Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230 – 8776

JISM 2026: 16 (1)
https://doi.org/10.6025/jism/2026/16/1/21-30

Optimizing the Value Chain Using A Fuzzy Gmdh Algorithm

Lian Qian
School of Physical Education
Nanjing University Nanjing, Jiangsu
210000. China
slxbr615765@163.com

ABSTRACT

The paper explores the application of the Group Method of Data Handling (GMDH) algorithmenhanced with fuzzy inference to analyze and optimize the value chain in the sports industry. The value chain framework encompasses enterprise relationships, value creation processes, and structural dynamics, all of which are critical to sustainable industrial growth. Given the sports industry's limited historical data and inherent uncertainty, traditional modeling approaches face challenges. The GMDH algorithm offers advantages such as self organization, global optimization, and effectiveness with small datasets. However, its susceptibility to overfitting and structural instability prompted the integration of fuzzy logic to handle imprecise and noisy data better. The fuzzy GMDH model improves prediction accuracy by incorporating fuzzy reasoning into the basic processing units, enabling more robust modeling of nonlinear and uncertain systems. The study applies this optimized model to the sports industry in "M Province," comparing performance between standard GMDH and fuzzy GMDH. Results indicate that the fuzzy variant provides superior fit and predictive capability, supporting more accurate evaluation of the industry's value and development potential. Despite these improvements, the authors acknowledge room for further refinement of the algorithm's architecture. The research underscores the importance of advanced data mining techniques, such as fuzzy GMDH, in enabling evidence based decision making in emerging sectors, including sports, where data scarcity and volatility are common challenges.

Keywords: Value Chain, GMDH Algorithm, Fuzzy Inference, Sports Industry development, Data Mining, Self-Organizing Modelling, Uncertainty Modelling, Predictive Analytics

Received: 30 August 2025, Revised 10 November 2025, Accepted 23 November 2025

1. Introduction

The value chain in the sports sector typically comprises three main components: the value chain itself, the relationships within the value chain enterprises, and the process of value production [1]. The value chain indicates the performance of the various enterprises involved, illustrating how well they align with the anticipated value and the overall value chain environment. Within the value chain, businesses experience both collaboration and competition in their interactions [2]. The interplay between the components of the value chain and the inter-enterprise relationships forms a network for industrial development. The value production process represents the realization of enterprise value, encompassing how production value is allocated through resource utilization, thereby demonstrating the connection between the input and output of value for enterprises. The relationships among value chain enterprises convert a traditional industry network into a dynamic value network [3]. Optimizing the industrial value chain typically occurs across various subcategories within three dimensions: the formation of the value chain, relationships among value chain enterprises, and the value production process [4]. Different industries must have a unique understanding and application of big data, as substantial datasets require enhanced visualisation tools for effective utilisation. Data mining technology primarily involves key areas such as mathematical statistics, artificial intelligence models, and database technology, along with visualization research. This research is comprehensive and represents a new discipline characterized by multi-disciplinary integration and interaction. Currently, it is essential to extract potentially valuable information from vast datasets filled with noise and to identify avenues for value chain optimization through scientific methodologies, leading to the inevitable choice of data mining technology.

2. Early studies

Many data mining technologies impose greater demands on users, and human factors can lead to variability in the quality of mathematical models [5]. The Group Method of Data Handling (GMDH) employs crossover, mutation, and selection operations to automatically establish and validate the model structure in accordance with evolutionary principles [6]. The GMDH algorithm model serves as a self-organizing data mining technique, achieving an optimal balance between memory and generalization functions. It stands out as one of the most widely utilized and versatile modeling algorithms [7]. The GMDH algorithm possesses four primary characteristics. Firstly, it has a robust self organizing capability, eliminating concerns about multicollinearity among data variables. It automatically filters input variables based on their best correlation, resulting in a relatively straightforward production model (Parsaie A, et al. 2015) [8]. Secondly, it offers global optimization advantages. Under identical conditions, the system's model is optimal, effectively addressing minor sample issues [9]. Thirdly, the model selection process is objective. Supported by computer-aided technology, the GMDH network requires only input data, an operational sequence, and calculation standards. The computer can directly manage other structural and variable parameters, making the model selection process more objective and precise [10]. Lastly, it showcases inductive features that integrate both dynamic and static elements.

The Fuzzy GMDH Network Algorithm provides a powerful, flexible approach to value chain analysis by enabling sophisticated modeling of complex, uncertain systems.

