

Studying the Effect of Activation Function on Classification Accuracy Using Deep Artificial Neural Networks

Serwa A*

Faculty of Engineering in El-Mataria, Helwan University, Cairo, Egypt

Abstract

Artificial Neural Networks (ANN) is widely used in remote sensing classification. Optimizing ANN still an enigmatic field of research especially in remote sensing. This research work is a trial to discover the ANN activation function to be used perfectly in classification (landcover mapping). The first step is preparing the reference map then assume a selected activation function and receive the ANN fuzzified output. The last step is comparing the output with the reference to reach the accuracy assessment. The research result is fixing the activation function that is perfect to be used in remote sensing classification. A real multi-spectral Landsat 7 satellite images were used and was classified (using ANN) and the accuracy of the classification was assessed with different activation functions. The sigmoid function was found to be the best activation function.

Keywords: Remote sensing; Classification; ANN; Activation function; Landcover mapping; Fuzzy accuracy assessment

Introduction

Automation in remote sensing systems is a challenging field of research work due to the need of reducing cost. Artificial neural networks (ANN) approach is coming back to the research focus after it is extended to deep learning (DL). Although it is not a novel these days but some of its behavior stills enigmatic specially in remote sensing classification [1]. DL techniques have been proven useful in landcover classification. So, we need to advance the state of the art by learning not only weights between neurons or the network structure itself but also the activation functions. Current deep learning literature largely focuses on improving architectures and adding regularization to the training process. Remote sensing, particularly satellite imagery, is perhaps the only cost-effective technology able to provide data at a global scale. Within ten years, commercial services are expected to provide sub-meter resolution images everywhere at a fraction of current costs [2]. A little number of researches discussing the activation function (AF) which transfers the signal from the input units to the next hidden units. Recently due to the invention of DL researchers start to investigate the effect of AF on accuracy e.g., Michael Xie [3], who studied the transfer learning from deep features for remote sensing and poverty mapping. His study did not deal with multispectral images (MS) but only single band to study the light as a sign for richness or poverty areas. This research dealing with low level deep feature classification using MS images. The standard sigmoid reaches an approximation power comparable to or better than classes of more established functions investigated in the approximation theory [4,5]. Jordan presented the logistic function which is a natural representation of the posterior probability in a binary classification problem [5,6]. Özkan and Erbek [7] made similar study but using shallow learning in addition to examine only three AFs (linear, sigmoid and tanh). His study focusing on hard classification not fuzzy one like this research.

The novelty of this research work is make a sharp decision about selecting AF in remote sensing classification. This research makes a spotlight inside the black box of DANN because of severity in performance and accuracy obtained concerning with classification.

Deep Artificial Neural Network (Dann)

ANN are computational approach that simulate the microstructures of a biological nervous system depending on the signal transference.

Basically, all ANNs have a similar topological structure. Some of the neurons interface with the real world to receive its input and other neurons provide the world with the network's output. All the remaining neurons are hidden from view. So, there are three types of neurons; input, hidden and output neurons. One must note that input neurons have single input and multiple output while the hidden neurons have multiple input and multiple output. On contrast to input neurons is the output neurons that have multiple inputs and single outputs. Figure 1 shows a typical multi-layer perceptron (MLP) architecture with shallow learning. In Case of DANN the number of hidden layers is larger so that the weight updates is very slow and gradually moves toward the optimized solution carefully as shown in Figure 2.

Using MLP in DANN is widely used epically in supervised classification in the field of remote sensing. Both ANN and DANN must be optimized in architecture and performance; the first is concerned with number of hidden layers and the second concerned with selecting activation function (AF) in addition to achieving the correct weights.

