CHAPTER (6)

Experimental Results

Chapter 6 Experimental results

6.1 Overview

This section presents experimental results detailing the performance of the back propagation neural network intrusion detection system (BPNNIDS) and radial basis function neural network intrusion system (RBFNNIDS) that are trained by:

- 10% version of the KDD 1999 Cup dataset.
- The selected sample from 10% version of the KDD 1999 Cup dataset according to the proposed reduction algorithm.

A Pentium 4 (2.33 GHz) laptop, with 2 GB of memory is used to implement the systems.

6.2 Five categories system using 10% version of KDD Cup 99 (FCS 10 KDD)

A four layer neural network (Figure 6-1) is used. It has three hidden layers. The size of the input, three hidden layers and the output layer are 114, 50, 25, 13, and 5 respectively. 114 represent the number of features used in the training and 5 represent the number of categories.

We first tried network architecture of two hidden layers. But, it did not converge to a solution. Then we increased it to three hidden layers. The number of neurons in each hidden layers is chosen by trial. We started with a size of the first hidden layer to be 41 neurons. This is the original number of features in the dataset. Then, this size is increased until the optimum results

are obtained. Similarly the size of the hidden two and hidden 3 layers are chosen. The parameters used are:

- The mean square error (MSE) in the training step is 0.001, transfer sigmoid.
- Learning rule momentum.
- Step size 1.0.
- Momentum 0.7

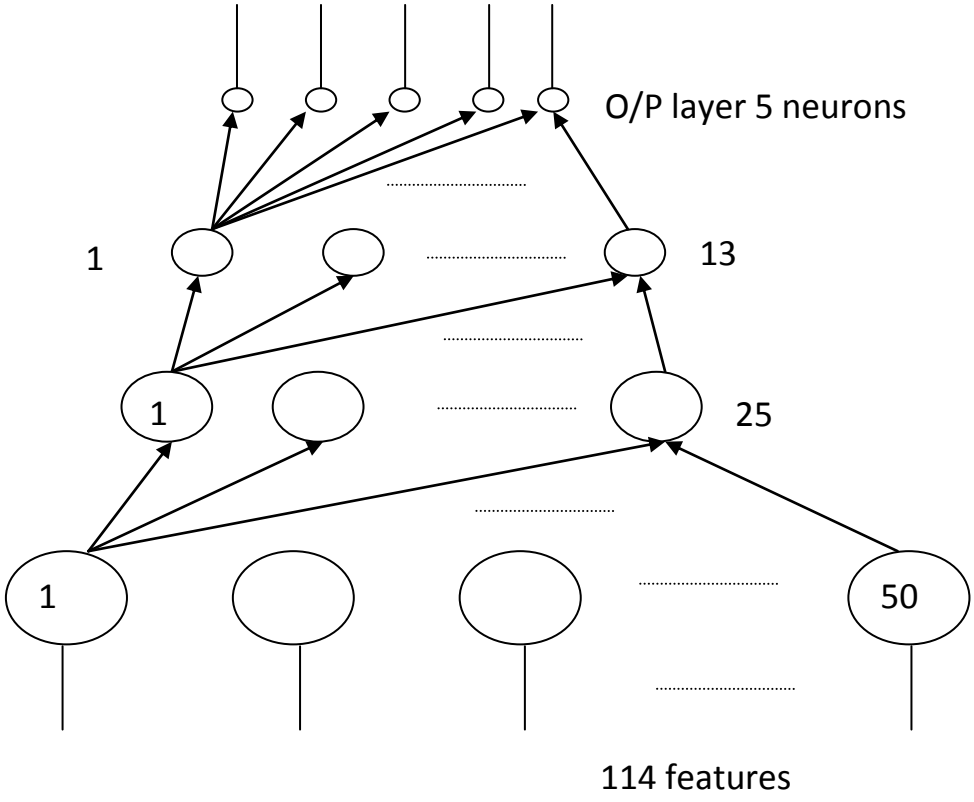

Figure 6-1 Five categories system neural networks architecture.

The suggested network architecture is trained using 10% version of the KDD 1999 Cup dataset (Table 5-1) in 40 minutes and 18 seconds, and tested using the KDD Corrected testing set (Table 5-4). The resulting confusion is matrix shown in Table 6-1.

