Research Article

A Clustering Application Scenario Based on an Improved Self-Organizing Feature Mapping Network System

Qian Cao [iD]

College of Information Engineering, Chaohu College, Chaohu 238000, China

Correspondence should be addressed to Qian Cao; 19875069@qq.com

Categorizing national football teams by level is challenging because there is no standard of reference. Therefore, the self-organizing feature mapping network is used to solve this problem. In this paper, appropriate sample data were collected and an appropriate self-organizing feature mapping network model was built. After training, we obtained the classification results of 4 grades of 16 major Asian football national teams. As for the classification results, it is different to normalize the input data and not to normalize the input data. The classification results accord with our subjective cognition, which indicates the rationality of self-organizing feature mapping network in solving the classification problem of national football teams. In addition, the paper makes a detailed analysis of the classification results of the Chinese team and compares the gap between the Chinese team and the top Asian teams. It also analyses the impact of the normalization of input data on the classification results, taking Saudi Arabia as an example.

1. Introduction

As the biggest sport in the world, football is widely popular in the world, and all the national football teams in the world are fighting for the honor of their country in the international arena, so the level of the national team is widely concerned. However, there has always been disagreement about which level a football team should belong to. In order to objectively and fairly reflect the actual ranking of a football team, the neural network is used to analyse the level of each football team [1–4].

The main contributions of this research are mainly reflected in the following three aspects:

(i) We build a model based on improved self-organizing feature mapping network with the aim to cluster teams more reasonably. To make it clear, we give the specific model parameters and build process, and it includes the collection of data sets, the division of data sets, the normalization and inverse normalization of data, the selection of model parameters, the determination of training functions, the determination of error tolerance, the determination of the number of iterations, the completion of test experiments, and the analysis of results.

(ii) In order to reflect the latest level situation of teams, we trained and tested the proposed model with the latest data set, namely, the last eight major international competitions.

(iii) We quantitatively analysed the performance of the model with a variety of mathematical tools and error analysis methods.

The rest of this paper is organized as follows. Section 1 reviews and summarizes the related work, on this basis, to clarify the significance of this study. In Section 2, the motivation of this research is expounded. Section 3 is preliminaries. In Section 4, the overall scheme of neural network modeling is proposed. In Section 5, an experiment

is designed and carried out and the results of the classification of 16 football teams were obtained. Finally, Section 6 concludes this paper.

2. Motivation

The Chinese national football team carries the expectations of hundreds of millions of Chinese fans. However, the Chinese national football team has performed poorly in recent years. So, where does China's national soccer team rank in Asia? Some think it belongs to the Asian second-tier team, others think it belongs to the fourth-tier team. With the help of historical data and the self-organizing feature mapping network, we can make an objective judgment on the level of the Chinese team. We not only hope to accurately reflect the real level of the Chinese team but also hope to verify the rationality of the algorithm.

3. Preliminaries

Self-organizing feature mapping, namely, self-organizing feature mapping network (SOFM or SOM), was proposed by Finnish neural network expert Kohonen in 1981. The biological basis of SOM is as follows. (1) The biological basis of lateral inhibition is like the lateral inhibition between nerve cells which brings out the competition, a degree of excitement from which the strongest nerve cells have obvious inhibitory effect on the peripheral nerve cells, and the excitement in peripheral nerve cells decreases as a result; thus, the neural network is the "winner" of the competition and other nerve cells fail in the competition. (2) When the biological neural network receives specific spatial and temporal information from the outside world, the specific region of the neural network is excited, and similar external information is continuously mapped in the corresponding region. After training, the competing layer neurons of SOM are close to each other with similar functions and far from each other with different functions, which is very similar to the tissue structure of biological neural network.

Each input pattern of self-organizing feature mapping corresponds to a localized region on a two-dimensional grid, and the location and properties of the region vary with the different input patterns. Therefore, there must be a sufficient number of input patterns to ensure that all neurons in the grid are trained and that the self-organizing process converges correctly. An important feature of SOFM is its topological conformal property, that is, the resulting feature map described by the output weight vector can reflect the distribution of the input pattern.