The back propagation neural networks (BPNN) algorithm is a generalized least squares algorithm that adjusts the connection weights between units to minimize the mean square error between the network output and the target output, its architecture is MLP form. The target output is known from training data it is classes' values (Experimental work Results). In the first data entered to input unit are multiplied by the connection weights and is summed to result the nets input to the unit in the hidden layer as shown in Figure 1 and given by:

$$\text{net}_s = \sum X_i W_{is} \quad (1)$$

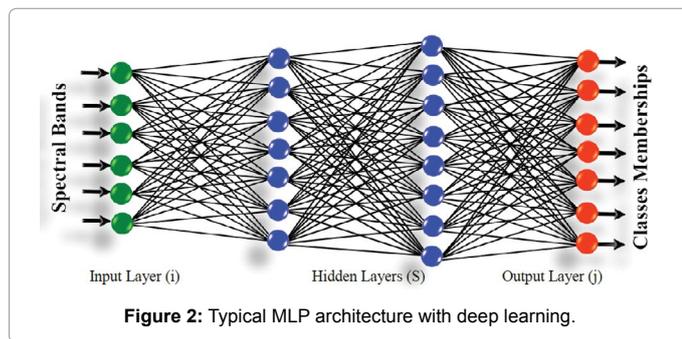
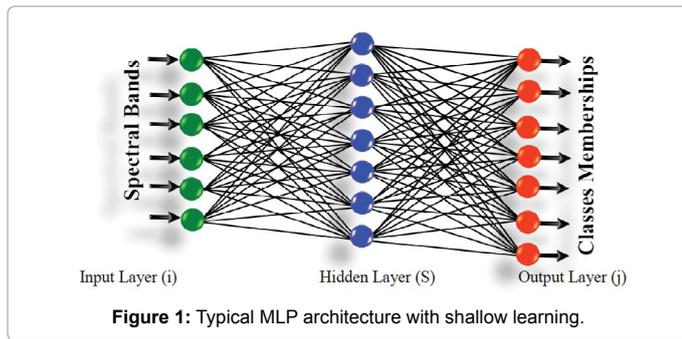
Where: X_i is a pixel vector of the input image of i^{th} input layer; W_{is} is matrix of the connection weights between the i^{th} input layer unit to s^{th} hidden layer unit. Each unit in s^{th} hidden layer calculates a weighted

***Corresponding author:** Serwa A, Faculty of Engineering in El-Mataria, Helwan University, Cairo, Egypt, Tel: 201003808123; E-mail: Dr.A.Serwa@m-eng.helwan.edu.eg

Received June 27, 2017; Accepted July 24, 2017; Published July 26, 2017

Citation: Serwa A (2017) Studying the Effect of Activation Function on Classification Accuracy Using Deep Artificial Neural Networks. J Remote Sensing & GIS 6: 203. doi: 10.4172/2469-4134.1000203

Copyright: © 2017 Serwa A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



sum of its inputs and passes the sum via an activation function to the units in the j^{th} output layer through weight vector W_{sj} . There are a range of activation functions to transfer the data from hidden layer unit to an output layer unit. These include linear, tangent hyperbolic, sigmoid functions etc. Although, the use of these functions may lead to difference in accuracy of classification, it can be defined as:

$$O_s = F(\lambda, \text{nets}) \quad (2)$$

where: F is called activation function (will be explained later), O_s is the output from the S^{th} hidden layer unit and λ is a gain parameter, which controls the connection weights between the hidden layer unit and the output layer unit. Outputs from the hidden units are multiplied with the connection weights and are summed to produce the output of j^{th} unit in the output layer as:

$$O_j = O_s W_{sj} \quad (3)$$

where: O_j is the network output for j^{th} output unit (i.e., the land cover class) and W_{sj} is the weight of the connection between s^{th} hidden layer unit and j^{th} output layer unit.

An error function (E), determined from a sample of target outputs (reference training data) and network outputs, is minimized iteratively. The process continues until E converges to minimum allowed value and the adjusted weights are obtained. E is given by:

$$E = 0.5 \sum_{j=1}^c (T_j - O_j)^2 \quad (4)$$

where T_j is the target output vector, O_j is the network output vector and c is the number of classes. The target vector is determined from the known class allocations of the training pixels, which are coded in binary form. The collection of known class allocations of all pixels will form the target vector. After computing the error of the network, it is compared with the limiting error E_L of the network. If $E < E_L$, the network training is stopped otherwise E is back propagated to the units in the hidden and the input layers. The number of iterations may vary from one dataset to the other, and is generally determined by trial and error. The process of back propagation and weight adjustment is explained in the following:

First the error vector at each unit of the output layer is computed as:

$$E_j = O_j (1 - O_j) E \quad (5)$$

Then the error vector for each unit at the hidden layer is computed as:

$$E_s = O_s (1 - O_s) \sum_{j=1}^c O_j E_j \quad (6)$$

Thereafter, the net error in connection weights between output layer and hidden layer is computed as,

$$E_{sj} = O_s E_j \quad (7)$$

And error in connection weights between hidden layer and input layer as,

$$E_{is} = F_m X_i E_s \quad (8)$$

where: F_m is the momentum factor which controls the momentum of the connections between the hidden layer unit and the input layer unit.

