

Solar Storm Type Classification Using Probabilistic Neural Network compared with the Self-Organizing Map

Gregorius S. Budhi¹, Rudy Adipranata², Bambang Setiahadi³, Adrian H. N.⁴

^{1, 2, 4)} Information Engineering Department, Petra Christian University,
Siwalankerto 121 – 131, Surabaya

³⁾ Aerospace Observer Station (SPD) LAPAN,
Watukosek, Gempol, POBox 04, Pasuruan

E-Mail: greg@petra.ac.id, rudya@petra.ac.id, bambangsetiahadi@rocketmail.com

Abstrak

Salah satu tugas LAPAN adalah melakukan pengamatan dan peramalan gangguan badai matahari. Gangguan ini dapat mempengaruhi medan elektromagnet bumi sehingga mengacaukan peralatan elektronik dan navigasi yang ada di bumi. LAPAN menginginkan adanya aplikasi komputer yang secara otomatis dapat mengklasifikasi tipe badai matahari, yang menjadi bagian dari sistem early warning yang akan dibuat. Klasifikasi terhadap citra digital badai matahari / bintik matahari ini berbasis pada sistem klasifikasi "Modified - Zurich Sunspot Classification". Metode klasifikasi yang kami gunakan disini adalah Jaringan Saraf Tiruan Probabilistik. Hasil dari pengujian cukup menjanjikan karena memiliki akurasi sebesar 94% untuk data testing. Akurasi ini lebih baik dari akurasi aplikasi serupa yang dibangun dengan kombinasi metode *Self-Organizing Map* dan *K-Nearest Neighbor*.

Kata kunci: Klasifikasi Tipe Badai Matahari, Sistem Klasifikasi *Modified - Zurich Sunspot Classification*, Jaringan Syaraf Tiruan Probabilistik.

Abstract

*One of the task of the LAPAN is making observation and forecasting of solar storms disturbance. This disturbances can affect the earth's electromagnetic field that disrupt the electronic and navigational equipment on earth. LAPAN wanted a computer application that can automatically classify the type of solar storms, which became part of early warning systems to be created. The classification of the digital images of solar storm / sunspot is based on "Modified - Zurich Sunspot Classification System". Classification method that we use here is the Probabilistic Neural Networks. The result of testing is promising because it has an accuracy of 94% for testing data. The accuracy is better than the accuracy of similar applications we've built with a combination of methods *Self-Organizing Map* and *K-Nearest Neighbor*.*

Key words: *Solar Storm Type Classification, Modified - Zurich Sunspot Classification, Probabilistic Neural Network.*

INTRODUCTION

Disruption from the sun that recently known as solar storm is strongly concerned. The disruption can affect the Earth's electromagnetic field, in addition to others dangers that can be happening. Today, human life is highly dependent on modern technologies especially in the field of electrical, electronics, computer and navigate the sea and air, that are susceptible to interference from electromagnetic field changes. Therefore, the emergence of 'interference' from solar storms should be anticipated quickly and precisely.

Actually the solar storm term is a great explosion (Flare) and Coronal Mass Ejection / CME on the surface of the sun. Before the famous term solar storm, this two type of disruptions is more commonly called a group of sunspots, because that's the kind seen in the image of the sun taken by the Michelson Doppler Image instruments (MDI) Continuum / SOHO (Solar and Hemispheric Observatory). Sunspots evolved from a tiny spot with low activity evolved into a very complex configuration with the possibility of having high activity, which issued a mass explosion and hurling corona. To determine the level of complexity of sunspots and sunspot activity Modified - Zurich classification is used in order to classify the types of sunspots.

This study is part of a larger study entitled "Automated Sunspot Group Classification for Analyzing Space Weather Conditions" we've done in collaboration with researchers from the National Aeronautics and Space Agency (LAPAN). Classification of sunspots groups is based on Modified - Zurich Sunspot Classification system. In this paper we publish some of the results of our study that is the use of probabilistic neural network to the classification of sunspots groups / solar storms type, which looks at the digital image of the sun.

Before the classification process, it is necessary proceeding digital image enhancement of the sun images. It should also detect the location of sunspots and sunspot groupings. All of our process has been published previously [1, 2].

