Universal Approximation of Nonlinear System Predictions in Sigmoid Activation Functions Using Artificial Neural Networks

R. Murugadoss[1]
[1]Sathyabama University, Research Scholar
Department of Computer Science and Engineering,
Chennai, St Ann's College of Engineering and Technology,
Chirala-523157. Andhra Pradesh.
murugadossphd@gmail.com
mdossresearch@gmail.com

Dr. M. Ramakrishnan[2]
[2]Professor & Chair Person, School of Information Technology
Madurai Kamaraj University, Madurai.
ramkrishod@gmail.com

Abstract: **The sigmoid activation function cast-off to convert the equal of activation of units (neurons) in the output indicator. There are a numeral of mutual tasks in activation with the use of artificial neural networks (ANN). The maximum communal use of manifold functions to Multi Layered Perceptron (MLP) and the transmission of professions in research and engineering. However, given the wide range of problematic fields are applied in the MLP, it is interesting to suspect that the detailed difficulties that require one or exact activation utilities of the group. The aim of this paper is to consider the presentation of buildings MLP generalized who appeared deployment algorithm by numerous dissimilar functions to activate the sigmoid neurons of the hidden and output layers.**

Keywords— Sigmoid Activation Functions, Multi-Layered Perceptron, Artificial Neural Networks, Performance Analysis

I. INTRODUCTION

Adaptive classifications Artificial Neural Network (ANN) [1] category was primarily interested by the equivalent dispensation competences of actual intelligences, but the dispensation essentials and structures used in artificial neural networks have slight in communal with organic structures. Neural industrial networks are systems of elements modest treatment with come across center adaptable parameters W. Amongst these adaptable limitations authorizations the system to study uninformed vector outline of the universe input *X* to the universe outputs *Y=FW(X)*. The perspective of probability structures must adapt to the convergence joint probability density *p(X, Y)* or rear probability *p (Y|X)* values of input and output. The flexibility of the associated transfer functions used to estimate the limits of the decision parameters strongly with the number of jobs (and thus with a number of parameters to adapt available for training) needed to model complex shapes of the border decision. In approximation theory and used many of the functions [2], while using a neural network simulation on almost sigmoid or Gaussian functions exclusively devices. This paper presents a study suitable for the transfer of functions and neural networks in an attempt to show the hidden potential in their choice.

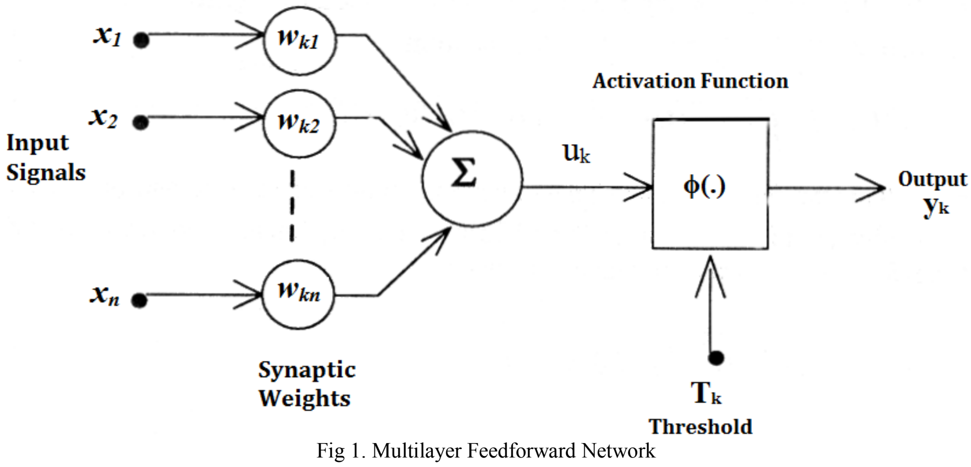

Fig 1. Multilayer Feedforward Network

Neural networks are industrial adjustment through the strength of a global mainframe systems, that is, they dismiss appreciate arbitrary scheme of space and unique vector universe to additional carriers. Neural networks with single hidden layer by means of sigmoid functions stand universal approximations, that is, they can be rounded arbitrary constant function on contracting with arbitrary exactness assumed a necessary numeral of neurons. These sports results do not mean that the sigmoid functions always provide the best option or the nervous good approximation is easy to find. With networks of neurons that give the Gaussian output instead of the sigmoid output is also approximations [3].

