

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,000

Open access books available

148,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Pre-Informing Methods for ANNs

Mustafa Turker

Abstract

In the recent past, when computers just entered our lives, we could not even imagine what today would be like. If we look at the future with the same perspective today, only one assumption can be made about where technology will go in the near future; Artificial intelligence applications will be an indispensable part of our lives. While today's work is promising, there is still a long way to go. The structures that researchers define as artificial intelligence today are actually programmed programs with limits and are result-oriented. Real learning includes many complex features such as convergence, association, inference and prediction. It has been demonstrated with an application how to transfer the input layer connections in human neurons to the artificial learning network with the pre-informing method. When the results are compared, the learning load (weights) was reduced from 147 to 9 with the proposed pre-informing method, and the learning rate was increased between 15–30% according to the activation function used.

Keywords: ANN, pre-informing, AHP, modified networks, interfered networks

1. Introduction

The learning mechanism makes human beings superior to all other creatures. Despite the fact that today's computers have much more processing power, the human brain is still much more efficient than any computer or any artificially developed intelligence.

Building a perfect learning network requires more than just cell structures and its weights. The human brain has a very complex network, and each brain is unique for itself. Today's technology is not enough to explain all the details of how our brain works. My observation of how our brain works starts from defining items. Every item has a key cell in our brain. Defining process is done by visuals, smell, feeling, linguistic name, hearing its sound. If these key cells match any of these information from body inputs, thinking and learning continues, if there is no key cell defined before, new cell is assigned for this item. Then, your brain wants to explore these item's behavior. You start to take this item in your hand and start the psychical observation. When the psychical observation is satisfied, your brain starts to categorize it. After categorization, your brain checks other items for same categorization and determines what other information can be learned. Whenever you see someone has more knowledge from you, then you want to speak about this newly learned item, or you want to do research on it. This key cell started to develop itself with explored information. Each key cell

and its network can also connect to each other in any part, if there are logical connections that exist.

Today's artificial intelligence studies are a little simple compared to reality. Mathematical modeling of learning in an artificial cell and solving the problem with an optimization mechanism has resulted in success in most areas. However, this success is due to the fast processing capacity of computers rather than the perfect modeling of machine learning. In this case, researchers need to work on developing artificial neural networks close to the real learning.

In this study, the pre-informing method and rules in artificial neural networks are explained with an example in order to establish a more conscious and effective learning network instead of searching for relationships in random connections.

2. ANN structure

In the literature of ANN design, the first principles were introduced in the middle of the 20th century [1, 2]. Over the following years, network structures such as Perceptron, Artron, Adaline, Madaline, Back-Propagation, Hopfield Network, Counter-Propagation Network, Lamstar were developed [3–10].

The complex behavior of our brain artificially imitated through layers is most network configuration. Basically, an artificial neural network has 3 types of layer group: input layer, hidden layers, output layer (See **Figure 1**). And all cells in these layers connected each other with artificial weights [1, 2].

Input layer is the cluster of cells present the data that has influence on learning. Each cell represents a parameter with a variable data value. These values are scaled