For the testing, we also compared the results of this sub-system testing with other classification sub-systems that we have made

previously by using a combination of artificial neural network method Self-Organizing Map and data mining methods K-Nearest Neighbor [3]. This comparison purposes so that we can obtain the most appropriate classification method to be used in the classification system of solar storms that will be created.

Sunspot Classification

Sunspots are the intersection of the solar magnetic flux tube with the photosphere. Sunspots appear dark because the magnetic field has the effect of sunspot cooling so that the temperature is lower than the surrounding area. The magnetic field of sunspots proved to be a source of disturbance energy from the sun, such as explosives, or hurling coronal mass [4]. The latter phenomenon known as solar storms can disrupt space environment around the Earth and technology facilities and life on the Earth's surface [5, 6, 7].

Because of the sunspot's magnetic field is the intersection of the tube with the photosphere, the appearance of sunspots begins with a small black spot. Over time, if the tube magnetic field is constantly out of the sun due to the buoyant force, it would appear two spots with opposite magnetic polarity (bipolar). Sunspots evolved into more complex configurations, in example, the number of spots and area increases. The more complex a configuration of sunspots are, the greater the possibility of instability magnetic field that triggers a flare or CME events [8, 9].

Stages of the evolution of sunspots is expressed in the "Modified - Zurich Sunspot Classification", namely class A, B, C, D, E, F (the level of complexity increases) and then gradually decreased until it became a class H. Image of the sunspot evolution classification found in Figure 1 [10].

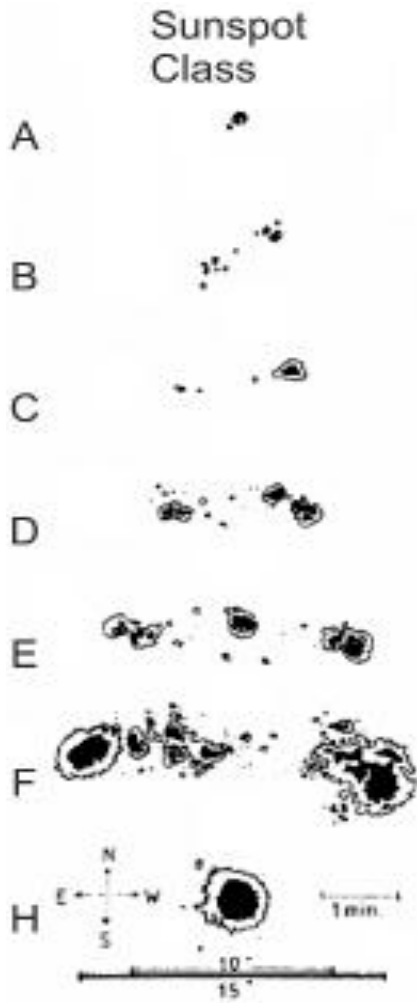


Figure 1. Modified - Zurich Sunspot Classification

Artificial Neural Networks

Artificial neural network is a processing model that is based on the workings of the human brain [11]. Just like the human brain, neural networks composed of cells interconnected neurons. Each cell is a neuron receives input signals are processed to produce an output signal. Neural networks can learn to like the human brain by giving weight for each input on neuron cells. By using weights, the neural network can learn a given input.

Probabilistic Neural Network

Probabilistic Neural Network (PNN) is a method of artificial neural networks using the principles of statistical theory Bayesian Classification to replace heuristic principles used in the Back-propagation algorithm [12].

That is why PNN is usually used to perform pattern classification [13].

Architecture of the PNN consists of four layers, namely input layer, pattern layer, summation layer and decision layer / output layer as can be seen in Figure 2 [13].

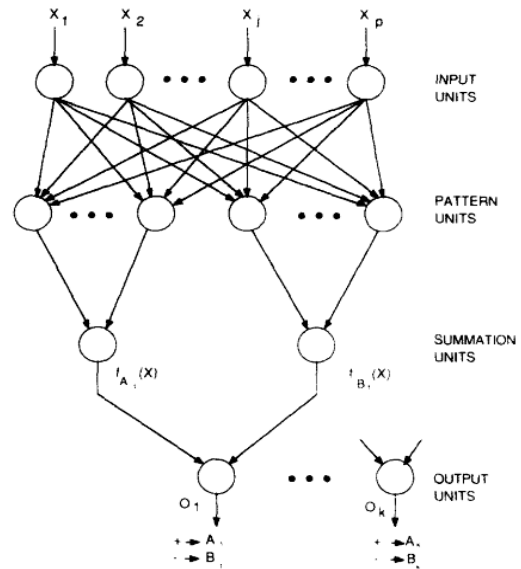


Figure 2. Detail Architecture of PNN [13]