A fuzzy GMDH-type neural network model is developed for forecasting in mobile communication systems, with potential applications to complex systems like value chain analysis. The algorithm offers unique advantages in handling nonlinear relationships and uncertainty [11] Specifically for value chain analysis, researchers have demonstrated its capability to optimize network structures and evaluate enterprise performance [12].

The key strengths include:

- Automatic model generation
- Handling of noisy and limited input data
- Ability to produce explicit mathematical models
- Flexibility across multiple domains (communication, finance, manufacturing)

The method's core innovation lies in introducing fuzzy reasoning into traditional modeling techniques, allowing more nuanced analysis of interconnected business processes [13] (Y. Zaychenko et al., 2019). Fuzzy GMDH is used for forecasting in macroeconomics and finance, enabling automatic model construction and interval estimates, but its direct application to value chain analysis is not explicitly addressed. [14] (Mikhail Z., 2016)

A fuzzy network data envelopment analysis (FNDEA) approach was developed and applied to 10 R&D projects in Iran to evaluate the value chain network structure of complex products and systems under data uncertainty. [15] (P Peykani, 2020) A multistage fuzzy comprehensive evaluation model using AHP for value chain analysis is developed to address vagueness and subjectivity, thereby supporting the evaluation of management efficiency and effectiveness. [16]

The value of an industry can be assessed from two perspectives: intrinsic value and market value. Intrinsic worth refers to the net present value of the industry's near-term cash flows, indicating its profitability. Within the framework of a socialist market economy, a free market system is promoted; hence, market value reflects the industry's intrinsic value. Ultimately, the sector's value depends on the future profitability of its asset portfolio, which corresponds to the industry's anticipated cash flows. Consequently, this paper investigates the value chain within the sports industry. It primarily begins with evaluating the sports industry's worth, followed by the development of a GMDH model using data processing grouping methods to assist the sports industry in optimizing and enhancing its value chain, fostering the integration of sports resources, and encouraging the swift and sustainable growth of the sports industry.

3. Methodology

3.1 GMDH Algorithm

Numerous publications on user surveys exist in the simulation literature. An earlier survey conducted by Kleine 1, 2 explored user opinions on eleven distinct simulation languages [4]. The survey findings indicated that interpreting the results was challenging, primarily because a limited number of respondents were proficient in multiple languages. Additionally, it was difficult to determine the expertise of some participants.

GMDH algorithm	
Parameter GMDH algorithm	Nonparametric GMDH algorithm
Combinatorial COMBI	Objective Computer Clusterization OCC
Multilayered Iteratiomal MIA	
Harmonial	Analogues Complexing AC
Objective System Analysis OSA	

Table 1. Classification of GMDH Algorithms

The GMDH algorithm is a heuristic, self-organizing mathematical model introduced based on multilayer neural network theory in the late 1960s. It represents a fusion of data processing techniques, relying on polynomial functions and utilizing continuous screening combinations to differentiate nonlinear systems, showcasing strong identifiability for such systems. The GMDH organizational structure features a feed forward neural network, predominantly serving a predictive role in applications, and is commonly referred to as a GMDH neural network. It primarily serves a predictive function in applications and is also known as the GMDH neural network. The main principle is that in a system, if there are m variables *Xi (1, 2, ...m)* and one output variable *y*, the K-G polynomial expression results of each fitting trajectory can be obtained according to formula (1). Among them, *a0, ai. aij. aijk,...(i, j, k=1, 2,...m)* are the coefficients. This function can depict historical data paths and approximate any linear function. From the formula, it is evident that the contribution of each term in the polynomial to the fitting accuracy varies. The system leverages the self-organizing traits of the GMDH neural network, continuously eliminating terms with minimal contribution to the fitting accuracy, thereby reducing computational complexity, resulting in a more streamlined and effective model.

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} a_{ijk} x_i x_j x_k + \cdots \qquad (1)$$

As a feedforward neural network, the primary distinction between a GMDH neural network and other neural network types lies in its dynamically formed training data, which is subject to continuous change. When a neuron progresses to the subsequent level to form a new neuron, the network's self-organization feature automatically discards neurons that contribute poorly. This makes the number of neurons in each layer not fixed. Fig.1 is the GMDH network diagram. $\hat{y}$ represents the network's evaluation value. X_{il} is the input variable i of the first sample, i=1, 2,... *m*. The *m* is the number of input variables in the sample. $\hat{y}_{jkl}$ is the evaluation value of the k neuron in layer *j* of the first sample, *k=1, 2,...m.* r_{jk}^2 is the evaluation index of the k neuron in layer *j*. The threshold in layer *j* is R_j.