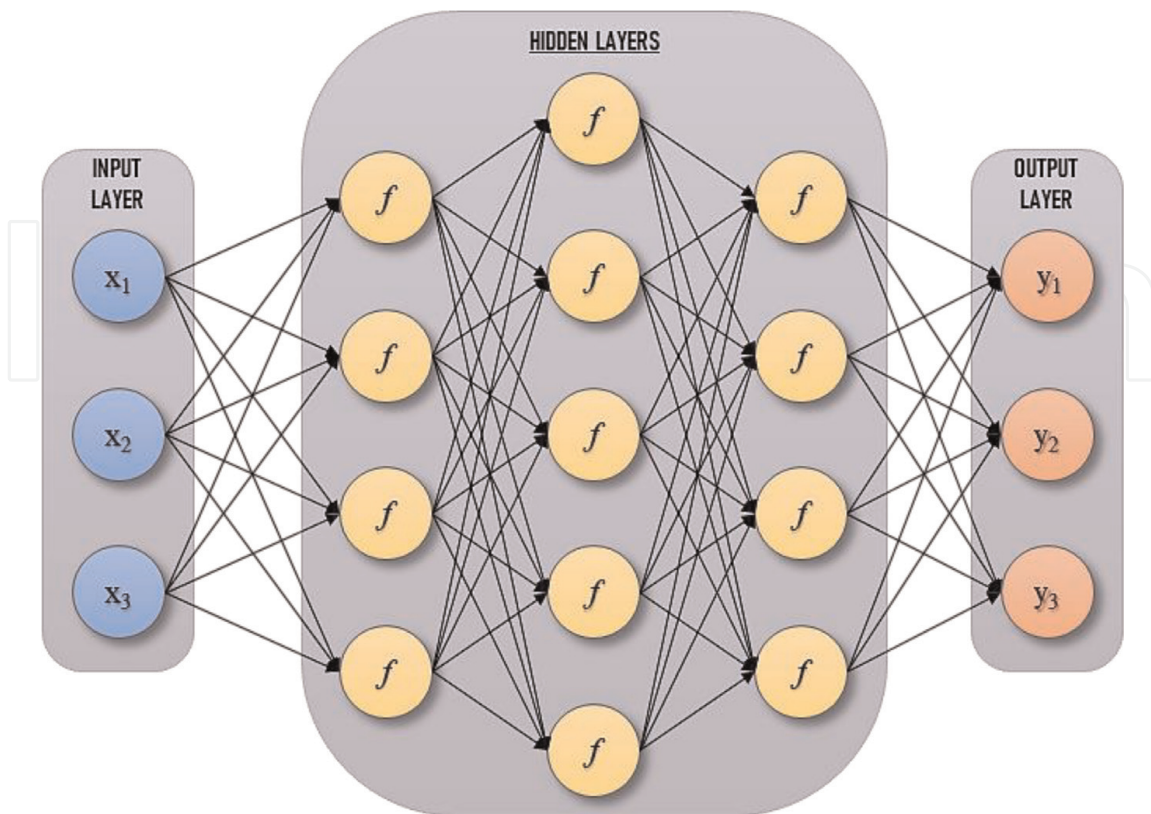


Figure 1.
Basic ANN structure.

according to the limits of the activation function used in the next layers. The selection of input parameters requires knowledge and experience on the subject to be created artificial intelligence. In fact, this process is exactly the transfer of natural neuron input parameters from our brain to paper. However, this is not so easy because a learning activity in our brain is connected by a huge number of networks managed subconsciously. To explain this situation, sometimes our minds make some inferences even on subjects we have no knowledge of, and we can make correct predictions about this subject. In some cases, we feel the result of an event that we do not know, but we cannot explain it. In fact, the best example of this is falling in love. No one can tell why you fall in love with a person, it happens and then you look for the reason. This is proof that the subconscious mind plays a major role in learning. This means that there may also be some input parameters that we did not notice. Therefore, it is necessary to focus on this layer and define the input parameters.

Hidden layer(s) is the layer where the data of the input parameters are interpreted, and the learning capability of the network is defined. Each cell in these layers transfers the data from the input layer cells or previously hidden layer cells with the defined activation function and sends it to all cells in the next layer. Learning of nonlinear behavior takes place in this layer. Increasing the number of layers and cells in this group does not always work, but provides memorization, not learning. This also increases the number of connections and thus highly increases the required experienced data to determine the weight values of these connections.

In general, the basic mechanism of an artificial neuron consists of two steps: summation and activation [1]. Summation is the process of summing the intensities of incoming connections. Activation, on the other hand, is the process of transforming the collected signals according to the defined function (See **Figure 2**).

There are many activation functions. The purpose of these functions is to emulate linear or non-linear behavior. The sigmoid function is one of the most commonly used activation functions.

Mathematically, the summation and activation process of an artificial neuron is expressed as below (See Eqs. (1) and (2)).

$$u = \sum_{n=1}^m x_i * w_i - \theta \quad (1)$$

$$y = f(u) \quad (2)$$

In these equations,

- x_i : Input value or previous cell output value for previous layer cells,
- w_i : Weight value of the connection for previous layer cells,

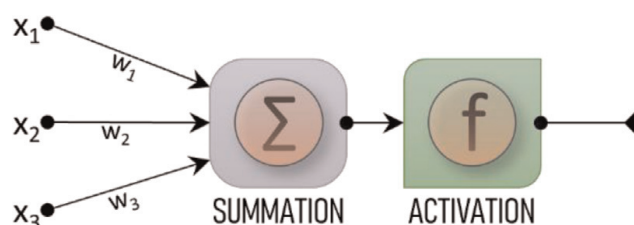


Figure 2.
Artificial neuron structure.

- θ : Bias value,
- u : Net collected output value of the cell,
- y : Activated output value of the cell.

In some cases, the learning network cannot find a logical connection between the results and the inputs, so that this does not stop learning, a bias value can be used for each cell. A high bias coefficient means that learning is low, and memorization is high.