The input layer does not perform any calculations, just transfer the data input to each neuron in the pattern layer. Each neuron in the pattern layer will perform probability calculations (distance) between the inputs to the data stored in the pattern layer neurons. Furthermore, the summation layer will receive input from neurons of pattern layer and sum it. And last, the output layer will produce classification results based on the results of the summation neuron that has the greatest value. In general, probabilistic neural network architecture is shown in Figure 3.

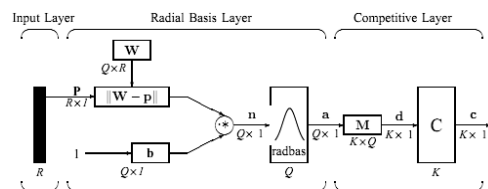


Figure 3. PNN Architecture [14]

Here is the training algorithm of probabilistic neural network [14]:

1. Initializing initial weight (radial basis layer weight) W obtained from the transpose of the matrix $R \times Q$ where R is the dimension of input and Q is the number of training data.

2. Initializing constant spread (s) of PNN.

3. Initializing the bias weight (b).

$$b = \frac{\sqrt{-\ln 0.5}}{s} \quad (1)$$

4. Initializing final weight (competitive layer weight) M which is a matrix of size $K \times Q$ where K is the number of the classification results and Q is the number of training data. I -th row of the matrix M represents the i -th training data and the j -th column matrix values will be worth 1 if the training data into the group, otherwise it will be 0.

5. Calculate the distance between the vectors of input data P with vector of each row in the initial weight W (Euclidean Distance between vectors P with vectors W_i) resulting distance matrix $||W - P||$ that have $Q \times 1$ dimension.

6. Calculating the activation value of the distance between the initial weight of the input data (vector a), using $\text{radbas}()$ function.

$$\text{radbas}(n) = e^{-n^2} \quad (2)$$

$$n = ||W - P|| * b \quad (3)$$

7. Multiplying vector a by matrix M to produce output vector d .

8. Looking for an output of PNN with competitive function C where the competitive function will locate the largest value in the vector d . Index of greatest value is to indicate the results of the classification of data input P .

9. Save the initial weight, spreads, and the final weight.

Min - Max Normalization

Min-max normalization is a method of normalization that transforms the data in a linear fashion into a new range [15]. Formula of this method is shown in Equation 4.

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (4)$$

DESAIN OF SOLAR STORM CLASSIFICATION SUB-SYSTEM

Probabilistic Neural Network Architecture used

Probabilistic neural network architecture used is shown in Figure 4.

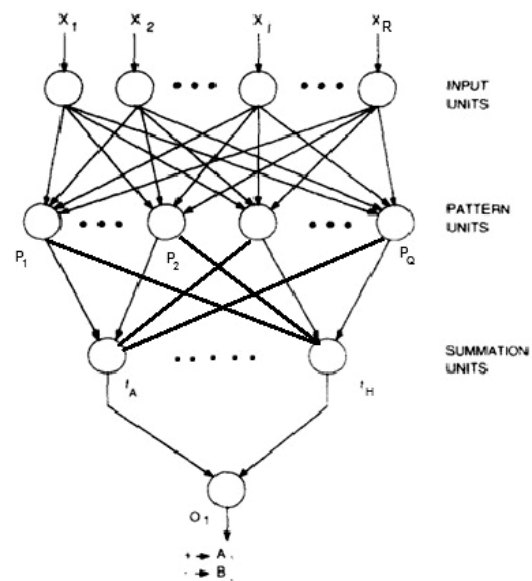


Figure 4. PNN architecture used

For the number of input units will be obtained from the large dimension of the data input (R), the processed data from the database input diameter, area, perimeter, number of sunspots each group, and the average black pixel for each region of the input image. After that, each input unit is connected to all the pattern units amounting to a weight Q , where Q is the amount of training data. So the initial weight matrix size $Q \times R$.

