Automatic Valuation of Jordanian Estates Using A Genetically-Optimised Artificial Neural Network Approach

Mousa AL-Akhras
Computer Information Systems Department
King Abdullah II School for Information Technology
The University of Jordan
Amman 11942, JORDAN
mousa.akhras@ju.edu.jo

Maha Saadeh
Computer Science Department
King Abdullah II School for Information Technology
The University of Jordan
Amman 11942, JORDAN
saadeh.maha@yahoo.com

Abstract: Estate appraisal can be defined as valuing a land or property as of a given date using common data utilising standardised methods and statistical testing. Estate appraisal has several applications such as asset valuation for lenders, property tax estimation, insurance estimation, and estate planning, grant new mortgages to new home buyers, and purchase mortgage packages, which can contain thousands of mortgages, as investments. It is also used to guide potential buyers and sellers with making purchasing decisions. The drawbacks of on-site manual property valuation include: it is time-consuming, costy, based on subjective judgments and sometimes it is based on validation using a negotiated price rather than estimating the true market value of a property. In this paper, the authors build an Artificial Neural Network model for the purpose of automatic appraisal of Jordanian estates to avoid the drawbacks of manual appraisal. The proposed Artificial Neural Network model is built in two stages: In the first stage, a Genetic Algorithm optimiser is used to determine the best topology for the Artificial Neural Network. In the second stage, the optimised Artificial Neural Network topology is trained for the best values for the weights. In evaluating the property price, several factors are taken into consideration such as area size, location, lot area, establishment year, price for building a square meter, and type of building (commercial or agricultural) and others. Records from Jordanian Department of Lands & Survey are used for training the Artificial Neural Network model and for testing its performance to find the best model that represents the underlying relation between a land/property and its characterising features. Statistical tests were performed to validate the effectiveness of the proposed method.

Key–Words: Artificial neural network, Genetic algorithm, Estate appraisal, Automatic valuation model, Computer-assisted mass appraisal, Cost approach, Sales comparison approach, Income approach.

1 Introduction

Estate appraisal can be defined as valuing a land or a property as of a given date using common data, standardised methods and statistical testing [1]. Many specialised, professional organisations were established worldwide for the purpose of property valuation, such organisations include: Appraisal Foundation, Appraisal Institute, International Association of Assessing Officers (IAAO). Such organisations produce many standards of appraisal best practices [2, 3, 4]. Mass appraisal methods of a group of lands and/or properties are applied in such organisations to assess land or property value for use, for example, in determining local government rates and for the possible imposition of land taxes on residential properties [1].

Lands are usually classified into either agricultural, commercial, industrial or residential. When a land has premises built on it, the premises (real properties) are usually classified as either commercial, industrial or residential properties. Residential properties are usually of special interests for human evaluators (appraisers). A residential property is a place aimed to be owned and used as a residence for a single family. Residential property valuation (appraisal) is the process of identifying the price of a residential property based on its characteristics such as size and location according to the current market conditions. This appraisal may change with changes in market conditions, therefore the estimate needs to be updated periodically, also the estimate must be supported by similar appraisal for real recent estate transactions [2, 5].

Taxes on properties is legalised in many countries and is considered an important source of tax revenue for local governments, several countries like China are continuously developing such legislations [1]. Usually taxes on properties are imposed based on the values of properties, therefore, a proper, objective, consistent, cost-effective and accurate valuation model should be employed. This is especially needed for mass appraisal of hundreds and thousands of properties.

Residential property appraisal should be performed accurately as such appraisal is required for tasks such as asset valuation for lenders, property tax estimation, insurance estimation, and estate planning. Accuracy of asset valuation is of critical importance to users such as commercial lenders whose slim margins leave little room for error [6]. Also when financial institutions like banks and governmental housing departments grant new mortgages to new home buyers, and purchase mortgage packages, which can contain hundreds or even thousands of mortgages, on the secondary market as investments, a property appraisal process is usually needed. Property appraisal is also used to guide potential buyers and sellers with making purchasing decisions, and are needed for variety of insurance purposes [5, 7, 8].

Enormous efforts are going on in cadastral departments worldwide, workers in Jordan cadastral department which is called Jordanian Department of Lands & Survey, in addition to their usual clerk work, are usually asked to do on-site visits for valuation of lands and properties. The disadvantages of on-site manual property valuation, including [5, 6, 7, 8]:

- It is time-consuming as it may require several days of preparation and it may take several site visits to give an estimation.

- It costs a lot of money per subject property (estimated of 100-200 Jordan Dinar (JD) in Jordan and $500 in the US market) and the appraiser is charging the customer, therefore it may become an expensive process.

- Lack of comparable properties forces the appraiser to make subjective judgments (appraiser bias) regarding the differences in value created by the differences in the original and the comparative properties. Also appraisers with different levels of experience may have different evaluations.

- Some appraisers may apply an easier approach by validating a negotiated price rather than estimating the true market value of a property.

According to the US census bureau, it is estimated that there are 129,065,264 houses in the US [9]. The number of land pieces exceeds this number by several times. According to some sources there are 4,055,755 real estate listings, homes for sale and rental properties [10]. If we consider the huge number of available properties and the efforts needed to evaluate their prices by human appraisers, the need come to automate this process. Automation could reduce the length and costs of the home buying process by as much as 20%-50% with the increased speed of processing and approval leading to savings of up to $1,000 [6].