The output layer is the last layer in the connection and receives inputs from the last set in the hidden layer. In this layer, data is collected and as a result, output data is exported in the planned method.

The learning process of the network established with the input, hidden and output layers is actually an optimization problem. The connection values between the cells of the network converge to reach the result depending on the optimization technique. A training set consisting of a certain number of input and output data is used for this purpose. If desired, a certain amount of data set is also tested to measure the consistency of the network. When the learning is complete, the values of the weights are fixed, and the network becomes serviceable. If desired, the mathematical equation of the network can be derived by following the cells from back to forward.

3. Pre-informing of ANNs

Pre-information, unlike pre-training, is the processing of a certain information or rule into the structure of the network. In reality, a person learns under some prejudices while learning something. These prejudices are a mechanism that allows us to make predictions about the event that will occur, and they make these inferences by utilizing similar events. With these prejudices, the number of training data required for learning decreases by a considerable ratio. As a result, you have a clean and efficient way of learning.

For example, for a child who goes out for the first time, his mother advises never to talk to strangers, and he guesses that if the child talks to a stranger, the result may be bad. In this case, the people to talk to are the input parameters, the possibility of something bad happening as a result of the conversation is the output parameter. If the mother did not give advice to her child, the child would talk to everyone and eventually learn that talking to a stranger is bad and dangerous. As a result of the mother's suggestion, the weight of strangers among the input parameters (people to talk to) increased before they even experienced it.

In order to transfer prejudices to artificial neural networks, some rules must be followed:

1. The pre-Informed network structure consists of 3 layers; input layer, hidden layer, output layer. The hidden layer consists of a single sublayer.
2. Input parameters should be grouped, if possible. For example, in a learning network that predicts heart attack, personal characteristics are one group, bad habits are another group, genetic diseases are another group. If there is no group, it should be considered as 1 group. These inputs should be scaled according to the activation function that will be used in the hidden layer.

3. The information to be processed (pre-informing) should be in the weights between the input layer and the hidden layer.
4. An artificial neuron cell is placed for each input group in the hidden layer to represent each group. This cell consists of 3 steps: summation, scaling, activation. Two or more different activation functions can be used in cells in the hidden layer. In this case, for each input group, same number of representation cells should be defined in the hidden layer.
5. The connections of cells in the input layer to the representation cells of other groups other than their own are considered 0.
6. The representation cells in the hidden layer are directly connected to the output layer.
7. Optimization optimizes the weights of the connections between the hidden layer cells and the output layer cells.
8. The connection values of the input layer groups to the representation cells in the hidden layer are determined and fixed for each group using the techniques in the literature.

In **Figure 3**, a total of 23 input parameters belonging to 3 input groups, these three groups are represented by two separate cells with hyperbolic tangent and sigmoid activations, and a hidden layer consisting of a total of 6 cells, and finally an output layer are described.

After the network structure is established, the next step is pre-informing the network. This stage is the transfer of information from the subconscious to network weights. This stage should be done for each group, and each group should be considered separately. The best method of this process is using AHP (Analytic Hierarchy Process) evaluation methods. In AHP evaluation methods, each parameter is compared with the other using verbal expressions. A simple superiority scale is used in this comparison. This means you can prepare a questionnaire and get the superiority information of parameters from an expert mind. After some calculations you will have the weights. These weights will be used in the network directly. The beauty of using this technique is consistency analysis can be done. In the end, if the input parameters are defined correctly, you will have 100% academically proofed subconscious information extraction.

AHP is a multi-criteria decision making (MCDM) method. The earliest reference to AHP is from 1972 [11]. Afterwards, Saaty [12], fully described the method in his article published in the Journal of Mathematical Psychology. AHP makes it possible to divide the problem into a hierarchy of sub-problems that can be more easily grasped and evaluated subjectively. Subjective evaluations are converted into numerical values and each alternative is processed and ranked on a numerical scale. Schematic AHP hierarchy is given in **Figure 4** below.

At the top of the hierarchy is the goal/purpose, while at the bottom there are alternatives. Between these two parts are the criteria and their sub-criteria. The most important feature that makes AHP important is that it can make comparisons both locally and globally when comparing the effect of sub-criteria at any level on alternatives.