	PRB	R2L	DoS	U2R	normal
PRB	97.17	0.00	2.02	0.02	0.80
R2L	4.02	70.02	15.04	0.00	10.92
DoS	0.00	0.00	100.00	0.00	0.00
U2R	0.00	38.46	0.00	0.00	61.54
normal	0.01	0.02	2.05	0.05	97.91

Table 6-1 Confusion matrix for BPNNIDS trained by 10% version of dataset (FCS).

The training of the neural networks is stopped at 20 epochs, with minimum MSE of 8.63806E-05, as shown in Figure 6-2, Table 6-2 and Table 6-3.

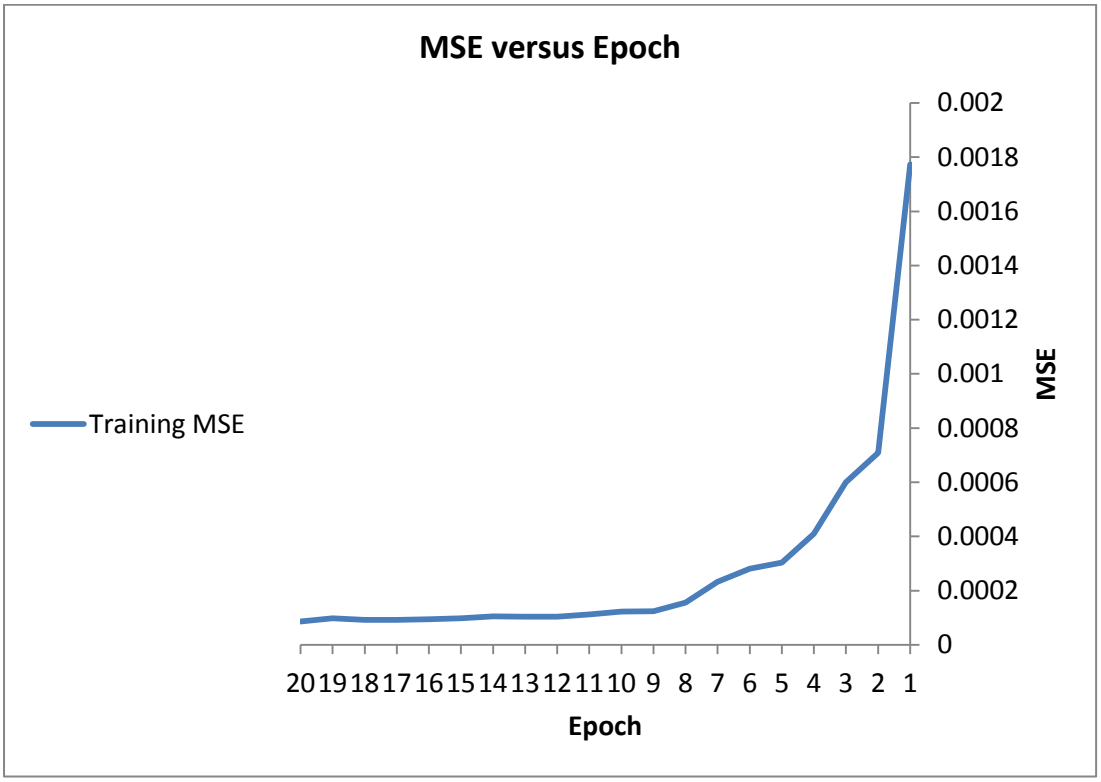

Figure 6-2 MSE versus Epoch (FCS 10 KDD).

Training	*Best Network*
20	Epoch #
8.63806E-05	Minimum MSE
8.63806E-05	Final MSE

Table 6-2 Minimum and Final MSE (FCS 10 KDD).

Training MSE
0.00177297
0.00070931
0.00060025
0.00040975
0.00030312
0.00028106
0.00023239
0.00015603
0.00012352
0.00012258
0.00011252
0.00010345
0.00010376
0.00010505
9.7735E-05
9.4642E-05
9.2704E-05
9.1803E-05
9.7917E-05
8.6381E-05
Minimum Training MSE
8.6381E-05

Table 6-3 MSE during each epochs (FCS 10 KDD).