The basic principle of SOFM is that when a certain type of mode is input, a node in the output layer wins by getting the maximum stimulus, and the nodes around the winning node are also stimulated by lateral action. At this time, the network performs a learning operation, and the connection weight vector of the winning node and the surrounding nodes is modified in the direction of the input mode. When the category of input pattern changes, the winning node on the two-dimensional plane also moves from the original node to other nodes. In this way, the network uses a large number of sample data to adjust its connection weight through self-organization, and finally, the network output layer feature graph can reflect the distribution of sample data [3–7].

The SOFM network is a two-layer network consisting of an input layer and an output layer. The output layer establishes the topology of the network to better simulate the phenomenon of lateral inhibition in biology [6]. Figure 1 shows a simple SOFM network in which the output layer is a two-dimensional topology. Of course, the output layer of the SOFM network can also be a higher-dimensional topology. In SOFM networks, input and output neurons are connected by weights, and neighboring output neurons are also connected by weights. The transfer function of the output neuron is usually a linear function, so the output of the network is a linear weighted sum of the input values, as shown in the following formula:

$$Y_j = f\left(\sum_i x_i w_{ij}\right), \tag{1}$$

where w_{ij} represents the weight value, x_i is the input value, and Y_j is the output value [8, 9].

4. Supposed Model

Actually, there are other methods of classification by the level that do exist. In general, classification by machine learning method needs to be defined in advance. However, this method does not apply to the classification of football teams because the classification is supervised learning, that is, the classification of certain football teams must be specified, and then, the other football teams must be evaluated on the basis of those football teams. That is, we have to have standards first, but because of the uncertainty of the football game, it is difficult to find some football teams in the football world as standards. Even the top teams lose sometimes. Even top teams sometimes lose games, and if this top team is used as the standard, the results will be inaccurate.

Therefore, we must consider using the unsupervised clustering method. Self-organizing feature mapping network, as a good unsupervised clustering method, is applied to our research. In this way, as long as the number of categories N that need to be classified is set, the algorithm will convert all samples to N according to the principle of similarity.

This study intends to classify 16 major Asian football national teams. In order to complete this classification, we need to set the number of categories. If the division is too detailed, many teams may be classified into a single category, which is of little significance. If the division is too thick, such as only two categories lose the meaning of classification, so it would be a reasonable choice to divide the 16 teams into four categories.

Since the category number is 4, the competition layer will be set to a 2×2 hexagon structure in the self-organizing feature mapping network.
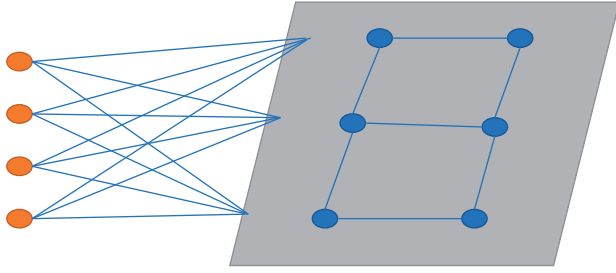
Figure 1: A SOFM network.

4.1. Sample Set Construction. Good classification experiment needs to be supported by good samples. There is no doubt that the results of the World Cup and the Asian cup can fully reflect the real level of a football team. Therefore, we selected the results of the 16 football teams in four international competitions, namely, the 2006 World Cup, the 2010 World Cup, the 2007 Asian cup, and the 2011 Asian cup, as samples. The results of the 16 football teams in the four international competitions are shown in Table 1. In the experiment, the sample data of each team can be represented by an eight-dimensional vector:

$$x = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]. \qquad (2)$$

The components of the vector represent its score in the World Cup 2006, the World Cup 2010, the Asian cup 2007, the Asian cup 2011, the World Cup 2014, the Asian cup 2015, the World Cup 2018, and the Asian cup 2019. Take the Chinese team as an example, the sample data of the Chinese team can be represented by the following 8-dimensional vector:

$$x = [43, 43, 9, 9, 43, 7, 33, 8]. \qquad (3)$$

The scores of each team shown in Table 1 are calculated according to the following rules.

Firstly, for the Asian cup 2007 and Asian cup 2011, if a team reaches the final four, its final ranking is its score. If a team reaches the last eight, its score is 5. If a team reaches the last 16, its score is 9. If a team does not reach the final stage of the Asian cup, its score is 17. For the Asian cup 2015 and Asian cup 2019, we use the official final league table as a score for a team.