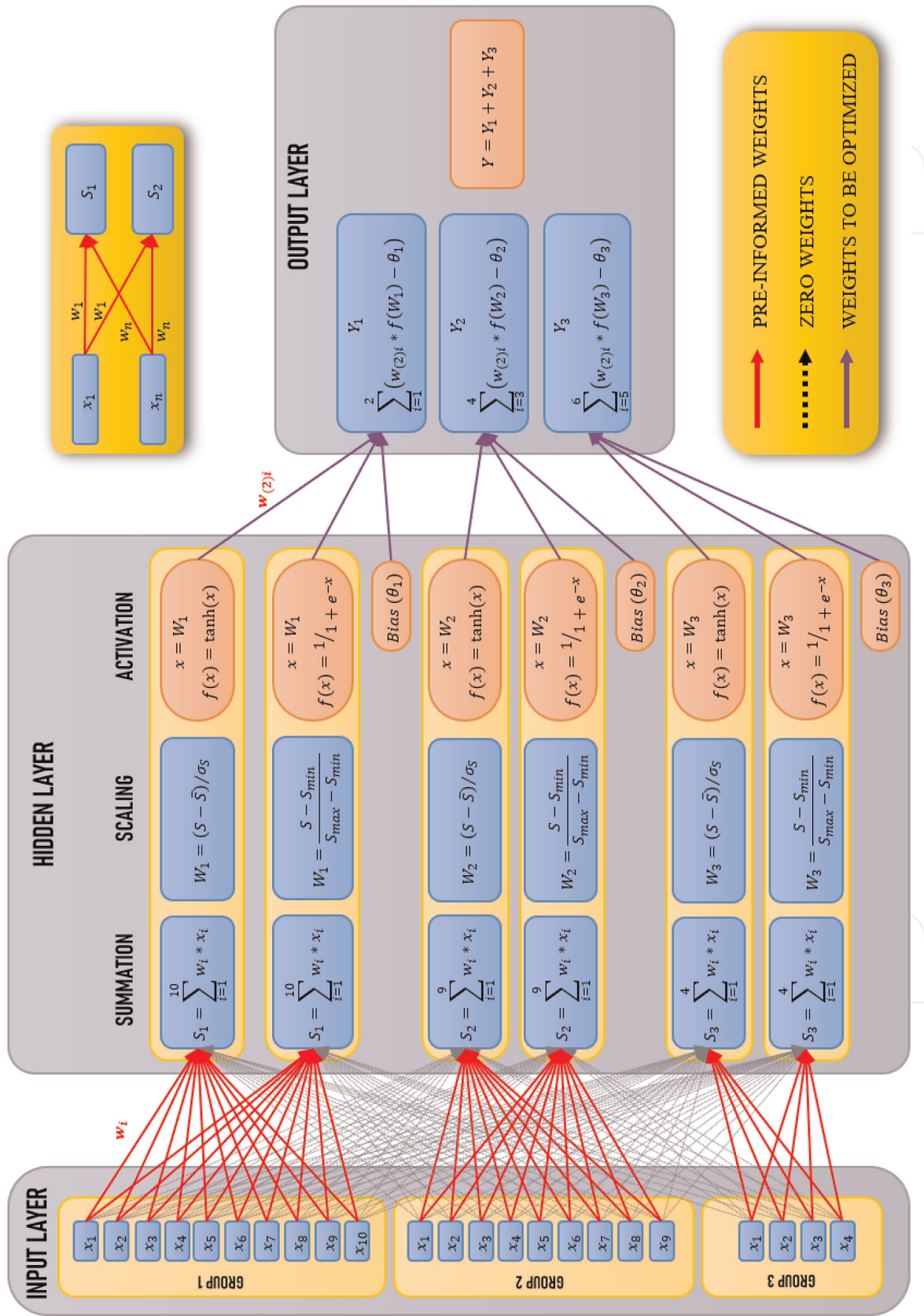


Figure 3. Pre-informed ANN structure.

