four numerical value terms assigned as: no precipitation = 0, light rain = 1, snow = 2, and rain = 3. Wind speed: average wind speed for eight working hours a day and included in the calculation in terms of km/h. Height: related to the floor being worked on and included in terms of the number of floors. Work type: three different types of formwork installation included as slabs = 1, walls = 2, and columns = 3. Work method: covers two techniques built in place (BIP) and flying forms (FF). BIP was coded as 1 and FF was coded as 2. Gang size: number of persons in a crew. Labour percentage: the ratio of labour to gang size obtained from the following equation:  $(\text{labour}\% = \text{labour size}/\text{gang size}) \times 100$ .

TABLE 2: Collected data descriptive statistics.

Variable	Mean	SE mean	Std. dev	Min.	Median	Max.
Temperature	4.08	0.81	12.03	-26	3	25
Humidity	66.34	1.05	15.67	18	67	97
Precipitation	0.28	0.04	0.6	0	0	3
Wind speed	15.42	0.57	8.46	3	14	43
Gang size	16.03	0.34	5.07	8	18	24
Labour percentage	35.49	0.26	3.79	29	36	47
Work type	1.43	0.03	0.51	1	1	3
Floor level	11.38	0.25	3.75	1	12	17
Work method	1.44	0.03	0.5	1	1	2
Productivity	1.57	0.02	0.35	0.82	1.51	2.53

Pearson correlation, an absolute value of 1 specifies a perfect linear relationship and a value of 0 indicates nonlinear relationship between variables. Table 3 shows that the

TABLE 3: Pearson correlation matrix for input and output parameters.

Variables	T	H	P	WS	GS	LP	WT	FL	WM	Productivity
T	1	0.151	-0.093	-0.122	0.390	-0.120	-0.122	0.358	0.078	0.589
H	0.151	1	0.338	0.026	-0.167	0.219	-0.110	0.048	0.176	0.090
P	-0.093	0.338	1	0.076	0.065	-0.063	-0.037	-0.288	-0.057	-0.175
WS	-0.122	0.026	0.076	1	-0.415	0.051	0.065	0.235	0.030	-0.202
GS	0.390	-0.167	0.065	-0.415	1	-0.310	-0.175	-0.352	-0.036	0.183
LP	-0.120	0.219	-0.063	0.051	-0.310	1	-0.135	0.142	0.177	-0.053
WT	-0.122	-0.110	-0.037	0.065	-0.175	-0.135	1	-0.052	-0.761	-0.353
FL	0.358	0.048	-0.288	0.235	-0.352	0.142	-0.052	1	0.225	0.301
WM	0.078	0.176	-0.057	0.030	-0.036	0.177	-0.761	0.225	1	0.328
Productivity	0.589	0.090	-0.175	-0.202	0.183	-0.053	-0.353	0.301	0.328	1

correlation between parameters most of the time is an approximate near 0. Consequently, none of the correlates very much and there is no over-estimation phenomenon. The Pearson  $p$ -values and  $R$ -squared prove the same the inputs and output behaviour.

**2.2. Backpropagation Productivity Modeling.** In this section, BNN was applied to model labour productivity. BNN is mostly used for unknown function approximation. As described in the literature review, a key BNN feature is its learning ability. It can be trained by historical datasets to find the accurate relation between inputs and outputs as well as predict the output(s) for new inputs. In this research, BNN models were developed, trained, validated, and tested in MATLAB 2017a with 221 data points. The dataset was randomly divided into 80% and 20% groups used for training and testing results, respectively. Several BNN models were developed which were different in three aspects: number of neurons in hidden layer varied between five and 100, random groups of datasets, and the number of hidden layers of one and two.

Bayesian Regularization (BR) algorithm was used for data training. BR is commonly used in noisy and small problems. The algorithms attempted to minimize the sum of the squared errors by updating the network's bias and weight. Training sets were used to adjust the network structure based on the associated errors until the best structure was reached. Validation sets were utilized to measure network generalization capabilities and to pause training when generalization stopped improvement. After training, testing sets provided an independent network performance index. For each BNN model, trials were performed to reach lowest error. Model performance was assessed based on  $R$ -squared and MSE developed through MATLAB coding according to the following equations, where " $t_i$ " is the target value while " $o_i$ " is the output value:

$$R^2 = 1 - \frac{\sum_i (t_i - o_i)^2}{\sum_i (t_i - (1/n)\sum_i t_i)^2} \quad (1)$$

$$MSE = \frac{1}{n} \sum_i (t_i - o_i)^2 \quad (2)$$

$R$ -squared is often used in statistical analysis since it is easy to calculate and understand. It fluctuates between [0, 1]

and evaluates the percentage of total differences between estimated and target values with respect to the average. Several BNN models with different numbers of neurons and hidden layers were developed to find the best model for identifying labour productivity. The number of hidden layers varied between one and two and the models were trained by five, ten, . . . , 100 neurons. Considering the differences in the number of neurons and the hidden layer, 32 different models were developed and their results compared.