Secondly, for the World Cup 2006, World Cup 2010, and World Cup 2014, if a team makes it to the finals, its score is its actual ranking in the finals. For a football team that did not make it to the finals, there are two situations. If it enters the top 10 of the qualifiers, then we consider its score to be 33, and if it does not enter the top 10 of the qualifiers, then its score is 43. For the World Cup 2018, if a team makes it to the finals, its score is its actual ranking in the finals. For a football team that did not make it to the finals, there are two situations. If it enters the top 12 of the qualifiers, then we consider its score to be 33, and if it does not enter the top 12 of the qualifiers, then its score is 45.

4.2. Construction of Elman Neural Network. Once we have the data, we can design the experiment. The process of the experiment is first to build the SOFM model, then to train, and finally to test. See Figure 2, for details [10–14].

For Model Creation, the selforgmap function in the matlab neural network toolbox can be directly used to create. The size of the competition layer of the model can be set to 2×2. So, the matlab code to create the model is as follows:

$$\text{Net} = \text{selforgmap} \left([2 \times 2] \right). \qquad (4)$$

Figure 3 shows the constructed Elman network structure.

5. Experiments

5.1. Experiments without Normalization. As for the test, the training data and the test data are the same, that is, the sample data. Therefore, 16 football teams can be classified by inputting the sample data into the model. Figure 4 is a matlab screenshot of the test results, and it clearly shows the categorization of the 16 teams. Among them, we find that the Chinese team belongs to the Asian third-tier team, indicating that the strength of the Chinese team is not satisfactory.

From this classification, we can see that the top teams in Asia are Japan, Korea, Iran, and Australia. In fact, this is in line with the actual situation, Japan, Korea, and Australia have reached the World Cup finals four times, and Iran has reached the World Cup finals three times.

In general, the selforgmap function tends to divide categories with more elements into finer categories so that categories with fewer elements may therefore merge with other categories so that each category tends to have the same number of elements. Saudi Arabia is the only second-tier team, and it shows that this level is not like other teams.

The Chinese team is classified as the 3rd tier Asian team. In fact, this is in line with the actual situation; after all, the Chinese team's performance in recent years is really very bad, and this is a well-known fact.

In order to see the level of the Chinese team more directly, we drew together the results of the Chinese team and four top Asian teams in the World Cup and compared them. As shown in Figure 5, there are 5 curves in the figure, representing 5 teams. The blue curve at the top represents the Chinese team. We can clearly see that there is a clear gap between the Chinese team and the other 4 teams, while the other 4 curves are intertwined, indicating that the level of the 4 first-class teams is very close. This also fully demonstrates the accuracy of the model we built for this classification.

If the results of the Asian cup are included, the conclusion is still the same. As shown in Figure 6, the five curves in the figure represent the five teams. The blue curve representing the Chinese team is still high, while the other four curves are intertwined. It shows that the gap between the level of Chinese team and the first-class team in Asia is

TABLE 1: Sample data (the results of the 16 football teams in the four international competitions).

| Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
China	43	43	9	9	43	7	33	8
Japan	28	9	4	1	29	5	45	2
Korea	17	15	3	3	27	2	19	5
Iran	25	33	5	5	28	6	18	3
Saudi Arabia	28	33	2	9	43	10	26	14
Iraq	43	43	1	5	33	4	45	14
Qatar	43	33	9	5	33	13	33	1
United Arab Emirates	43	33	9	9	43	3	33	4
Uzbekistan	33	33	5	4	33	8	33	10
Thailand	43	43	9	17	43	17	33	13
Vietnam	43	43	5	17	43	17	45	5
Oman	43	43	9	17	33	12	45	16
Bahrain	33	33	9	9	43	11	45	12
North Korea	33	32	17	9	43	14	45	24
Indonesia	43	43	9	17	43	17	45	25
Australia	16	21	4	2	30	1	30	5
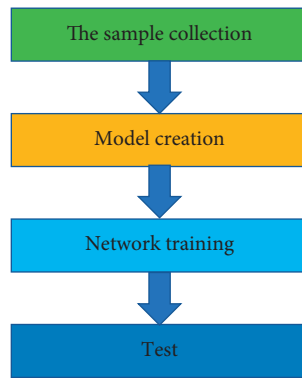
FIGURE 2: The process of the experiment.
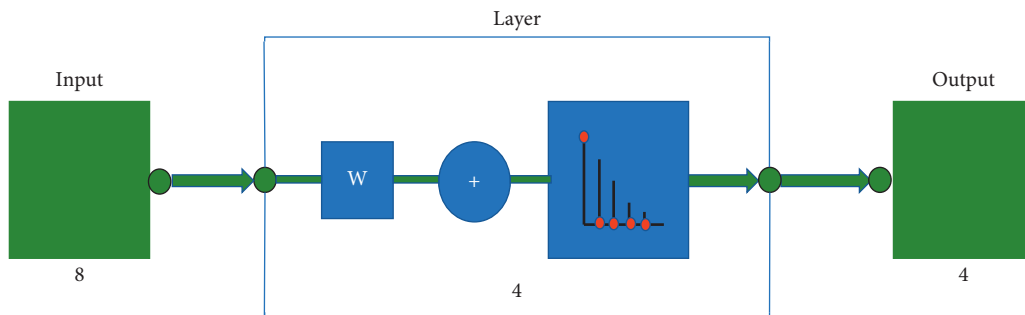
FIGURE 3: Constructed Elman network structure.

Football teams	Power level
China	3–tier
Japan	1–tier
Korea	1–tier
Iran	1–tier
Saudi Arabia	2–tier
Iran	3–tier
Qatar	3–tier
United Arab Emirates	3–tier
Uzbekistan	3–tier
Thailand	4–tier
Vietnam	4–tier
Oman	4–tier
Bahrain	4–tier
North Korea	4–tier
Indonesia	4–tier
Australia	1–tier

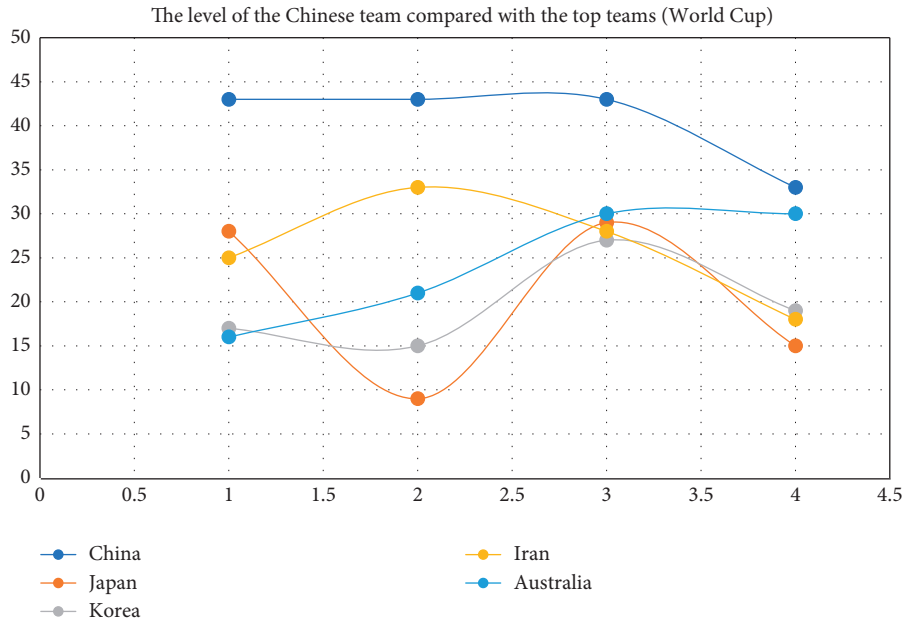FIGURE 4: The test results (classification result).

FIGURE 5: The level of the Chinese team compared with the top teams (World Cup).

relatively large, and it also shows that the classification effect of the model we established is accurate.

The number of fourth-tier teams is the largest, indicating that these teams have a poor track record, as can be seen from Table 1.

5.2. Experiments after Normalization. It needs to be emphasized here that the above results are obtained without normalization of input data. If the input data is normalized by mapminmax function, the results may change. Figure 7 shows the final experimental results obtained after normalization of the input data.

After comparing with the result of the last time, we find that Japan, South Korea, Iran, and Australia are still divided into the category of top Asian teams, which shows the super level of these four teams.