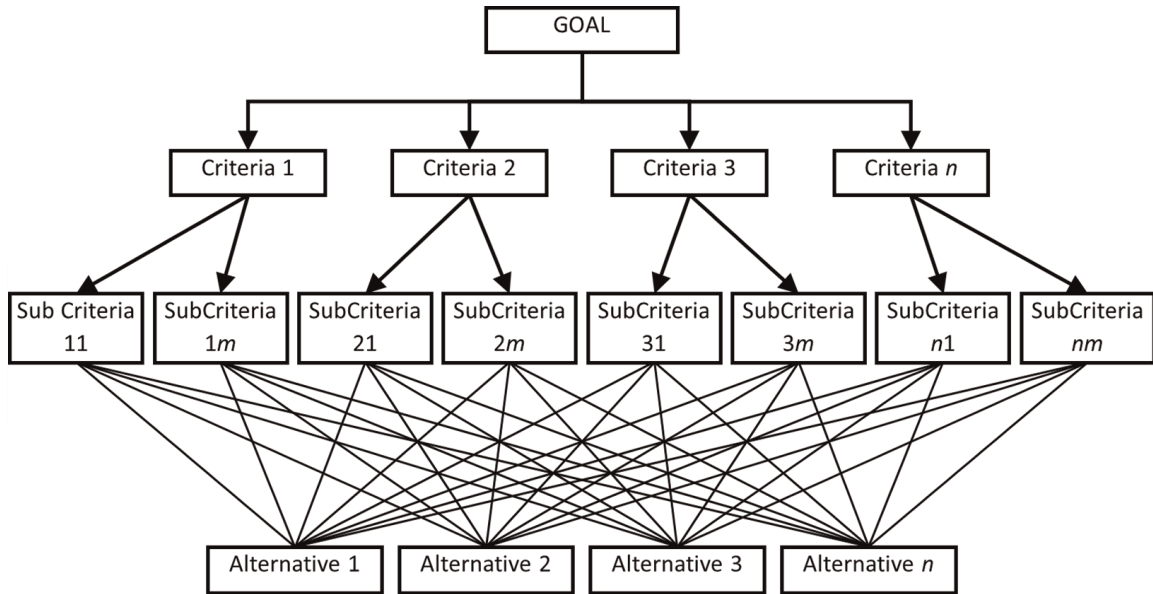


Figure 4.
 AHP hierarchy.

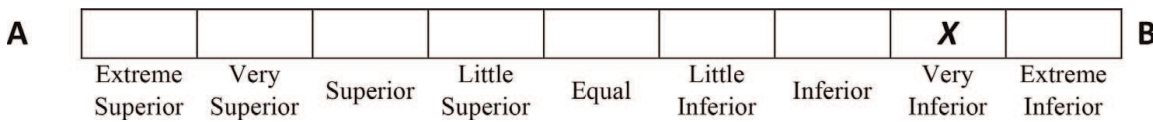


Figure 5.
 Pairwise comparison chart of alternatives A and B. B is very inferior compared to A.

Data corresponding to the hierarchical structure is collected by experts or decision makers by pairwise comparison of alternatives within the scope of a qualitative scale. Experts can rate the comparison as equal, less strong, strong, very strong and extremely strong. A general table, as shown in **Figure 5**, is used for expert evaluation of pairwise comparisons and data collection. This design can be customized for purpose, method and user usage.

Comparisons are made for each criterion and converted to quantitative numbers according to **Table 1**.

The pairwise comparison values of the criteria arranged in a matrix shown in **Table 2**.

Scale	Definition	Description
1	Equal	The two criteria are equally important.
3	Little Superior	One of the criteria has some superiority based on experience and judgment
5	Superior	One of the criteria has many advantages based on experience and judgment.
7	Very Superior	One criterion is considered superior to the other
9	Extreme Superior	Evidence that one criterion is superior to another has great credibility
2, 4, 6, 8	Intermediate values	Intermediate values to be used for reconciliation

Table 1.
 Comparison scales and explanations.

	C_1	C_2	C_3	C_n
C_1	$a_{11} = 1$	a_{12}	a_{13}	a_{1n}
C_2	$1/a_{12}$	$a_{22} = 1$	a_{23}	a_{2n}
C_3	$1/a_{13}$	$1/a_{23}$	$a_{33} = 1$	a_{3n}
C_n	$1/a_{1n}$	$1/a_{2n}$	$1/a_{3n}$	$a_{nn} = 1$
$\sum a$	$S_1 = \sum_{i=1}^n a_{i1}$	$S_2 = \sum_{i=1}^n a_{i2}$	$S_3 = \sum_{i=1}^n a_{i3}$	$S_n = \sum_{i=1}^n a_{in}$

Table 2.
Pairwise comparison matrix of criteria.

	K_1	K_2	K_3	K_n	w_i
K_1	a_{11}/S_1	a_{12}/S_2	a_{13}/S_3	a_{1n}/S_n	w_1
K_2	a_{21}/S_1	a_{22}/S_2	a_{23}/S_3	a_{2n}/S_n	w_2
K_3	a_{31}/S_1	a_{32}/S_2	a_{33}/S_3	a_{3n}/S_n	w_3
K_n	a_{n1}/S_1	a_{n2}/S_2	a_{n3}/S_3	a_{nn}/S_n	w_n
$\sum a$	S_1/S_1	S_2/S_2	S_3/S_3	S_n/S_n	$w_i = \left[\sum_{j=1}^n \left(\frac{a_{ij}}{S_j} \right) \right] / n$

Table 3.
Obtaining the weights of the normalized comparison values of the criteria.