The values at the diagonal of the matrix in Table 6-1 represent the correct detected records. So, the detection rate can be calculated from Table 5-4 and Table 6-1 by using Equation 6-1 as follow:

$$\text{dectection rate (DR)} = \frac{\text{correct detected attack}}{\text{total number of attack}}$$

Equation 6-1

$$= \frac{0.97 * 4166 + 0.7 * 16189 + 229853 + 0}{4166 + 16189 + 229853 + 228}$$

$$= 97.92\%.$$

Similarly, the values at the last row of the confusion matrix Table 6-1 represent the records detected to be normal. The first four values are obviously represented the percentage of the normal behavior that are misclassified to be an attack. The false positive rate is calculated as follow:

$$\text{FPR} = \ 0.01 + 0.02 + 2.05 + 0.05 = 2.13\%$$

6.3 Five categories system using the proposed reduced data set (FCS P KDD)

The suggested network architecture is trained using the thesis proposed reduced sample from 10% version of the KDD 1999 Cup dataset (Table 5-2) in 3 minutes and 27 seconds and tested using the KDD Corrected testing set (Table 5-4). The resulting confusion matrix is shown in Table 6-4 .

	PRB	R2L	DoS	U2R	normal
PRB	99.85	0.00	0.15	0.00	0.00
R2L	0.40	92.10	6.50	0.00	0.00
DoS	0.00	0.22	99.54	0.00	0.24
U2R	7.69	53.85	26.92	0.00	11.54
normal	0.00	0.00	0.40	0.00	99.60

Table 6-4 Confusion matrix for BPNNIDS trained by the thesis proposed reduced KDD dataset (FCS).

The training of the neural networks is stop at 22 epochs, with minimum MSE of 0.000531693, as shown in Figure 6-3, Table 6-5 and Table 6-3.

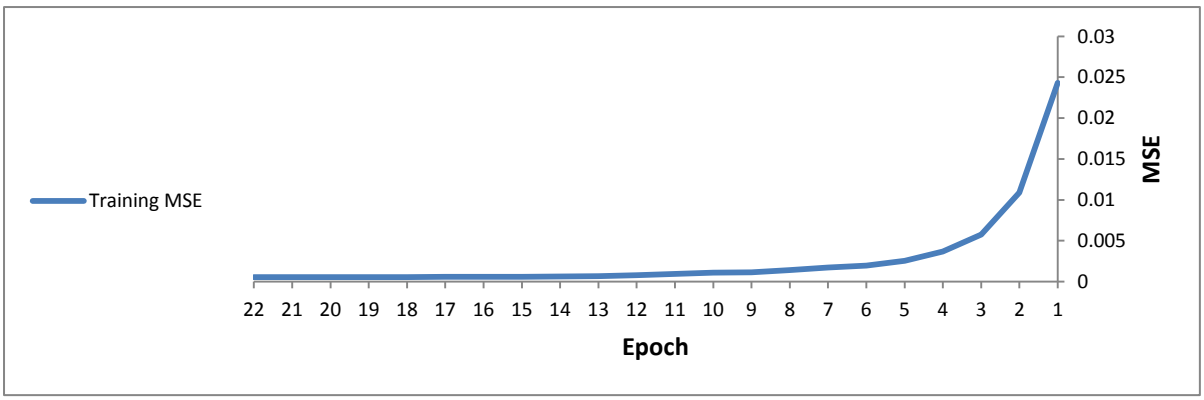

Figure 6-3 MSE versus Epoch (FCS P KDD).

Training	*Best Network*
22	Epoch #
0.000531693	Minimum MSE
0.000531693	Final MSE

Table 6-5 Minimum and Final MSE (FCS P KDD).

Training MSE
0.02429938
0.01089762
0.00573438
0.00366661
0.00253899
0.00197192
0.00172264
0.00139453
0.00113005
0.00110059
0.00095251
0.00079253
0.00066941
0.00062261
0.00060178
0.00058876
0.00057443
0.00056139
0.00055221
0.00054657
0.00053905
0.00053169
Minimum Training MSE
0.00053169

Table 6-6 MSE during each epochs (FCS P KDD).

Similarly, the DR can be calculated from Table 5-4 and Table 6-4 by using Equation 6-1 as follow:

$$= \frac{4159 + 14910 + 228795 + 0}{250436}$$

$$= 98.97\%.$$

Similarly, $\mathrm{FPR} = 0 + 0 + 0.4 + 0 = 0.4\ \%$.

6.3.1 Comparing Performance under the two sets (FCS)

	DR (%)	FPR (%)	Training time
10% KDD	97.92%	2.13 %	40 min 18 sec
The proposed Date set	98.97%.	0.4 %	3 min 27 sec

Table 6-7 Performance under the two sets (FCS).

