Classification of Dryland Salinity Risk using Artificial Neural Networks

[1]Spencer, M., [1]Whitfort, T., [1]McCullagh, J. and [2]Clark, R.

[1]La Trobe University, [2]Department of Primary Industries, E-Mail: j.mccullagh@latrobe.edu.au

Keywords: *Soil Salinity; Artificial Neural Networks; Artificial Intelligence.*

EXTENDED ABSTRACT

Soil salinity is a problem that affects millions of hectares of land across Australia. The costs associated with salinity have been increasing in agricultural production and infrastructure. To manage the effects of soil salinity, land owners, the government and catchment management groups need to know the extent and severity of soil salinity and how it changes over time. This paper investigates the use of artificial neural networks (ANN's) to map dryland salinity. ANNs have been inspired by biological neural networks, and are popular tools in the application of classification, prediction and recognition based problems. They use a network of interconnected processing units to estimate the outputs. The most common type of ANN is the backpropagation neural network (BPNN), which uses layers of interconnected processing units and a supervised learning approach.

The data used in this research comes from the Wimmera region of Victoria (see Figure 1), and includes airborne geophysical measurements and satellite imagery. The use of such data has been considered to be complete and accurate, and previous studies have indicated a relationship exists between these input variables and the presence of salinity. A large proportion of the study area was non-saline (98.5%). To help minimise network bias towards a particular class, non-saline and saline examples were randomly selected in equal numbers for experiments.

The experimentation in this research examined the potential of neural networks to map dryland salinity. The experiments were divided into two sections: individual salinity set experiments using the three types of salinity separately (salinity due to local groundwater, regional groundwater and waterlogging), and combined salinity set experiments where the three salinity types were combined into one class. The results achieved in the combined salinity set experiments showed an overall accuracy of 78.4% when using the BPNN approach. The saline examples were correctly classified 83.1% of the time and non-saline examples were correctly classified with 73.7% accuracy. The results for the individual salinity set experiments demonstrated that the performance of the neural networks on salinity due to regional groundwater and waterlogging was reasonably successful with a percentage correct of 83.0% and 76.0% respectively. The performance of the neural network on salinity due to local groundwater was less accurate with a percentage correct of 68.8%.

Overall the results achieved were promising and indicate the potential for further research in this area. Potential research areas include: (a) finding other cost-efficient inputs that influence salinity risk; (b) modifying relative data calculations to better identify significant readings and show the difference to nearby cells; (c) investigating the right balance of saline and non-saline examples to incorporate the large proportion of non-saline cases without causing imbalance to neural network training; and (d) investigating the application of expert neural networks.

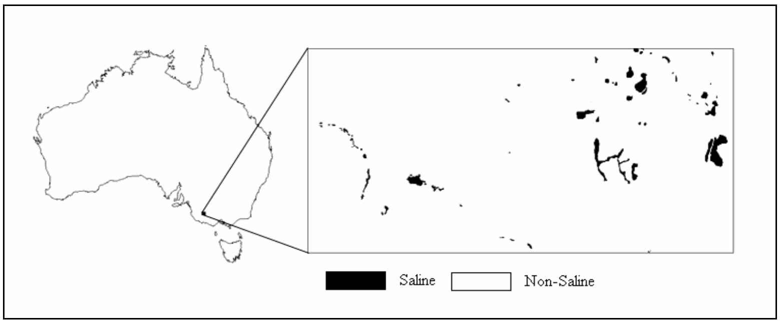

Figure 1. Map of Australia showing region under study

1. INTRODUCTION

Since European settlement, significant environmental changes have occurred in Australia. Trees were cleared to promote agricultural practices and many of the deeper–rooted native grasses and shrubs were not able to withstand grazing by the introduced European livestock. The vegetation community that evolved following the changed land management regime was dominated by relatively shallow-rooted annual vegetation and was unable to use as much rainfall as the pre–existing vegetation. This resulted in the water balance being altered as the unused water drained past the rootzone to the groundwater, causing watertables to rise. As watertables rise towards the earth's surface, the salts carried within the groundwater also rise, causing non-salt tolerant plants to become stressed and in severe cases die. The extent of the soil salinity problem within Australia includes the degradation of millions of hectares of agricultural land and loss of flora affected by rising watertables (AusStats 2002). In addition, dryland salinity increases maintenance costs to infrastructure such as road and rail. Due to the devastating effects associated with this growing problem, steps have been taken to identify regions at risk of developing dryland salinity.