Automatic Valuation Models (AVM) for land/property valuation can be used by evaluating the price of a land/property based on set of features for a set of lands/properties with similar features to the one to be estimated, taking into consideration the changing conditions of the market [5]. An automatic valuation system may be easily embedded in a larger processing system, such as one for loans. Once the system is setup, it can also be integrated inside tax valuation systems to calculate the tax deserved on properties independence from experts [6].

In this paper, an attempt is made to automatically obtain an evaluation of the prices of lands/properties in the Jordanian market using an Artificial Neural Network (ANN) function approximation model. the ANN model is first optimised using Genetic Algorithm (GA) to find the best topology of the network. Then the best ANN topology is further fine-tuned using standard feedforward training. The ANN model approximates the price of a land/property based on several features. The used records for lands/properties is obtained from Jordanian cadastral department (Jordanian Department of Lands & Survey) from the north and the south of the capital Amman, the Jordanian capital. To the best of our knowledge no such effort was made to automatically evaluate the prices of properties in the Jordanian market.

The rest of this is paper is organised as follows: section 2 reviews current approaches used for properties' valuation. Section 3 reviews previous attempts to automatically evaluate the price of residential properties. Section 4 introduces preliminary information about ANN and GA which are used in the approach proposed in this paper. Section 5 details the technical details of the proposed system. Section 6 summarises the main findings and the results while section 7 draws conclusions and presents for future work.

2 Approaches for Property evaluation

The usual approach of property appraisal it to do the appraisal by an on-site visit by a human expert (appraiser). The first step in evaluating the price of a land with a property built on it is to evaluate the price of the land, the evaluation is based on the area of the land to be sold and the basic price for the square meter. By multiplying these two values, the basic price of the land can be obtained. Several indicators also affect the price of a land, some of the indicators have positive effect while others have negative effect. Each indicator either increase or decrease the price of the land by a certain percent. The price of the land is basic price plus the price corresponding to the sum of positive indicators minus the price corresponding to the sum of negative indicators.
After evaluating the price of the land, the price of the property can be evaluated, common methods used by human appraisers to evaluate a property based on mathematical models, which illustrate the relationship between the value and the features of the concerned property, such methods include: Cost Approach, Sales Comparison Approach and Income Approach [1, 5, 11, 12, 13] [Jordanian Department of Lands & Survey, personal communications]:

2.1 Cost Approach

It is reasonable to assume that no rational investor will pay more for an existing property than it would cost to buy the land and to build a new building on it. The cost approach if used properly can produce highly accurate valuations. In the cost approach, the cost structure of a real property being estimated is analysed, and try to get the replacement price of the property. Then the reasonable value of the property can be assessed by considering depreciation rate where in Jordan it is estimated to be 2% per year for residential properties (Jordanian Department of Lands & Survey, personal communications). This approach requests to assess the value of the land and property separately, the theory is that the value of a property can be estimated by summing the land value and the depreciated value of any improvement. Cost approach calculation depends on land and depreciated values, the properties sale area, per meter price, and the advantage and defects values. Due to the depreciated value, the cost approach is considered reliable when used on newer structures, but the method tends to become less reliable for older properties. According to the Jordanian Department of Lands & Survey, the Cost Value (CV) is calculated using the following equation:

$$\text{CV} = (A * P) + AV - DV - D + LV \quad (1)$$

Where

A	Property area
P	Property price per unit (square meter)
AV	Summation of positive indicators
DV	Summation of negative indicators
D	Depreciation

The Depreciation is calculated by:

$$\text{Depreciation} = 0.02 * \text{property age} * (A * P) \quad (2)$$

The appraiser needs to adjust the calculated value to reflect market conditions.

2.2 Sales Comparison Approach

The economic rationale of the sales comparison approach is that when the general market conditions are the same, no investor would pay more for a property than other investors have recently paid for similar properties [1]. The main idea of the sales comparison approach is to simulate the price that would have been paid if each comparable sale was identical to the subject property. If the adjustment to the comparable is superior to the subject, a downward adjustment is necessary. Likewise, if the adjustment to the comparable is inferior to the subject, an upward adjustment is necessary [13]. Sources of comparable data include real estate publications, public records, buyers, seller, appraisers, and others. The basic steps an appraiser do in the sales comparison approach [1, 5, 6]:

1. Finds comparable properties the same as the property being valued; i.e. recent sales that are located near the subject property (using sales records). Finding the most similar properties is performed by contrasting the subject property with the comparables. This process assumes that the items market value can be derived by the prices demanded by similar items in the same market.

2. Selects a balanced subset of the most promising comparables to derive the final estimate.

3. Adjusts the comparables sales price to reflect their differences from the subject property (using heuristics and personal experience).

4. Average the comparables adjusted sales prices (using any reasonable averaging method) and derives the value for the subject property.