By comparing the two results (Table 6-7), we can conclude that the thesis reduced data set will take an excellent training time only 3 minutes and 27 seconds with a better detection rate (DR) and false positive rate (FPR).

6.4 Two categories system using 10% version of KDD Cup 99 (TCS 10 KDD)

A four layer neural network is used. It has three hidden layers. The size of the input, three hidden layers and the output layer are 114, 50, 25, 15, and 2 respectively, 114 represent the number of features used in the training and 2 represent the number of categories, as shown in Figure 6-4

Similarly, we chose the number of hidden layers and the number of neurons in each layer as five categories system.

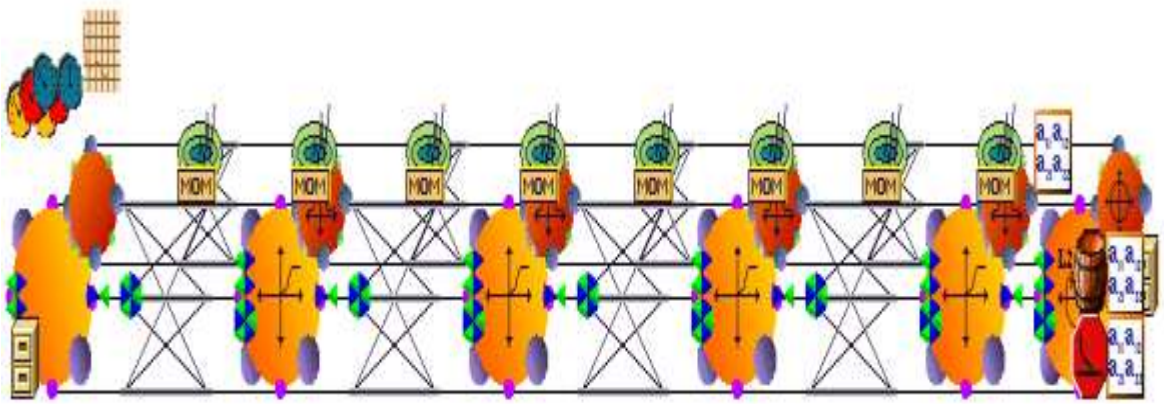

Figure 6-4 Two categories system neural network architecture.

The parameters used are:

- The mean square error (MSE) in the training step is 0.001, transfer sigmoid.
- Learning rule momentum.
- Step size 1.0.
- Momentum 0.7

The suggested network architecture is trained using 10% version of the KDD 1999 Cup dataset (Table 5-1) in 48 minutes, and tested using the KDD Corrected testing set (Table 5-4). The resulting confusion matrix is shown in Table 6-8.

	Attack	Normal
Attack	99.96	0.0398
Normal	2.124	97.876

Table 6-8 Confusion matrix for BPNNIDS trained by 10% version of dataset (TCS).

The training of the neural networks is stop at 21 epochs, with minimum MSE of 0.000204713, as shown in Figure 6-5, Table 6-9 and Table 6-10.