Traditionally, soil salinity was mapped by airphoto interpretation or mapped in the field by either identifying salt tolerant vegetation, electromagnetic induction (EMI) surveys or soil sampling and analysis. Although accurate, these techniques are considered to be time consuming, expensive and difficult to implement at landscape scale. As a consequence, there is a growing interest in models that can produce accurate salinity maps or assessments of salinity risk areas at scales ranging from the individual paddock to landscape scale.

Remotely sensed data offers a cost-efficient, complete and accurate means of obtaining information about the environment at a variety of scales. Data is commonly collected by sensors located on aircraft or satellite based platforms by active or passive sensors. Passive sensors record naturally occurring radiation that is reflected or emitted from the terrain. Active sensors such as microwave radar transmit electromagnetic energy and record the amount of radiant flux returning to the sensor (Jensen 1996).

In this study, data from the Enhanced Thematic Mapper+ (ETM+) sensor carried on board the Landsat7 satellite is combined with airborne gamma radiometric data. The ETM+ sensor measures electromagnetic radiation (from 0.45µm to 12.5µm) reflected from the earth's surface, while the airborne gamma radiometric data provides a record of gamma radiation emitted from the top 0.5 m of the Earth's surface (around 85% of the signal comes from the top 0.3 m). This study focuses on the ability of a neural network to process remotely sensed data for mapping soil salinity.

2. BACKPROPAGATION NEURAL NETWORKS

The study of Artificial Neural Networks has been inspired by the biological neural system. The most commonly used ANN is the backpropagation neural network. A BPNN is a multi-layered non-linear feed-forward network trained by the backpropagation learning algorithm. It is composed of a series of artificial neurons (see Figure 2).

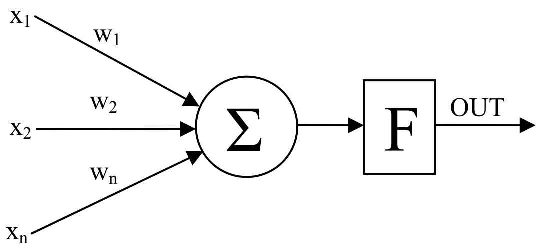

Figure 2. An artificial neuron.

An artificial neuron generates an internal activation which is termed NET. NET is calculated by summing all of the input weight products as outlined in the following equation.

$$NET = \sum_{j=1}^{n} w_j x_j + w_B$$

where:

$X_1 .. X_n$ are the inputs
$W_1 .. W_n$ are the connection weights
W_B is the bias weight.

A bias acts exactly as a weight on a connection from a neuron whose activation is always one. After NET is calculated, an activation function F, commonly sigmoid or hyperbolic tangent, is applied to modify it thereby producing the signal OUT.

The structure of the BPNN including the number of neurons used, their organisation into layers and the connections between them is referred to as its architecture. The most common architecture is the fully interconnected multi-layered network which is outlined in Figure 3. It contains an input layer, one or more hidden layers and an output layer.

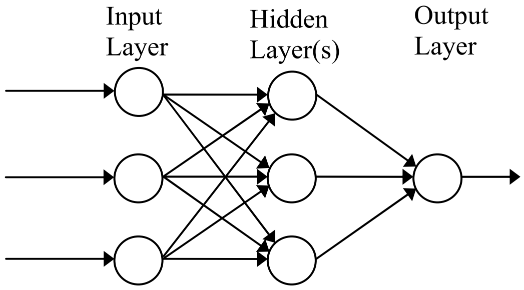

Figure 3. A simple fully interconnected multi-layered neural network.

