PalArch's Journal of Archaeology of Egypt / Egyptology

A DATA MINING APPROACH TO CROP YIELD PREDICTION USING MACHINE LEARNING

Dr. Kamakshaiah Kolli[1], B. Neeraja[2], V. Shiva Narayana Reddy[3]

[1,3]Associate Professor, Department of CSE, Geethanjali college of Engineering and Technology, Hyderabad, Telangana, India
[2]Department of CSE, Geethanjali college of Engineering and Technology, Hyderabad, Telangana, India

Dr. Kamakshaiah Kolli, B. Neeraja, V. Shiva Narayana Reddy,A Data Mining Approach To Crop Yield Prediction Using Machine Learning, Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(12). ISSN 1567-214x.

Abstract:
One of the essential industrial sectors in India is Agriculture and the economy of a country is relied on it highly for sustainability in rural areas. The Indian agriculture level reduces gradually owing to some factors like excessive utilization of pesticides, water level decrement, climate changes, and unpredicted rainfall, etc. On the agriculture data, descriptive analysis is performed to understand the production level. Because of lack of ecosystem control technologies deployment, most of the agriculture fields are under developed. The production of crops is not increased owing to these issues that affects the economy of agriculture. Based on the prediction of plant yield, agricultural productivity is improved. By using machine learning techniques, the crop from given dataset have to predict by agricultural sectors for preventing this issue. To capture the information, the dataset analyses by supervised machine learning technique (SMLT). The effectiveness of proposed method of machine learning algorithm can compare with best accuracy based on the results. In this paper use ANN with cascade-forward backpropagation and Elman backpropagation for yield prediction. To determine input variables that maximize the interested neurons' activation, the positive gradients backpropagate by Cascade-Forward backpropagation method. A recurrent connection exists from the hidden layer output to its input is included in Elman backpropagation network which is a two layer backpropagation network. These two techniques are better prediction techniques compared to ANN with backpropagation.

Introduction:
In the agricultural sector, it is important that strategy management begins to prosper based on information, bearing in mind the great uncertainty of the factors that influence the crop production and the enormous amount of data that is and can be captured during the whole process of production.
Currently, India has the opportunity to become one of the main paddy exporters. However, there is a problem regarding the index of yield (expressed in tons per hectare) of its

production compared to other exporting countries because it does not know the factors that influence the production of the crop.

The agricultural sector cannot be the exception in starting to adopt the best practices and tools that help the farmer in making decisions and encourage the agricultural investment and the growth of the economy in the sector. Having said the above and having taking into account that different public and private entities have been collecting different types of structured, semi-structured and unstructured data at different scales, such as, For example, meteorological stations, where they produce large volumes of hourly data, daily, monthly, among others, of various variables [1]. Reason why, they allow the possibility of start a path of data analytics in a sector that has the potential to grow and be each more profitable if tools, methodologies and best practices are incorporated in a preventive measures to mitigate financial risks, increasing the profitability of the agriculture crops.

In the agricultural sector, production agriculture depends on many biological factors, climatic, economic and human that interact in complex ways. Agricultural producers and companies in the agricultural industries must make countless decisions every day that impact on performance and supply chain operation respectively [2]. Therefore, decision making requires to be better grounded in various sources of information that are made progressively more difficult to manage outside the data paradigm.

Remote sensing has great potential as a source of information for the predictionof agricultural production, both at the regional and the global scale, because itprovides data at a level of consistency, repeatability, timeliness and scalability that isunmatched by any other data source. The costs of collecting the information and altering them into a reliable alternative for costly ground-based surveys are reduced markedly over the last decades due to the continuous improvements in remote sensing techniques [3]. In the agricultural production estimation field, satellite and areal images have become essential data sources rapidly as a consequence.

The importance of yield estimation gain and crop monitoring includes on the scientific and the political agenda as the rapid growth of global population and a (negative) impact of climate change on global crop production becomes probable.

One of the principal determinants of crop yield is the percentage of solar irradiation intercepted by the plants' foliage. One way to assess the productivity of crops depends on the 'fraction of Absorbed Photosynthetically Active Radiation' (fAPAR) and the efficiency with which that energy is converted into new biomass [4]. The value of fAPAR is largely determined by the crop's foliage, which in turn can be related to the value of vegetation indices (VIs). VIs are numerical transformations of measured reflectance that are related to plant and canopy characteristics in a crop-specific and nonlinear way. The Normalized Difference Vegetation Index (NDVI) is improved by Deerin and it is used extensively.

These requirements have been fulfilled by Machine Learning (ML) methods and suitable for large scale applications specifically. Data mining software is used as an analytical tool which allows the users for assessing the information from various angles or dimensions, summarize, and categorize the identified relationships. Data mining is technically described as the process of determining the patterns or correlations among dozens of fields in large relational databases. The information provides by the associations, patterns, or relationships among all this data. The knowledge about future trends and historical patterns can be converted from information. The crop losses identification and prevention in future can make easier for farmers based on the summary data about crop production. An important problem of agricultural field is a crop yield prediction. So that, each farmer can always get to know how much yield will receive from the expectation. Based on the assessment of previous experience of farmers on a particular kind of crop, prediction of yield was determined. By relying on the agricultural pests, weather conditions, and harvest operation planning, the yield of agricultural products is obtained. To make decisions regarding the risk management of agricultural field, accurate information about the crop yield is playing an important role. The crop yield prediction is focused in this paper based on the proposed idea. The yield of the crop per acre will check by farmer before cultivating onto the field.

Machine learning:

ML is the design and study of program artifacts computational systems that use past experience to make future decisions. Without being programmed explicitly, the ability to learn by using computers which is studied in ML. The process of ML is performed as follows: if your performance of a task in T, a program learns from an experience E regarding a class of tasks T and performance measure is P, improves experience E. Within ML we find 2 approaches, supervised learning and non-supervised learning.

According to [5], learning Supervised consists of predicting the values of a set of output data, from a set of input data. It is called supervised because according to the model predicts the outputs for test data, the error between what the algorithm predicted is calculated and the real value. The objective is to minimize the error, adjusting the density function of probability relating inputs to outputs.

On the other hand, in unsupervised learning, you only have data sets of input, without knowing its relationship with output variables, so you have no way to verify easily if the model performance is adequate.

Literature Survey:

Priya et al., 2018 [6] was predicted the crop yield based on machine learning algorithms from existed data using Random Forest algorithm. To establish the models, real data of Tamil Nadu were utilized and the testing of models was done with samples. For getting accurate results in prediction of crop yield, Random Forest Algorithm can utilize.

Balakrishnan et al., 2016 [7] has been proposed ensemble model based on AdaSVM and AdaNaive for predicting the production of crops over a certain period of time. By using AdaNaive and AdaSVM, its implementation is accomplished. AdaBoost increases efficiency of SVM and Naive Bayes algorithm. Based on the climate conditions, the crop yield is predicted using machine learning approach. In order to forecast the effect of climatic parameters on the crop yields, a software tool called as Crop Advisor that has been improved as a user friendly web page in the current research. For producing the most influencing climatic conditions on the selected crops' yield, C4.5 algorithm utilizes in selected district areas of Madhya Pradesh. It is implemented using Decision Tree.