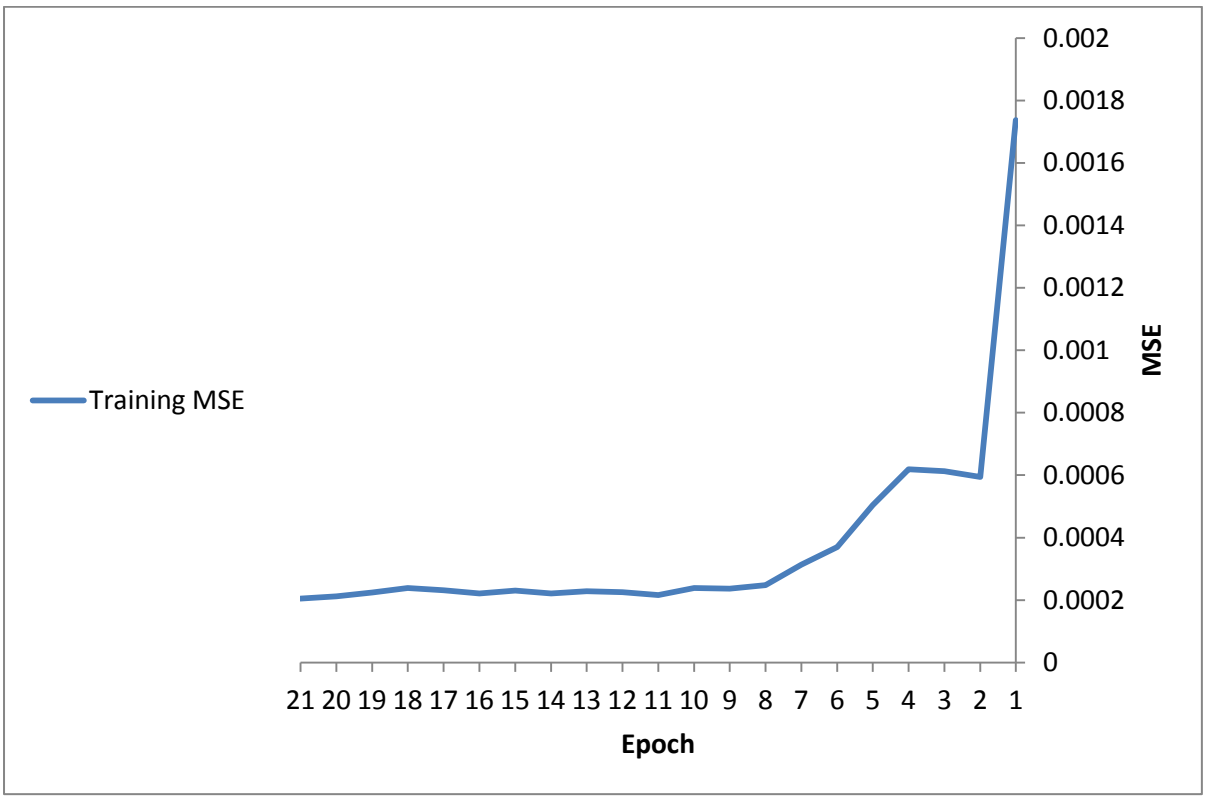

Figure 6-5 MSE versus Epoch (TCS 10 KDD).

Training	*Best Network*
21	Epoch #
0.000204713	Minimum MSE
0.000204713	Final MSE

Table 6-9 Minimum and Final MSE (TCS 10 KDD).

Training MSE
0.0017373
0.0005944
0.0006124
0.000619
0.0005041
0.00037
0.0003137
0.0002479
0.0002367
0.0002383
0.0002156
0.0002258
0.0002281
0.0002209
0.0002307
0.0002217
0.0002318
0.0002385
0.0002238
0.000212
0.0002047
Minimum Training MSE
0.0002047

Table 6-10 MSE during each epochs (TCS 10 KDD).

Similarly, the DR and FPR can be calculated from Table 5-4 and Table 6-8 by using Equation 6-1 as follow:

$$\mathrm{DR} = 99.96\ \%.$$

$$\mathrm{FPR} = 2.124\ \%.$$

6.5 Two categories system using the proposed reduced data set (TCS P KDD)

The suggested network architecture is trained using the thesis proposed reduced sample from 10% version of the KDD 1999 Cup dataset (Table 5-2) in 3 minutes 4 seconds, and tested using the KDD Corrected testing set (Table 5-4). The resulting confusion matrix is shown in Table 6-11.

	Attack	Normal
Attack	99.91	0.0929
Normal	1.5997	98.4003

Table 6-11 Confusion matrix for BPNNIDS trained by the thesis reduced KDD dataset (TCS).

The training of the neural networks is stop at 22 epochs, with minimum MSE of 0.000556344, as shown in Figure 6-6, Table 6-12 and Table 6-13.