Each pattern unit will generate a value of the distance between the weight and the input, the result is a matrix dimension $Q \times 1$ which will then be used for further calculations using $\text{radbas}()$ function.

Each pattern unit will be connected with all summation units through the final weight, which is a matrix of size $K \times Q$, where K is the number of the classification results, which is a class A, B, C, D, E, F, and H. The magnitude

of Q is the amount of training data. I -th row of the matrix M represents the i -th training data and the j th column matrix values will be worth 1 if the training data into the group, otherwise it will be 0.

Then each summation unit will be connected to an output unit whose function is to seek the greatest value of summation unit and take the index of the summation unit of as a result of the classification.

Sub-System Design

In Figure 5, 6 and 7 can be seen a block diagram and flowchart of the design sub-systems that are built for the solar storms type classification using probabilistic neural network method.

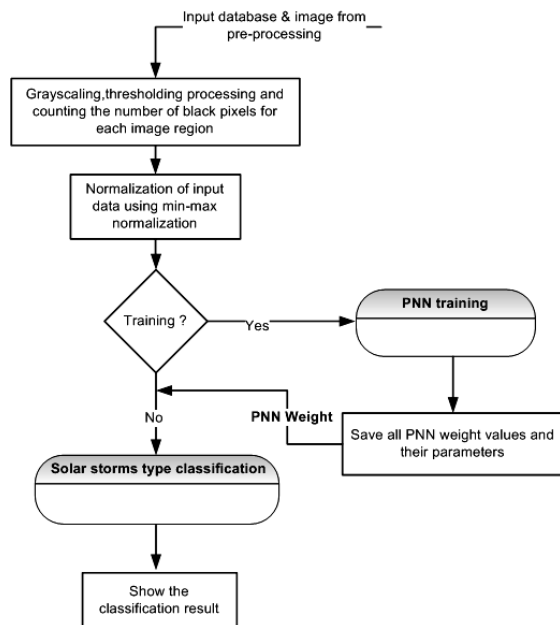


Figure 5. Block diagram of solar storm type classification sub-system using PNN

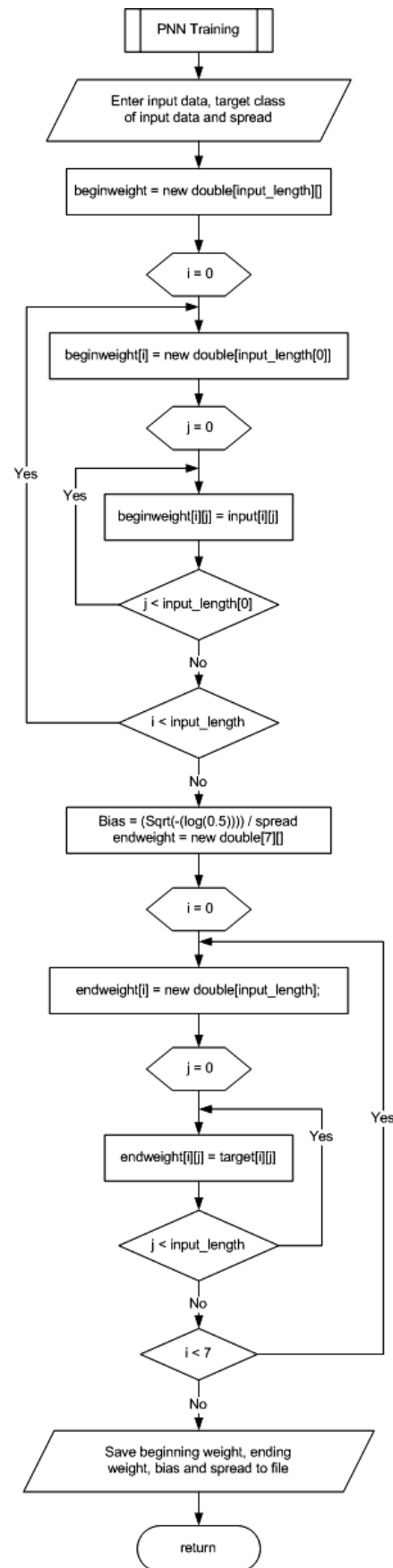


Figure 6. Flowchart of the PNN Training process

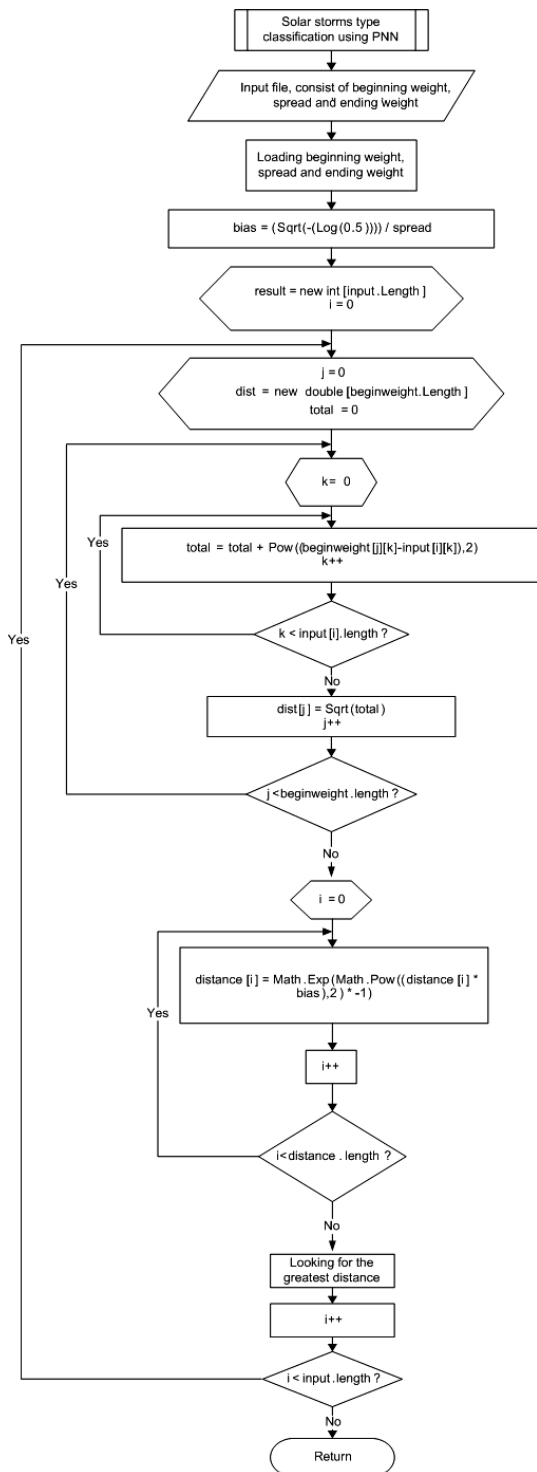


Figure 7. Flowchart of the solar storm type classification process using PNN

Input from the sub-system is derived from the sub-system that we created earlier and have been published [1, 2]. Any data that goes into the sub-type classification system solar storm consists of 4 kinds of values and a digital image. It consists of four different values: area, perimeter, diameter, and the