The BPNN is trained by being presented with numerous examples where each example consists of a set of inputs and their corresponding output(s). The "learning" that takes place in backpropagation neural networks occurs by the adjustment of the connection weights. Each presentation of an example results in a small change in the weights, which reduces the error on that example. This process is repeated thousands of times until the weights have been adjusted to represent a mapping between inputs and outputs to within a specified error. The network then has the potential to predict the output from a new set of inputs (Rumelhart *et al.* 1989). Backpropagation neural networks are capable of handling incomplete, noisy or partially incorrect data with a minimal reduction in performance and have been successfully applied to many areas of the business, industrial and scientific world (Smith *et al.* 2002).

3. MAPPING SALINITY

Research has previously been conducted into the problem of mapping salinity using remotely sensed data. Statistical methods were used in a study at Kakadu National Park, Australia to map soil salinity using Airborne Synthetic Aperture Radar (AirSAR) and Landsat TM satellite data (Bell *et al.* 2001). Variables used in this study included Electrical Conductivity (EC), percentage ground cover, vegetation height, species and leaf litter depth. The salinity map produced by this study classified salinity presence into 9 different classes, based on density of vegetation and salinity. The best level of accuracy on test data was 82%. The accuracy of this study was quite high, however several of the classes that contributed a small portion of the population were not classified accurately using this model. Hocking (2001) used fuzzy modeling to map dryland salinity in the Wimmera Plains, Australia. Data collected for this study included airborne gamma radiometrics, DEMs (Digital Elevation Models) and DEM derived data. Factors used included the slope of the land, elevation, potassium and thorium content. Rather than assessing soil salinity as either present or absent, output from this study estimated the probability of soil salinity occurring in each grid cell. The results from the fuzzy techniques were poor, with a correlation of 0.32 between the outputs of the fuzzy system to the expected result. Evans (1998) used Conditional Probabilistic Networks to examine if probabilistic relationships could be identified for mapping dryland salinity. The focus of this study was to demonstrate that the accuracy of mapping salinity could be increased by incorporating prior knowledge of relationships between attributes and salinity. Previous years' attributes were combined with the assessed year to estimate salinity risk. The overall results of the study were poor, with many non-saline examples being classified as saline.

Due to the difficulty in mapping soil salinity using statistical techniques, there has been growing interest in exploring a variety of other methods including Artificial Intelligence techniques. Research was conducted to investigate the use of decision trees for mapping salinity in the Wimmera region (Walklate 2002). The decision tree was developed using C4.5, a software package used for the generation of rule sets (Quinlan, 1993). The data used was a combination of gamma radiometrics, satellite imagery and DEMs. The variables included potassium and thorium content, curvature, elevation and slope. For a localised area of the study space, the results were 91% on the testing set for the classification of salinity. However, when this study space was increased to a wider area, the accuracy was reduced to 63.7%. A study performed by Evans (1998) in Western Australia investigated the use of decision trees and neural networks for mapping dryland salinity. Data collected for this study included satellite imagery, DEMs, drainage models and ground truth data. The salinity maps produced using the decision tree approach appeared promising with a relatively low number of misclassifications. Neural networks were also used to determine whether the technique was a more effective method for the mapping of dryland salinity. The neural networks demonstrated similar results to the decision tree approach indicating the potential for these techniques in this problem domain.

Artificial Neural Networks have been studied in other salinity-based problems including the prediction of groundwater activity (Clarke *et al.* 2002), river salinity (Maier *et al.* 1998, Rajkumar *et al.* 2001) and dryland salinity (Evans *et al.* 2000).

4. DATA

The data used in this study was ETM+imagery from the Landsat7 satellite and airborne gamma radiometric data. Data was collected over a 34km x 17km area of the Western Wimmera region in Victoria, Australia (see Figure 1). The airborne gamma radiometric data was measured continuously along lines spaced either 200m or 250m apart. Standard processing of this data produced an interpolated grid surface with 50m x 50m cells. For the purposes of this study the data was resampled a second time to produce a grid surface with 20m x 20m cells. The variables used in the study are summarised in Table 1 below.

