

Training Wave-Net based on Extended Kalman filter

Hayder Mahdi Abul-Ridha
drenghaider@yahoo.com
University of Babylon

Hilal A. Hussain Abood
Hilal_hussain@uobabylon.edu.iq
University of Babylon

PhD. In Electrical Engineering

MSc. In Electrical Engineering

Abstract—Wave-Net is a promise network which utilizes the wavelet transform in building a new structure of neural network. Training is the major problem facing researchers with the neural networks. In this paper, we use the extended Kalman filter as an efficient tool in training the Wave-Net. The results show that the ability of using extended Kalman filter as training algorithm for wave-net and use it in classification problem and come up with good results especially in reducing the number of iterations in training phase. A comparison between extended Kalman filter and conventional back propagation, which are used in a classification problem, show that the training with Kalman filter is better than training with gradient descent.

Index Terms: wave-net, Extended Kalman filter, wavelet, Neural network.

الخلاصة:

ان شبكة الموجة العصبية هي شبكة واعدة تستخدم محول الموجة لبناء تركيب جديد للشبكة العصبية. ان اهم مشكلة تواجه الباحثين هي مشكلة تدريب الشبكة العصبية. تم استخدام مرشح كالمان المطور في هذا البحث كاداة كفوءة في تدريب شبكة الموجة العصبية. حيث بينت النتائج نجاح تطبيق مرشح كالمان المطور في امكانية تدريب شبكة الموجة العصبية في حل مسألة التصنيف والحصول على نتائج جيدة وخصوصا في تقليل عدد خطوات الزمن اللازمة للتدريب. بينت المقارنة بين اثنين من خوارزميات التدريب ، وهي خوارزمية مرشح كالمان المطور وخوارزمية الانتشار العكسي التقليدية التي تم استخدامهما في مسألة التصنيف افضلية استخدام خوارزمية مرشح كالمان المطور على خوارزمية الانحدار التدريجي من حيث عدد خطوات الزمن اللازمة للتدريب.

1. INTRODUCTION

Since the emergence of neural network, this field gains a great interest by the researchers and new structures of neural networks have been proposed. Wave-Net is one of the attractive structures of neural networks that combine the wavelet with neural network to give a new model [Deghani , Ahmadi , and Eghtesad 2008].

Many structures of neural networks have been introduced, beginning with the perceptron through Hopfield network, Feed forward neural network, radial basis neural network, etc. A new direction trying to combine two fields or more to give a new model, and all that toward building efficient neural network such as the hybrid neural fuzzy network and other hybrid structures[Cheng-Jian and Cheng-Hung 2005]. All the combinations of neural networks have been used extensively in classification, recognition, control, and other problems [Lin ,Quan, and Yao 2006] ,[Puskorius , Feldkamp 1994].

Wavelet transform (WT) is a good tool for the analysis of signals and images. The main merit of wavelet transform over Fourier transform is the ability of specifying the time-frequency position [Kannan, Martin 1996]. Recent feature extraction methods utilize WT to get the best components to be used as an input vectors to the classifiers [Michael, Perz, Black, and Sammer 1993].

Learning process is the major problem with neural network where by which the effective parameters are specified. While the back propagation represent a nice procedure in the training of neural networks, but it carries a drawbacks, one of which is the problem of local minima [Dan 1996]. Many algorithms have been proposed for training neural networks in a trial to find a best and fast way for defining the weights and other parameters.

Kalman filtering plays an essential role in systems theory and has found a wide range of applications in engineering and non-engineering systems.[Xiao, Xie and Minyue, 2008],[Dan 2002]..

This paper presents a way for training Wave-net by using extended Kalman filter and a comparison with another learning algorithm which is the Gradient descent method.

2. WAVE-NET STRUCTURES

Wave-net is a hierarchal structure network formed from artificial neural network (ANN) with mother wavelet function as its bases functions to produce powerful computational system for complex mapping and identification problems.

Fig. 1 shows the network design to solve identification problem with input layer and one hidden layer with variable neurons and depend on Mexican hat mother wavelet as a base of the hidden layer to produce outputs as a summation through the weights of network.