Siju, H. L., & Patel, P. J. et al. [8] were reviewed on the prediction of crop yield based on Data Mining focusing on Groundnut crop using the technique of Naïve Bayes. In horticulture, various applications of information mining has been illustrated and examined. The detection of crop yield prediction has been explored and actualization of Naïve Bayes technique has been done for different applications. The research works were demonstrated the expectation of Groundnut crop yield. By using different procedures of data mining, the model of exact groundnut crop yield expectation can be improved.

Bhanumathi, S., Vineeth M., &Rohit N. et al. [9] were focused on breakdown various relevant properties like location and esteem from which the dirt alkalinity resolves in addition to the supplements level like Potassium (K), Phosphorous (P), and Nitrogen (N). By using the outsider applications like APIs, the Location has been taken for soil type, climate and temperature, estimation of supplements in the dirt, and precipitation measure in the area, and creation of soil can resolve. The information properties will break down and train it based on various appropriate AI calculations such as Random Forest Algorithm (RFA) and Backpropagation Algorithm to design a model. In prediction of crop yield, the framework with a model to be provided the precise and exact results. Based on the soil and barometrical metrics of the land, the end client with appropriate proposals is conveyed about the required compost apportion that helps to improve the establishment of harvest yield and increment in rancher income. By considering this philosophy, the future work is extendible to build the web applications and allow the client to use this effectively for comprehending the harvest yield.

T. Giri Baby, Dr. G. AnjanBabu et al., [10] were made a conclusion that the proposed method will provide solutions for the problems of fertilizer and water. The yield production will be more with the proposed technique. Agro algorithm is used in this paper. The accuracy in crop production didn't provide by this method properly.

B. Vishnu Vardhan, D. Ramesh et al., [11] were proposed multiple linear regression technique that can implement on existing information and assists in assessment and verification of data. It is also providing less accurate results which is the drawback of proposed method.

E. Manjula, S. Djodiltachoumy et al., [12] have been focused on proposing and implementing the rule based system which should forecast the production of crop yield using the previous collected data. The utilized algorithms include clustering method and K-means algorithm. The drawback of the system is only suitable for association rule and less data is considered.

Chlingaryan A., Sukkarieh S., and Whelan B. et al. [13] have accomplished an audit which considered the AI techniques predominantly, estimation of yield, and nitrogen accuracy on the board. The method of back proliferation significance and harvest yield expectation precision for various lists of vegetation have been explored by the survey. To determine and foresee various qualities of plant leaves, the Gaussian procedure is valuable by them. For evaluating the expectations of numerous yield productions, the most appropriate device is considered as the M5-Prime Regression Tress based on the auditing of significance. Additionally, this survey explores the Fuzzy Cognitive Map (FCM) that will utilize for expectation of crop yield for model and portrayal of master information.

Priya et al., 2018 [14] was predicted the crop yield based on the algorithm of machine learning. By using Random Forest algorithm, the crop yield is forecasted from the existing data. Here, the real-time data of Tamil Nadu were utilized to establish the models and test them with samples. To get the crop yield prediction accurately, Random Forest Algorithm can be utilized.

Hunt et al., 2018 [15] has been investigated on the Precision Agriculture for determining yield for crop insurance based on an Aerial Platform. Precision Agriculture (PA) is utilized for identification of field variations and dealing with them based on various strategies as it is the application of remote sensors and geospatial methodologies. Owing to irrigation practices, crop stress, and incidence of pest and disease, etc., the crop growth variability might be caused in an agricultural field. By using Ensemble Learning (EL), PA is implemented.

K Kamakshaih et al., [16] has proposed objective of agriculture not only lies in enhancing the cultivation but also to satisfy the end users with high quality goods. Rice pests and diseases are stipulating more importance day by day with the global changes. It leads concern on both global food safeties along with security of a major food crop worldwide. The small changes in availability may have heavy impact on prices as the rice global market is very low. Technical innovations need to be explored to meet the quality standards and ever growing rice demands. To achieve this, digital agricultural domain is focusing on enabling distinct applications by combining emerging technologies with the traditional techniques. The Proposed system mainly focuses on predicting the diseases in rice crop for pesticide management by utilizing machine learning algorithms such as J48, Naïve-Bayes and SVM. This study also presents the comparative results of these algorithms for disease prediction in terms of classification, prediction accuracy, time complexity and space complexity.

Existing Methods:

Operational forecasting of crop production over large areas from remotely sensed data can only be efficient if highly automated methods are used. Automation is needed in order to significantly reduce the amount of human interpretation that makes these forecasts expensive, time-consuming and difficult to repeat in time and space. As fully automated methodologies have not yet been developed, the best methods available today are semi-automated ones. Machine learning is a well-established subclass of semiautomated models that have shown their merits in a wide range of application domains.

Machine learning for crop yield prediction

The establishment of empirical models for estimating crop yield has since long been recognized as an important and challenging task for the remote sensing community. Since the late 1960s, researchers in the US, Europe and Asia made great research efforts on this subject but, due to the complexity and nonlinearity of the systems, traditional (linear) statistical modelling proved problematic. In the 1990s, the research community started to realize that nonlinear models could offer a more realistic and potentially more accurate solution to the challenging task of empirical yield modelling. Artificial neural networks (ANNs) and decision trees (DTs) made their appearance in the research area of empirical crop yield prediction.

Artificial Neural Network:

An Artificial Neural Network is one of the popular models in neural networks (NN for Neural Network) in ML and it is a computing model which includes a number of basic computational units known as neurons that are connected each other. A communication network is formed by this model that allows the complex computations (Shaley-Shwartz& Ben-David, 2014). The below image figure 12 is showed the neural network basic form.

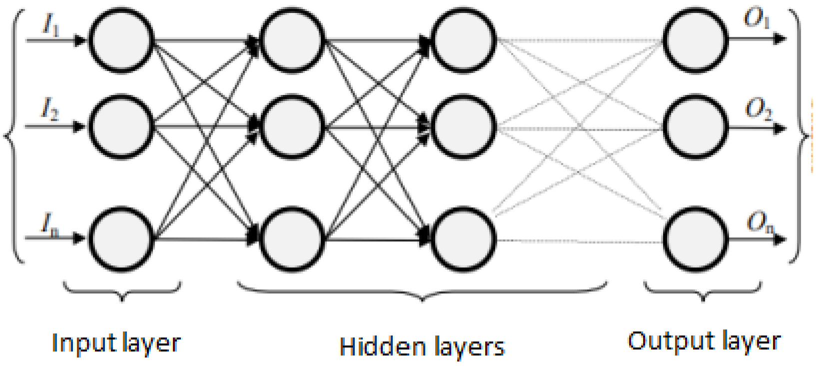

Figure 12. Diagram of a neural network. Source: (Matich, 2001)

Each node is considered as a neuron and the connections between them are represented by the arcs. The grouping of neurons is done in layers. From input model, the information receives by an input layer that can pass through the next group known as hidden layers. For problem complexity, as many layers can determine to train it. The values to the resulting functions from the network processing will obtain in an output layer finally.