It has been suggested there is a new type of job transfer, and called on the rails camouflage, by Hartmann and Keller [4]. In interconnection networks and functional Bao used a blend of dissimilar tasks, by way of polynomial functions, rotating, sigmoid and Gaussian. Using balanced transferal functions by Haykin Wong with actual respectable consequences [5]. Networks in conical section function Dorffner presented jobs that change smoothly from the sigmoid to like camouflage. And used transfer functions Lorentzian, which can be considered a simple Gaussian functions, through Giraudet God. [14].

II. LITERATURE SURVEY

One of the most attractive in Artificial Neural Networks features is the prospect to adjust their performance to the altering features of the system showed. Post periods, several scholars have made a diversity of approaches to progress ANN performance through improved training approaches, and learning parameters, or network infrastructure, and is relatively little business with the use of some functions activation. Liu Yao and improve the configuration of Generalized Neural Networks (GNN) with two diverse types of playback function, which is the basis of the functions of the sigmoid and Gaussian [6]. Sopena and others. A number of experiments (with standard problems used on a large scale) showing that the grids forward multilayer with the function of activating the condition learn two applications of the fastest size while increasing the ability to generalize (compared with Artificial Neural Networks with the function of logistics activated). Dorffner Conic Section Function Neural Networks (CSFNN) which is a united background for the MLP and RBF networks to sort instantaneous advantages of together networks used in [7]. Showing Bodyanskiy double neuronal structure new wavelets obtained by modifying the typical wavelet neurons, and suggests learning algorithms. The proposed structural design allows to improve the convergence properties between neurons wavelet. And use all of these activation functions are known in the contract of separate layer of the MLP to resolve the various difficulties of nonlinear. However, no studies comparing the performance of these functions using activation.

III. METHODOLOGY

The neural network are parallel, and the distribution of the signal mainframe consists of unified totaling elements called neurons [8]. It is commonly granted that the neural network is similar to the anthropological mind in two ways: (a) it obtains data concluded the learning development. Use the synaptic weightiness linked with the interconnection between neurons to encode the above data as fit as data gained concluded explanations by devices considered for the environment in which the neural network functions.

In this paper are multilayer neural prior feeding network on the source of receptors networks initially established by Rosenblatt [8]. Are presented non-linear model of neurons and three-layer grid in Figure 3. The distinguishing comparison for all neurons are matching. Resulting from neurons in layer k^{th} m^{th} signals defined by

$$y_k^m(t) = \phi[u_k^m(t)] \tag{1}$$

$$u_k^m(t) = \sum_{j=1}^{p} w_{kj}^m(t) y_j^{m-1}(t) + T_k^m \tag{2}$$

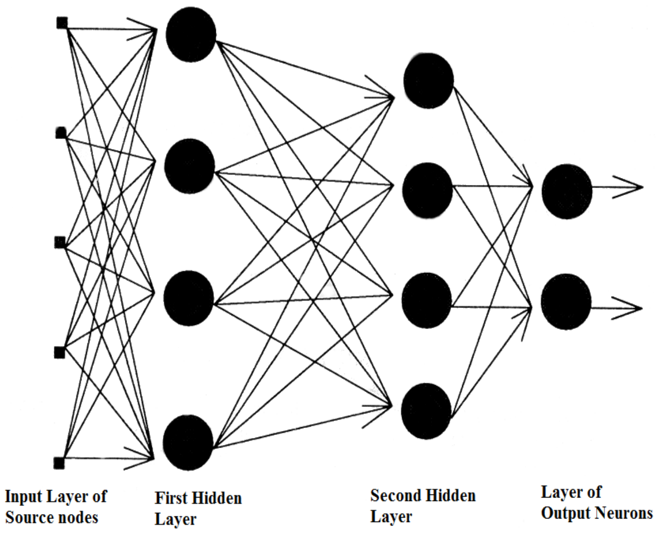

Fig 2: A neuron model and a multilayer feedforward neural network.