This process assumes that the items market value can be derived by the prices demanded by similar items in the same market. Equation (3) represents the Sales Comparison Value (SCV) calculated according to the Jordanian Department of Lands & Survey.

$$\text{SCV} = (A * CAP) + LV \tag{3}$$

Where

A	Property area
CAP	Average meter price of comparables
LV	Land value

2.3 Income Approach

The rationale behind the income approach is that no investor will pay more for a property than s/he will retrieve by holding the property. The income approach is based on the idea that the current value of a property is the present worth of future benefits to be derived through income production by the property over the remainder of its economic life. The income approach is based upon the principle of anticipation; i.e., the expectation of future benefits such as the market rent that a property can be expected to earn, and the reversion (resale) when a property is sold. It converts (through capitalization) the anticipated benefits of the ownership of a property into an estimate of present value. The following equations characterise the Income Value (IV) of a property in the income approach according to the Jordanian Department of Lands & Survey where the Interest Rate (IR) equals 0.07.

$$\text{IV} = (PI - LI) * MF + LV \tag{4}$$

Where

PI	Property Income
LI	Land Interest
MF	Multiplying Factor
LV	Land Value

The parameters in equation (4) are characterised by the following equations:

$$\text{PI} = A * AAR - (MMC + TF) \tag{5}$$

Where

A	Property area
AAR	Property average annual rent per meter
MMC	Maintenance and Management Cost
TF	Taxes and Fees

and

$$\text{LI} = IR * LandValue \tag{6}$$

and

$$\text{MF} = \frac{(1 + IR)^{RemainingAge} - 1}{(1 + IR)^{RemainingAge} * IR} \tag{7}$$

Where $RemainingAge$ is calculated as:

$$RemainingAge = 50 - (CurrentYear - YearofConstruction) \tag{8}$$

The summation of Maintenance and Management Cost (MMC) and Taxes and Fees (TF) in equation (5) can be calculated as

$$\text{MMC+TF} = 0.19 * A * AAR \tag{9}$$

3 Previous Attempts for Automatic Property Appraisal

Timely asset valuation and pricing is a big challenge in the commercial and financial world. The value of stocks and bonds can be easily determined, on the other hand, the same precise valuation for real-estate assets cannot be performed, which could have been used as collateral for mortgage-backed securities [14]. Real property mass appraisal theory was developed in 1920 by the American John A. Zangerle [1]. Appraisal for multiple properties is currently estimated by sampling techniques. Sampling is less accurate when used for large batch of properties, therefore automatic valuation process is needed.

The most commonly used quantitative valuation model in real estate is Multivariate Linear Regression (MLR). Other Multivariate Regression Analysis (MRA) techniques are occasionally used, and any model is known as an Automated Valuation Model (AVM) within the industry. Linear regression assumes that variables affecting real estate values are linearly related to the price. It may be easily shown, however, that a number of variables affecting prices exhibit nonlinear relationships with the price. Among these variables are age, distance to city centre, number of bathrooms, and square footage among others [6].

AVMs are computer programs that analyse real property market to provide estimates of value, sometimes this is called Computer Assisted Mass Appraisal (CAMA) techniques. Computer assistance allows thousands of calculations per second and gets a very high degree of accuracy. Also automatic valuation ensures the objectivity of the appraisals and facilitates quality control. Depending on automatic valuation, human appraisers can spend more time gathering the information necessary to value the property. AVMs also reduce costs, permit the use of larger, more representative samples [1].

Several approaches can be used for the purpose of automatic property appraisal. The simplest one would be the use of a Case-Based Reasoner Reasoning (CBR) [15] as proposed by Gonzalez and Laureano-Ortiz [16]. The proposed CBR approach by gonzales did not use the intuitive imprecision used in sale comparison approach (manual approach) which include: finding the most similar house(s), located near to the subject property, sold not too long ago; and selecting a balanced subset of the most promising comparable properties to derive the final estimate.

Bonissone and other researchers [5, 14] proposed the use of a fuzzy CBR system in which the input features of the properties (address, date of sale, living area, lot area, number of bathrooms and bedrooms) in the data set of previous properties are assigned a fuzzy membership value (between 0 and 1) based on their closeness to the subject property to be evaluated. The closest 4-8 properties are then selected for further computations like averaging and adaptation to estimate the price of the subject property. The price is adjusted according to: construction quality, conditions, pools, fireplaces, etc. [14].

Khedkar et al. [7] proposed a method in which a generative Artificial Intelligence (AI) method trains a fuzzy-neural network using subset of cases from the a case-base to produce a run-time system to provide an estimate of the subject's property value. The system implements a fuzzy system with five-layer neural network. High level human-plausible if-then fuzzy inference rules are used to map raw input to an ANN input by using the output of the rule (then part). Then the new input is used to train a supervised BP trained ANN to map these values to an evaluated property price.

Anderson et al. [6] pointed that there are a large number of complexly inter-related variables that function to determine a value for a property. Possible features are "quality" of the property, the neighborhood, and the location of the property within the neighborhood [6]. Among the features studied by Anderson et al. are: Level of quietness (absence of noise from road traffic), distance to railways or airports, distance to public transportation, distance to city centre, quality of unobstructed view to surroundings, distance to shopping facilities, distance to nature places like forests, open areas, or lakes, distance to primary and secondary schools, and social standing of the area. Anderson et al. also indicated that it is also possible to include economic factors that influnce the price of a property in valuation. While one can hypothesise the general state of the economy, the local rate of unemployment, the average rate for a 30 year fixed mortgage, the patience of the buyer, the month or quarter in which the houses are sold, only experimentation can determine the feasibility of including these factors. It may be that the influence of these types of factors cannot be distinguished from noise in the data set [6].

Lasota et al. [11] utilised a tool called KEEL (http://sci2s.ugr.es/KEEL/) to test the performance of several learning algorithms for the purpose of evaluating the prices of properties. They include features like usable area of premises, number of rooms, floor, year of construction, number of storeys in the building, number of residential and non-residential premises in a building, and geodetic coordinates Xc and Yc of a building.