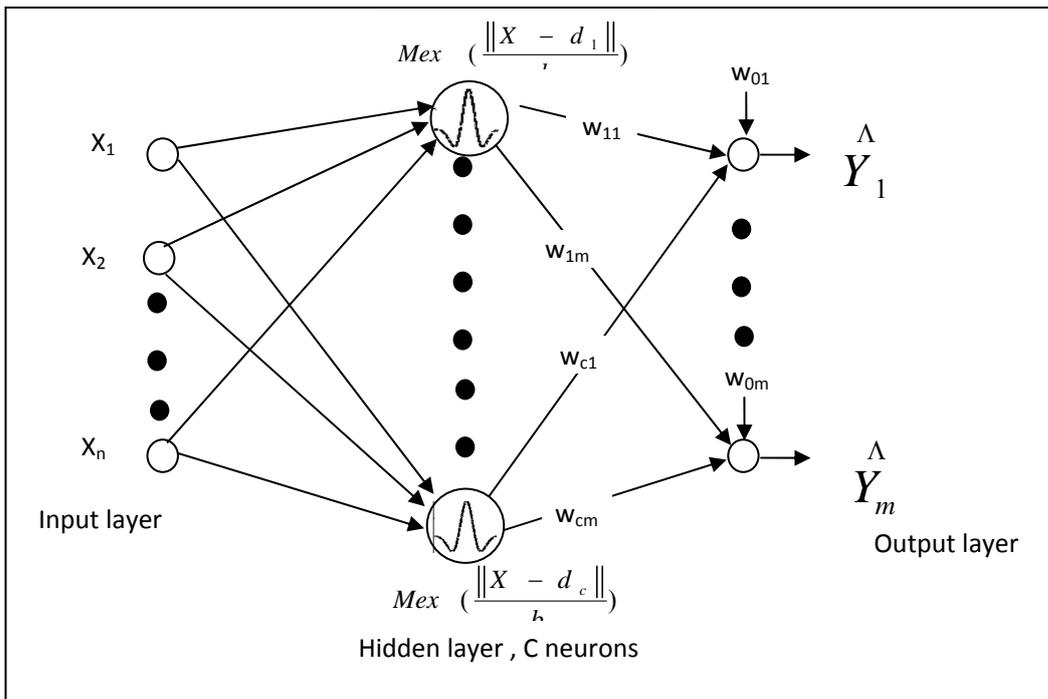


Fig. (1) Wave-Net structure

3. Network and Training Procedures

A notation of (W) will be used to describe the weight matrix of the network which is a matrix of ($c \times m$) dimensions, where c is the number of neurons in the hidden layer and m is the number of outputs as in Eq. (1)

$$W = \begin{bmatrix} w_{01} & w_{11} & \dots & w_{c1} \\ w_{02} & w_{12} & \dots & w_{c2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & & \cdot \\ w_{0m} & w_{1m} & & w_{cm} \end{bmatrix} \quad (1)$$

The output of this network will be the sum of multiplication of the weights with the response of the wavelet function in hidden layer as cleared in Eq. (2) for a set of L input- out training vectors

$$\begin{bmatrix} \hat{Y}^1 & \dots & \hat{Y}^L \end{bmatrix} = W \begin{bmatrix} 1 & 1 & \dots & 1 \\ \text{Mex}\left(\left\|\frac{X^1 - d_1}{b}\right\|\right) & \text{Mex}\left(\left\|\frac{X^2 - d_1}{b}\right\|\right) & \dots & \text{Mex}\left(\left\|\frac{X^L - d_1}{b}\right\|\right) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{Mex}\left(\left\|\frac{X^1 - d_c}{b}\right\|\right) & \text{Mex}\left(\left\|\frac{X^2 - d_c}{b}\right\|\right) & \dots & \text{Mex}\left(\left\|\frac{X^L - d_c}{b}\right\|\right) \end{bmatrix} \quad (2)$$

Where

\hat{Y}^i : The i th output vector

X^i : The i th input vector

$\text{Mex}(t) = (1 - t^2)e^{-t^2/2}$ is the Mexican hat mother wavelet function which is work here as a base function of the hidden neurons,

$t_k^i = \left\|\frac{X^i - d_k}{b}\right\|^2$: is the input the Mexican hat function of the i th input vector and k th

neuron of the hidden layer

d_k : the translation parameter of the mother wavelet of the k th hidden neuron

b : the dilation parameter of the mother wavelet.