In the weights of connection, the acquired knowledge by networks is contained and each neuron has included within the network by considering an assumption of its values (weights connection) to the training phase (Gallo, 2015).

The differentiation of two kinds of architecture (Gallo, 2015) is allowed by the way where neurons are interconnected and those architectures are described below:

- The feedback architecture where the connections existed between neurons from the same layer or from previous one;
- Advance architecture (after Hornik, Stinchcombe and White, 1989 cited by Gallo 2015) in which the signals go to neurons only in the next layer without any feedback connections.

In various ways (the infinite possibilities), the configuration of neural networks can be made. The primary objective function of the application is to choose the optimal settings for dealing with the problem or data.

ANN with Backpropogation:

Machine Learning is a booming technology that operate recursively for various applications (i.e. Artificial Neural Networks) and it was introduced for non-linear functions. To obtain a reliable model, the most important aspect is to train a Neural Network properly in this context. The term “Back-propagation” is relevant to this training and is highly vague for most people to getting into Deep Learning.

Back-propagation is the core part of the neural net training. Based on the error rate i.e. loss, the weights of a neural net is fine-tuned and are retrieved in the previous epoch i.e. iteration. The model becomes reliable by ensuring the proper tuning of weights that helps to obtain lower error rates and increasing the generalization.

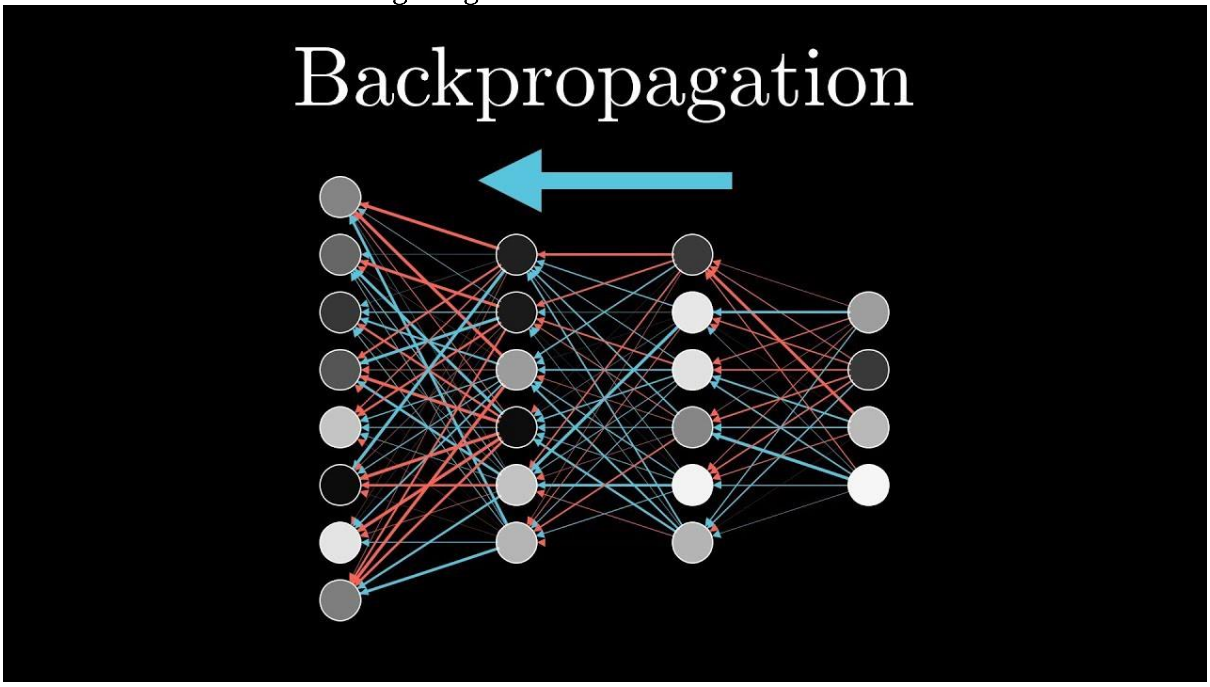

Figure : ANN with Back propagation

A model is established under training that performs the functionality of **XOR** (exclusive OR) using three hidden units and two inputs. The **training set** or truth table is as follows:

X1	X2	Y
0	0	0
0	1	1
1	0	1
1	1	0

In the neural net, the activation value determines at each node by an **Activation Function**. To simplify the calculation, an identity activation function is chosen:

f(a) = a

A **hypothesis function** is determined what the input to the activation function is. However, the function is typical and popular one:

h(X) = W0.X0 + W1.X1 + W2.X2 or

h(X) = sigma(W.X) for all (W, X)

The **loss function** is relevant to the logistic regression cost function and it looks complex a little bit but is simple actually:

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\log((h_\Theta(x^{(i)}))_k) + (1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{j,i}^{(l)})^2$$

To get a lower loss than the one we currently have, the adjustment of weights in what direction is determined by using the Batch Gradient Descent **optimization function**. All weights will initialize to 1 and the **learning rate** will be 0.1.

Performance measures:

To validate the performance of Machine Learning models, different measures are used depending on the type of problem being analyzed. Table 5 shows the confusion table, which makes it possible to understand the different measures based on the cross between the real condition and the predicted condition. Among the most used are Accuracy, Precision, Recall, F1-score, for classification models and R2, Mean Squared Error and Mean Absolute Percentage Error, for regressive models (Powers, 2007).

Table 5. Confusion table (or error matrix) with the different metrics available to validate the Machine Learning algorithms. Source: (Wikipedia, 2019)

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

Below is a brief description of the performance measures listed above:

Accuracy: is the percentage of tuples (ordered lists of elements) of the test set that were correctly classified by the model.

Precision: It is related to the classifier capacity and not label as positive. A sample is negative and is given by tp / (tp + fp). Here, fb is referred to the number of false positives and tp is indicated the number of true positives with 0 is being the worst value and is 1 the best value.

Recall: It indicates the relation tp / (tp + fn) where fn is the number of false negatives and tp is the number of true positives. To determine all positive samples, it can indicate the classifier capacity.

F1-score: It can use as a weighted average of recall and prevision. The best value at 1 and the worst score to 0 are reached by an F1 score. The average of F1 score of each class with a weight is resulted in case of multiple classes and labels based on the average parameter.

R2: It is the coefficient of determination or regression score function where the best score is 1.0 and it can also be negative in case the model is worse arbitrarily. The expected value is always predicted by a constant model and would obtain the R2 score of 0.0 no matter the input characteristics.

Mean Squared Error: The MSE refers to the measurement of mean squared error of predictions. The square difference between the target and the predictions is calculated for each point and average those values.

Mean Absolute Percentage: The MAE is indicated the linear score where the error is computed based on the average of absolute differences between the predictions and the target values. Here, the averaging of all individual differences are equally weighted.