Table 1. Variables available for this study.

Variable	Description
aspect	Direction of the downward slope (degrees).
curvature	Curvature of the Earth.
elevation	Elevation above sea level (metres).
potassium	Potassium (% by volume).
slope	Indicates the gradient from one cell to its neighbouring cells. Calculated as a percentage.
soil	Soil class.
thorium	Thorium (ppm).
TWI	Topographic Wetness Index. The natural log of the total area from which water would flow through a given cell, divided by the slope of that cell.
target	Expert on ground assessment of soil salinity based on the presence of salt tolerant vegetation or absence of salt sensitive vegetation.

Training, testing and validation data sets were created to be used by the neural networks. The training, testing and validation data sets are three independent sets. The training set is used to train the neural network while the testing set is used to evaluate the performance of the neural network at regular intervals. When the neural network has been fully trained, the network is evaluated on the validation data set which has been totally unseen by the network during the entire training phase. A large proportion of the study area was non-saline (98.5%). To help minimise network bias towards a particular class, non-saline and saline examples were randomly selected in equal numbers for experiments.

5. EXPERIMENTATION

The experimentation conducted in this research examined the potential of neural networks to map dryland salinity. The experiments were divided into two sections: individual salinity set experiments and combined salinity set experiments.

5.1. Individual Salinity Set Experiments

Three types of salinity were recorded: salinity due to local groundwater, salinity due to regional groundwater and salinity due to waterlogging. Each type of salinity was in localised regions, with no overlap between regions. Preliminary experiments were conducted to classify each of these types of salinity separately. Three different data sets were created for each salinity type.

The data was selected from a region of the study that had a specific type of salinity, with saline and non-saline examples selected in equal proportions. For each of the salinity types the data was divided equally into training, testing and validation sets. There were a total of 10,114 examples selected for the local groundwater experiments, with approximately 3,370 in each of training, testing, validation sets. The regional groundwater sets had a total of 3,076 examples and waterlogged a total of 3,174 examples

Before being presented to a network for training, data was preprocessed to aid network performance. In addition to the original values for inputs, relative values were used for the following inputs: curvature, elevation, potassium, slope and thorium. Relative measurements were calculated based on readings of nearby cells. These were used to indicate changes of land characteristics such as elevation. The soil type was the only discrete input. A separate input was created for each of the 8 soil types. Based on the distribution of inputs, appropriate scaling techniques were applied to the data. Non-linear scaling was applied to the slope. The other continuous inputs were scaled linearly. In total, 21 inputs were supplied to the network (7 continuous, 6 continuous relative, and 8 discrete soil). The neural network architecture and parameters used in the individual salinity set experiments are outlined in Table 2.

Table 2. Neural network parameters and architecture

Parameters	Values
No. of hidden layers	1
Hidden layer neurons	2 – 18
NN passes	100,000 – 400,000
Learning rate	0.1 – 0.9
Momentum	0.9
Epoch size	10
Initial weights	± 0.1

Results and Discussion

Table 3 shows the performance of the neural networks on the individual salinity sets. The results indicate that good classification accuracies can be achieved using this technique. However the neural network's performance on the local groundwater data set is lower than the network's performance on the other two data sets. This may indicate that different factors are required for the mapping of salinity due to local groundwater. It is also possible that the reduced performance on the local groundwater salinity data set is due to a more complex relationship existing between the input factors and this salinity type or that not all the areas affected by soil salinity were identified during the on–ground survey. Variations on the backpropagation neural network architecture may assist in improving this performance.

Table 3. Training and testing performance on individual salinity set experiments

	Training Accuracy	Testing Accuracy
Local groundwater	67.8%	68.8%
Regional groundwater	89.0%	83.0%
Waterlogged	78.8%	76.0%

In order to analyse the results further, the performance on the saline and non-saline examples were examined. Figure 4 shows the percentage of saline examples which have been correctly classified for each of the three salinity types.