4 Preliminaries

Artificial Neural Network (ANN) are chosen to model the relation between the land or property features and the price for that land or property. The reason for choosing ANN for this task is its ability to model underlying relations and patters in data. By choosing the right topology and right parameters for the ANN, it can model linear and non-linear relations in patters. ANN was classically used in enormous number of studies for the task of function approximation [17, 18].

The task of designing an ANN faces the problem of deciding the appropriate size, structure and parameters of the ANN. The problem of optimising the parameters of an ANN for obtaining high classification accuracy is a hard problem in the literature. Empirically optimising the ANN may need a large number of experiments due to the large number of parameters that need to be decided. Several methods were proposed for the task of determining the best parameters for the ANN to achieve high classification accuracy. Lappas proposed a mathematical formula for priori calculating the size of an ANN for achieving high classification accuracy [19]. The formula estimates ANN size based only on the number of available training samples. Using this formula in hard classification data sets aims to decide the size of an accurate ANN and allows researchers to concentrate on other aspects of their experiments. On the other hand, Genetic Algorithm (GA) is used in optimisation problems to find solutions that minimises or maximises certain output. GA is this paper is used to help in the task of selecting the best parameters for the ANN structure used to model the relation between the input features of a land/property and the associated price. Next both ANN and GA are introduced.

4.1 Artificial Neural Network

Artificial Neural Network (ANN) is a mathematical model that attempts to simulate the structure and functional aspects of biological neural networks. It consists of interconnected artificial neurons that process information using a connectionist approach to computation. An ANN is an adaptive system that changes its structure and/or connection weight based on external or internal information that flows through the network during the training phase. Usually an ANN model consists of an input layer, followed by one or more hidden layers leading to an output layer. Each layer contains a set of neurons where each neuron accepts input and computes the output. ANN obtained widespread use in several applications including digital image processing. In the mid 1980s the back-propagation learning algorithm for neural networks was introduced by the PDP group together with others researchers [17, 20, 21, 22]. The back-propagation learning algorithm made it possible to train a non-linear neural network equipped with hidden layers. To train a multi-layer backpropagation networks, Levenberg Marquardt (LM) algorithm, for instance, can be used for function approximation problems using moderate-sized feedforward ANNs up to several hundred weights. Multiple-layer networks can be powerful. Since the development of back-propagation algorithm, ANNs networks with one or more hidden layers can, in theory, be trained to perform virtually any regression or discrimination task. Moreover, no assumptions are made as with respect to the type of underlying (parametric) distribution of the input variables, which may be nominal, ordinal, real or any combination hereof [17, 22].

A two-layer networks rained using the backpropagation algorithm with biases, a sigmoid (tansig or logsig) transfer function in the first layer and a linear transfer function in the output layer can be used as general function approximator networks as they are capable of approximating any function arbitrarily well, given sufficient neurons in the hidden layer. Neurons in the first layer with non-linear transfer functions allow the network to learn non-linear and linear relationships between input and output vectors. The sigmoid functions squeeze the output into a limited range, between 0 and 1 or between -1 and +1, the linear output layer lets the network produce values outside the squeezed range [17, 20]. ANNs are chosen to model the relation between the land or property features and the price for that land and property due to its ability to model complex relationships between inputs and outputs or to find patterns in data. They are able to model underlying relations and patters in data whether these relations are linear or non-linear.

4.2 Genetic Algorithm

Genetic Algorithm (GA) is a search technique used to find solutions to optimisation problems. This paper utilises GA to find the best topology for the ANN model. The best ANN structure will then be fine-tuned for the best weights with the aim of obtaining the best weights for the best network structure.

To measure how successful are different ANN structures in predicting the price for lands and properties, a fitness function is used. The higher the prediction accuracy, the higher the fitness value for the ANN. In GA, a set of initial solutions (ANNs in this case) called initial population is generated. Each population is used to produce a new population to find better solutions. When a new population is generated, new solutions are born while other solutions die. The generated solutions are selected according to their fitness, the higher the fitness, the higher the probability a given solution would survive for coming generations. Each individual is represented by a binary string which is called a chromosome, the length of the string depends on the problem's representation.

Several attributes characterise GA which are: elitism, cross-over, and mutation. Encoding is the way in which solutions (chromosomes) are represented to be input for the fitness function. There are many types of encodings, including: binary encoding, value encoding, and permutation encoding. Elitism is a process that copies the best chromosome(s) in the old population to the new population directly to ensure that the new generation is not worse than the old one. Cross-over operation starts by changing some genes (bits) in parent chromosomes to create children (offspring). By mixing the characteristics of parents, new children may combine the best features of their parents. Mutation operation randomly mutates bits in individual parents to create new children with modified genes (characteristics) [15, 23, 24, 25].

5 The Proposed Methodology

The cadastral system in Jordan, called the Department of Lands & Survey, comprises uniform for the country, set of information on the land, buildings and premises, their owners and other physical and legal persons using these lands, buildings and premises updated on a regular basis and created and kept according to the rules as stipulated under the act.

The data obtained from the Department of Lands & Survey include features about 891 real properties recently sold and are located in different areas of the Jordanian capital Amman. The real properties were constructed over 35 years ranging from 1975 to 2010 distributed as following: 7 properties between 1975-1979, 46 properties between 1980-1989, 129 properties between 1990-1999, 640 properties between 2000-2009, and 69 properties in 2010.