number of spots available to a group of sunspots. Meanwhile, the input digital image is a digital image of the solar storm as shown in Figure 8. Furthermore, the solar storm digital image is converted to data that can be received by the input neurons of PNN. To change, the image is processed into black and white image using the grayscale processing and thresholding method. After becoming a black-and-white image, it is proceeded to divide of regions in the image, where the number of regions specified by user. Examples of the process can be seen in Figure 8. Furthermore, the number of black pixels in each region is counted.



Figure 8. Examples of solar storm digital image and the process of region dividing

Next is the process of normalization of all the input data, using the Min - Max Normalization formula. These data are the area, perimeter, diameter, number of spots and values derived from counting the black pixels of each region. The values that have been normalized are used for input neurons PNN.

Once all the input data has been prepared, the user can select the process to be executed, namely: ANN training or classification. For training the neural network, as in general the process of training the PNN method, all sample data will be processed to update all the weight until convergent. After converging, the weight is stored and can then be used in the classification process. The process of training for a sample data is shown in Figure 6.

For the classification process, input required is a set of PNN weights that have been convergent, PNN architecture used during training, and image the solar storms that will be classified. With passing the solar storms image to be classified in the sub-system of PNN (with a weight that has converged), we will get the value of the classification based on Modified - Zurich Sunspot Classification, which is a class A, B, C, D, E, F or H. The process of classification for a data can be seen in Figure 7.