The weights between output layer and hidden layer are updated as:

$$(W_{sj})_{\text{new}} = (W_{sj})_{\text{old}} + E_{sj} \quad (9)$$

And the weights between input layer and the hidden layer are updated as:

$$(W_{is})_{\text{new}} = (W_{is})_{\text{old}} + E_{is} \quad (10)$$

The gain parameter λ is also updated as:

$$(\lambda)_{\text{new}} = (\lambda)_{\text{old}} + L_R E_s \quad (11)$$

where: L_R is the learning rate which controls the time of the learning process.

Methodology

The research is focused on the optimization of AF for remote sensing image classification so, Landsat 8 MS image is used. The reference of the study area is prepared pixel by pixel in order to achieve best training and testing performance. ENVI 5.3 software is used to achieve the necessary tasks such as: spatial subletting, preparing regions of interest (ROIs), geofencing classification etc. Another self-developed software called advanced digital image processor for remote sensing (ADIPRS) is used to carry out the DANN with different AFs. This software was developed by Serwa [8]. It is modified to cover DL task.

System overview

In this research Landsat 8 MS image is used with its spectral 30 m bands resolution. The reference of the study area is prepared pixel by pixel to achieve best training and testing performance. ENVI 5.3 software is used to achieve the necessary tasks such as: spatial subletting, preparing regions of interest (ROIs), geofencing classification etc. Another self-developed software called advanced digital image processor for remote sensing (ADIPRS) is used to carry out the DANN with different AFs. This software was developed by Serwa [8,9]. It is modified to cover DL task. System overview is indicated in Figure 3 in the form of block diagram. The reason for using ADIPRS in classification is the lack of selection of AF in any other remote sensing software.

Research data

The data used in this research work concerned with a part of great Cairo (Giza governate) in Egypt and it contains: 1-Eight bands of Landsat 8 satellite image. 2- AutoCAD maps with scale 1:5000. The

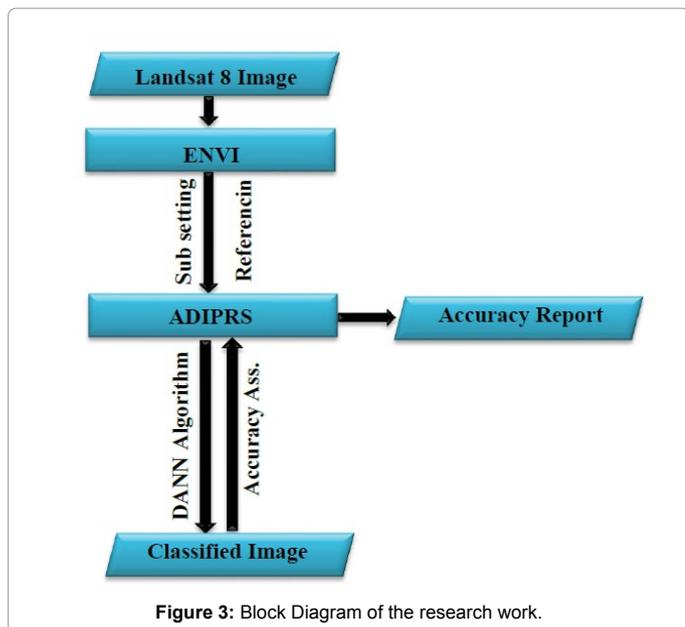


Figure 3: Block Diagram of the research work.

reference map was produced using AutoCAD maps (mainly) and SPOT 5 satellite image (secondary) of the study area in addition to suite visits with Garmin GPS map 62s navigator (3 meters accuracy). Figure 4 indicates the study area on ENVI environment in before spatial and spectral subset. The study area is represented by 500×500 pixels (225 km²) for the southern part of great Cairo (Giza governate). Figure 5 shows the reference that is built pixel by pixel to obtain a best training and testing performance.