In next step, each a_{ij} value is normalized by dividing by the corresponding column sum, and the weights shown in the table above are obtained with the corresponding equation shown in the **Table 3** above.

Network connections of input parameters using AHP are explained as shown above. Next step is how to assign weights. **Figure 6** shows how the AHP weights are defined to the network.

In this way, a large number of connections are canceled and a fast, efficient and less data-needing network is obtained.

4. Estimation of the severity of occupational accidents with using pre-informed ANN

The pre-informed neural network method was used by Turker [13] to predict the severity of occupational accidents in construction projects. In this study, it has been estimated how the accidents will result if they happen instead of the possibility of their occurrence. The scope of the study was made for the 4 most common accident types in the world. These are falling from high, hit from a thrown/falling object, structural collapse, electrical contact. In this study, 23 measures to be taken in occupational accidents are discussed in 3 groups. These measures have been associated with occupational accident severity in the artificial intelligence network (**Table 4**).

First of all, defined measures in occupational accidents, which are the input parameters, were turned into a questionnaire by creating paired comparison questions for comparison within their own groups. Occupational health and safety experts working professionally in the sector were reached through a professional firm. The questionnaires were administered online and recorded. Survey results were taken and converted to weights with AHP matrices. Weights are shown in **Tables 5–7**.

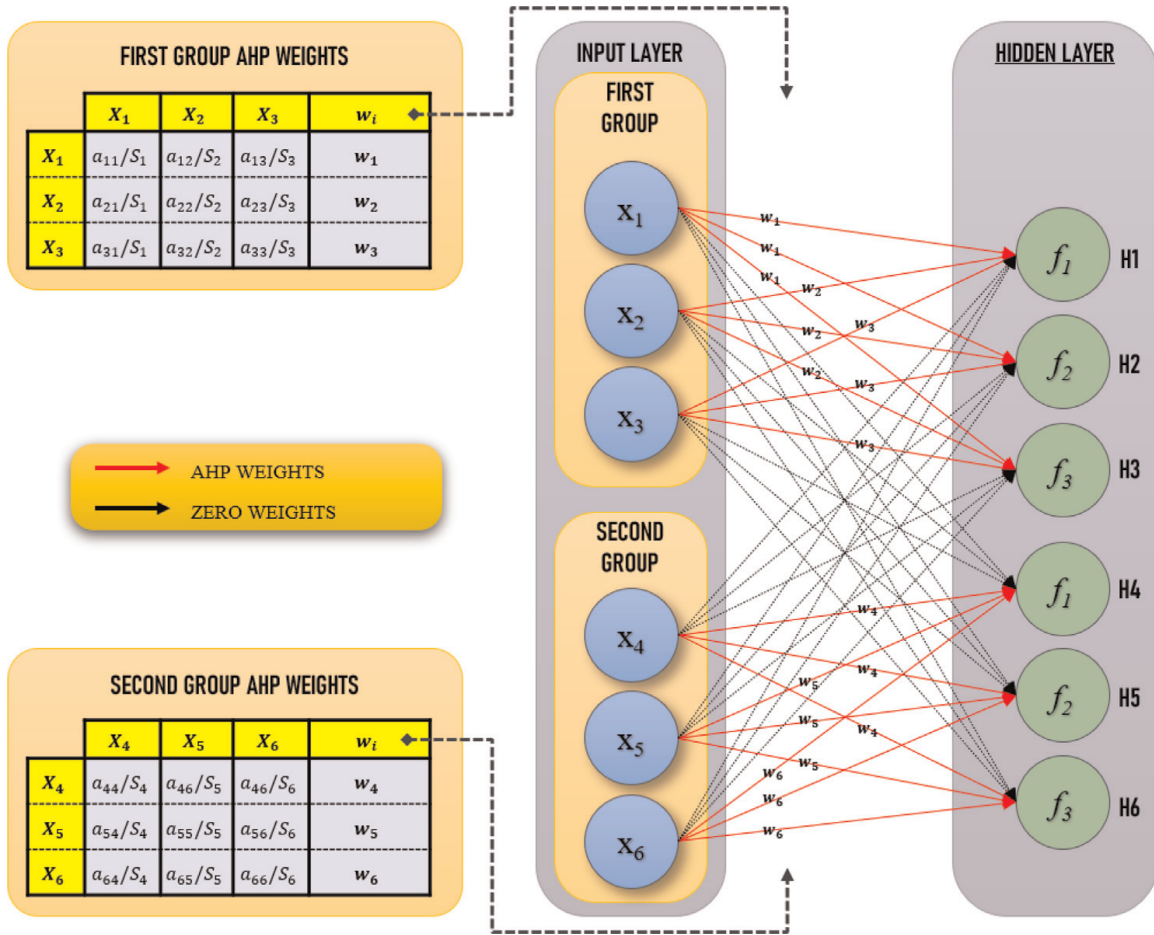


Figure 6. Connections of two input groups to three different types of representation cells and implementation of AHP weights.