Eq. (3) shows the network (plant) description

$$\hat{Y} = WZ \quad (3)$$

$$\text{Where } Z = \begin{bmatrix} 1 & 1 & 1 \\ z_{11} & z_{12} & z_{1L} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ z_{c1} & z_{c2} & z_{cL} \end{bmatrix} = [z_1 \quad z_2 \quad \dots \quad z_L]$$

and $z_{ik} = \text{Mex}(t_k^i)$

Now to train this network we have to find the error function and try to minimize the error as cleared in Eq.(4)

$$Er = \left\| Y - \hat{Y} \right\| \quad (4)$$

Where Y is the desired output of the network, to get the best performance from network we should optimize it against its parameters (the weights and translation parameter of mother wavelet). Here we took the dilation parameter to be constant (b =1). For the first parameter it has shown in eq.(5) [Karayiannis, 1999].

$$\frac{\partial E}{\partial w_i} = \sum_1^L (y_{ik} - \hat{y}_{ik}) z_k \quad (i=1, 2, \dots, n) \quad (5)$$

While for translation parameter for Mexican function mother wavelet as described in eq(6-8)

$$\text{Mex}(t) = (1 - t^2) e^{-t^2/2} \quad (6)$$

where: $t = \|x - d\|^2$

$$\frac{\partial \text{Mex}(t)}{\partial t} = -t * (1 - t^2) e^{-t^2/2} - 2t e^{-t^2/2} = \text{Mex}(t)' \quad (7)$$

$$\frac{\partial y}{\partial d} = -W * 2(x - d) * \text{Mex}(\|x - d\|)' \quad (8)$$

The Kalman filter addresses the general problem of trying to estimate the state of a discrete-time linear system process that is governed by the linear stochastic difference equations as in Eq.(9 & 10) [Kalman, 1960].

$$x_k = Ax_{k-1} + w_{k-1} \quad (9)$$

And measurements equation:

$$y_k = Hx_k + v_k \quad (10)$$

The random variables w_k and v_k represent the process and measurement noise (respectively). They are assumed to be independent (of each other), white, and with Gaussian probability distributions

$$p(w) \sim N(0, Q)$$

$$p(v) \sim N(0, R)$$

Q : the process noise covariance
R: the measurement noise covariance

but in our work the wavelet function is non linear, so here we can use extended Kalman filter to overcome the non-linearity problem. We can linearize the estimation around the current estimate using the partial derivatives of the process and measurement functions to compute estimates even in the face of non-linear relationships as in Eq. (11 & 12).

$$\begin{aligned} x_k &= f(x_{k-1}) + w_k \\ y_k &= h(x_k) + v_k \end{aligned} \quad (11)$$

Eq. (13) shows that Extended Kalman filter can be derived by using first-order Taylor series approximations (neglecting higher order terms)

$$\left. \begin{aligned} f(x_k) &\approx f(\hat{x}_k) + F_k \times (x_k - \hat{x}_k) \\ h(x_k) &\approx h(\hat{x}_k) + H_k^T \times (x_k - \hat{x}_k) \end{aligned} \right\} \quad (13)$$

where

$$F_k = \left. \frac{\partial f(x_k)}{\partial x} \right|_{x=\hat{x}_k}$$

$$H_k = \left. \frac{\partial h(x_k)}{\partial x} \right|_{x=\hat{x}_k}$$

We can update the error and obtain the best prediction by the recursion as in Eq. (14 & 15)

$$\hat{x}_k - x_k = f(\hat{x}_{k-1}) - f(x_{k-1}) - w_k \quad (14)$$

$$\left. \begin{aligned} x_k &= f(\hat{x}_{k-1}) + F(x_{k-1} - \hat{x}_{k-1}) + w_k \\ y_k &= h(\hat{x}_k) + H_k^T (x_k - \hat{x}_k) + v_k \end{aligned} \right\} \quad (15)$$

Now the desired estimation \hat{x}_k can be obtained by the recursion described in Eq.(16)

$$\left. \begin{aligned} \hat{x}_k &= f(\hat{x}_{k-1}) + K_k [y_k - h(\hat{x}_{k-1})] \\ K_k &= P_k H_k (R + H_k^T P_k H_k)^{-1} \\ P_{k+1} &= F_k (P_k - K_k H_k^T P_k) F_k^T + Q \end{aligned} \right\} \quad (16)$$

K_k : Kalman gain.

P_k : the covariance matrix of the state estimation error.

\hat{x}_k : the state estimation

Now to achieve the optimization problem in a form fit with Kalman filtering, we let the elements of the weight matrix W and the translation parameter of the mother wavelet d to be the state of a nonlinear system, and we constitute the output of the

wavelet network with the output of the nonlinear system to which the Kalman is applied.