For an M-layer complex, the synaptic weightiness, $w_{ji}^m(t)$ linking neuron i of layer $(m-l)$ to neuron j of layer m, $l \le m \le M$, is efficient rendering to

$$w_{ji}^m(t+1) = w_{ji}^m(t) + \Delta w_{ji}^m(t), \tag{3}$$

$$\Delta w_{ji}^m(t) = \eta \delta_j^m(t) y_i^{m-1}(t), \tag{4}$$

$$\delta_j^m(t) \equiv -\frac{\partial E_t}{\partial u_j^m} \tag{5}$$

where $0 < \eta < 1$ is the learning proportion, and $\delta_j^m(t)$ is the indigenous error incline. The output layer $(m = M)$, the limited error gradient has the method.

$$\delta_j^M(t) = [d_j(t) - y_j^M(t)]\dot{\phi}[u_j^M(t)] \equiv e_j(t)\dot{\phi}[u_j^M(t)] \tag{6}$$

$$\delta_j^m(t) = \dot{\phi}[u_j^m(t)] \sum_{k=1}^{N_{m+1}} \delta_k^{m+1}(t) w_{kj}^{m+1}(t) \tag{7}$$

Towards progress the back-propagation learning algorithm, an impetus period is added to level the weightiness variations

$$\Delta w_{ji}^m(t) = \eta \delta_j^m(t) y_i^{m-1}(t) + \alpha \Delta w_{ji}^m(t-1) \tag{8}$$

where $0<\alpha<1$ is the energy element. The general burden modification for the consignment learning algorithm is agreed by

$$\Delta w_{ji}^{m}=\sum_{t}\Delta w_{ji}^{m}(t) \tag{9}$$

The present request develops an entire of *60* signal illustration standards (patterns) for respective iteration.

IV. PROPOSED METHOD

Consider the network is regarded as by mass vector w represents free. The definition of the function of the typical total that would be the minimum during the physical activity stage in expressions of N input and output patterns are as surveys:

$$E_{av}(w)=\frac{1}{2N}\sum_{n=1}^{N}\sum_{j\in C}\left[d_j(n)-y_j(n)\right]^2 \tag{10}$$

Focus on neurons arbitrary i, which may be situated to found everywhere in the system, and can be considered conduct throughout the training segment as a dynamic non-linear, and that in the background of the theory of Kalman filter can be labelled by measuring the subsequent state calculations [9], [10] :

$$w_i(n+1)=w_i(n) \tag{11}$$

$$d_i(n) = y_i(n) + e_i(n) \tag{12}$$

$$y_i(n) = \varphi\,(\mathbf{x}_i^{T}(n),\mathbf{w}_i(n) \tag{13}$$

Where consistent iteration n to provide pattern entry thousand, atheist and $y_i(n)$ is the input vector and output of neurons i correspondingly and $e_i(n)$ is the dimension in the output blunder of neurons i, are given an approximation of instantaneous them by [11]:

$$e_i(n)= -\frac{\partial E(n)}{\partial y_i(n)} \tag{14}$$

$$E(n)=\frac{1}{2}\sum_{j\in C}\left[d_j(n)-y_j(n)\right]^2 \tag{15}$$

The differentiation in calculation (14) resembles toward the back-propagation of the universal inaccuracy to the output of neuron *i*.

Sigmoid Activation function *φ(•)* is blamable for the non-linearity in the nerve cells. W_i vector of the optimal model of neurons i should be "predictable" over training with illustrations. It is supposed that the activation function to be differentiable. Consequently, we can custom the Taylor sequence expansion of equation (15) for the present approximation of the mass vector and thus a linear equation becomes as surveys [12]:

$$\varphi\left(x_i^T(n)w_i(n)\right)\cong$$
$$q_i^T(n)w_i(n)+\left[\varphi\left(x_i^T(n)\hat{w}_i(n)\right)-q_i^T(n)\hat{w}_i(n)\right] \tag{16}$$

where

$$q_i(n)=\left[\frac{\partial\varphi\left(x_i^T(n)w_i(n)\right)}{\partial w_{i(n)}}\right]_{w_i(n)=\hat{w}_i(n)}$$
$$=\hat{y}_i(n)\left[1-\hat{y}_i(n)\right]x_i(n) \tag{17}$$