Collective protection measures (TKY)	Personal protective equipment (KKD)	Control, training, inspection (KEM)
(TKY-1) Constr. site curtain system	(KKD-1) Safety Helmet	(KEM-1) OHS specialist
(TKY-2) Colored excavation net	(KKD-2) Protective Goggles	(KEM-2) Occupational Doctor
(TKY-3) Safety rope system	(KKD-3) Face Mask	(KEM-3) Examination
(TKY-4) Guardrail systems	(KKD-4) Face Shield	(KEM-4) OHS trainings
(TKY-5) Facade cladding	(KKD-5) Working Suit	
(TKY-6) Safety Field Curtain	(KKD-6) Reflector	
(TKY-7) First aid kit, fire extinguisher	(KKD-7) Parachute Safety Belt	
(TKY-8) Facade safety net	(KKD-8) Working Shoes	
(TKY-9) Mobile electrical dist. panel	(KKD-9) Protective Gloves	
(TKY-10) Warning and info signs		

Table 4. Risk reduction measures in occupational accidents.

After obtaining the preliminary information weights, 3 different artificial intelligence networks were created (Table 8). 140 historical accident data were collected on selected accidents within a company. These data include the precautions taken at the time of the accident and how the accident resulted. Accident results are divided into 4 categories: near miss, minor injury, serious injury, death. For each accident,

Code	Structural collapse	Falling from high	Object hit	Contact w/ Electricity
TKY-1	0,000	0,000	0,000	0,000
TKY-2	0,000	0,000	0,000	0,000
TKY-3	0,555	0,398	0,109	0,000
TKY-4	0,000	0,185	0,109	0,000
TKY-5	0,000	0,102	0,000	0,000
TKY-6	0,252	0,099	0,109	0,107
TKY-7	0,097	0,039	0,406	0,120
TKY-8	0,000	0,126	0,000	0,000
TKY-9	0,000	0,000	0,000	0,411
TKY-10	0,097	0,052	0,269	0,361

Table 5.
AHP weights of collective protection measures group.

Code	Structural collapse	Falling from high	Object hit	Contact w/ electricity
KKD-1	0,195	0,243	0,225	0,076
KKD-2	0,095	0,050	0,080	0,098
KKD-3	0,044	0,044	0,035	0,039
KKD-4	0,072	0,050	0,091	0,100
KKD-5	0,071	0,093	0,106	0,179
KKD-6	0,039	0,044	0,050	0,058
KKD-7	0,337	0,388	0,252	0,086
KKD-8	0,081	0,045	0,087	0,142
KKD-9	0,066	0,045	0,074	0,222

Table 6.
AHP weights of personal protective equipment group.

Code	Structural collapse	Falling from high	Object hit	Contact w/ electricity
KEM-1	0,481	0,167	0,399	0,426
KEM-2	0,210	0,167	0,161	0,134
KEM-3	0,098	0,167	0,083	0,067
KEM-4	0,210	0,500	0,357	0,372

Table 7.
AHP weights of control, training, inspection group.

35 datasets were collected and a total of 120 datasets were used in training the network and 20 datasets were used in testing the network.

Three alternative network structures were trained with the same data. As a result, the pre-informed neural network provided a better learning rate of 5% in the training

Network	Regular ANN	Pre-informed ANN	Pre-informed ANN
Software	SPSS – Neural Networks Engine	EXCEL VBA + SOLVER	EXCEL VBA + SOLVER
Network Structure	Multilayer Perceptron (MP)	MP	MP
Number of Hidden Layers	1	1	1
Cells in Hidden Layer	6 (Cells) + 1 (Bias)	6 (Cells) + 3 (Bias)	6 (Cells) + 3 (Bias)
Activation Function in Hidden Layer Cells	Hyperbolic Tangent (6 Cells)	Hyperbolic Tangent (6 Cells)	Parabolic Functions 3 Cells; $f(x) = x^2$ 3 Cells; $f(x) = x$
Output function	$f(x) = x$	$f(x) = x$	$f(x) = x$
Scaling Method	$(x - \bar{x}) / \text{Standard Dev.}$	$(x - \bar{x}) / \text{Standard Dev.}$	$(1-x) * 10$
Optimization Algorithm	Gradient Methods	Gradient Methods	Gradient Methods
Randomizer	Mersenne Twister algorithm	Mersenne Twister	Mersenne Twister
Initial Value	10	10	0,1