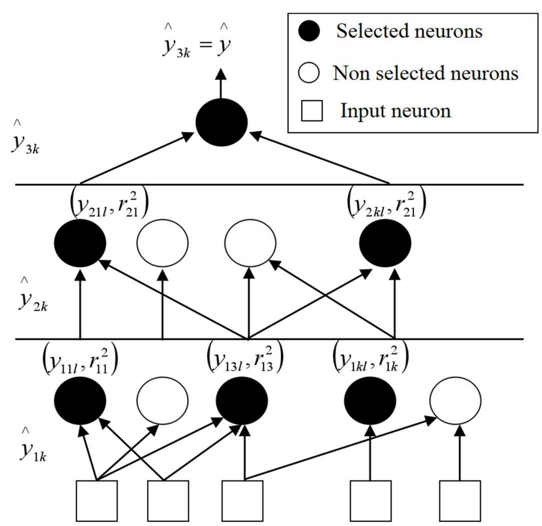

Figure 1. An iterative design model

The basic neuron in the GMDH neural network is a dual input, single output structure. A polynomial of two variables and two degrees represents the relationship between the input and the output. Among them, the coefficients *a, b, c, d, e, and f* are mainly determined by the least square method. The unit output here is the same as the content of the lower units. The production of the upper two different units is the input content of the lower unit, and the output variable of the processing unit is the polynomial with two degrees of the input variable. When the network has more than one layer, the polynomial order increases by 2. Finally, the network is a multinomial expression of 2k orders. The GMDH neural network has shortcomings in practical applications, primarily because certain structures are unstable. This is because the intermediate variable increases with iteration, and the goodness of fit with the output variable is higher. This makes the correlation between the middle variables too high. The recognition rate for the two least components in a nonlinear system is low. This is because its retrieval principle is the use of gradient information, which will have the possibility of going to a local minimum to bring limitations and misleading to searching. To address the shortcomings of the GMDH neural network, a fuzzy inference model was proposed to optimise the primary processing unit.

3.2 Optimize GMDH Algorithm

Reasoning is the process of drawing potentially imprecise conclusions from an inaccurate dataset. The reasoning in human thinking is often possible and approximate. That is to say, there is a certain degree of fuzziness in humans' natural thinking. Fuzziness is a transitional stage that reflects the intersection and integration of elements across performance, attributes, and other parameters. This is the natural state of existence, and it is also an important research topic in artificial intelligence. Fuzzy reasoning generally uses two valued logic or multi-valued logic to complete. Among them, the most crucial theory is the composition rule that converts the conditional statement "if x is A, y is B" into a fuzzy relation, which Zade proposed in the 70s of the last century. On this basis, fuzzy theory, employing various fuzzy logic methods and fuzzy truth values, is introduced. Fuzzy inference is introduced to replace the basic processing unit in the GMDH algorithm to enhance the algorithm's effectiveness.

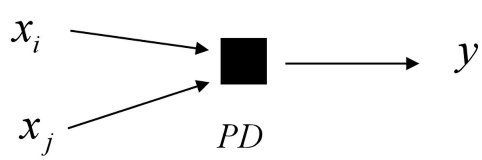

Figure 2. GMDH network basic processing unit

Fig. 2 shows a neuron of fuzzy inference GMDH model. The fuzzy model comprises k fuzzy rules. The expression of rule k ($1 \le k \le K$) is R_K: *If* x_1 *is* A_1^k *and* x_2 *is* A_2^k *Then* y *is* y^k. Among them, x_1, x_2 stand for input, and y is the input of the model. A_1^k, A_2^k are the membership functions of the input variables x_1, x_2. The Gauss function is chosen here, which is shown in formulas (2) and (3). Among them $a_{1k}, a_{2k}, b_{1k}, b_{2k}$ are the parameters of the model.

$$u_{1k}(x_1) = \exp\left\{-\frac{(x_1 - a_{1k})^2}{b_{1k}}\right\} \qquad (2)$$

$$u_{2k}(x_2) = \exp\left\{-\frac{(x_2 - a_{2k})^2}{b_{2k}}\right\} \qquad (3)$$

Fig. 3 shows a typical distribution graph of the Gauss function shaped like an inverted clock. Parameter a is the peak value of the Gauss curve, *b* is the corresponding abscissa, and *c* is the RMS width of Gauss, which controls the width of the function.