Figures 2 and 3 display the effect the number of neurons has on the  $R$ -squared for one and two hidden layers. As can be seen from Figure 2,  $R$ -squared values are mostly between 90 and 100% for the training phase and 70–90% for the testing phase. The model with one hidden layer and 50 neurons shows maximum accuracy. For the two hidden layers' models, the model with 20 neurons in each layer showed the best performance in predicting labour productivity.

Increasing the number of hidden layers resulted in better performance; however, this approach takes more computational time and does not change model accuracy in any significant way. In this research, the maximum number of neurons that could be considered in the ANN model was set at 60 due to the extreme computational time needed for what is an insignificant improvement to model accuracy. Figures 4 and 5 show that the MSE value was the smallest in the models with two hidden layers and 20 neurons and one hidden layer and 50 neurons.

It should be mentioned that no performance index was available during the validation phase while the given datasets were trained by the BR algorithm because the algorithm does not validate data and the datasets were randomly divided into trained and tested datasets only.

Model performances were assessed based on  $R^2$  and MSE, as summarized in Table 4 for one hidden layer and Table 5 for two hidden layers. As can be seen, the final model was the one with two hidden layers and 20 neurons, which showed the highest accuracy for identifying labour productivity. The model had MSE and  $R^2$  performance value indices of 0.0054 and 0.949, respectively. Therefore, this model was considered for comparison with the ANIFS model in the next sections.

The error histogram of the final model with two hidden layers and 20 neurons demonstrates that most of the errors oscillate between  $-0.55$  and  $0.75$  in all training and testing



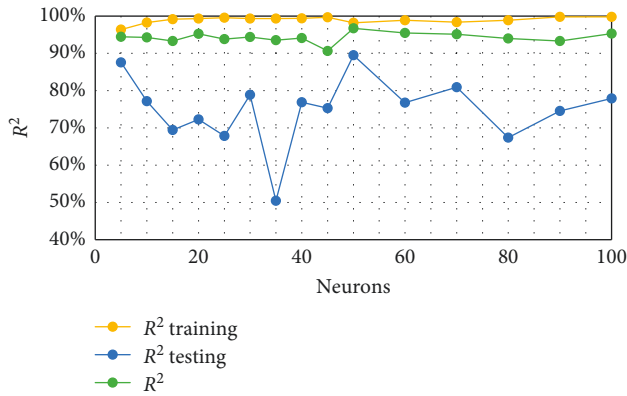


FIGURE 2:  $R^2$  values of ANN models using BP algorithm with 1 HL.

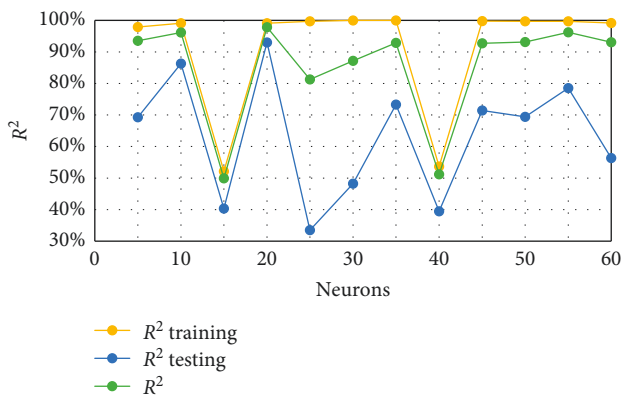


FIGURE 3:  $R^2$  values of ANN models using BP algorithm with 2 HL.

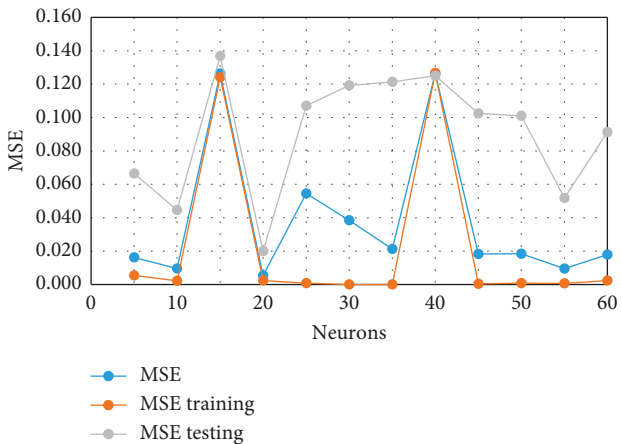


FIGURE 4: MSE values of ANN models in 1HL for distinct phases of ANN.

phases. The concentration of errors was 0.003, which is a small error for prediction. The  $R$ -squared in the selected model shows the fitted line for all data as output =  $0.97 \times$  target + 0.099 and an  $R^2$  value of 97.68% in the training dataset, demonstrating that the outputs are very close to target values. The  $R^2$  value for testing was 83.27%, proof that the model is able to predict 83% of future outcomes accurately. The applied method is able to rank predictor variables