From the results of this classification can be determined quickly whether a solar storm (sunspot groups) is dangerous or not. Class types to be concerned are D, E and F.

RESULT AND DISCUSSION

There is some testing that has been done on the sub-systems, namely:

a. Testing the speed of the training and qualification process. It would be tested with a maximum total data that we have (214 kinds of sample data). Test results state that each process has a speed less than 0.1 milliseconds. It can be concluded that in terms of processing speed, the sub-system is very good.

b. Tests with constant variation spread of PNN. The test is performed with the number of training data = 150 data and digital images of solar storms that is divided into 5 x 5 regions. The test results can be seen in Table 1. So, we could conclude that the higher spread value will be, the decrease of classification accuracy value would be produced.

Table 1. The constant variation spread testing

<i>Spread</i>	classification accuracy of the training data	classification accuracy of the testing data
0.2	97%	94%
0.5	85%	89%
0.9	72%	86%

Note: The testing data are samples data that did not participate in the training process.

c. Tests on the amount of training data variation. Setting the test is spread = 0.2 and a digital image of solar storms is divided into 5 x 5 regions. The test results can be seen in Table 2. From the results it can be concluded that if we tested with the data had been trained, the accuracy of sub-systems did not have a clear pattern. However, when tested with data testing, it can be seen that the more training data is used, the higher the accuracy of the sub-systems.

Table 2. Testing variations in the amount of training data

The number of training data	classification accuracy of the training data	classification accuracy of the data testing
50	98%	76%
100	96%	86%
150	97%	94%

d. The next test is to vary the number of regions of the solar storm digital image that are processed. The test is performed with the number of training data = 150 data and the constant spread = 0.2. The test results are shown in Table 3. The test results stated that for the classification of the training data, the best regions number is 7x7. However, if the classification performed on the data testing, the best regions number is 5x5.

Table 3. Regions number variations testing

The number of region	classification accuracy of the training data	classification accuracy of the data testing
3 x 3	88%	92%
5 x 5	97%	94%
7 x 7	99%	91%

Further, we will compare this sub system with similar sub-system that we have built previously used a combination of methods Self-Organizing Neural Network and K-Nearest Neighbor [3]. Testing variations spread constants cannot be compared because the constant spread only on PNN and not used in the SOM. In comparison of the speed of the training process, the PNN method was much better. Speed the process of training the SOM-KNN is still above 1 second for as many as 50 pieces of training data, while the PNN has speed training under 0.1 milliseconds for maximum training data (214 training data). Furthermore, the results of comparison of the classification accuracy of PNN and SOM-KNN can be seen in Figure 9 and 10. These accuracy comparisons are performed only on the data testing (sample data that not used during training).

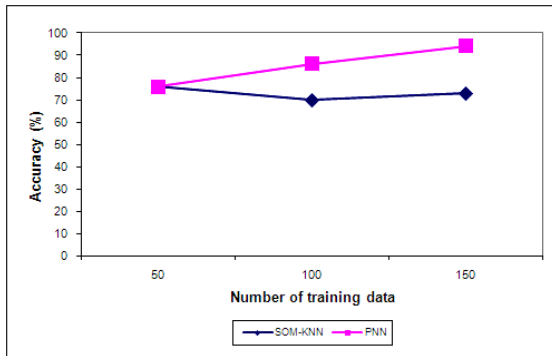


Figure 9. Comparison of classification accuracy for variation in the amount of training data with the distribution of digital image region = 5 x 5

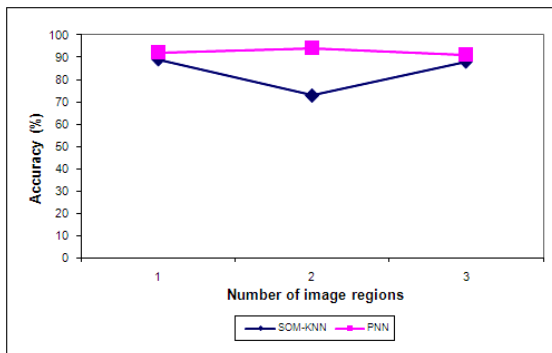


Figure 10. Comparison of classification accuracy for variations of digital image regions with the number of training data = 150 data

From the results of the comparison between the solar storms type classification sub-systems built with PNN and SOM-KNN, it appears that the performance of the PNN sub-system was better.