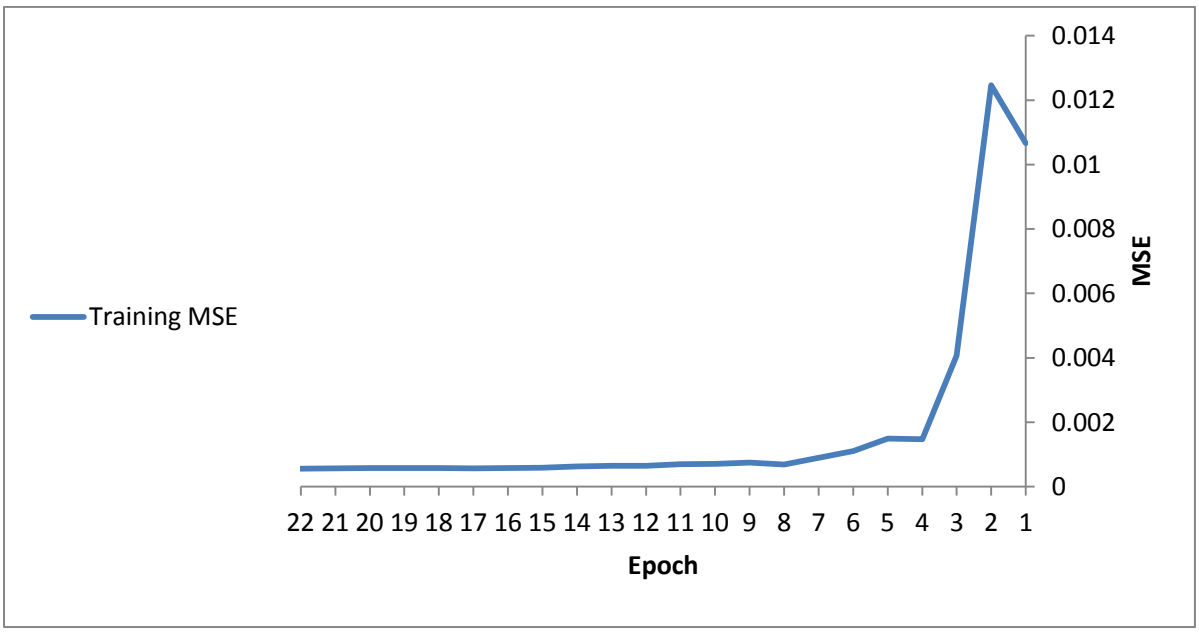

Figure 6-6 MSE versus Epoch (TCS P KDD).

Training	*Best Network*
22	Epoch #
0.000556344	Minimum MSE
0.000556344	Final MSE

Table 6-12 Minimum and Final MSE (TCS P KDD).

Training MSE
0.01066416
0.01246526
0.00406252
0.00147369
0.00149332
0.00110564
0.00089731
0.00068758
0.00074889
0.0007115
0.0006949
0.00065223
0.00064997
0.00062765
0.00058694
0.00057459
0.00057052
0.00057493
0.00057831
0.00057896
0.00057242
0.00055634
Minimum Training MSE
0.00055634

Table 6-13 MSE during each epochs (TCS P KDD).

Similarly, the DR and FPR can be calculated from Table 5-4 and Table 6-11 by using Equation 6-1 as follow:

$$\mathrm{DR} = 99.91\%.$$

$$\mathrm{FPR} = 1.5997\ \%.$$

6.5.1 Comparing Performance under the two sets (TCS)

	DR (%)	FPR (%)	Training time
10% KDD	99.96 %	2.124 %	48 minutes
The proposed Date set	99.91%	1.5997 %	3 min and 4 sec

Table 6-14 Performance under the two sets (TCS).

By comparing the two results (Table 6-14), we can conclude that the thesis proposed reduced data set will take an excellent training time only 3 minutes and 4 seconds with a better detection rate (DR) and false positive rate (FPR).

6.6 Alternative approach

6.6.1 Five categories system using the proposed reduced data set and RBF (FCS P KDD).

A four layer neural network (Figure 6-7) is used. It has one RBF layer, two hidden layers and one output layer. The size of the input layer, the RBF layer, two hidden layers and the output layer are 114, 114, 25, 13, and 5 respectively.

We first tried network architecture with the number of center less than the number of features. But, it did not converge to a solution. Then we increased it to be equal to the number of features. The number of neurons in each hidden layers is chosen by trial. We started with a size of the first hidden layer to be 20 neurons. Then, this size is increased until the optimum results are obtained. Similarly the size of the hidden two layers is chosen.

The parameters used for an unsupervised learning are:

- The number of centers is 144.
- Maximum epochs are 100.
- Termination – weight change is 0.001
- Learning rate is 0.01 to 0.001

The parameters used for the supervised learning are:

- The mean square error (MSE) in the training step is 0.001, Transfer sigmoid.
- Learning rule momentum.
- Step size 1.0.
- Momentum 0.7