Several features that are available for selection for each land including: location, pointer to indicate if the land is shared or not, number of shares, type of land (residential, commercial, agricultural or industrial), area, and estimated price per meter. The basic price is calculated by multiplying the area of the land to be sold by the basic price for the square meter. Several indicators also affect the price of a land, some of the indicators have positive effect while others have negative effect. Each indicator either increase or decrease the price of the land by a certain percent. The price of the land is the basic price plus the price corresponding to the sum of positive indicators minus the price corresponding to the sum of negative indicators. Among the positive indicators along with the percentage of increase they affect the basic price of the land, including: If the land affords a view (5%), if the sold area can be used on its own separated from the rest of the land (5%), the main streets surrounding the land are paved (5%), if the main street in front of the land is 12 meters in width (0%), the main street in front of the land is between 14-16 meters (3%), if the main street is 20 meters in width (5%), the main street is 30 meters in width (8%), and if the land is over two streets (5%). Among the negative indicators along with the percentage of decrease they affect the basic price of the land, including: The slope of the land (varies between 0% and 100% depending on the degree of slope), the main street is less than 6 meters in width (10%), the main street is less than 10 meters in width (5%), if the shape of the land is trapezoid (10%), if the shape of the land is triangle (15%), if the land is on closed street (10%), the land is close to a high-voltage station (10%), and the land is located near a graveyard (10%) (Jordanian Department of Lands & Survey, personal communications). Other positive and negative features are used as indicators that may increase or decrease the price of the land, these are: quietness of the area for residential living which may increase the cost of living in that area, not obstructed view, existence of nearby schools, hospitals and other facilities.

For properties, several features are available including: building class (villa, commercial-offices, commercial-stores, residential building high quality, or residential building low quality), lot area, year of construction, price for building a square meter, shares of the building. The price for the lands is separated from the price of the property; i.e. to calculate the overall price, both prices should be accumulated. Also the land may not be affected by the depreciation rate, with properties in Jordan suffer a 2% depreciation rate.

Given that the underlying functional model describing real estate values in Jordan is not clearly defined, and that large amounts of data are available, an ANN model could be build and trained to perform a proper, objective, consistent, cost-effective and accurate valuation in comparison with other traditional, manual methods by finding the underlying relation between the attributes of the land, property and the estimated price. Valuation approaches used in Jordanian Department of Lands & Survey are: cost approach, sales comparison approach and income approach. These 3 approaches are separately modelled where 3 prices are available for estimation (targets) according to the 3 models. In the context of ANN, there are 3 output variables.

The aim here is to let the ANN discover the underlying relation between the land/property features to produce an estimation of the land/property price. ANNs are capable of approximating very complex linear and nonlinear relationships, one would expect ANNs to perform better than regression in the presence of data with clearly demonstrable non-linear relationships [6]. The proposed system avoids the time-consuming and costy manual appraisal and it helps human appraisal by finding the estimation of property

value using the cost approach, sales comparison approach and income approach.
Ten features were chosen as input features to the ANN which are: Code to represent the area in the city, the sold area of the land, the price for the square meter, code to represent the type of the land (residential, commercial, agricultural or industrial), code to represent the building class (villa, commercial-offices, commercial-stores, residential building high quality, or residential building low quality), lot area, price for building a square meter, year of construction, summation of positive and negative indicators, overall indicator sign (positive or negative).
First of all, data are randomised and then it is divided into training data and testing data. The structure of ANN model will be first constructed using an evolutionary computation/GA approach, then the best ANN structure will be fine-tuned for the best weights. The aim here is to obtain the best weights for the best network structure. ANN and GA, among other techniques, are usually classified as SoftComputing (SC) methods. By using GA to find the best network structure, several SoftComputing (SC) methods are combined. The main reason for the popularity of SC is the synergy derived from its components. SC's main characteristic is its intrinsic capability to create hybrid systems that are based on loose or tight integration of constituent technologies such as ANN, GA, and Fuzzy Logic (FL) [14, 26, 27].

5.1 Genetically Optimising the ANN Topology

Finding the best structure for the ANN is usually a trial and error process. Utilising optimisation methods such as GA, different possible structures are attempted and the best structure can be tuned further for better performance. The work of Montana & Davis (do coarse search but not fine search) [28] elaborated upon by the work of Kitano [29] (fine tuning but may let ANN fall in local minima) and then is further improved by McInerney and Dhawan [30] (avoided the local minima problem) were among the first researchers to propose the use of GA to train ANN and to find the best topology. This operation corresponds to coarse search to find the best region that contains the initial weights for the ANN. After finding the initial weights, local search method like backpropagation (BP) is used to fine-tune these parameters through a learning algorithm.
Using Pythia which is an ANN designer tool, an ANN was designed to model the price of the land with a property constructed on it taking into consideration several positive and negative indicators that may affect the final price. The ANN model is built in two stages: In the first stage a GA optimiser is used to determine the best structure for the ANN (coarse-tuning). In the second stage, the selected ANN structure is trained (fine-tuned) for the best values for the weights [31].
The features of the lands and properties form the training data that is ready to be fit into an ANN structure to be trained. During the training phase sample data containing both - inputs and desired outputs - are processed to optimise the network's output, meaning to minimise the deviation. After training the ANN structure, the resulted ANN structure can aid in accurate prediction of properties' prices. The data are first normalised to have 0 mean and 1 standard deviation for faster convergence during the training . To choose the best parameters for the GA optimiser, each possible ANN structure is represented as a chromosome.
The following GA parameters are used in building the ANN model: Generate up to 1000 generation, 50 ANNs (chromosomes) are generated in each population, elitism count of 5, cross-over rate of 0.2, mutation rate of 0.4. For the number of generations produced, maximum number of generations is set to 1000 as the experiments we conducted showed that no improvement is achieved after this number of generations. The population size determines how many possibilities (ANNs) will be examined in each generation, a population size of 50 is chosen. Several values were attempted and it was noticed that no performance gain was achieved with values greater than 50. The best 5 ANNs of the current generation are taken to the next generation (elitism) to ensure the performance of the new generations is not worse than the performance of the current generation. After experimenting with many values for crossover, it is chosen to produce 20% of the children (other than children produced with elitism) using crossover, while the mutation rate was chosen to be 40%. Three factors were considered to determine the fitness of an ANN in the production process, the first factor is the mean deviation between the desired output and the output produced by a chromosome. In our case the difference is between the price of a property provided by the training data and the output produced by an ANN chromosome, this factor is set to the target of 0.1. The second factor is the maximum deviation between the training data output and the output produced by an ANN. The third factor is the number of neurons in an ANN model, the fewer the number of neurons, the fitter the ANN. We set this factor to the maximum of 40 neu-