Proposed Method:

ANN Architecture

Over 40 years ago, the ANN was pioneered and a great interest towards ANN is increased recently in neural network as it shares some of the behavioural and physical features of a biological system [3]. To determine the solutions for complex problems, the ANN structure is utilized and it is a parallel system based on the biological neural process of human brain. Here, the problems are imitated as mathematical models [6]. The ANN architecture had demonstrated briefly by a few authors [1, 3, 7]. To improve the system of ANN (Fig.1), a minimum of three layers is required such as the output layer, the hidden layer, and the input layer. The number of hidden nodes can extend to more hidden layers as they rely on particular problem of the study. The input layer comprises of nodes that corresponding to the input variables whereas the output layer nodes are referred to the output variables [7].

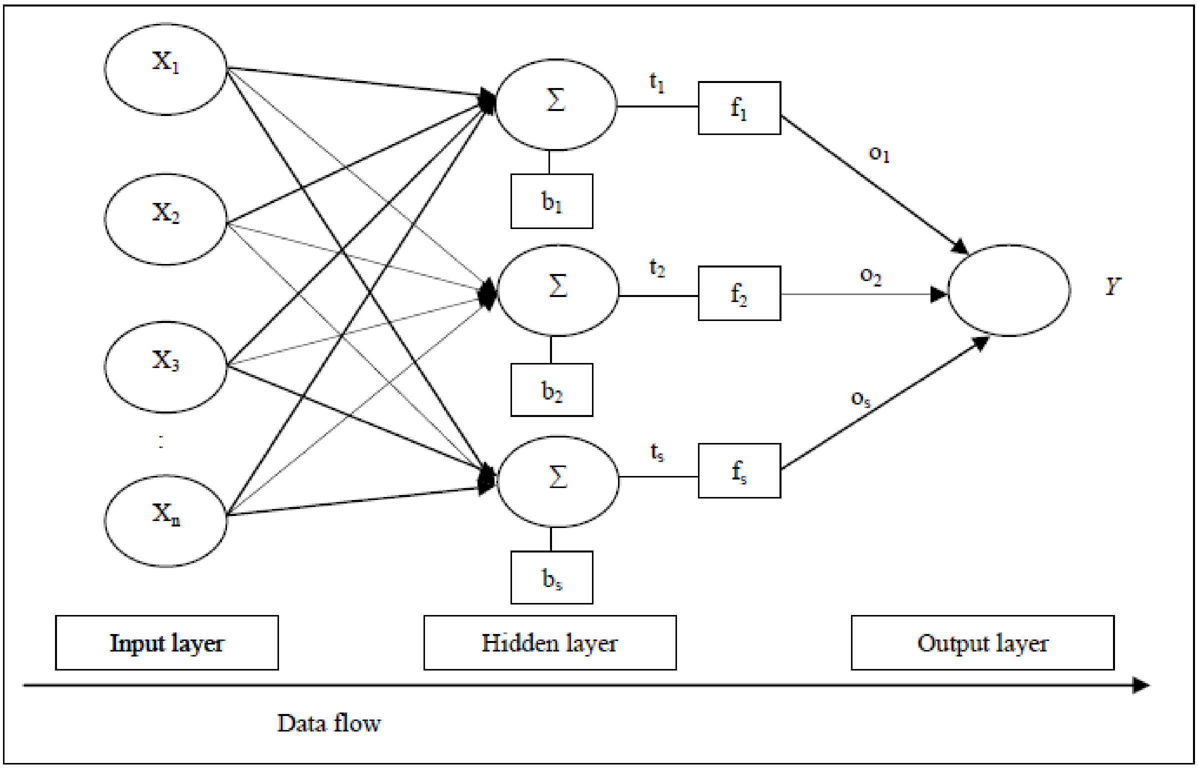

Fig. 1 Layers and connection of a feed-forward back propagation ANN.

For distributing the inputs to a number of hidden layers, the input layer uses and the output of which is connected to an output layer in which the units of outputs are connected to the inputs of the next layer through the connection weight [3]. To move between layers, the weighted connections allow data in a simpler way. From previous layer, the data is accepted by the node and a weighted sum of all net inputs have been determined:

$$t_i = \sum_{j=1}^{n} (w_{ij} x_j + b_i)$$

Where, x is the input from node j, w is referred to the connection weights between node I and j, n is the number of inputs, and bi is a bias. A transfer function fi applies to the weighted value to compute the node output oi:

$$o_i = f_i(t_i)$$

For the hidden and output layers, the widely used activation or transfer function is sigmoidal function [3, 7]. For transmission of information to hidden layers [7], a linear transfer function is utilized by an input layer most commonly.

Based on training [3], the dataset can be “learnt”. The definition of learning process can be described as a process which contains the adjusting of weights associated to the transfer functions between neurons by comparing output of ANN with observed information according to the view of Alvarez [8]. The back-propagation is the most common training technique which utilizes for training the feed-forward neural network to reduce the error [1]. The difference between determined value of output and the target value is represented the error in the training [9]. The less accurate predictions [10] and consumption of more memory [1] will be resulted by memorizing the training data which causes by a large networks that involve higher number of nodes. The repetition of process is considered until the total number of training cycles (epochs) has been accomplished or a particular error limit is achieved.

ANN with cascade-forward backpropagation:

The number of neurons in the output layers and the disposal of an entry’s number R are required in the modelling process of a neural network. Based on resolving the specifications of the problem, these neural networks have been defined. To predict the new output data, the CFBP or Cascade Forward Back Propagation model utilizes as it is one of the artificial neural network types.

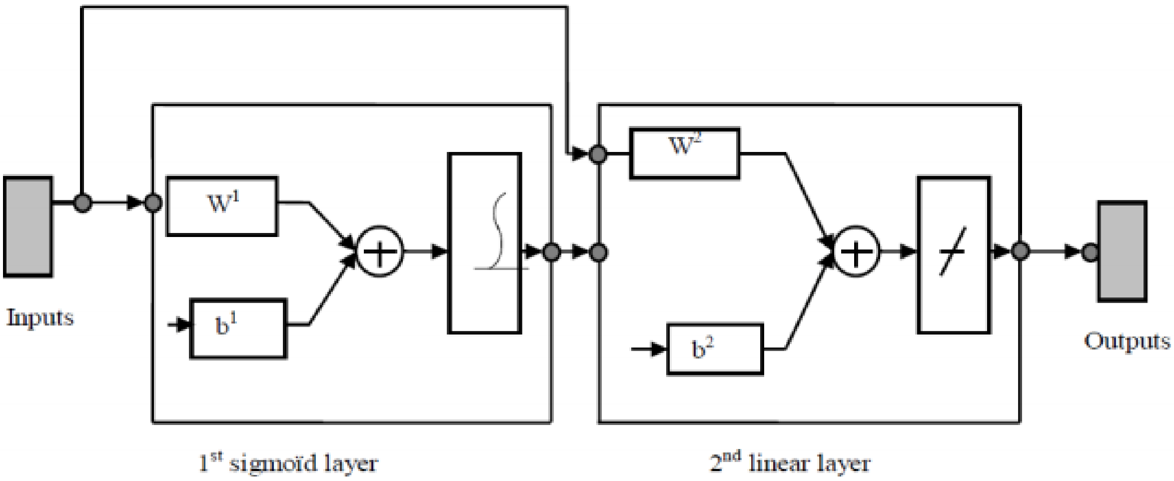

Figure 5. Model of Cascade Forward Back Propagation (CFBP).