Tools

Tools include two software; ENVI 5.3 for spatial subset and for all necessary tasks (reference preparing, all corrections) beside ADIPRS for applying DANN and for accuracy assessment. Figure 6 shows a flowchart diagram for DANN algorithm that ADIPRS apply according to the explained equations in section 2. In Figure 7 the interface of the DANN module of ADIPRS software. It is developed basically to achieve the research objective by selecting different AFs and apply DANN algorithm. The optimal architecture is indicated where number of input neurons equals number of image bands while the output neurons equals number of landcover classes. The hidden structure is:4 hidden layers with neurons 5, 8, 7 and 7.

The AF must be selected before run ADIPRS and the network architecture is fixed to study the effect of the variance of AF on the accuracy. After reaching the converging condition the classification report can be printed as shown in Figure 8. User accuracy, producer accuracy, overall accuracy and Kapa value (K-hat statistic) is included in the report. Both, user and producer accuracies for each class is also computed. A series of twelve AFs are studied they are; Linear, step, Piecewise Linear, Sigmoid, Complementary loglog, Bipolar, Bipolar Sigmoid, Tanh, Hard Tanh, Absolute, Rectifier and Smooth Rectifier. Figure 9 indicates the definition of each AF in the form of equation and graph.

The reason for selecting these twelve AFs is that they are the most mentioned in the literature. A heroic effort is made to develop and test each AF especially in testing its effect on classification accuracy. AF is affecting the classification results in remote sensing but rarely researches handle its effect on accuracy using real data. Performance accuracy (signal to noise ratio) was used to assess the accuracy.

Results

Each AF is selected to be used in classification then the accuracy assessment is carried out. Overall accuracy is chosen to express the accuracy gauge. The network architecture is fixed to examine only the effect of selecting AF on classification accuracy. Table 1 shows the end results numerically in the form of overall accuracy and number of iterations. Figure 10 shows the final abstracted results graphically.

Number of iteration is necessary because of its pointing to the cost function even if the cost is out of our research scope. We cannot neglect

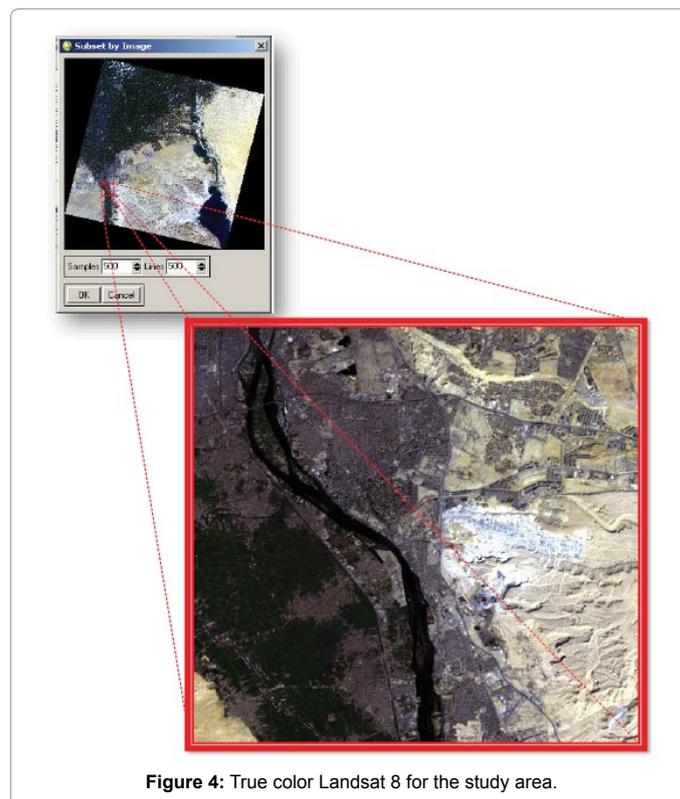


Figure 4: True color Landsat 8 for the study area.