The state of the nonlinear system can then be written as in eq. (16)

$$x = [w_1^T \quad \dots \quad w_n^T \quad d_1^T \quad \dots \quad d_m^T]^T \quad (16)$$

The block diagram of our proposed system is outlined in figure (2). The state of the system has been modeled depending on the weights of wavenet and the translation parameter of the Mexican hat base function. The initial value of the weights has been given the same value, while the translation parameter is set to zero. After that a system noise (uncertainty) is added.

The system will predict the output from input of current state and compare with training set output, the error will correct the state of the system and the system error covariance by changing the Kalman gain and system error covariance

This process will continue until we get acceptable error, Then the network will be ready to use and solve the classification problem

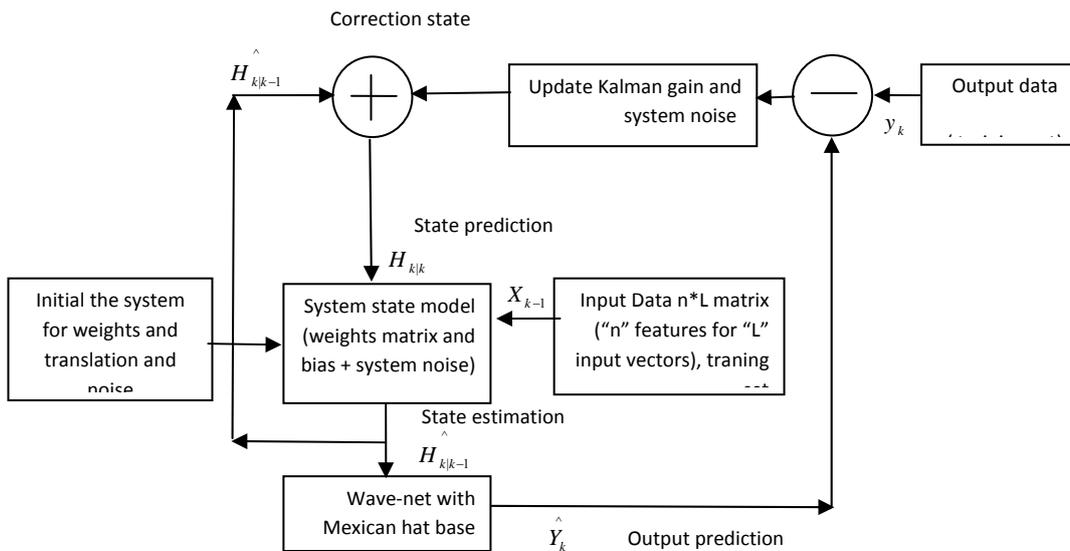


Fig. (2) Wavenet Training block diagram

4. RESULTS

We use Fisher's Iris data set [Fisher,1936] to train and test our wave-net, the dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample, they are the length and the width of sepal and petal, in centimeters.

A program of MATLAB R2008a is used to train the network using 50% of the data chosen randomly then it tested with the randomly chosen data, and the weights of the wave-net were chosen to be zeros initially. After some experiments we found that the best performance for the network was with in error rate of 0.001 to terminate the training.

Here we test the network to check its performance. We change the number of hidden layer and start with multi values of covariances P,Q and R and the result was reported by measuring the average of the CPU time for 10 trials also get the average of corrected outputs for the input vectors as shown in Fig. (3 & 4)