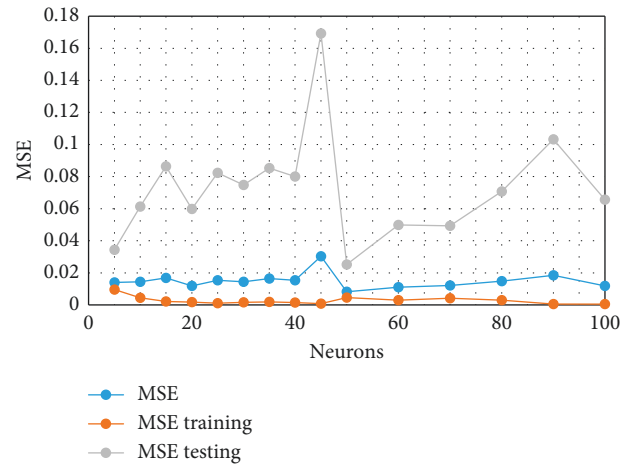


FIGURE 5: MSE values of ANN models in 2HL for different phases of ANN.

for the selected model as shown in Figure 6. The model shows that temperature and floor level are the most important factors in labour productivity.

**2.3. RBF Productivity Modeling.** RBF is a three-layer forward network applied for modeling science and engineering problems fast and precisely. RBF is a branch of ANN first introduced in the late eighties. RBFNN architecture is simple and includes one hidden layer and output. The RBF was selected to model labour productivity because of its feed-forward training done on a layer-by-layer basis in default of having input signals going through convoluted and time consuming multihidden layer developments. Thus, as compared to other ANN techniques, RBFNN is faster and has application flexibility [18].

Because of the abovementioned advantages, an effort was made in this research to model the nine predictor variables' convoluted relations and target based on actual datasets gathered from two high-rise buildings using RBFNN. The RBFNN model was trained using a BP algorithm to minimize MSE with selected predictor variables. The RBF neural network included three layers: input, hidden, and summation.

In this model, there is one neuron in the input layer for each predictor variable, where  $N$  is the number of categories and  $N-1$  the neurons used. Input neurons normalize a range of the values by subtracting the median and dividing it by an interquartile range. The input neurons feed the values to each of the neurons in the hidden layer. The hidden layer has a changing number of neurons (the optimal number is determined by the training process). Each neuron contains a radial basis function centered on a point with the dimensions equal to predictor variables. The radius of a RBF function is different for each dimension. Here, the training process determined the centers and spreads. A hidden neuron measured the Euclidean distance of the test case from the neuron's center point and then applied a RBF kernel function to this distance using the spread values when presented with the  $x$  vector of input values from the input

TABLE 4: Performance indices for models with one hidden layer.

Neurons	5	10	15	20	25	30	35	40	45	50	60
MSE	0.014	0.014	0.017	0.012	0.015	0.014	0.017	0.015	0.030	0.008	0.011
MSE train	0.010	0.004	0.002	0.002	0.001	0.002	0.002	0.002	0.001	0.005	0.003
MSE test	0.034	0.061	0.086	0.060	0.082	0.075	0.085	0.080	0.069	0.025	0.050
$R^2$	0.94	0.94	0.93	0.95	0.94	0.94	0.94	0.94	0.91	0.97	0.96
$R^2$ train	0.96	0.98	0.99	0.99	1	0.99	0.99	0.99	1	0.98	0.99
$R^2$ test	0.88	0.77	0.69	0.72	0.68	0.79	0.5	0.77	0.75	0.89	0.77

TABLE 5: Performance indices for models with two hidden layers.

Neurons	5	10	15	20	25	30	35	40	45	50	60
MSE	0.0162	0.0096	0.1264	<b>0.0215</b>	0.0545	0.0384	0.0212	0.1264	0.0183	0.0184	0.0096
MSE train	0.0055	0.0022	0.1242	<b>0.0200</b>	0.0008	0.0000	0.0000	0.1267	0.0005	0.0008	0.0007
MSE test	0.0665	0.0446	0.1369	<b>0.0230</b>	0.1071	0.1193	0.1213	0.1251	0.1024	0.1010	0.0518
$R^2$	0.935	0.942	0.499	<b>0.949</b>	0.813	0.872	0.929	0.512	0.927	0.932	0.962
$R^2$ train	0.979	0.981	0.522	<b>0.976</b>	0.997	1	1	0.537	0.998	0.997	0.997
$R^2$ test	0.693	0.693	0.403	<b>0.832</b>	0.335	0.482	0.733	0.395	0.714	0.694	0.785