$\hat{y}_i(n)$ is the productivity of the neurons that result as of using estimate weight i. In equation (8) we consume expected the usage of the utility of logistics. Sigmoid activation functions other, such as hyperbolic refraction, and can also be cast-off. During the initial period of the rightward side of equation (7) is the rectilinear term is mandatory while lingering period modeling error.

$$d_i(n)= q_i^T(n)w_i(n)+ e_i(n) \tag{18}$$

Due to a couple of calculations, we can sort use of Recursive Least Squares (RLS) algorithm comparisons, a special Kalman filter case, make an estimate of the Wi-vector weight (n) of the neuron i. And knows the solution resulting system following repeated equations [13], which describes multiple extended Kalman algorithm (Mika) [14]:

(3)

$$r_i(n) = \lambda^{-1}P_i(n-1)q_i(n) \tag{19}$$

$$k_i(n) = r_i(n)\left[1+r_i^T(n)q_i(n)\right]^{-1} \tag{20}$$

$$w_i(n+1) = w_i(n) + e_i(n)k_i(n) \tag{21}$$

$$P_i(n+1) = \lambda^{-1}P_i(n) - k_i(n)r_i^T(n) \tag{22}$$

Wherever, n=1,…,N is the repetition numeral and N is the entire number of occurrences. The vector $\boldsymbol{q}_i(n)$ characterizes the linearized neuron initiation utility assumed in calculation (6), $\boldsymbol{P}_i(n)$ is the present approximation of the converse of the covariance background of $\boldsymbol{q}_i(n)$ and $\boldsymbol{k}_i(n)$ is the Kalman improvement. The limitation λ is a disremembering factor which receipts principles in the sequence (0,1], and $e_i(n)$ is the confined quantity of the universal error. Equation (13) is entitled the Riccatti modification equation.

V. EXPERIMENTAL RESULT

The Bayesian suggestion outline functional effectively in the enterprise of multi-layer perceptron (MLP) in the effort of Mackay.

Bayesian neural networks to resolve the problems of reversion and classification approaches have been proposed. These approaches assertion to overwhelm some of the complications bump into in such a unified approach over fitting.

In customary techniques, based on the training of the neural networks industrial reduce the error purpose, and are frequently driven by approximately fundamental attitude, such as the extreme probability [15]. The drawback of this approach is that the design of networks can agonize as of a number of inadequacies, counting the problematic of decisive the suitable level of typical complexity. Further complex models gives the best episodes of the training data, but if it is a very complex model might give poor oversimplification (overfitting).

And offers the perspective of Bayesian common framework and dependable data pattern acknowledgement and statistical exploration. In the background of neural networks, and follow the Bayesian methodology proposals several significant structures, comprising the subsequent [16]:

- ✓ The organization of style rises in an accepted method in the context of Bayesian. And can be treated with the corresponding parameters are constantly organizing within Bayesian setting, deprived of the essential for methods such as cross-authentication.
- ✓ For classification difficulties, and the propensity of conservative methods to sort it can be avoided excessive expectations in confidence in different parts of the data.
- ✓ Bayesian methods provide a framework and objective of the initial deal with the issue of the complexity of the model and circumvent a lot of glitches of over fitting that rise when using extreme likelihood.

TABLE 1: Accuracies for the five training /test data sets with no offset and 25 hidden nodes.

Set	*Training Data Accurateness*	*Test Data Accurateness*	*Bootstrapped Accurateness*
0	82.5%	66%	75.11%
1	77.2%	59%	71.75%
2	84.3%	60%	74.33%
3	82.9%	59%	75.64%
4	84.4%	64%	73.85%
Mean:	82.26%	61.60%	74.13%

TABLE 2: Precisions for the five training data set/test data sets with no offset and 45 hidden nodes.

Set	*Training Data Accurateness*	*Test Data Accurateness*	*Bootstrapped Accurateness*
0	84%	54%	72.96%
1	81%	62%	74.64%
2	77%	65%	75.11%
3	82%	67%	73.32%
4	80%	64%	74.11%
Mean:	80.80%	62.40%	74.03%

ANN representations are algorithms intelligent responsibilities such as learning, classification, recognition, and appreciation of optimization centered on the notion of how the human brain workings. [17] ANN consists of a huge number of dispensation fundamentals of the model, called neurons. For each neuron is associated to supplementary neurons concluded associates, each with attendant weight. Neurons deprived of links headed for them neuronal input and individuals who have nope connection to leave left from them, entitled neurons output. And characterized the relationship between the input and output of the shift in the neurons by sigmoid activation task. Grouping of input neurons, the nerve cells and output, and the associations concerning neurons with weights connected form ANN structure.