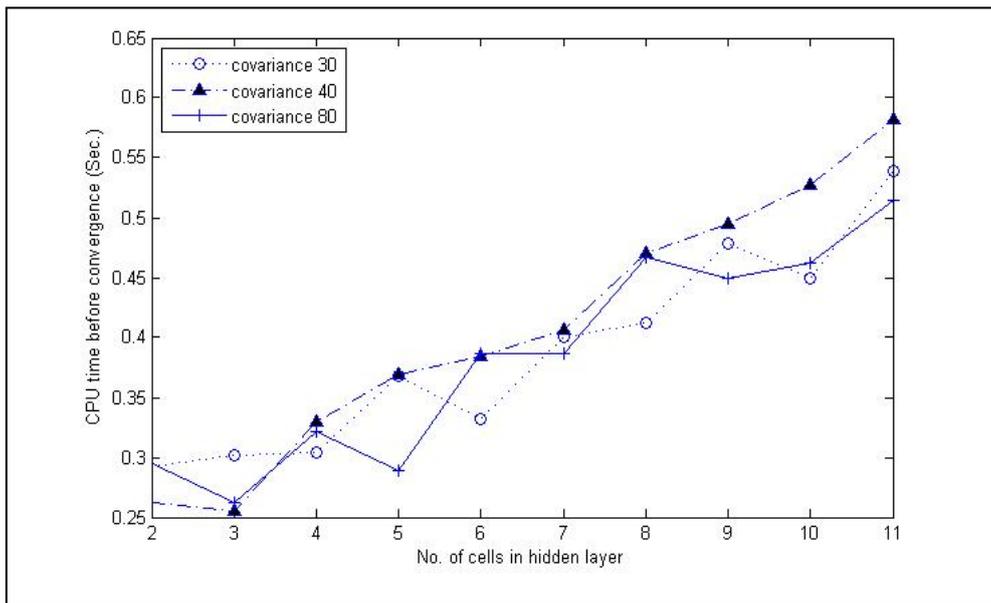


Fig. (3) average CPU time before convergence against No. of cells in hidden layer with different initial covariances

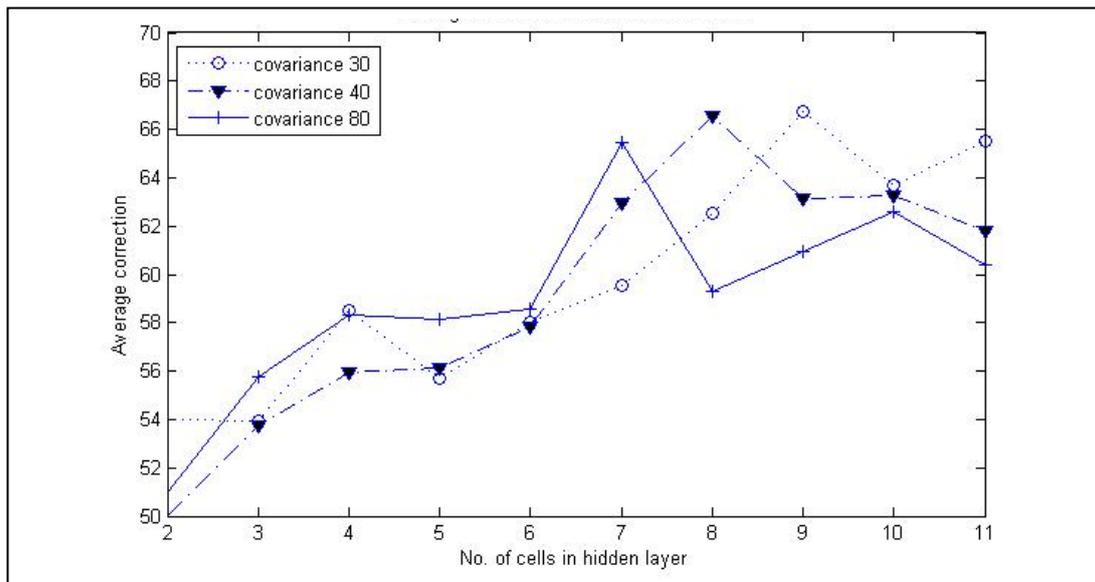


Fig.(4) Performance of the network against No. of cells in hidden layer with different initial covariances

As we notice from results the network succeed to overcome classification problem with acceptable performance and with good time for CPU before the training converge or locked in local minima which can not improve its error anymore.

Also its clear that initial covariances didn't affect the training process with in the limits we took here in our work (30, 40 and 60) while the number of hidden layers usually improve the performance as increase till reach 6or 7 cells and the improvement will not be noticeable.

The success of Kalman training is very clear in reducing the computational efforts in compare with conventional back propagation algorithm Resilient back propagation, Multilayer networks typically use sigmoid transfer functions in the hidden layers. The results are shown in Fig. (5).