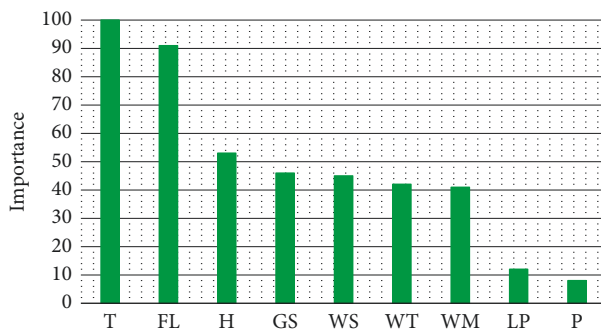


FIGURE 6: BNN model relative variable importance.

layer. The resulting value was handed to the summation layer. The coming out value of a neuron in the hidden layer was multiplied by a weight associated with the neuron ( $W_1, W_2, \dots, W_n$ ) and passed to the summation which added up the weighted values and presented this sum as the output of the network. A bias value of 1.0 was multiplied by a weight ( $W_0$ ) and fed into the summation layer. For classification reasons, there was one output along with a separate set of weights and summation unit for each target category. The output value of a category equaled to the probability that the evaluated case has that category.

In order to develop a reliable model, the dataset was randomly divided into two separate training and testing subsets. Eighty percent of the dataset was considered for network training and 20% of data was used for network reliability and to avoiding overfitting. It should be noted that RBF was developed using DTREG predictive modeling software. One of the key advantages of using DTREG for RBF development is that DTREG uses an evolutionary method called Repeating Weighted Boosting Search (RWBS) for building neural networks. In DTREG, a population of candidate neurons is first built with random centers and spreads which is limited by the minimum and maximum specified radius. The population size parameter controls how many candidate neurons are created. If there are many

variables, increasing the population size is recommended. Increasing the population size helps to prevent local minima and find the optimal global solution. In addition, DTREG lets the user modify the network and neuron parameters as well as the testing and validation percentage, select how to handle missing predictor variable values, and select one of four options for target categories' prior probabilities. Another interesting option available to the user is that the software can compute predictor variables' importance [19].

To train the DTREG algorithm, sequential orthogonal training developed by [20] was used. This algorithm uses an evolutionary approach to determine the optimal center points and spreads for each neuron. It also determines when to stop adding neurons to the network by monitoring the estimated Leave-One-Out (LOO) error and terminating when the LOO error begins to increase due to overfitting. Optimal weight computation between the neurons in the hidden layer and the summation layer was done using ridge regression. An iterative procedure was used to compute the optimal regularization Lambda parameter that minimizes generalized cross-validation (GCV) error [20]. During training, it was found that model errors can be reduced sufficiently to a lower level after incorporating 11 neurons and the model reached a steady state with 47 neurons. Thus, 47 neurons were used to model labour productivity. The RBF network with 47 neurons developed in this research has  $R$ -squared values of 0.91 and 0.67 for training and testing, respectively. Table 6 shows the developed model performance indicators. The RBF network algorithm ranked humidity, floor level, and temperature as the most important variables, as shown in Figure 7.

**2.4. GRNN Productivity Modeling.** GRNN, proposed by Donald F. Specht in 1990, is often used for nonlinear function approximation. It has a special linear and radial basis layer which makes it different from radial basis networks. GRNN is a neural network model that mimics nonlinear relations between a target variable and a set of

TABLE 6: Performances indices for RBF.

Performance index	$R^2$	MSE	RMSE	MAE
Training	91.30%	0.0103	0.1026	0.0756
Testing	85.84%	0.0471	0.1531	0.1421

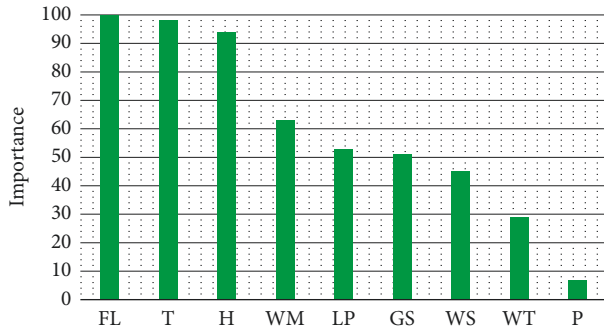


FIGURE 7: Relative importance of variables in the RBFNN model.

predictor variables. GRNN falls into the class of probabilistic neural networks and requires less training samples in comparison to a BNN. A GRNN's main advantage is that since available datasets for developing neural networks are not usually sufficient, probabilistic neural networks are more attractive for modeling. In other words, GRNN can solve any function approximation problem in case sufficient data is available in abbreviated time.