In the learning process, the used methodology considers different analyses which have been implemented.

1. The weights initialization is done based on small random values;
2. In the learning sample for each combination (p_q, d_q):
    - The entries $p_q$propagate forward through the neural network layers:

      $$a^0 = p_q; a^k = f^k\left(W^k a^{k-1} - b^k\right), \quad k = 1, \cdots, M$$

    - Through the neural network layers, back propagate the sensitivities:

      $$\delta^M = -2F'^M\left(n^M\right)\left(d_q - a^M\right); \delta^k = F'^k\left(n^k\right)\left(W^{k+1}\right)^T \delta^{k+1}, k = M-1, \cdots, 1$$

    - Modify the weights and biases:

   $$\Delta W^k = -\eta\, \delta^k\left(a^{k-1}\right)^T, k = 1, \cdots, M,$$

   $$\Delta b^k = \eta\, \delta^k, k = 1, \cdots, M,$$

3. The process stops when the stopping criteria is reached; if not, the presentation is built in order of the combination from the learning database and start again from step 2.

Figure 6 illustrates the learning process of neural network with different steps.

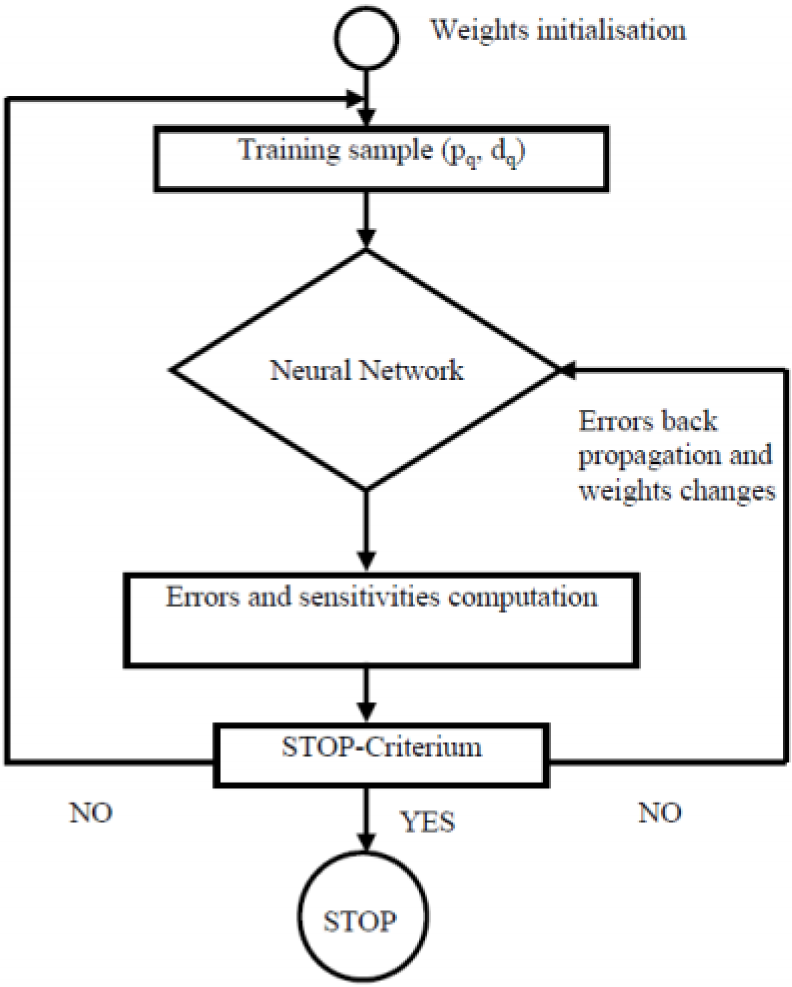

Figure 6. Diagram of the learning process of the neural network.

ANN with Elman backpropagation:

Elman network type is a three-layered feedback architecture of neural network (Fig. 1b). It has feedback connection from the output of hidden layer to its input which includes the context representation layer. To identify and generate the spatial or temporal patterns, the learning of Elman networks allow this feedback loop. A recirculation of information inside the Elman ANN's hidden layer is ensured by this internal looping. The values from previous time step are stored in this connection by the delay that can be used in the current time step [5]. The context layer and input layer are activated the hidden layer simultaneously in this kind of networks. The output of Elman ANN is communicated with the hidden layer processing at its output and the result is stored in the context layer. Before taking any final decision by a human brain, the hesitations will result by this delayed data based on the analogy of human brain thinking.

The hidden layer's activation functions are relevant to the sigmoid type and is given as follows:

$$f(x) = \frac{1 - e^{-2\alpha x}}{1 + e^{+2\alpha x}}$$

Where a refers to a parameter that controls the steepness of function near x= 0. The outputs with a reasonable discriminating power can produce by using this function which is essential for the errors of back-propagation. Based on the desired optimum performance, the number of neurons could be chosen according to the trial and error in the hidden layer.

From the hidden layer output to its input, a recurrent connection exists in a two layer backpropagation network which displays in figure 2.

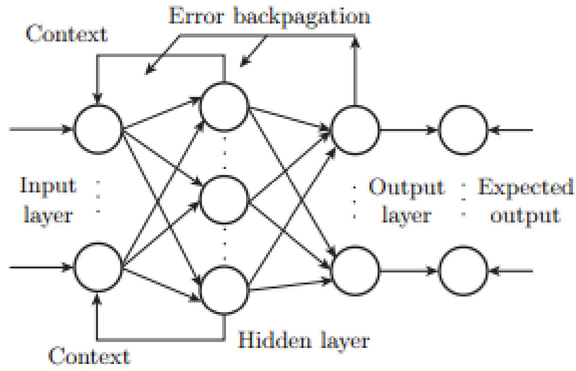

Figure 2: Elman backpropagation network

By adaptive learning rate, the training of network is done with Gradient Descent backpropagation. The output vector size T and input pattern vector P are similar in this experiment. Table 1 shows the used parameters for network architecture and training.

Table 1: Parameters used for creating the Elman backpropagation neural network

Parameter	Value
Number of hidden layers	1
Number of neurons in hidden layer	10
Number of neurons in output layer	10
Transfer function for layer	Hyperbolic Tangent Sigmoid
Training function	Gradient Descent backpropagation
Maximum number of epochs	10000
Performance function	Mean squared error
Error goal	0.00001
Adaption rate	1.0
Back-propagation learning rate	0.1
Initial weights and biased term values	Values generated randomly between 0 and 1

Results and Discussion:

ANN with Backpropagation:

The Multilayer Perceptron Feed Forward Fully Connected Neural Network is implemented with a Sigmoid activation function. Based on different options for Learning Rate Decrease, Momentum Backpropagation, and Resilient Gradient Descent, the training is completed using the algorithm of Backpropagation. If the Mean Square Error (MSE) is reached to zero or a predefined number of epochs is reached to a maximum value, the training process is halted.