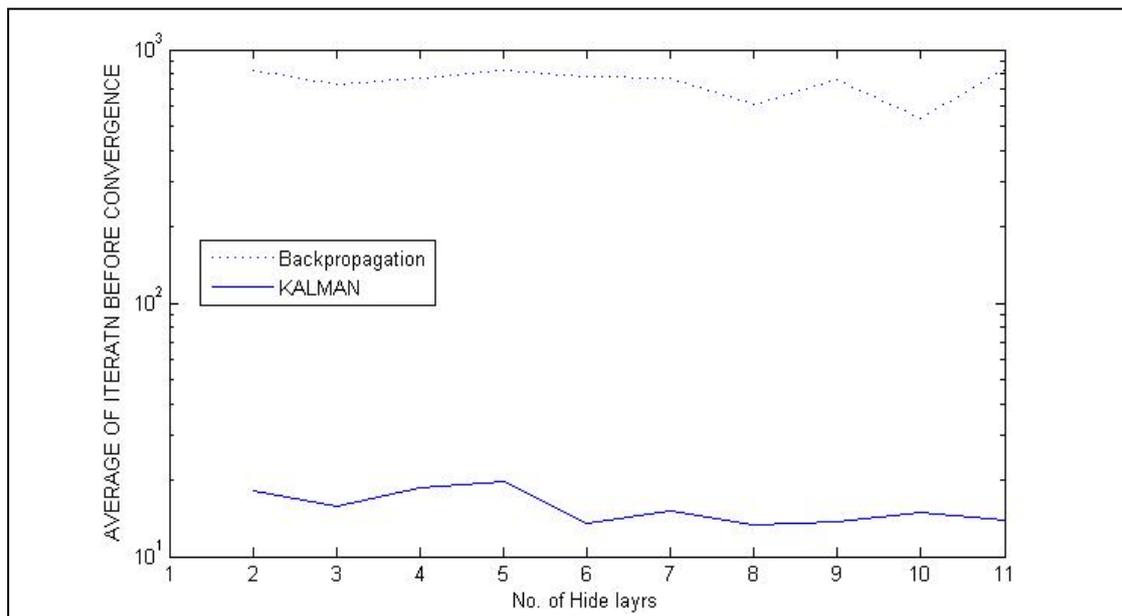


Fig. (5) average No. of iteration for learning process in logarithmic scale for both Kalman and back propagation algorithms training against No. of cells in hidden layer

4. CONCLUSIONS

Our work here shows that we can apply Kalman filter in training wave-net network and how this network succeed in classification problem.

Furthermore the results we have got show the big role of Kalman filter in reducing the load on computer processing by reducing the number of iterations in learning phase when compare with back propagation learning algorithm.

It's suggested for future work to improve the performance of network to use the dilation parameter in addition to translation parameter and weights to minimize the error in training process to solve classification problem,

References

Cheng-Jian Lin and Cheng-Hung Chen, "A self-constructing compensatory neural fuzzy system and its applications", Elsevier, Mathematical and computer modeling, Vol. 42, pp. 339-351, 2005.

Dan Simon, "Training radial basis neural networks with the extended Kalman filter", Neurocomputing Vol. 48, Issues 1-4, pp. 455-475, 2002

Dan W., "Artificial neural networks", Prentice Hall, 1996.

Deghani M., Ahmadi M., and Eghtesad M., "Wavelet Based Network Solution for forward kinematics problem of HEXA parallel robot", IEEE conference in intelligent Engineering systems, pp. 63-70, 2008.

Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME—Journal of Basic Engineering, pp. 35-45 ,March 1960.

Karayiannis N., "Reformulated radial basis neural networks trained by gradient descent, IEEE Trans. Neural Networks". Vol.3 pp. 657–671, 1999.

Kannan Ramchandran, Martin Vetterli, and Cormac Herleg," Wavelets, Subband Coding, and Best bases", IEEE Transaction, Vol. 84, pp. 541-560, 1996.

Lin Mei, Quan Taifan, and Yao Tianbin, "Tracking maneuvering target based on neural fuzzy network with incremental neural learning", journal of systems engineering and electronics, Vol. 17, No. 2, pp. 343-349, 2006.

Michael M, Perz S., Black C., and Sammer G., "Detection and classification of P waves using Gabor wavelets", IEEE, Computer in cardiology, pp. 531-534, 1993.

Nan Xiao¹, Lihua Xie and Minyue,²⁰⁰⁸, Fu², "Kalman filtering over unreliable communication networks with bounded Markovian packet dropouts", *Int. J. Robust Nonlinear Control* (2008), Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rnc.1389

Puskorius G., Feldkamp L "Neuro control of nonlinear dynamical systems with Kalman filter trained recurrent networks", IEEE Trans. Neural Networks: pp. 279–297, 1994.