Table 8.
 3 alternative ANN structures.

set and 15% in the test set compared to the neural network without a pre-informed stage. The configuration using parabolic activation function from pre-informed artificial neural networks provided 1% better learning rate in the training set and 15% better in the test set compared to the configuration using hyperbolic tangent. Other configurations with activation functions were not included in the comparisons because of their low learning rates. As a result, it has been seen that the preliminary information phase significantly increases the learning performance in artificial neural networks. In addition, it has been observed that the parabolic activation function performs better than the hyperbolic tangent in relation to the prevention methods in occupational accidents and the result of the accident (**Table 9**).

	Network	Regular ANN	Pre-informed ANN	Pre-informed ANN
STRUCTURAL COLLAPSE	Training Set	26/30 (87%)	29/30 (97%)	30/30 (100%)
	Test Set	2/5 (40%)	4/5 (80%)	4/5 (80%)
CONTACT w/ ELECTRICITY	Training Set	27/30 (90%)	30/30 (100%)	30/30 (100%)
	Test Set	2/5 (40%)	4/5 (80%)	5/5 (100%)
OBJECT HIT	Training Set	30/30 (100%)	30/30 (100%)	30/30 (100%)
	Test Set	4/5 (80%)	3/5 (60%)	5/5 (100%)
FALLING FROM HIGH	Training Set	30/30 (100%)	30/30 (100%)	30/30 (100%)
	Test Set	4/5 (80%)	4/5 (80%)	4/5 (80%)
TOTAL	Training Set	113/120 (94%)	119/120 (99%)	120/120 (100%)
	Test Set	12/20 (60%)	15/20 (75%)	18/20 (90%)

Table 9.
 3 alternative ANN structure results.

5. Conclusions

In this study, how the learning ability of artificial neural networks should be increased with the pre-informing method is explained with rules and demonstrations. It is not possible to implement this method with the existing ready-made ANN software on the market. Instead, ANN should be expressed mathematically, and pre-informing method should be applied using programming languages such as MATLAB, Excel VBA, Python.

In this section, the application of this method has been demonstrated in an artificial neural network in which the precautions in occupational accidents are associated with the results of the accident and high performance has been achieved. With the application of the specified rules, this method can be used to solve many problems. In future studies, it can be investigated which other methods such as AHP can be used for the preliminary information phase.

Conflict of interest


The authors declare no conflict of interest.

Author details

Mustafa Turker
Gorkem Construction Company, Ankara, Turkiye

*Address all correspondence to: mustafaturker@yahoo.com

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Graupe D. Principles of artificial neural networks. 3rd ed. In: Advanced Series in Circuits and Systems. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2013. DOI: 10.1142/8868
- [2] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. 1943;5(4): 115-133
- [3] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958; 65(6):386
- [4] Graupe D, Lynn J. Some aspects regarding mechanistic modelling of recognition and memory. *Cybernetica*. 1969;12(3):119
- [5] Hecht-Nielsen R. Counterpropagation networks. *Applied Optics*. 1987;26(23): 4979-4984
- [6] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*. 1982;79(8): 2554-2558
- [7] Bellman R, Kalaba R. Dynamic programming and statistical communication theory. *Proceedings of the National Academy of Sciences*. 1957; 43(8):749-751
- [8] Widrow B, Winter R. Neural nets for adaptive filtering and adaptive pattern recognition. *Computer*. 1988; 21(3):25-39
- [9] Widrow B, Hoff ME. *Adaptive Switching Circuits*. Stanford, CA: Stanford University; 1960
- [10] Lee RJ. Generalization of learning in a machine. In: Preprints of Papers Presented at the 14th National Meeting of the Association for Computing Machinery (ACM '59). New York, NY, USA: Association for Computing Machinery; 1959. pp. 1-4. DOI: 10.1145/612201.612227
- [11] Saaty TL. *An Eigenvalue Allocation Model for Prioritization and Planning*. Pennsylvania, USA: University of Pennsylvania; 1972. pp. 28-31
- [12] Saaty TL. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*. 1977;15(3):234-281
- [13] Turker M. *Estimation of the Severity of Occupational Accidents in the Building Process with Pre-informed Artificial Learning Method*. Gazi: Gazi University; 2021