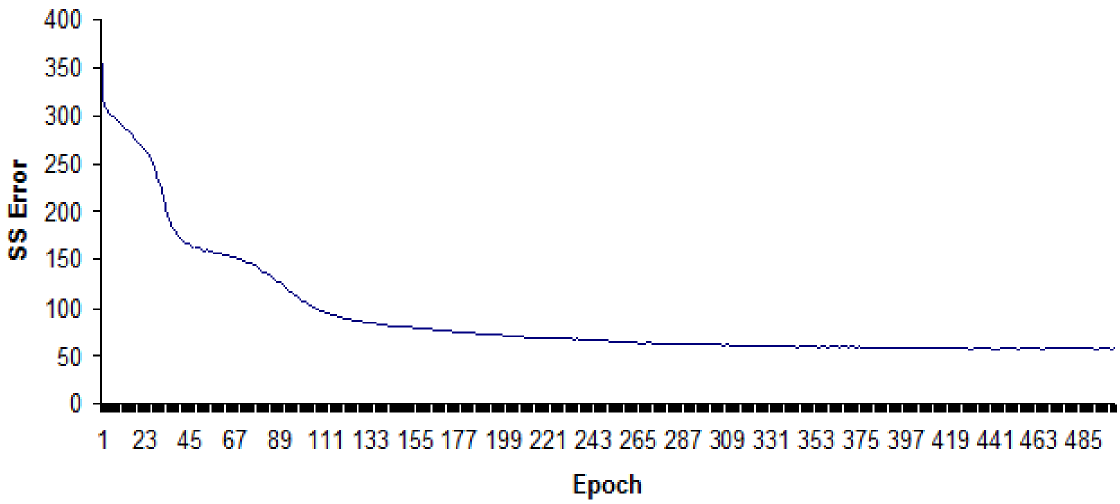

Fig 3: Sum of Squares Error of Training Set

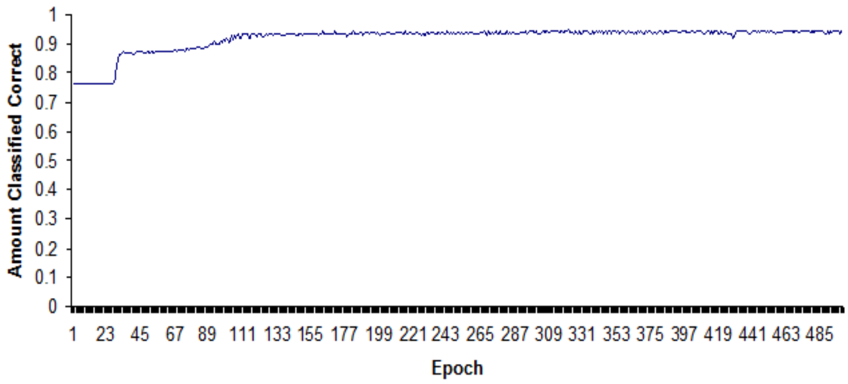

Fig 4: Test Set Classification percent

Some of the gains of using artificial neural networks is that they can citation patterns and notice trends that are frequently very complex to be observed by either humans or other computer performances [18]. Neural networks are suitable for industrial to pick up the patterns and trends of the current blaring data. The technique includes the training of the ANN with a huge representative illustration of the data and testing ANN using the data that was not comprised in the training in order to predict the outcomes of the new ANN. The training procedure comprises various statistics of layers and nerve cells and connections among neurons with weights related with them. Preceding layer signifies the output. The sum of hidden layers remains distinct by the manipulator.