rons. These three factors are combined to determine the overall fitness.
After 1000 generations, the best network produced by the GA optimiser is selected for further training and tuning of the values of its weights to better model the relation between the input and the output. The final ANN model has 10 inputs corresponding to the input data, it has also 2 hidden layers consisting of 10 neurons each with tansig transfer function. Linear transfer function is used in the output layer which has 3 neurons corresponding to the predicted price according to the cost approach, sales comparison approach and income approach. The total number of neurons equal 23.

5.2 Training the ANN Structure

Multi-layer feedforward ANN, trained using the backpropagation algorithm is used to model the relation between the input features and the target price. Levenberg Marquardt (LM) algorithm was used to train multi-layer backpropagation networks as it performs well for function approximation problems, it also has a very efficient MATLAB implementation [17]. The network used in this paper was allowed to be trained for up 10000 epoch although in most cases it stabilised after less than 50 epoch. Mean Square Error (MSE) is used as performance function. The goal was to achieve goal of up to 0.01, by stopping the training process before achieving 100% accuracy prevents the network from overfitting the training data and performing bad with testing data.

6 Results and Analysis

A standard 10-fold cross validation is performed where in each fold 10% of the data are chosen as testing subset and the remaining data are divided as 80% training subset and 10% as validation subset. This corresponds to 713 training, 89 validation and 89 testing vectors. For each fold, an ANN is trained using the training set where the input for the training was lands and properties features to predict lands and properties prices according to the cost approach, sales comparison approach and income approach. The performance is tested using the testing subset.
To check whether there is a difference between the empirical values and the values obtained through ANN, a two paired-t statistical test is performed for the produced classifier in each fold to compare between target prices and prices produced by ANN. The produced t-values are listed in Table 1 where the degree of freedom is 88 as there are 89 testing pair of empirical and ANN produced values in each fold.

From Table 1, it can be concluded that there is no significant difference between the target prices and the ANN predicted prices at the 1% significance level (99% confidence level) for the 10 ANNs as all the t-values are included in the interval [-2.5758, 2.5758] which confirms the effectiveness of the 10 trained ANNs in different folds in predicting accurate prices that corresponds to the target prices obtained from the records of the Jordanian Department of Lands & Survey. When different test sets (890 vector) were grouped together to compare the target values with the predicted values produced by the corresponding ANN, the produced t-values were 0.0008, 0.0003, and -0.0007 for the cost approach, sales comparison approach and income approach, respectively, which also indicates that there is no significant difference between the two models at 1% significance level.
Multiple correlation coefficient (R) values between the values obtained from Jordanian Department of Lands & Survey and the predicted prices for the cost approach, sales comparison approach and income approach in each of the 10 folds are listed in Table 2. The average R values over the 10 folds are also computed in the last column for the three approaches.
Performing comparison between the target and ANN values over the whole data sets yields R values of 0.9372, 0.9436, and 0.7941 for the cost approach, sales comparison approach and income approach, respectively, which indicates strong positive correlation and a good fit. The R^2, the coefficients of determination have the values of 93.72%, 94.36%, and 79.41% which indicates that 93.72%, 94.36%, and 79.41% of the time the variation in the independent variable is explained by the model for the cost approach, sales comparison approach and income approach, respectively. The correlation value according to the income approach equals 0.7941. The reason for the relatively low value is the absence of an important field in the training data which is the average rent value for the square meter. The value for this field is not available in the obtained data and it is highly recommended to store this value with the record of the property.
Boxplots of differences between prices obtained from Jordanian Department of Lands & Survey and prices predicted from ANN over the whole data sets are shown in Figure 1 (a), (b), and (c) for the cost approach, sales comparison approach and income approach, respectively. It appears that the values of prediction error are clustered in the lower range. The

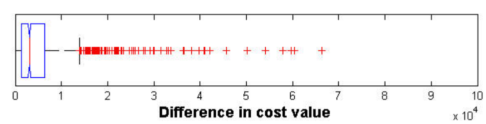
	Fold										
Approach	1	2	3	4	5	6	7	8	9	10	**Average**
Cost	-0.0012	0.0206	-0.0196	-0.0045	0.0069	0.0038	0.0189	0.0041	0.0146	0.0052	0.0049
Sales	0.0022	0.0144	-0.0239	0.0065	-0.0024	-0.0015	0.0204	-0.0021	0.0137	0.0121	0.0039
Income	0.0033	-0.0162	-0.0053	0.0166	0.0135	-0.0092	0.0215	-0.0027	0.0165	-0.0079	0.0030

Table 1: t-values between the target and the ANN predicted prices in each of the 10-folds.