In GRNN, the target value of the predictor is achieved by considering the weighted average of the values of its neighbouring points. Target neighbour variable distance plays a key role in predicting target value. Neighbouring points close to target points have a greater impact on target value; distant points, on the other hand, are not influential as much as close neighbouring points. A radius base function is used for calculating the neighbouring point influence level. As mentioned, GRNN is able to build a model with a relatively small dataset and has the capability to handle outliers [21]. There are two main disadvantages associated with GRNN; it needs considerable calculations to evaluate new points and is not able to ignore unrelated inputs without assistance and needs major algorithm modifications. Consequently, this method is not a choice for problems with a substantial number of predictor variables. A GRNN algorithm can be enhanced by advancing GRNN in two ways: using clustering versions of GRNN and applying parallel calculations to take advantage of GRNN structure characteristics [22].

In addition to the abovementioned drawbacks, GRNN models can be large due to having one neuron for each training row. In the developed GRNN model, DTREG was utilized; thus, after the model was constructed, DTREG provided an option for facilitating the removal of unnecessary neurons from the model. By removing unnecessary neurons, computational time was reduced and it became possible to improve model accuracy. DTREG was utilized in order to select the best possible model. Three criteria available for guiding the removal of neurons are

minimizing error, minimizing the number of neurons, and limiting the number of neurons to a certain number.

The developed GRNN model's accuracy was compared with other models' using the same dataset. Therefore, 80% of the dataset was selected randomly as the training set, which corresponded to 177 input-output pairs. Twenty percent of the data were kept unused for testing, which corresponds to 44 input-output pairs. Note that all techniques used for modeling labour productivity in this study used the same training and testing datasets for a proper comparison approach.

In addition, a Gaussian kernel function type was selected as it is the best function among other kernel functions, and a single sigma for the whole model was selected to reduce computational time. Using a trial and error approach to select the best model, three models were developed based on three options provided by the software: remove unnecessary neurons, minimize error, and the constant number of neuron. Table 7 summarizes various statistical indices for the developed models.

The models with 34 and 10 neurons had overfitting in their training process since there was a large difference between the  $R^2$  in the training and testing phases. Therefore, the best GRNN model was found to have 107 nodes with the  $R^2$  value of 87.87%, which is higher than the two other approaches. Like the RBF neural network, GRNN is able to rank predictor variables for the selected model as shown in Figure 8. Here, temperature and floor levels were the significant factors found for modeling productivity.

**2.5. ANFIS Productivity Modeling.** ANFIS is used in various engineering fields such as environmental, civil, electrical, etc. [23–25]. ANFIS utilizes a hybrid learning algorithm which can model the relationship between predictor variables and respond variables based on expert knowledge by using neural network capabilities. It represents expert knowledge in the form of fuzzy "if-then" rules with an approximation of membership functions from given predictors and response datasets. Fuzzy logic handles the vagueness and uncertainty associated with the system being modeled, whereas the neural network provides model adaptability. By combining the learning abilities of a neural work with the reasoning capacities of fuzzy logic into a unified platform, ANFIS can be considered an enhanced prediction tool in comparison with a single methodology one. ANFIS can adjust membership function (MF) parameters and linguistic rules directly from neural network training capabilities with respect to refining model performance. ANFIS is able to capture expert knowledge regarding a nonlinear system and its behaviour in a qualitative model without quantitative descriptions of the system. Fuzzy inference system (FIS) is a knowledge interpretation technique based on the concept of fuzzy set theory, fuzzy "if-then" rules, and fuzzy reasoning, where each fuzzy rule characterizes a state of the system. ANFIS uses a Sugeno FIS for a structured approach to generate fuzzy rules by using a given dataset [26]. Training and testing data are matrices with ten columns where the

TABLE 7: Performances indices for GRNN.

Performance index	Number of neuron	$R^2$ train (%)	$R^2$ test (%)	MSE train	MSE test	RMSE train	RMSE test	MAE train	MAE test
Minimize error	107	87.87	75.32	0.0103	0.0475	0.1108	0.2219	0.0651	0.1421
Minimize number of neurons	34	85.84	48.70	0.0172	0.0716	0.1311	0.2676	0.0984	0.1785
Number of neurons	10	64.82	41.07	0.0427	0.0823	0.2067	0.2868	0.1471	0.2164

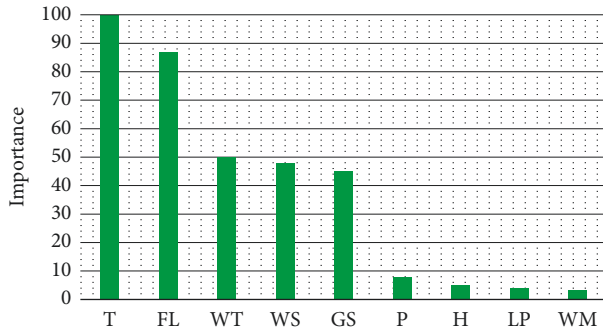


FIGURE 8: Relative importance of variables in the selected GRNN model.

first nine columns contain data for each FIS predictor variable and the last column contains the response data. It should be noted that the same 177 data points were used for training and the other 44 for testing purposes. Then, the FIS model structure was generated by choosing the subtractive clustering technique. Subtractive clustering technique is faster than grid partitioning and with a satisfactory result for justifying use. Based on the selected parameters of subtractive clustering, 63 clusters were detected as the most suitable MF number. Here, the number of clusters and MFs were equal. The hybrid method was selected for training the MF because it generates better results rather than BP. The hybrid method includes BP and least squares for MF parameter estimation, BP for estimating input MF parameters, and least square for MF parameters. ANFIS parameters were selected to reach higher accuracy with less computational time, as shown in Table 8.