The parameters of code configuration are mentioned below:

1. The variable nbrOfNeuronsInEachHiddenLayer is represented the number of neurons per hidden layer. The variable is set to [4 10 5] to design a neural network with 3 hidden layers that consist number of neurons 4, 10, and 5 correspondingly. N
2. Number of output layer is nits. The number of classes is equivalent to the number of output units usually but it can be less (≤ log2(nbrOfClasses)). The variable nbrOfOutUnits represents the output layer. From the dimension of training samples, the number of input layer units is retrieved.
3. If the sigmoid activation function is polar or unipolar, the selection is completed and it is represented by the variable unipolarBipolarSelector.
4. The learning rate η.
5. The variable nbrOfEpochs_max represents the maximum number of epochs at which the training is stopped until MSE reaches to zero.

6. The enable_learningRate_momentum variable represents an option for enabling or disabling the Momentum Backpropagation.
7. The variable of momentum_alpha represents the rate of Momentum Backpropagation α.
8. The enable_resilient_gradient_descent variable represents the option to disable or enable the Resilient Gradient Descent.
9. The parameters of Resilient Gradient Descent $\eta+$, $\eta-$, Δmin, Δmax, are represented using the variables of learningRate_plus, learningRate_negative, deltas_min, and deltas_max respectively.
10. The variable of enable_decrease_learningRate is used to disable or enable the Learning Rate Decrease.
11. The varaibles of min_learningRate and learningRate_decreaseValue are represented the parameters of Learning Rate Decrease.

To draw the decision boundary that separates the classes and the MSE curve, a parameter is also contained in the code. The figure is drawn and saved on the machine is specified after the number of epochs. The storing of figures is done in a folder known as Results along with the m files. The variable of draw_each_nbrOfEpochs is represented this parameter. The Sharky input points file known as string is considered by the variable dataFileName.

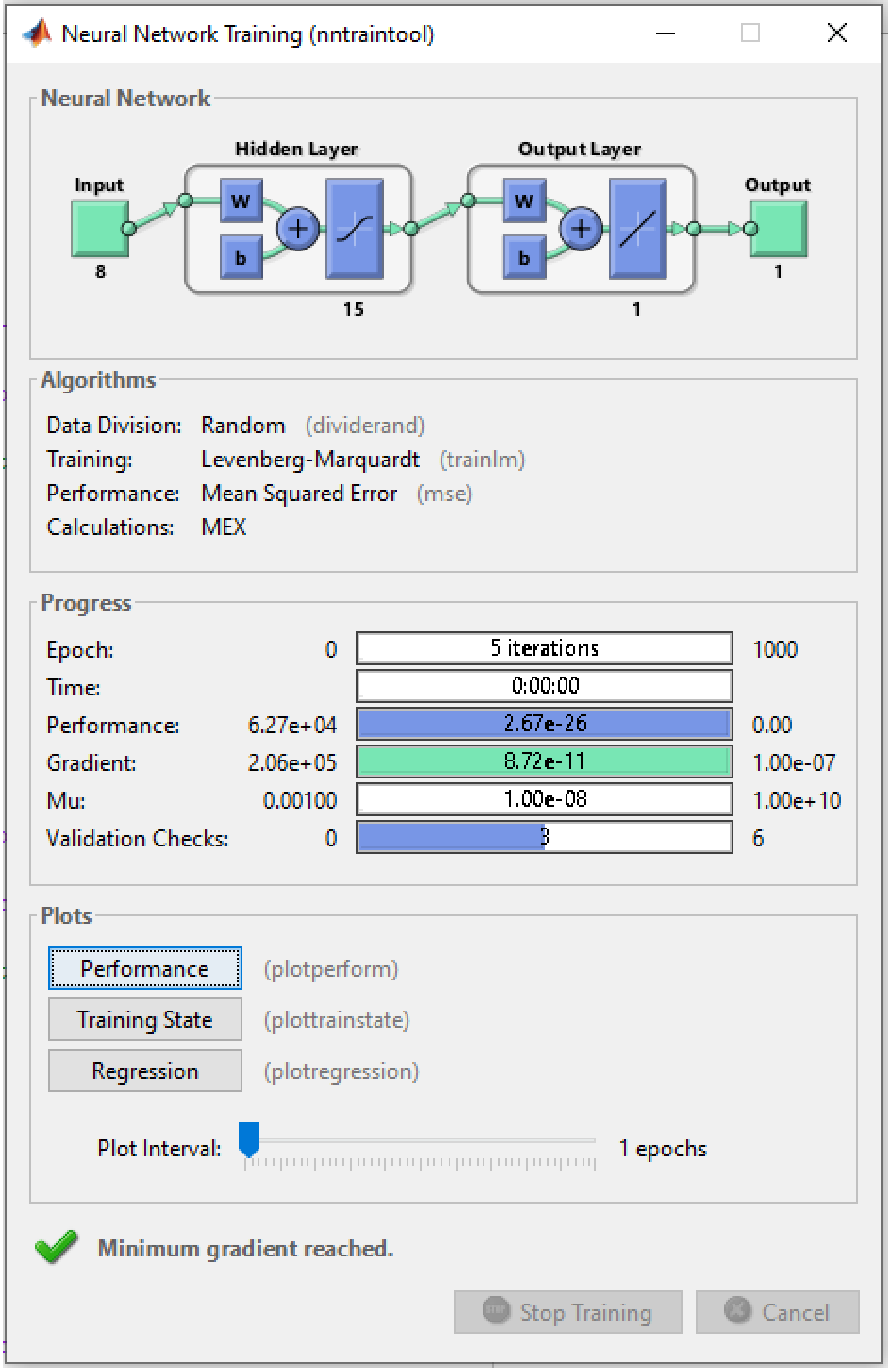

Figure 9: Training process of ANN with Backpropogation

ANN with cascade-forward backpropagation:

net = newcf creates a new network with a dialog box.

newcf(PR,[S1 S2...SNl],{TF1 TF2...TFNl},BTF,BLF,PF) takes,

- PR -- R x 2 matrix of min and max values for R input elements

Si -- Size of ith layer, for Nl layers

TFi -- Transfer function of ith layer, default = 'tansig'

BTF -- Backpropagation network training function, default = 'traingd'

BLF -- Backpropagation weight/bias learning function, default = 'learngdm'

PF -- Performance function, default = 'mse'

and returns an N layer cascade-forward backprop network.

TFi or transfer function can represent any differentiable transfer function such as purelin, logsig, or tansig.

Any of the backprop training functions like traingd, trainrp, trainbfg, trainlm, etc. can include the training function BTF.

Caution: As trainlm is very rapid but consumes more memory for execution, it is the default training function. Try any one of these when you get an “out-of-memory” error in the training process:

1. Through the setting of net-trainParam.mem_reduce to 2 or more, slow the training of trainlm but decrease the requirements of memory.
2. It’s better to use trainbfg which is a slower process but more efficient than trainlm in terms of memory utilization.
3. Or else, consider trainrp which is also slower but efficient in terms of memory than trainbfg.

The BLF or learning function can be either of the learning functions of backpropagation such as learngdm or learngd.

Any of the differentiable performance functions like msereg or mse can be the performance function.