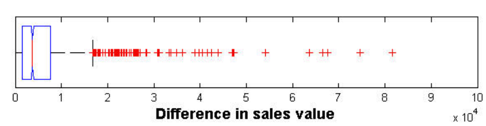
	Fold										
Approach	1	2	3	4	5	6	7	8	9	10	**Average**
Cost	0.9693	0.8845	0.9038	0.9832	0.9724	0.9766	0.9062	0.9819	0.9568	0.9814	0.9516
Sales	0.9655	0.8951	0.9070	0.9794	0.9689	0.9855	0.9004	0.9795	0.9570	0.9796	0.9518
Income	0.8645	0.8352	0.7769	0.8453	0.9680	0.7739	0.8933	0.9154	0.9166	0.5976	0.8387

Table 2: Correlation values between the target and the ANN predicted prices in each of the 10-folds.

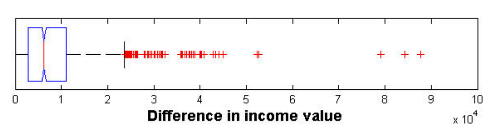
boxplot for the cost approach is shown in Figure 1 (a), the first quartile is below 1345 JD and the lower two quartiles (median value) are below 3065 JD. More than 75% of the data are below 6373 JD. Around 100% of the differences are below 13890 JD which indicates the ANN model gives high accurate estimation according to the cost approach. The average absolute difference in estimation is 6406.73 JD which is relatively a very small value knowing the range of prices in the considered properties varies from 570 JD to 589640 JD. There are few outliers (out of 890) with high prediction error.
The boxplot for the sales comparison approach is shown in Figure 1 (b), the first quartile is below 1600 JD and the median value is below 3747 JD. More than 75% of the data are located below 7666 JD. Around 100% of the differences are below 16780 JD which indicates the ANN model gives high accurate estimation according to the sales comparison approach. The average absolute difference in estimation is 7070.632494 JD which is relatively a very small value knowing the range of price in the considered properties varies from 560 JD to 574760 JD. Few outliers have high prediction error. The boxplot for the income approach is shown in Figure 1 (c), the first quartile is located below 2701 JD and the median value is located below 6228 JD. More than 75% of the data are below 11080 JD. Around 100% of the differences are below 23570 JD which indicates the ANN model gives high accurate estimation according to the income approach. The average absolute difference in estimation is 10236.55 JD which is relatively a very small value knowing the range of price in the considered properties varies from 560 JD to 574760 JD. Few outliers have high prediction error. The price in the income approach is deviated from the price obtained from the ANN model due to the absence of the average rent value for the square meter.



(a) BoxPlot of differences in cost approach.



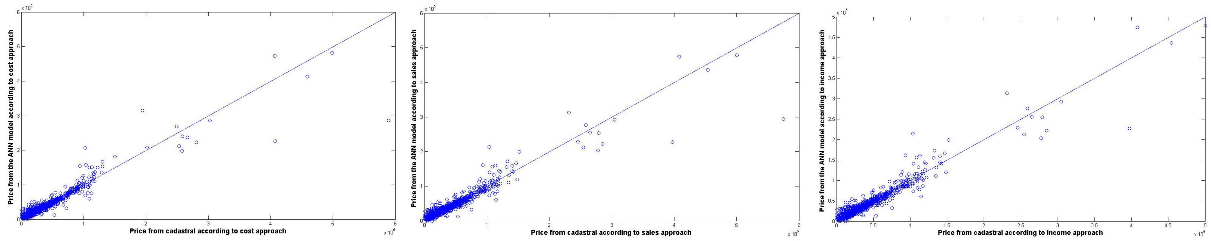
(b) BoxPlot of differences in sales comparison approach.



(c) BoxPlot of differences in income approach.

Figure 1: Boxplot of the error in price prediction between target prices and Artificial Neural Network prediction.

Figure 2 (a), (b), and (c) show the scatter diagram between prices obtained from Jordanian Department of Lands & Survey and prices predicted from ANN over the whole data sets from the ANN model to visualise the correlation between the corresponding values. Most of the points are located near the perfect fit line due to the very high correlation. Few outliers are located far from the perfect fit line.



(a) scatter diagram of price prediction according to the cost approach.

(b) scatter diagram of price prediction according to the sales comparison approach.

(c) scatter diagram of price prediction according to the Income approach.

Figure 2: Scatter diagram between of price prediction between target prices and Artificial Neural Network prediction.

7 Conclusions and Future Work

Timely asset valuation and pricing is a big challenge in the commercial and financial world. Mass properties appraisal is currently estimated by sampling techniques which are less accurate when used for large batch of properties, therefore automatic valuation process is needed. Property appraisal should be performed accurately as such appraisal is required for tasks such as asset valuation for lenders, property tax estimation, insurance estimation, and estate planning. Accurate appraisal of assets is of critical importance to users such as commercial lenders whose slim margins leave little room for error. Property appraisal is also used to guide potential buyers and sellers with making purchasing decisions and are needed for a variety of insurance purposes.

In this paper, ANN was used to model the relation between the land and properties attributes and the estimation of the prices according to the cost approach, sales comparison approach, and income approach for data obtained from the Jordanian Department of Lands & Survey. Several experiments were conducted with 10-fold cross validation. To check the accuracy of the proposed ANN structure in predicting the price of properties accurately, several statistical tests were performed to check if there is a difference between the values obtained using the proposed method and the values obtained using data from Jordanian Department of Lands & Survey. Experimental results indicate that the proposed method obtains accurate price prediction of properties.