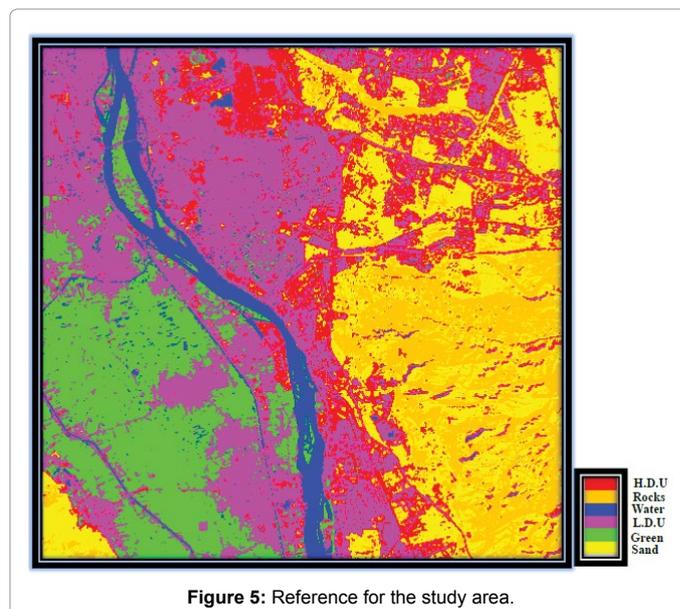
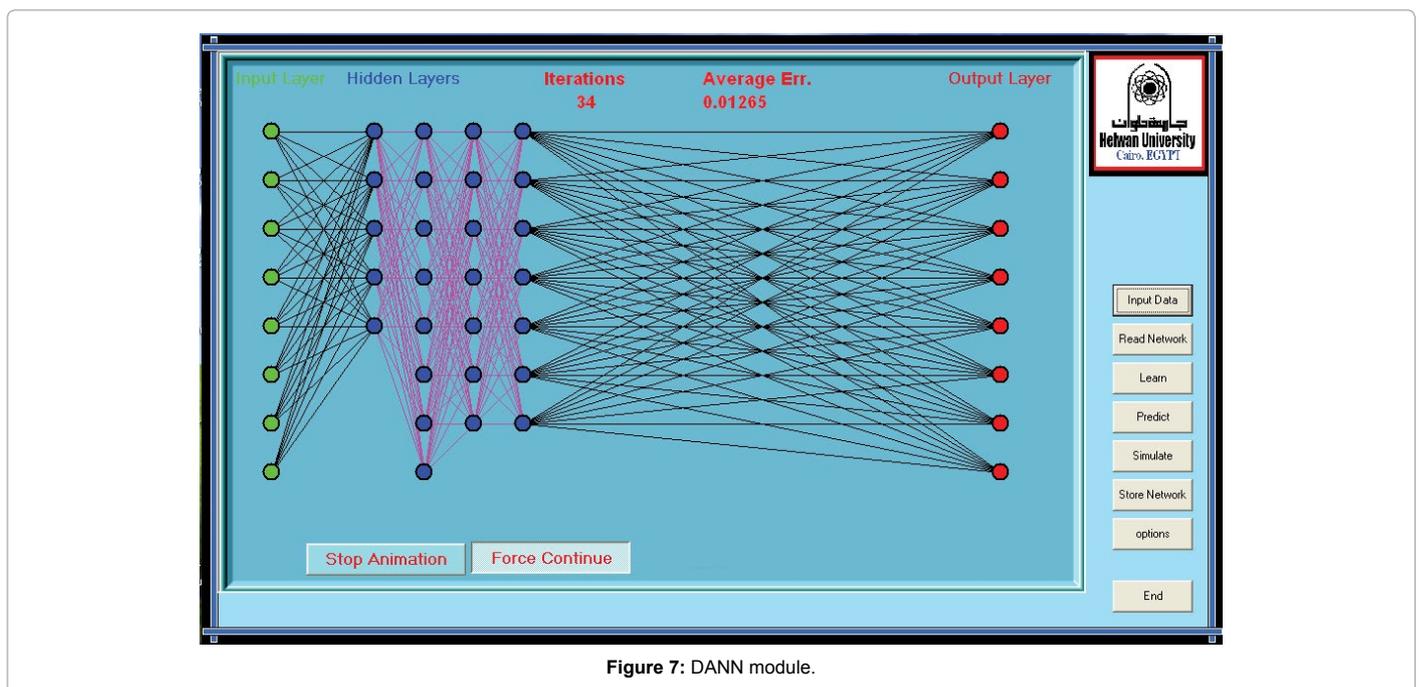