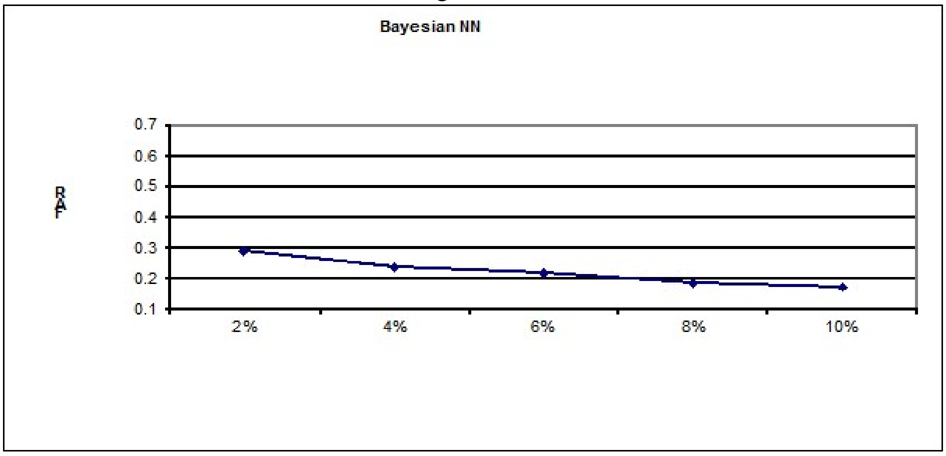

Figure 5: POD for different thresholds for different percentage of tornado in the test set.

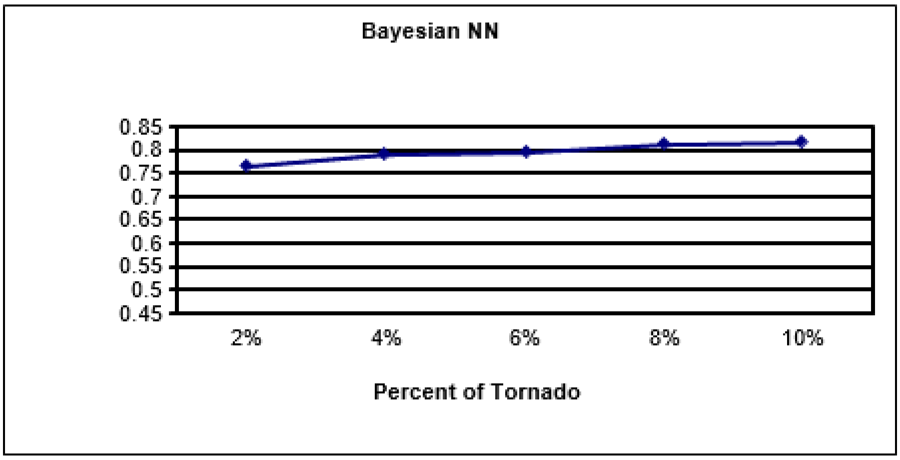

Figure 6: FAR for different thresholds for different percentage of tornado in the test set.

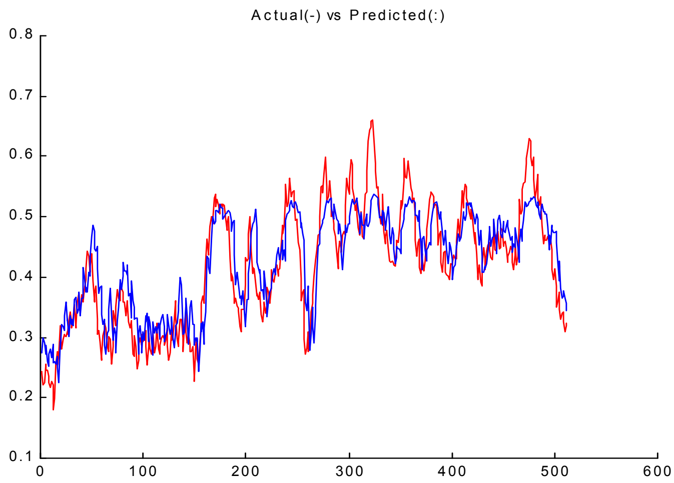

Fig 7: ANN Prediction with horizon = 5

The user can adapt the sum of neurons in respective layer to it. It can also be training and test error tolerances modified by the user. After being trained on the network and test the extent of user consummation, and it is standing by for use. It can display original sets of data entry on the network, then it will produce forecasts created on what it consumes erudite. ANN training and can be preserved as a practiced in information given to the analysis classification.