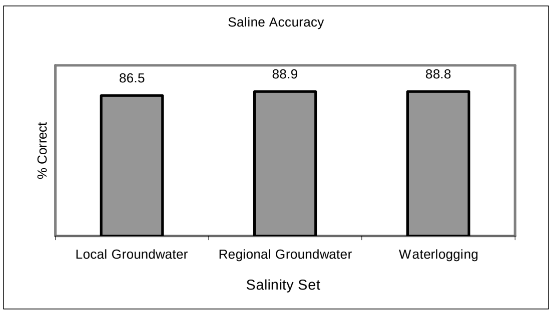

Figure 4. Percentage of correctly classified saline examples (testing set)

Figure 5 shows the percentage of non-saline examples which have been correctly classified.

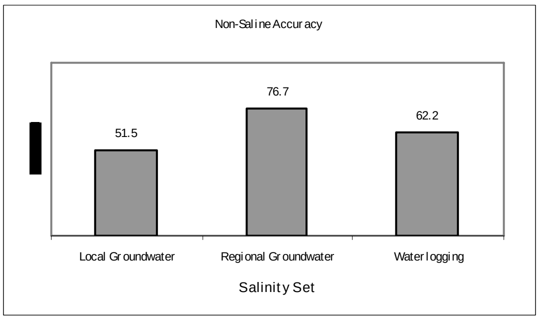

Figure 5. Percentage of correctly classified non-saline examples (testing set)

Figure 4 shows that a high percentage correct is achieved on the examples for the three saline types with an average of 88.0%. The results for the correct classification of the non-saline examples were not as high with an average percentage correct of 63.4%. The values for the local groundwater data and the waterlogged data sets are 51.5% and 62.2% respectively. It is possible that different neural network models including expert neural networks (McCullagh 2003) may assist in improving the performance on the non-saline examples. This has been identified as an area for future research.

5.2. Combined Salinity Set Experiments

A second series of experiments were conducted to evaluate the performance of a neural network when the three sources of salinity (local, regional, and waterlogged) were combined into a single class. Examples were classified as saline or non-saline. Each of the three salinity classes were equally represented in the data sets. The training, testing, and validation sets each comprised 6,612 examples: with 3,306 non-saline and 3,306 saline examples randomly selected (1,102 each due to local groundwater, regional groundwater and waterlogged salinity).

The data was preprocessed and presented to the network as specified in the previous experiment (Section 5.1).

Results and Discussion

Table 4 shows the performance of the neural networks on the combined salinity sets. The best network used 16 hidden layer neurons, a learning rate of 0.4, momentum of 0.9 and an epoch size of 5.

Table 4. Training, testing and validation performance on combined salinity set experiments

Training Accuracy	Testing Accuracy	Validation Accuracy
81.2%	79.1%	78.4%

The performance of the networks on the saline and non-saline examples were investigated to further analyse the results. The results are presented in Figure 6.

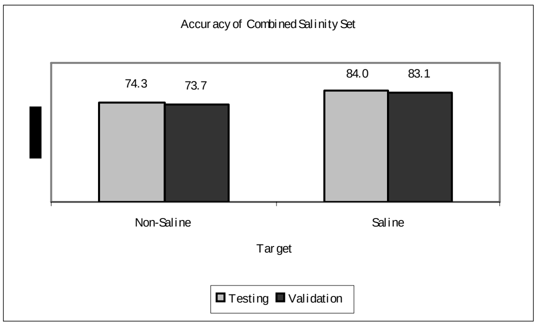

Figure. 6. Testing and validation performance for saline and non-saline classes.

Saline examples have been classified with an accuracy of 84.0% correct. Non-saline examples have not been classified as accurately, with 74.3% correctly classified. The distribution of the non-saline examples that are used in the training set may be a possible cause. Only 3,306 examples have been selected to represent the 1,395,784 examples from the dataset available for research. It is possible that the non-saline examples used for training the network may not accurately represent all the possible non-saline examples. Further research is required to investigate the data selection process.

The results achieved by the neural networks in this research were compared to decision trees and the k-nearest neighbour classifier. These results are presented in Table 5.