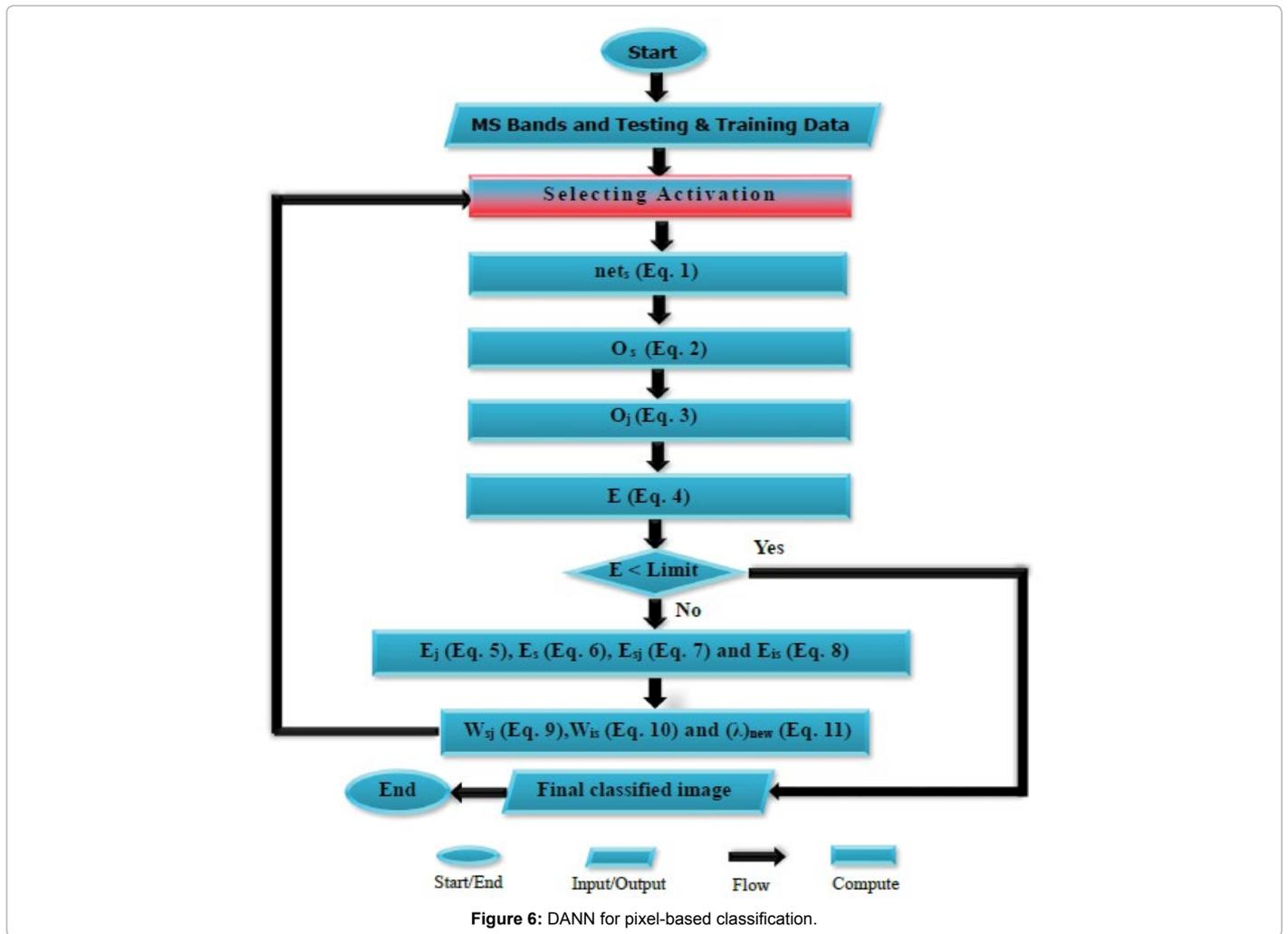