The proposed model is an attractive solution for property price appraisal as it offers an automatic proper, objective, consistent, cost-effective and accurate valuation for large number of properties. Therefore, it has wide applicability in estimating the properties prices in Jordan to aid workers in Jordanian Department of Lands & Survey.

In the future more data should be collected to confirm the findings of this paper, such data include physical condition of the premises and their building, and their equipment and facilities, and the neighbourhood of the building. Other variations to the proposed system here can be attempted to check their performance. For example fuzzy logic inference could be used in conjunction of neural networks to test the performance of the system where the membership of the input parameters is checked against certain intervals.

References:

[1] R. Zhang, W. Chen, A study on automated valuation model in mass appraisal system for real property tax, 2009, pp. 254 –257.

[2] Appraisal Institute, Appraising Residential Properties, Chicago IL, Part VI (1994).

[3] Appraisal Foundation, Uniform standards of professional appraisal practice and advisory opinions (USPAP), Washington, DC: Appraisal Foundation (2004).

[4] IAAO, Standard on automated valuation models (avm's), Assessment Journal (2003 fall) 109 – 154.

[5] P. Bonissone, W. Cheetham, Financial applications of fuzzy case-based reasoning to residential property valuation, Vol. 1, 1997, pp. 37 –44.

[6] C. Anderson, C. Kelley, B. Simpson, Neural networks in real estate valuation, Tech. rep., Southern Methodist University, Dallas, Texas (2002).

[7] P. Khedkar, P. Bonissone, Method and System for Automated Property Valuation-US Patent Number 20020147695 (October 2002).

[8] P. Bonissone, Soft computing systems: commercial and industrial applications, Vol. 2, 1999, pp. 580 –585 vol.2.

[9] U.S. Census Bureau, USA Quick Facts (2008).

[10] Realtor.com-The official site of the National Association of REALTOR, Real Estate Listings, Homes for Sale and Rental Property Listings (2010).

[11] T. Lasota, E. Pronobis, B. Trawinski, K. Trawinski, Exploration of soft computing models for the valuation of residential premises using the keel tool, Intelligent Information and Database Systems, Asian Conference on 0 (2009) 253–258.

[12] R. Schulz, Valuation of Properties and Economic Models of Real Estate Markets,, Ph.D. thesis, Humboldt-Univ, Berlin, Germany (2003).

[13] Department of Real Estate, A Real Estate Guide, State of California, Department of Real Estate, California, 2000, 1st edition.

[14] P. Bonissone, Y.-T. Chen, K. Goebel, P. Khedkar, Hybrid soft computing systems: industrial and commercial applications, Proceedings of the IEEE 87 (9) (1999) 1641 –1667.

[15] T. M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.

[16] J. G. A, R. Laureano-Ortiz, A case-based reasoning approach to real estate property appraisal, Expert Systems with Applications 4 (2) (1992) 229 – 246.

[17] M. AL-Akhras, H. Zedan, R. John, I. AL-Momani, Non-intrusive speech quality prediction in voip networks using a neural network approach, Neurocomputing 72 (10-12) (2009) 2595 – 2608.

[18] G. O. Tirian, C. B. Pinca, Applications of neural networks in continuous casting, WSEAS Transactions on Systems 8 (6) (2009) 693 – 702.

[19] G. Lappas, Designing neural networks for tackling hard classification problems, WSEAS Transactions on Systems 8 (6) (2009) 743 – 752.

[20] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, Upper Saddle River, NJ, 1999, 2nd edition.

[21] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation (1986) 318–362.

[22] Image processing with neural networks–a review, Pattern Recognition 35 (10) (2002) 2279 – 2301.

[23] L.-Y. Chang, H.-C. Chen, Tuning of fractional pid controllers using adaptive genetic algorithm for active magnetic bearing system, WSEAS Transactions on Systems 8 (1) (2009) 158 – 167.

[24] J. Min, W. Qin, X. Lifeng, Research and implementation on genetic algorithms for graph fitness optimization, WSEAS Transactions on Systems 7 (4) (2008) 321 – 331.

[25] S. Thede, An introduction to genetic algorithms, DePauw University Greencastle, IN 46135.

[26] N. Clara, Neural networks complemented with genetic algorithms and fuzzy systems for predicting nitrogenous effluent variables in wastewater treatment plants, WSEAS Transactions on Systems 7 (6) (2008) 695 – 705.

[27] H. Yue, L. Chun-Mei, Partitioning study of complex system, WSEAS Transactions on Systems 7 (12) (2008) 1516 – 1525.

[28] D. J. Montana, L. Davis, Training feedforward neural networks using genetic algorithms, in: IJCAI'89: Proceedings of the 11th international joint conference on Artificial intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989, pp. 762–767.

[29] H. Kitano, Designing neural networks using genetic algorithms with graph generation system, Complex Systems 4 (4) (1990) 461–476.

[30] M. McInerney, A. Dhawan, Use of genetic algorithms with backpropagation in training of feedforward neural networks, in: Neural Networks, 1993., IEEE International Conference on, 1993, pp. 203 –208 vol.1.

[31] RUNTIME SOFTWARE, Pythia v1.02 - the neural network designer, `http://www.runtime.org/pythia.htm` (2000–2004).