Figure 9 shows the predicted daily productivity against corresponding actual values, and Table 9 summarizes the dataset various statistical indices using the developed ANFIS model.

ANFIS is also able to rank predictor variables for the selected model as shown in Figure 10. Temperature and floor level are the important factors in modeling productivity. Table 10 summarizes the performance indices of  $R^2$  Train,  $R^2$  Test, MSE Train, and MSE Test for the four models. BNN shows the highest  $R$ -squared in the training phase followed by RBF, ANFIS, and GRNN. Moreover, BNN shows the highest  $R$ -squared in testing phases followed by RBF, ANFIS, and GRNN. GRNN is the least accurate technique among all the techniques for the given dataset in both training and testing phases. BNN has the lowest MSE of the techniques in the training phase followed by RBF. ANFIS and BNN show the lowest MSE in the testing phase followed by RBF and GRNN.

TABLE 8: ANFIS parameters for modeling labour productivity.

ANFIS parameters	
Number of input variables	9
Training data points	177
Testing data points	44
Number of layers	5
Operator	Subtractive
Number of membership functions (MF)	63
MF type	Bell shape
Transfer function of output layer	Linear
Training algorithm	Hybrid
Error tolerance	0
Number of epochs	1000

To recognize which algorithm is the best method to be utilized in modeling construction labour productivity, analysis of variance (ANOVA) was applied to demonstrate the significant superiority of the BNN algorithm over other algorithms, which had the lowest  $F$ -Value.

Furthermore, the results of the model were compared with the SOM model developed by [9] and results show that for formwork productivity prediction, BNN performs better in comparison to SOM. The correlation of coefficient and the MSE for the available database were 94.9% and 0.0215 for backpropagation method while it was 89.25% and 0.07 for SOM.

Both regression and AI techniques have merits and demerits. Three statistical methods, namely, best subset, stepwise, and Evolutionary Polynomial Regression (EPR), were applied to the available database, and the coefficient of correlation and MSE were calculated and are presented in Table 11. EPR predicted the data better than best subset and stepwise. However, BNN outperformed and achieved a better fit and forecast with the given dataset than the regression models due to the nonlinearity of the dataset in modeling labour productivity for formwork. The statistical performance of those models is far behind BNN. The analysis of variance for different techniques is presented in Table 12.

### 3. Conclusions

One of the major strategic components in determining the success or failure of a construction project is the productivity rate, which has a relationship with different factors. This paper attempts to demonstrate a way to use AI models to predict labour productivity. These are effective tools for quantifying loss of productivity and can be used as a support method for actual loss of productivity calculations. GRNN, BNN, RBFNN, and ANFIS were tested against Khan's



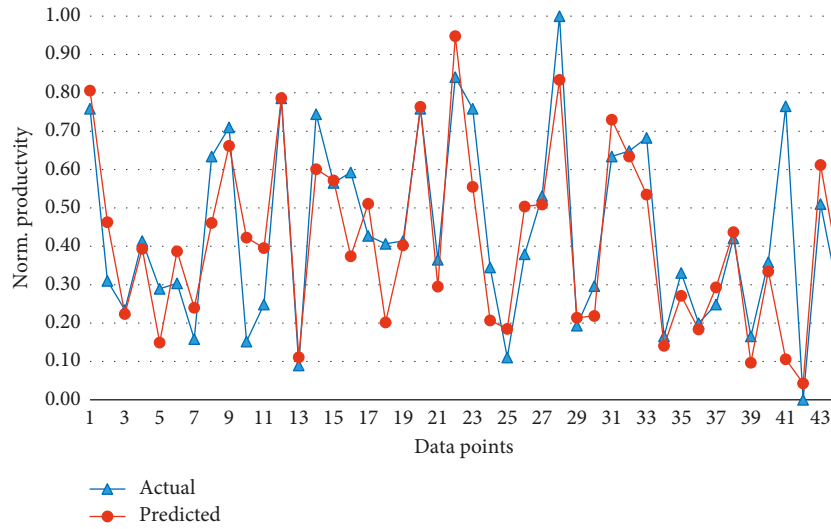


FIGURE 9: Actual vs predicted value using ANFIS.

TABLE 9: Statistical indicators of the developed ANFIS model.