Figure 5: Reference for the study area.



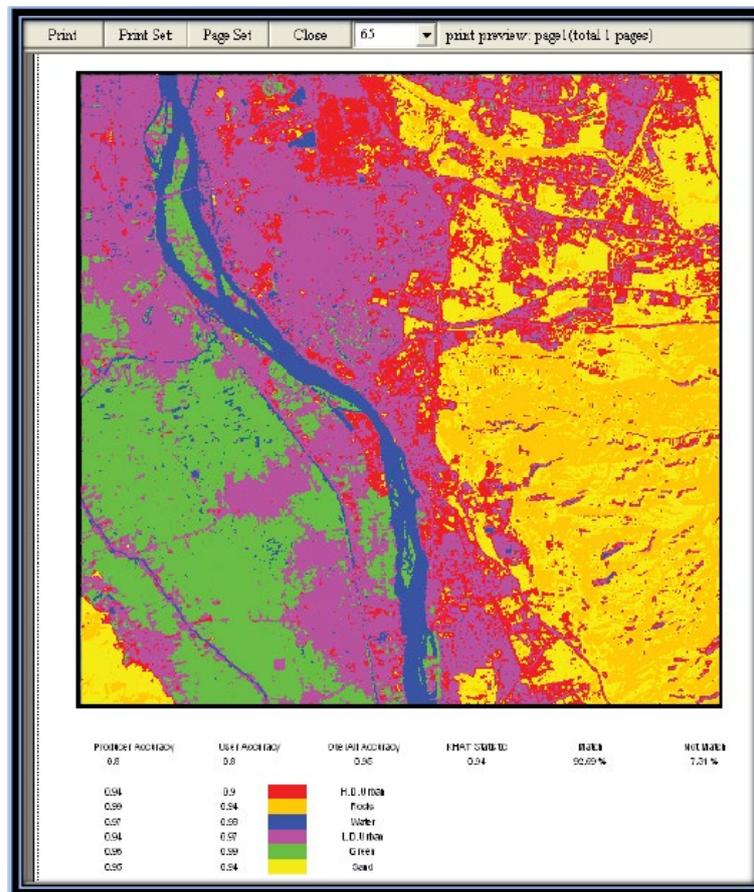


Figure 8: Classification report.

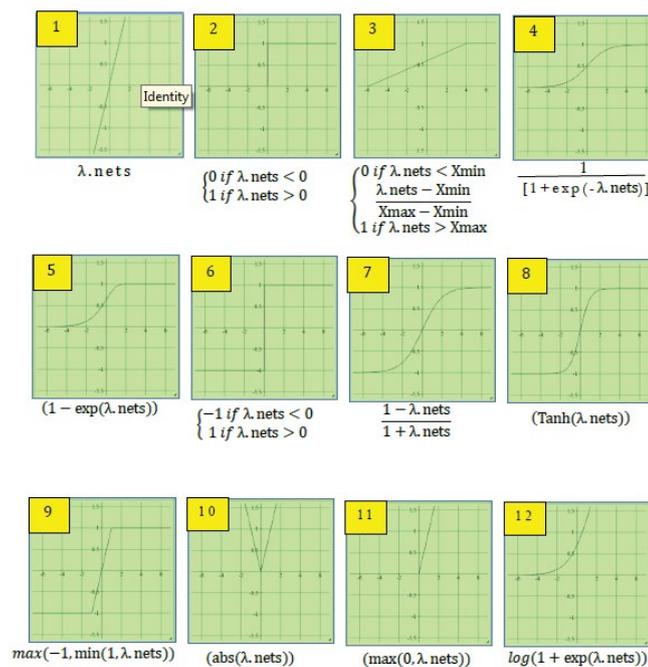


Figure 9: Studied AFs: 1) Linear. 2) Step. 3) Piecewise linear. 4) Sigmoid. 5) Complementary log log. 6) Bipolar. 7) Bipolar sigmoid. 8) Tanh. 9) Hard Tanh. 10) Absolute. 11) Rectifier. 12) Smooth rectifier.