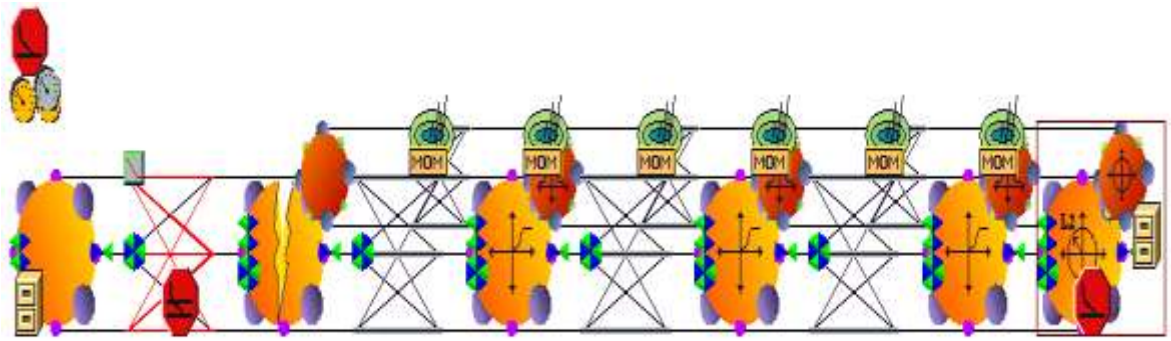

Figure 6-7 Five categories system RBF neural network architecture.

The suggested network architecture is trained using the thesis proposed reduced sample from 10% version of the KDD 1999 Cup dataset (Table 5-2) and we stopped the training process after 4 minutes when it exceeds the training time in BPNNIDS, and tested using the KDD Corrected testing set (Table 5-4), the resulting confusion is matrix shown in Table 6-15 .

	PRB	R2L	DoS	U2R	normal
PRB	65.90	0.00	34.10	0.00	0.00
R2L	0.50	30.50	69.00	0.00	0.00
DoS	0.88	0.00	95.02	0.00	4.10
U2R	0.00	0.00	100	0.00	0.00
normal	0.87	0.00	21.10	0.00	78.03

Table 6-15 Confusion matrix for RBF trained by the thesis selected KDD dataset (FCS) with force stop.

After that the suggested network architecture is trained using the thesis proposed reduced sample from 10% version of the KDD 1999 Cup dataset (Table 5-2) in 36 minutes 14 seconds without stopped the training process,

and tested using the KDD Corrected testing set (Table 5-4). The resulting confusion matrix is shown in Table 6-16.

	PRB	R2L	DoS	U2R	normal
PRB	71.01	0.00	28.99	0.00	0.00
R2L	0.42	36.58	63.00	0.00	0.00
DoS	0.05	0.00	96.95	0.00	3.00
U2R	0.00	0.00	100	0.00	0.00
normal	0.04	0.00	14.90	0.00	85.06

Table 6-16 Confusion matrix for RBF trained by the thesis selected KDD dataset (FCS) without force stop.

By comparing Table 6-15 and Table 6-16, we can note a small increase in the detection rate when leaving the network to finish training and reach the stopping condition.

By comparing Table 6-11, Table 6-15 and Table 6-16 , we can conclude that the thesis BPNNIDS will take an excellent training time only 3 minutes and 4 seconds (Table 6-14) with a better detection rate and false positive rate.

This thesis introduced two ways of machine learning in the supervised environment. The first one depends on the back-propagation neural networks and the second one depends on the radial basis function. By comparing the training time, the detection rate and the false positive rate it is concluded that the engine with the back-propagation neural networks produced better results than the one using the radial basis function.

We compare the performance of the thesis method BPNNIDS with some of the other neural-network-based approaches, such as K-means NN, SVM, SOM and PCA. For this purpose, we use the published results in [4; 6; 13; 23]. We compare the %DR and %FPR. Table 6-17 shows the experimental results. Some incomplete items in the published results are symbolized by '_'. The back-propagation neural networks intrusion detection system (BPNNIDS) achieves higher DRs and lower FPRs than all the other listed in less time.

Technique	DR	FPR	Time
K-means NN (Faraoun and Boukelif, 2006)	92%	6.21%	28 min 21 s
SVM (Esking et al., 2002)	98%	10%	-
PCA (RachidBeghdad, 2008)	93.83%	6.16%	26 min 56 s
SOM (Kayacik, H. G., Nur, Z.-H., & Heywood, M. I., 2007)	90.14 %	1.4%	-
BPNNIDS-FCS-10%	97.92%	2.13 %	40 min 18 s
BPNNIDS-FCS-P	98.97%	0.4 %	3 min 27 s

Table 6-17 Comparison with the previous work.