In this category of second-tier teams, Qatar, Iraq, and Uzbekistan have pushed Saudi Arabia, which ranked second in the previous category, into the category of third-tier teams in Asia, which is related to their excellent performances of these three teams in the last two Asian cups. It also shows the level of instability in Saudi Arabia.

For an in-depth analysis of the impact of normalization on the results, the differences between Saudi Arabia and four top teams before and after normalization were compared as examples. Table 2 shows the scores of Saudi Arabia and the four first-class teams before the normalization, and Table 3 shows the scores of Saudi Arabia and the four first-class teams after the normalization.
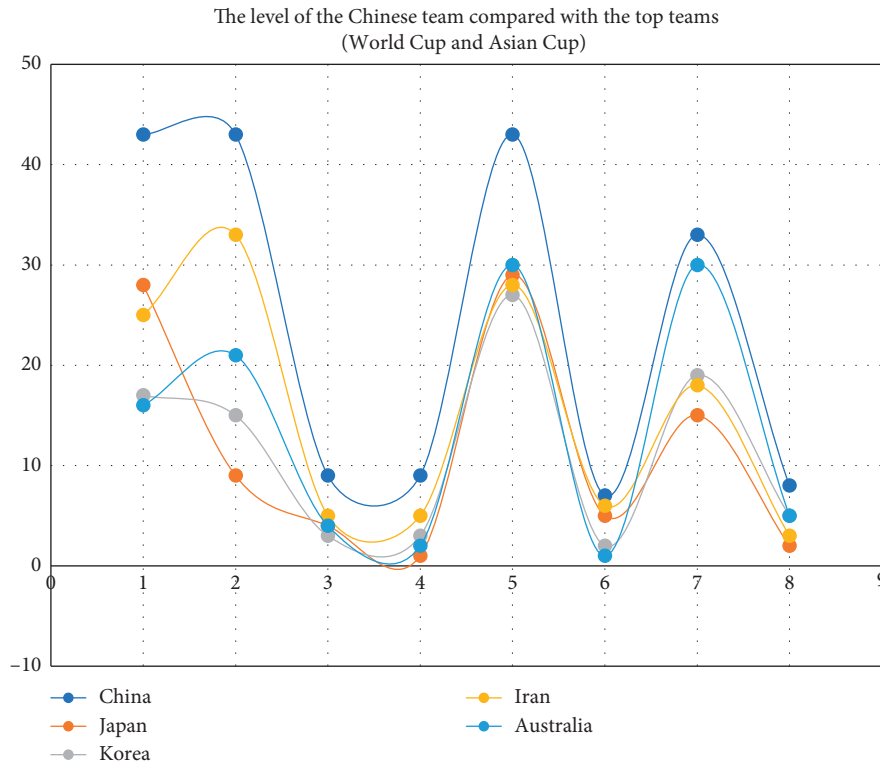
FIGURE 6: The level of the Chinese team compared with the top teams (World Cup and Asian Cup).

Football teams	Power level
China	3–tier
Japan	1–tier
Korea	1–tier
Iran	1–tier
Saudi Arabia	3–tier
Iran	2–tier
Qatar	2–tier
United Arab Emirates	3–tier
Uzbekistan	2–tier
Thailand	4–tier
Vietnam	4–tier
Oman	4–tier
Bahrain	3–tier
North Korea	3–tier
Indonesia	4–tier
Australia	1–tier

FIGURE 7: The final experimental results obtained after normalization of the input data.

TABLE 2: The scores of Saudi Arabia and the four first-class teams before the normalization.

Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
Saudi Arabia	28	33	2	9	43	10	26	14
Japan	28	9	4	1	29	5	15	2
Korea	17	15	3	3	27	2	19	5
Iran	25	33	5	5	28	6	18	3
Australia	16	21	4	2	30	1	30	5

TABLE 3: The scores of Saudi Arabia and the four first-class teams after the normalization.

Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
Saudi Arabia	−0.1111	0.4118	−0.875	0	1	0.125	−0.2667	0.0833
Japan	−0.1111	−1	−0.625	−1	−0.75	−0.5	−1	−0.9167
Korea	−0.9259	−0.6471	−0.75	−0.75	−1	−0.875	−0.7333	−0.6667
Iran	−0.3333	0.4118	−0.5	−0.5	−0.875	−0.375	−0.8	−0.8333
Australia	−1	−0.2941	−0.5	−0.875	−0.625	−1	0	−0.6667

To sum up, we can conclude that the relatively stable teams in the clustering of this study are as follows. Firstly, top teams in Asia are Japan, South Korea, Iran, and Australia. Secondly, Asian third-tier teams are China and United Arab Emirates. Thirdly, Asian fourth-tier teams are Thailand, Vietnam, Oman, and Indonesia.

6. Conclusion and Future Work

In order to classify the national football teams, this study took 16 major Asian national teams as samples and eight international competitions as sample features, built a self-organizing feature mapping network model, took matlab as the experimental platform, and finally achieved a reasonable classification result. In this paper, we also focus on the analysis of the situation of the Chinese team. In addition, we further compare and analyse the differences caused by the normalization of the input data. Of course, we know that our model is not perfect and that there may be improper classifications in some classification work; in the future, we will use the same approach to categorize other teams around the world and refine our model based on that. The results of this study are applicable to other scenarios, for example, the results can be used when we want to rank the income levels of a country's residents.

Data Availability

The underlying data supporting the results of this study can be found on the Internet.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

References

[1] P. Maji and S. Paul, *Neural Network Tree for Identification of Splice Junction and Protein Coding Region in DNA*, Springer International Publishing, New York, NY, USA, 2014.

[2] Y. Zhang, C. Yin, Q. Wu, Q. He, and H. Zhu, "Location-aware deep collaborative filtering for service recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3796–3807, 2021.

[3] R. Qie, X. Guan, X. Yan et al., "Method and its application of natural quality evaluation of arable land based on self-organizing feature map neural network," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 30, no. 23, pp. 298–305, 2014.

[4] Y. Zhang, J. Pan, L. Qi, and Q. He, "Privacy-preserving quality prediction for edge-based IoT services," *Future Generation Computer Systems*, vol. 114, pp. 336–348, 2021.

[5] Q. Rui-Qing, L. Xue, and M. Wang, "Self-organizing feature map (SOFM) neural network method in reservoir quality synthetic evaluation and its application in Jixi basin," *Ence of Surveying & Mapping*, vol. 39, no. 1, pp. 168–174, 2009.

[6] F. Shen and O. Hasegawa, "A fast nearest neighbor classifier based on self-organizing incremental neural network," *Neural Networks*, vol. 21, no. 10, pp. 1537–1547, 2008.

[7] Y. Zhang, K. Wang, Q. He et al., "Covering-based web service quality prediction via neighborhood-aware matrix factorization," *IEEE Transactions on Services Computing*, vol. 1, 2019.

[8] X. Fu, H. Chen, and R. Cai, "Comprehensive evaluation of power quality based on self-organizing feature mapping networks," *Huanan Ligong Daxue Xuebao/Journal of South China University of Technology (Natural Science)*, vol. 42, no. 6, pp. 7–11, 2014.

[9] L. Zhao, "Study on development mechanism of regional tourism system based on self-organizing theory: a theoretical analysis framework," *Journal of Beijing International Studies University*, 2010.

[10] S. Yuan and J. Bi, "Comprehensive evaluation approach of power quality based on neural network ensemble and subordinate degree," in *Proceedings of the International Conference on Information Science & Engineering*, IEEE, Harbin, China, January 2011.

[11] X. Wang and L. Wang, "Comprehensive evaluation of students based on SOFM neural networks," *Journal of Hebei Normal University (Natural Science Edition)*, vol. 35, no. 3, 2011.

[12] J. Zhao, H. Peng, J. Zhang et al., "Short-term load forecasting based on self-organizing feature mapping neural network," *Journal of Guizhou University of Technology*, vol. 86, no. 8, pp. 164–167, 2003.

[13] C. Senabre, S. Valero, and J. Aparicio, "Using a self organizing map neural network for short-term load forecasting, analysis

of different input data patterns," in *Proceedings of the Distributed Computing & Artificial Intelligence-International Symposium. DBLP*, pp. 397–400, Avila, Spain, June 2010.

[14] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study," *Electric Power Systems Research*, vol. 91, pp. 18–27, 2012.