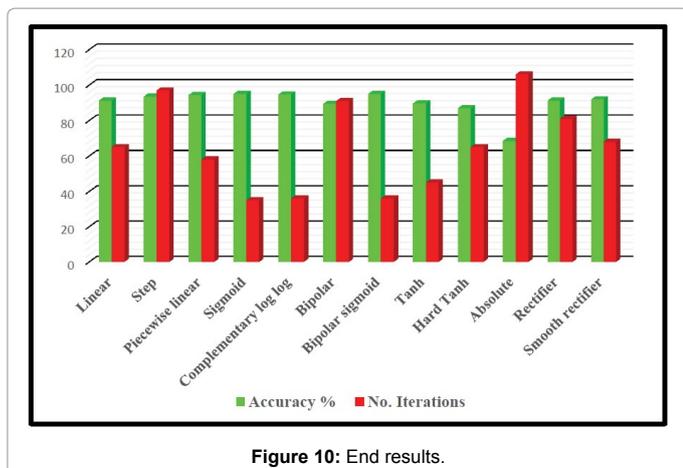


Figure 10: End results.

AF	Accuracy %	No. Iterations
Linear	91.23	65
Step	93.56	97
Piecewise linear	94.36	58
Sigmoid	94.98	35
Complementary log log	94.68	36
Bipolar	89.4	91
Bipolar sigmoid	95	36
Tanh	89.65	45
Hard Tanh	86.98	65
Absolute	68.45	106
Rectifier	91.21	81
Smooth rectifier	91.95	68

Table 1: Tabulation of end results.

the solution conditions such as number of iterations but it can be put as secondary criterion in the case of equality in accuracy. The maximum accuracy is achieved by using both bipolar sigmoid (95%) and sigmoid (94.98%) AFs. A test of hypothesis is carried out and found that it is not significant if we assumed that both results are equally. A statistical correlation test is made and the correlation is -0.6148 and it can be considered not correlated (overall accuracy vs. number of iterations). Sigmoid AF gives best accuracy in best number of iterations while the absolute AF gives the worst accuracy in max number of iterations. Some AFs such as step, piecewise linear and complementary log log gives unexpected results because they are not familiar in remote sensing. Their accuracies varied in the range of 93% -94%. The rest of the tested AFs can be considered moderate in accuracy and cost but we cannot

recommend it. The results show that accuracies of most of the tested classes behaves the same as the overall accuracy which means that the AF affect the general solution not a specific solution. That means the results can be generalized in behaviour.

In fact, the correlation study was not necessary the first dimension (AF) is variable and its order is not fixed but it is made just in case some researcher used to make it. Last issue, the accuracy varied from 68.45% to 95% due to the selection of AF and it is a meaningful change.

Conclusion

Little number of research works focused on optimizing AF for DANN. Remote sensing data behaves in different way as a random signal. AF must be selected carefully because it affects the classification results. One can conclude that AFs affect classification accuracy in remote sensing so that about 27% accuracy enhancement can be achieved by selecting a good AF. The classification time is affected by selecting a good AF. Both sigmoid and bipolar sigmoid AFs are recommended to be used in classification of remote sensing landcover features.

References

1. Serwa A (2016) Development of soft computational simulator for aerial imagery project planning. *Surveying and Land Information Science* 75: 65-75.
2. Murthy K, Shearn M, Smiley BD, Chau AH, Levine J, et al. (2014). SkySat-1: very high-resolution imagery from a small satellite. In: *Proceedings of SPIE*, 9241: 92411.
3. Xie M, Jean N, Burke M, Lobell D, Ermon S (2015) Transfer learning from deep features for remote sensing and poverty mapping. *Association for the Advancement of Artificial* pp: 3929-3935.
4. Das Gupta B, Schnitger G (1993) The Power of Approximating: A Comparison of Activation. In: *Advances in Neural Information Processing Systems*, Morgan Kaufmann Publishers, Burlington, USA.
5. Karlik B, Olgac AV (2010) Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. *International Journal of Artificial Intelligence and Expert Systems (IJAE)* 1: 111-122.
6. Jordan M (1995) Why the logistic function? A tutorial discussion on probabilities and neural networks. Massachusetts Institute of Technology, Massachusetts, USA.
7. Özkan C, Erbek FS (2003) The comparison of activation functions for multispectral Landsat TM image classification. *Photogrammetric Engineering & Remote Sensing* 69: 1225-1234.
8. Serwa A (2009) Automatic Extraction of Topographic Features from Digital Images. PhD thesis. Faculty of Engineering in Cairo-Azhar University, Cairo, Egypt.
9. Serwa A, Semary H (2016) Integration of Soft Computational Simulator and Strapdown Inertial Navigation System for Aerial Surveying Project Planning. *Spatial Information Research* 24: 279-290.