Performance index	$R$ (%)	$R^2$ train (%)	$R^2$ test (%)	MSE train	MSE test	RMSE train	RMSE test	MAE train	MAE test
Value	81.1	89.3	66.2	0.01	0.02	0.114	0.146	0.017	0.097

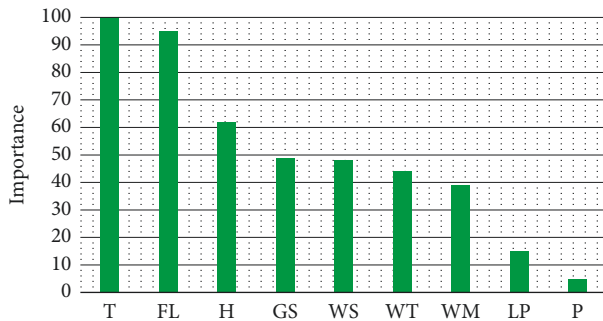


FIGURE 10: Relative importance of variables in ANFIS Model.

TABLE 10: Performance results comparison.

	$R^2$ train	$R^2$ test	MSE train	MSE test
BNN	0.98	0.83	0.0023	0.020
RBF	0.91	0.85	0.0096	0.018
GRNN	0.88	0.75	0.0103	0.047
ANFIS	0.89	0.66	0.0100	0.020

TABLE 11: Statistical performance of the regression methods.

Technique	$R^2$	MSE	Time (sec)	Number of variables
Best subset	46.8	0.259	~2	8
Stepwise	48.61	0.259	~2	7
EPR	52.69	0.057	1140	8

datasets of two high-rise buildings related to formwork operation. From the comparisons, BNN showed the best performance among the techniques. However, BNN can be considered a black box approach and is prone to overfitting, which can be the result of network architecture

TABLE 12: Analysis of variance for different techniques.

	Source	DF	Adj. SS	Adj. MS	$F$ -value	$P$ -value
BNN	Actual	37	2.14235	0.05790	4.42	0.034
	Error	6	0.07864	0.01311		
	Total	42	2.22099			
ANFIS	Actual	36	5.69202	0.15811	15.53	0.001
	Error	6	0.06108	0.01018		
	Total	42	5.75310			
RBF	Actual	34	3.4887	0.10261	6.67	0.004
	Error	8	0.1231	0.01539		
	Total	42	3.6118			
GRNN	Actual	37	5.59931	0.151333	31.89	0.002
	Error	4	0.01898	0.004745		
	Total	42	5.61829			

(i.e., decreasing the number of nodes), early training phase stopping, or weight decay use. Furthermore, the dataset used to develop the aforementioned models was a raw and unbalanced dataset. Studying the behaviour of the given datasets prior to feeding it to any AI techniques is required in order to have a robust model for modeling labour productivity.

Researchers have proved that supervised BNN models are more successful in predicting construction crew productivity in comparison to statistical methods like regression. Furthermore, if the causal relationship between input and output has a complex variability in areas other than construction, in most cases, the learning task is easier with unsupervised learning. Therefore, this study focuses on the application of supervised methods in formwork crew productivity data to compare the predicted results. This study focused on formwork installation operations since

they constitute a substantial part of the overall labour component of concrete framing in building construction. Results reveal that BNN shows superior performance in comparison to RFBNN, ANFIS, and GRNN for formwork productivity prediction in the following ways:

- (1) Selected input variables are those that cause variations in productivity in the short-term or daily basis. The developed models were compared based on statistical performances and BNN outperformed the techniques of RBF, ANFIS, and GRNN. The developed model of this research can be utilized in different ways.
- (2) The model can help to estimate formwork productivity by considering variables such as temperature, humidity, gang size, labour percentages, work type, etc. In addition, this study also found that productivity is not significantly correlated with precipitation, labour percentage, work method, and humidity. Within the scope of the conducted study, the number of parameters observed, and the range of their values, this study found temperature to have the most significant impact on productivity followed by floor level.
- (3) The model can also be useful for quantifying loss of productivity by considering the output of the developed model as the value for unimpacted productivity period since the identifying unimpacted period for quantifying loss of productivity is impossible to calculate sometimes. Therefore, BNN can help save time and cost associated with quantifying loss of productivity.