VI. DISCUSSION AND CONCLUSION

In this paper, we have the traditional five functions to activate and monotonic variation of the development of architecture MLP along with comprehensive learning Qaeda Delta. This is a known functions proposed effective activation and sigmoid bipolar, unipolar sigmoid, conic section, and the foundations of radial function (RBF). Based on the improvements in the performance of BNNs on Artificial Neural Networks, it is recommended that further research for the application of Bayesian networks to detect a tornado. A high level of skill showed an improvement over the previous research in terms of the ability to predict (the reduction of chaos) and can lead to a significant reduction in the loss of lives in the event of a practical implementation.

REFERENCE

[1] Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 8609-8613). IEEE.

[2] Sibi, P., Jones, S. A., & Siddarth, P. (2013). Analysis Of Different Activation Functions Using Back Propagation Neural Networks. *Journal of Theoretical and Applied Information Technology*, *47*(3), 1344-1348.

[3] Goodfellow, I. J., Mirza, M., Da, X., Courville, A., & Bengio, Y. (2013). An Empirical Investigation of Catastrophic Forgeting in Gradient-Based Neural Networks. *arXiv preprint arXiv:1312.6211*.

[4] Xiao, T., Zhang, J., Yang, K., Peng, Y., & Zhang, Z. (2014, November). Error-Driven Incremental Learning in Deep Convolutional Neural Network for Large-Scale Image Classification. In *Proceedings of the ACM International Conference on Multimedia* (pp. 177-186). ACM.

[5] Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *arXiv preprint arXiv:1404.7828*.

[6] Li, S., Chen, S., & Liu, B. (2013). Accelerating a recurrent neural network to finite-time convergence for solving time-varying sylvester equation by using a sign-bi-power activation function. *Neural processing letters*, *37*(2), 189-205.

[7] Xiao, L., & Lu, R. (2014). Finite-time solution to nonlinear equation using recurrent neural dynamics with a specially-constructed activation function.*Neurocomputing*.

[8] Miao, P., Shen, Y., & Xia, X. (2014). Finite time dual neural networks with a tunable activation function for solving quadratic programming problems and its application. *Neurocomputing*, *143*, 80-89.

[9] Balaji, S. A., & Baskaran, K. (2013). Design and Development of Artificial Neural Networking (ANN) system using sigmoid activation function to predict annual rice production in Tamilnadu. *arXiv preprint arXiv:1303.1913*.

[10] Nie, X., Cao, J., & Fei, S. (2014, May). Multistability and Instability of Competitive Neural Networks with Mexican-Hat-Type Activation Functions. In*Abstract and Applied Analysis* (Vol. 2014). Hindawi Publishing Corporation.

[11] Gowdru Chandrashekarappa, M. P., Krishna, P., & Parappagoudar, M. B. (2014). Forward and Reverse Process Models for the Squeeze Casting Process Using Neural Network Based Approaches. *Applied Computational Intelligence and Soft Computing*, *2014*.

[12] Ze, H., Senior, A., & Schuster, M. (2013, May). Statistical parametric speech synthesis using deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 7962-7966). IEEE.

[13] Ismail, A., Jeng, D. S., Zhang, L. L., & Zhang, J. S. (2013). Predictions of bridge scour: Application of a feed-forward neural network with an adaptive activation function. *Engineering Applications of Artificial Intelligence*, *26*(5), 1540-1549.

[14] Tan, T. G., Teo, J., & Anthony, P. (2014). A comparative investigation of non-linear activation functions in neural controllers for search-based game AI engineering. *Artificial Intelligence Review*, *41*(1), 1-25.

[15] Wu, S. F., Chiou, Y. S., & Lee, S. J. (2014). Multi-Valued Neuron with Sigmoid Activation Function for Pattern Classification. *Journal of Computer and Communications*, *2*(04), 172.

[16] Costarelli, D., & Spigler, R. (2013). Approximation results for neural network operators activated by sigmoidal functions. *Neural Networks*, *44*, 101-106.

[17] Castelli, I., & Trentin, E. (2014). Combination of supervised and unsupervised learning for training the activation functions of neural networks. *Pattern Recognition Letters*, *37*, 178-191.

[18] Yang, F., Tang, S., & Xu, G. (2013). Horseshoe chaos in a 3D neural network with different activation functions. *Discrete Dynamics in Nature and Societ y*,*2013*.