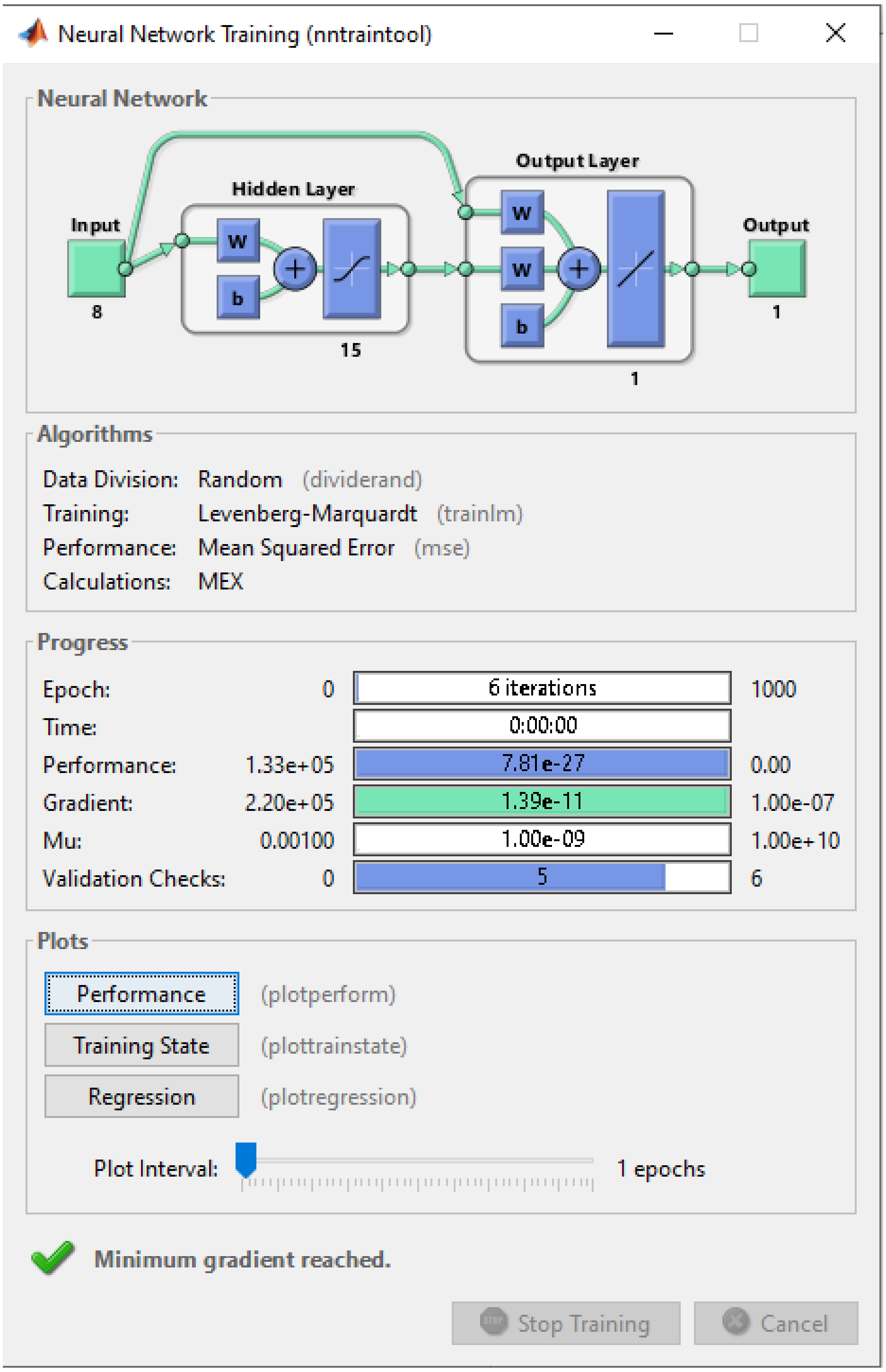

Figure 10: Training process of ANN with cascade-forward backpropagation

ANN with Elman backpropagation:

The initialization of threshold value and weight is done randomly when using newelm function. To assume the training function as traingdx, ' ' traingdxis represented that utilizes a back-propagation algorithm with momentum gradient-falling and adaptive learning spped to set the initial weight matrix on training and to learn it.

The required value is reached and maximum training time are achieved with the making of error function of the network and reaching maximum training times. The performance function gradient reduces until it reaches the minimum or proving the failure times in exceed of maximum times.