Ultimately, this study shows that the backpropagation model can be an alternative tool to supervised learning-based tools and can be used in various prediction applications. One limitation of this study is that the findings are limited to the collected data range and parameters considered in the study. It should be noted that the developed model does not involve any parameters that directly account for management strategies and skills or any project-specific conditions.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Z. U. Khan, "Modeling and parameter ranking of construction labor productivity," Doctoral Dissertation, Concordia University, Montreal, Canada, 2005.
- [2] W. Yi and A. P. C. Chan, "Critical review of labor productivity research in construction journals," *Journal of Management in Engineering*, vol. 30, no. 20, pp. 214–225, 2014.
- [3] M. Lu, S. M. AbouRizk, and U. H. Hermann, "Estimating labor productivity using probability inference neural network," *Journal of Computing in Civil Engineering*, vol. 14, no. 4, pp. 241–248, 2000.
- [4] S. AbouRizk, P. Knowles, and U. R. Hermann, "Estimating labor production rates for industrial construction activities," *Journal of Construction Engineering and Management*, vol. 127, no. 6, pp. 502–511, 2001.
- [5] O. Moselhi, I. Assem, and K. El-Rayes, "Change orders impact on labor productivity," *Journal of Construction Engineering and Management*, vol. 131, no. 3, pp. 354–359, 2005.
- [6] A. S. Ezeldin and L. M. Sharara, "Neural networks for estimating the productivity of concreting activities," *Journal of Construction Engineering and Management*, vol. 132, no. 6, pp. 650–656, 2006.
- [7] S. C. Ok and S. K. Sinha, "Construction equipment productivity estimation using artificial neural network model," *Construction Management and Economics*, vol. 24, no. 10, pp. 1029–1044, 2006.
- [8] L. Song and S. M. AbouRizk, "Measuring and modeling labor productivity using historical data," *Journal of Construction Engineering and Management*, vol. 134, no. 10, pp. 786–794, 2008.
- [9] E. L. Oral and M. Oral, "Predicting construction crew productivity by using self organizing maps," *Automation in Construction*, vol. 19, pp. 791–797, 2010.
- [10] S. Muqeem, M. Idrus, and F. Khamidi, "Construction labor production rates modeling using artificial neural network," *Journal of Information Technology in Construction*, vol. 16, pp. 713–725, 2011.
- [11] S. Mohammed and A. Tofan, "Neural networks for estimating the ceramic productivity of walls," *Journal of Engineering*, vol. 17, no. 2, pp. 200–217, 2011.
- [12] F. M. S. AL-Zwainy, H. A. Rasheed, and H. F. Ibraheem, "Development of the construction productivity Estimation model using artificial neural network for finishing works for floors with marble," *ARNP Journal of Engineering and Applied Sciences*, vol. 7, no. 6, pp. 714–722, 2012.
- [13] O. Moselhi and Z. Khan, "Significance ranking of parameters impacting construction labour productivity," *Construction Innovation*, vol. 12, no. 3, pp. 272–296, 2012.
- [14] G. Heravi and E. Eslamdoost, "Applying artificial neural networks for measuring and predicting construction-labor productivity," *Journal of Construction Engineering and Management*, vol. 141, no. 10, article 04015032, 2015.
- [15] G. Aswed, "Productivity estimation model for bricklayer in construction projects using neural network," *Al-Qadisiyah Journal for Engineering Sciences*, vol. 9, no. 2, pp. 183–199, 2016.
- [16] K. M. El-Gohary, R. F. Aziz, and H. A. Abdel-Khalek, "Engineering approach using ANN to improve and predict construction labor productivity under different influences," *Journal of Construction Engineering and Management*, vol. 143, no. 8, article 04017045, 2017.
- [17] O. Moselhi, T. Hegazy, and P. Fazio, "Neural networks as tools in construction," *Journal of Construction Engineering and Management*, vol. 117, no. 4, pp. 606–625, 1991.
- [18] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, "On the training of radial basis function classifiers," *Neural Networks*, vol. 5, no. 4, pp. 595–603, 1992.
- [19] P. Sherrod, "DTREG predictive modeling software," 2018, <http://www.dtreg.com>.
- [20] S. Chen, X. Hong, and C. J. Harris, *Orthogonal Forward Selection for Constructing the Radial Basis Function Network*

*with Tunable Nodes, in Lecture Notes in Computer Science, Vol. 3644, 2005.*

- [21] H.-I. Yip, H. Fan, and Y.-h. Chiang, "Predicting the maintenance cost of construction equipment: comparison between general regression neural network and Box-Jenkins time series models," *Automation in Construction*, vol. 38, pp. 30–38, 2014.
- [22] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.
- [23] H. Alasha'ary, B. Moghtaderi, A. Page, and H. Sugo, "A neuro-fuzzy model for prediction of the indoor temperature in typical Australian residential buildings," *Energy and Buildings*, vol. 41, no. 7, pp. 703–710, 2009.
- [24] A. Subasi, A. S. Yilmaz, and H. Binici, "Prediction of early heat of hydration of plain and blended cements using neuro-fuzzy modelling techniques," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4940–4950, 2009.
- [25] L.-C. Ying and M.-C. Pan, "Using adaptive network based fuzzy inference system to forecast regional electricity loads," *Energy Conversion and Management*, vol. 49, no. 2, pp. 205–211, 2008.
- [26] M. Negnevitsky, *ANFIS: Adaptive Neruo-Fuzzy Inference System, Artificial Intelligence-A Guide to Intelligent Systems*, Pearson Education Limited, Essex, UK, 2nd edition, 2005.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