Table 5. Validation performance compared to other studies

Technique	% Correct
Artificial Neural Networks	78.4%
Decision Trees (Walklate 2002)	74.0%
k- Nearest Neighbour (kNN)	67.1%

The neural networks demonstrated a 16.8% improvement in performance over the kNN classifier. In this comparison the same data sets from the Wimmera region were used. The neural networks were also compared to a study conducted in the Western Wimmera (Walklate 2002). This study used decision trees to determine if rule induction was a suitable tool for estimating the presence of salinity. The results from the neural networks demonstrated a 5.9% improvement over the decision tree technique. It should be noted that the results presented in this research are preliminary in nature, however they do indicate the potential for further work in this area.

6. CONCLUSION

Remote sensing techniques have the potential to provide a cost effective technique to collect data that can be used to map soil salinity. Neural networks were used in this study to process remote sensing data and produce salinity maps, with promising results. It is believed that, with further research, the accuracy of this technique could be improved. Potential areas for further research include: (a) finding other cost-efficient inputs that influence salinity risk; (b) modifying relative data calculations to better identify significant readings and show the difference to nearby cells; (c) investigating the right balance of saline and non-saline examples to incorporate the large proportion of non-saline cases without causing imbalance to neural network training; (d) investigating the application of expert neural networks; and (e) use sensitivity analysis to determine significant inputs.

7. REFERENCES

Australian Bureau of Statistics (2002), *AusStats: Land Degradation* [online] Available: http://www.abs.gov.au/ausstats/abs@.nsf/0/FAEB6096CDA4D9ADCA256BDC001223FF?Open&Highlight=0,Salinity [Accessed 27 March 2003].

Bell, D., C. H. Menges, R. E. Bartolo, W. Ahmad, and J. J. VanZyl (2001), A multistaged approach to mapping soil salinity in a tropical coastal environment using airborne SAR and Landsat TM data, *Geoscience and Remote Sensing Symposium,* 3, 1309-1311.

Clarke, C. J., R. J. George, R. W. Bell, and T. J. Hatton (2002), Dryland salinity in south-western Australia: its origins, remedies, and future research directions, *Australian Journal of Soil Research*, 40, 93-113.

Evans, F. H. (1998), An investigation into the use of maximum likelihood classifiers, decision trees, neural networks and conditional probabilistic networks for mapping and predicting salinity, *MSc thesis,* Department of Computer Science, Curtin University.

Evans, F. H., and P. A. Caccetta (2000), Broad-scale spatial prediction of areas at risk from dryland salinity, *Cartography*, 29(2), 33-40.

Hocking, M. (2001) Salinity Risk on the Wimmera Plains, *Technical Report No. 72, Centre for land protection research*, Bendigo, Dept. of Natural Resources and Environment.

Jensen, J. R. (1996), *Introductory digital image processing, A remote sensing perspective,* Second edition, Prentice–Hall, New Jersey.

Maier, H. R., and G. C. Dandy (1998), Understanding the behaviour and optimising the performance of back-propagation neural networks: an empirical study, *Environmental Modelling & Software*, 13(2), 179-91.

McCullagh, J. A. (2003), An investigation into the use of neural network ensembles for meteorological rainfall estimations, *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI2003)*, 8, Florida, USA, 190-194.

Quinlan, J. R. (1993), *C4.5: Programs for Machine Learning,* Morgan Kaufmann, San Mateo, CA.

Rajkumar, T., and M. L. Johnson (2001), Prediction of salinity in San Francisco bay delta using neural network, *Proceedings of IEEE SMC,* Arizona, USA, 329-334.

Rumelhart, D. E., and J. L. McLelland (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, I & II,* Cambridge: MIT Press.

Smith, K. A., and J. N. D. Gupta (2002), *Neural networks in business: techniques and applications,* Hershey, PA: Idea Group.

Walklate, J. (2002), *Machine learning using AI techniques,* BComp(Hons) Thesis, La Trobe University, Bendigo.