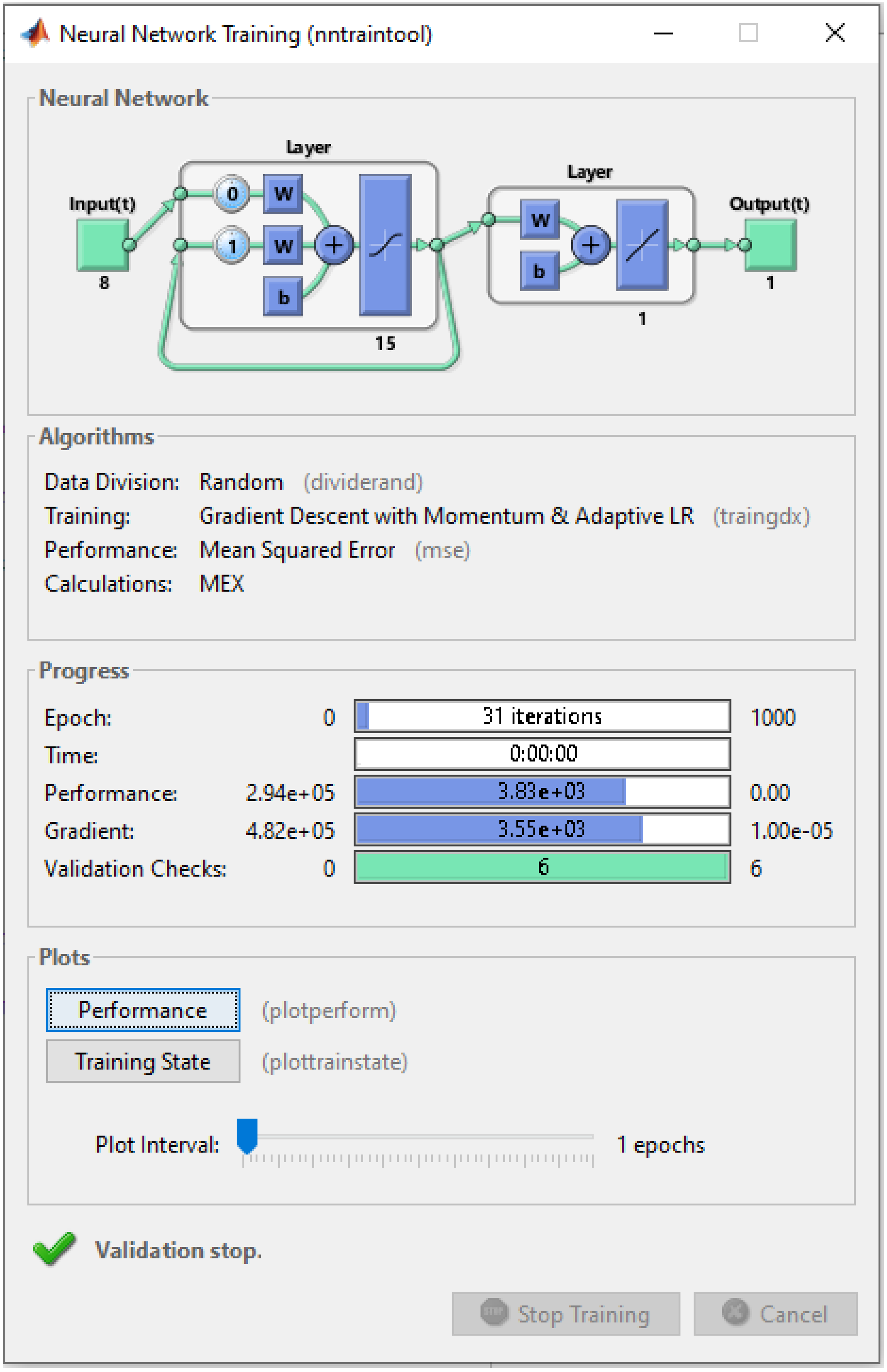

Figure 11: Training process of ANN with Elman backpropagation

The figure 12 represents the training results. In training process, backpropagation has the error 2577.26 with the accuracy of 69.55% and cascade-forward backpropagation has the error 2486.59 with the accuracy of 70.62% and Elman backpropagation has the error 1442.57 with the accuracy of 82.95% .

```
Command Window
 ------------- Training --------------
 ------------- ANN with Backpropogation --------------

  ANN with Backpropogation - Error = 2577.262130, Accuracy = 69.553903
 ------------- ANN with cascade-forward backpropagation ------------

  ANN with cascade-forward backpropagation - Error = 2486.598245, Accuracy = 70.624947
 ----------- ANN with Elman backpropagation --------------

  ANN with Elman backpropagation - Error = 1442.571479, Accuracy = 82.958400
fx >>
```

Figure 12: Training Results

The figure 13 representing the testing results. The actual value to predict 443 and 486, but ANN with back propagation predicts 385 and 776 respectively with the error and accuracy is 384.00 and 62.54%. ANN with cascade-backpropagation predicts 458 and 402 respectively with the error and accuracy is 99.00 and 89.34%. ANN with Elman-forward backpropagation predicts 469 and 422 respectively with the error and accuracy is 90.00 and 90.31%.

```
Command Window
 ------------- Testng --------------
 ------------- ANN with Backpropogation --------------

  Predicted Output1 - 385

  Predicted Output2 - 776

  Actual Output1 - 443

  Actual Output2 - 486

  ANN with Backpropogation - Error = 348.000000, Accuracy = 62.540366
 ------------- ANN with cascade-forward backpropagation ------------

  Predicted Output1 - 458

  Predicted Output2 - 402

  Actual Output1 - 443

  Actual Output2 - 486

  ANN with cascade-forward backpropagation - Error = 99.000000, Accuracy = 89.343380
 ----------- ANN with Elman backpropagation --------------

  Predicted Output1 - 469

  Predicted Output2 - 422

  Actual Output1 - 443

  Actual Output2 - 486

  ANN with Elman backpropagation - Error = 90.000000, Accuracy = 90.312164
fx >>
```

Figure 13: Testing Results

Conclusion:

For crop yield prediction, a machine learning approach is presented that provides superior performance in crop challenges based on large data. To achieve effective results in terms of crop yield, the deep neural networks have been utilized using previously produced information. However, the nonlinear and complex relationships in environmental conditions and their yield from historical data were learnt carefully by designed deep neural networks. Additionally, they can make accurate predictions in crop yields for new planted hybrids in new locations with known weather conditions. For the quality of weather prediction, the performance of model would be sensitive relatively that recommends the importance of weather prediction methods. The essential features are found out by the feature selection approach successfully and disclosed that environmental conditions had shown impact greatly on the crop yield when compared to the genotype. To overcome this limitation, the future work is based on researching more advanced models that should be more accurate and described clearly.

References:

1. Roche, Dominique. "Stomatal conductance is essential for higher yield potential of C3 crops." Critical Reviews in Plant Sciences 34, no. 4 (2015): 429-453.
2. Gornott, Christoph, and Frank Wechsung. "Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany." Agricultural and Forest Meteorology 217 (2016): 89-100.
3. Sujatha, R., and P. Isakki. "A study on crop yield forecasting using classification techniques." In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), pp. 1-4. IEEE, 2016.
4. Viña, Andrés, and Anatoly A. Gitelson. "New developments in the remote estimation of the fraction of absorbed photosynthetically active radiation in crops." Geophysical Research Letters 32, no. 17 (2005).
5. Hutter, Frank, Lars Kotthoff, and Joaquin Vanschoren. Automated machine learning: methods, systems, challenges. Springer Nature, 2019.
6. Priya, P., U. Muthaiah, and M. Balamurugan. "Predicting yield of the crop using machine learning algorithm." International Journal of Engineering Sciences & Research Technology 7, no. 1 (2018): 1-7.
7. Balakrishnan, Narayanan, and GovindarajanMuthukumarasamy. "Crop production-ensemble machine learning model for prediction." International Journal of Computer Science and Software Engineering 5, no. 7 (2016): 148.
8. Siju, H. L., & Patel, P. J. (2018). Review on Crop Yield Prediction using Data Mining Focusing on Groundnut Crop and Naive Bayes Technique.
9. Bhanumathi, S., Vineeth, M., &Rohit, N. (2019, April). Crop Yield Prediction and Efficient use of Fertilizers. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0769-0773). IEEE.
10. T.GiriBabu, Dr.G.AnjanBabu, Big Data Analytics to Produce Big Results in the Agricultural Sector, March 2016.
11. D Ramesh , B Vishnu Vardhan, Analysis Of Crop Yield Prediction Using Data Mining Techniques, International Journal of Research in Engineering and Technology, Jan-2015.
12. S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, March 2017.
13. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.

14. Priya, P., U. Muthaiah, and M. Balamurugan. "Predicting yield of the crop using machine learning algorithm." International Journal of Engineering Sciences & Research Technology 7, no. 1 (2018): 1-7.
15. Hunt Jr, E. Raymond, and Craig ST Daughtry. "What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture?" International journal of remote sensing 39, no. 15-16 (2018): 5345-5376.
16. Chalumuru Suresh , Kolli Kamakshaiah , SatishThatavarti …ect, Accurate And Timely Prediction Of Rice Crop Disease By Means Of Machine Learning Algorithms, ISSN: 2005-4238, Vol. 28, No. 13, (2019), pp. 662-671**.**
17. G. SuryaNarayana, Kamakshaiah Kolli, MohdDilshad Ansari and Vinit Kumar Gunjan A Traditional Analysis for Efficient Data Mining with Integrated Association Mining into Regression Techniques, ISBN : 978-981-15-7961-5, LNEE,Volume 698(2020),pp 1393-1404