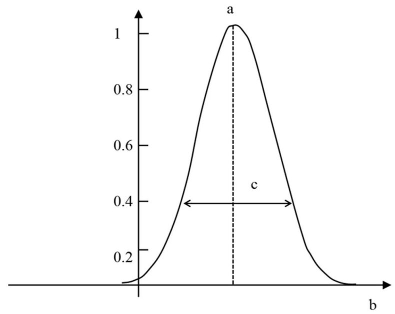

Figure 3. The standard distribution curve of the Gauss function

The topological structure of the fuzzy inference model comprises four layers. The first layer is the input layer, which directly connects the neurons to the input variables. The second layer is the fuzzification layer. All nodes in this layer are divided into k groups, and each node represents the first half of a fuzzy rule. Each node computes the membership value of the input variable. The third-layer fuzzy inference layer primarily applies the product principle to calculate the excitation intensity for each fuzzy rule. The formula is $u_k(x) = \prod_{i=1}^{2} u_{ik}(x_i)$. The defuzzification operation is performed on the fourth layer to produce the final results. The calculation formula of input value is $y = \frac{\sum_{k=1}^{k} u_k(x)\omega_k}{\sum_{k=1}^{k} u_k(x)}$. The mixed projection method is used to estimate the parameters of the $a_{1k}, a_{2k}, b_{1k}, b_{2k}$ model here.

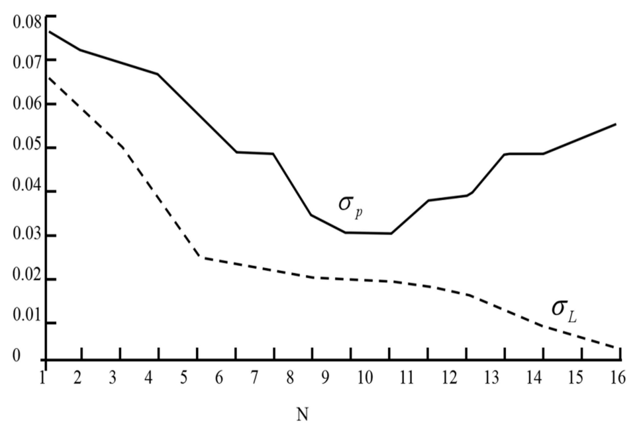

Figure 4. Minimum fitting variance

In the modelling of self-organising data mining, the training samples are typically partitioned into categories. This is to make the generated model have a certain balance in generalization ability and complexity. Prior knowledge must be used to control the model's complexity effectively. In the modelling of GMDH algorithm, based on this principle, it is necessary to use effective external information so that the best model can be selected from the given data sample. The external information here is data that has not been used in the model's parameter estimation. Fig. 4 shows the least-fit variance map. The abscissa represents the complexity, and the ordinate is the least fitting square error. The least-square error in the selection data set is the minimum square error in the training set. As the model complexity increases, the number of variables and the maximum polynomial degree will also increase. The minimum squared error on the training data set decreases. It is impossible to determine whether the model exhibits overfitting solely from this information; therefore, an external criterion is necessary to select the training samples to exclude, using the least-squares error. The lowest value corresponds to the minimum. In self-organizing data mining, each step will select many models. When it is not necessary to improve the external criterion at a certain stage, the model is the optimal complexity model can be considered, and the best balance of fitting generalization of training samples can be realized so as to end the modeling process of GMDH algorithm.

4. Result Analysis and Discussion

In this paper, based on the basic data of the sports industry in M Province, the GMDH algorithm based on an optimized value chain was simulated and tested. Construction of the fuzzy GMDH network model was carried out based on *Knowledge miner5.0* software. First, sample data were collected. In the value evaluation of the sports industry, the model's inputs represent indicators of the value driving factors, and the output is the value of the sports industry for the training samples. September 1, 2016 was selected as the basis point, and sports industry related enterprise data in *M* province were selected as a sample. The indicators were extracted to get the sample data for this paper. Subsequently, the sample data were preprocessed to map them to the (0, 1) range. The raw data were normalized to reduce the dynamic changes in the data processing process to increase the effectiveness of the evaluation results. Although the GMDH network has good data recognition and self-organization ability, data can also be trained without being processed; if the input data preprocessing can reduce the process and range of data recognition, the results are more practical and accurate. The formula used in data preprocessing is $x_i = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}$. Among them, the x_i represents the data of the i row in any column. $x_{i,\min}$ is the minimum value of all the data in the column in which is located and is the maximum value of all the data in the column in which is located.