CONCLUSION

Solar storm type classification sub-system using Probabilistic Neural Network has been created properly. It can be seen from the rapid process of training and classification, as well as the high level of classification accuracy. Besides that the sub-system has also proven to have a better performance compared to the similar sub-system that has been built previously using a combination of Self-Organizing Map and K-Nearest Neighbor methods.

However, we have not decided yet whether the sub-system will be used or not. The test results must be compared with the results of similar system test that we built using Feed-Forward Back-propagation method. From the comparison result, analysis and further testing, we will determine which sub-system to use.

ACKNOWLEDGMENT

This research was funded by Research Competitive Grant DIPA-PT Kopertis East Java region VII (0082/SP2H/PP/K7/KL/IV/2011), entitled "Automated Sunspot Group Classification for Analyzing Space Weather Conditions".

DEDICATION

We dedicate this paper to a co-researcher from the National Aeronautics and Space Agency (LAPAN), namely: Dr. Bachtiar Anwar M.Sc. who had died of illness at the end of 2011.

REFERENCES

- [1] Adipranata R, Budhi GS, Setiahadhi B, dan Anwar B, Segmentation of Sunspot Group using Watershed Method. Proceeding of National Conference of System and Informatics, Bali: 332-337, 2010.
- [2] Budhi, GS, Adipranata R, Sugiarto M, Anwar B, dan Setiahadhi B, Sunspot Grouping of Solar Digital Image using DBSCAN Clustering Method. Proceeding of National Seminar of Information Technolgy Application, Yogyakarta: F81 – F85, 2011.
- [3] Budhi GS, Adipranata R, Setiahadhi B, Anwar B, Hartanto A, Tjondrowiguno AN, Combination of Self-Organizing Map Neural Network and K-Nearest Neighbor for Automatic Classification of Sunspot Group Image. Proceeding of National Conference of System and Informatics, Bali: 369 - 374, 2011.
- [4] Setiahadhi B, Sakurai T, Miyazaki H and Hiei E, Research on Magnetohydrodynamic Transport Phenomena in Solar-Terrestrial Space at LAPAN Watukosek 2006, Proceeding of Seminar Antariksa Nasional III Pusat Pemanfaatan Sains Antariksa, LAPAN Bandung: 17 – 34, 2006.
- [5] Boteler DH, Space Weather Effects on Power Systems. Space Weather. Geophysical Monograph, vol. 125: 347 - 355, 2001.
- [6] Bothmer V and Daglis IA, Space Weather, Physics and Effects, Chicester UK: Springer - Praxis Publishing, 2007.
- [7] Lanzerotti LJ, Space Weather Effects on Technologies. Space Weather, Geophysical Monograph, vol. 125: 11 - 22, 2001.
- [8] Anwar B, Acton LW, Makita M, Hudson HS, McClymont AN and Tsuneta S, Rapid

- Sunspot Motion During A Major Solar Flare. *Solar Physics*, vol. 147: 287 - 303, 1993.
- [9] Setiahad B, Problems of Equilibria and Instabilities on Solar Coronal Magnetic Fields and Its Evolution Towards Energetic Energy Liberation: Effect to Interplanetary Space, *Proceeding of Seminar Nasional Matematika, FMIPA UNDIP*, E1: 1 - 7. 2005.
- [10] McIntosh PS, The Classification of Sunspot Groups, *Solar Physics*, vol. 125 no.2: 251 - 267, 1990.
- [11] Negnevitsky M, *Artificial intelligence: A guide to intelligence systems* 2nd Edition, New York: Addison Wesley, 2005.
- [12] Specht DF, Probabilistic Neural Network, *Neural Networks*, vol. 3 no. 1: 109 – 118, 1990.
- [13] Fausett L, *Fundamentals of neural networks: Architectures, algorithms, and applications*, New Jersey: Prentice - Hall International, 1994
- [14] Wu SG, Bao FS, Xu EY, Wang YX, Chang YF and Xiang QL, A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network, *IEEE 7th International Symposium on Signal Processing and Information Technology*, Cairo :12 – 17, 2007.
- [15] Han J, Kamber M and Pei J, *Data Mining: Concept and Techniques* 3rd Edition, Waltham USA: Morgan Kaufmann Publishers, 2012.