$$\begin{aligned} A:&(y_1; x_{11,} x_{12,} \cdots x_{1n,}) \\ &(y_2; x_{21,} x_{22,} \cdots x_{2n,}) \\ &\cdots\cdots \\ &(y_s; x_{s1,} x_{s2,} \cdots x_{sn,}) \end{aligned} \tag{4}$$

$$\begin{aligned} B:&(y_{s+1}; x_{s+1,1,} x_{s+2,2,} \cdots x_{s+1,n,}) \\ &\cdots\cdots \\ &(y_m; x_{m1,} x_{m2,} \cdots x_{mn,}) \end{aligned} \tag{5}$$

Based on the GMDH network architecture, the model is applied to fuzzy reasoning. The selected sample data are used as input and output to establish a fuzzy GMDH-based value-evaluation model. The processed sample data are divided into training samples and test samples. The sample data sequence is set as $\{y_i, x_{ij}\}(i = 1,2,\cdots,m),(j = 1,2,\cdots,n)$. The number of network input signals is n, the output variable is y, and the input variable is x. The training sample data set *A* and the test sample data set *B* are obtained, which are shown in formulas (4) and (5). The sequence $\{y_i, x_{ij}\}$ generates the sample set *W*, which contains data samples with m inputs and outputs, $N_W = m$. 80% of them are training samples, and the remainder are test samples. In sample *A*, any *Xi* in the *N* inputs of the sports industry data is selected. When the input variables are set, the corresponding y can be obtained as the output variable. In the basic processing unit fuzzy reasoning model, *n* (*n*-1) /2 neurons can be generated. This can get the first layer of the initial network. Screening criteria can be used in neuronal screening. Assuming that that the R_j is the threshold of layer j. When $r_{jk}^2 < R_j$, the neuron is retained as the following input, the neuron that does not satisfy the condition is deleted. Finally, the value evaluation model of the sports industry established in the form of self-organization based on fuzzy GMDH network can be obtained, which is connected with the output layer directly associated with the neurons retained by the output layer.

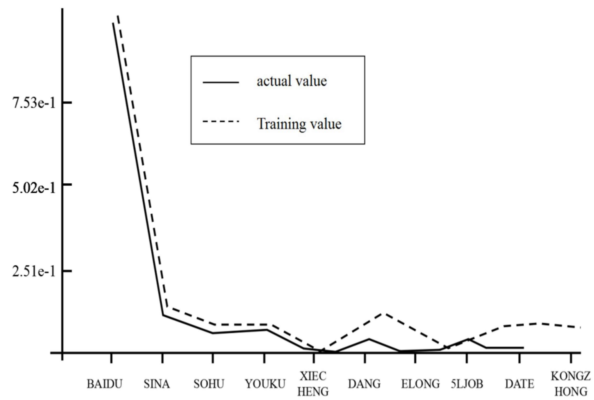

Figure 5. GMDH Comparison of data fitting process

To verify the practicability of the model, this paper selected the basic GMDH network and the fuzzy GMDH network for comparative analysis, providing a scientific basis for studying sports industry development and its industrial value. The basic GMDH network model is a linear output model. The result is that $y = -6.987e - 2X5 + 9.528e - 1X11 + 3.934e - 2$. The residual square sum of the prediction is 0.321, and the average absolute error percentage is 23.01%. The approximate error variance is 0.0187. The data-fitting process is shown in Fig. 5. The solid line represents the actual values, and the dotted line represents the training values. The data indicate that the fit is poor.

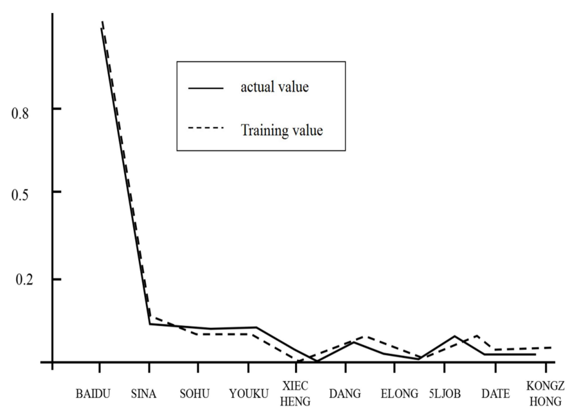

Figure 6. Comparison chart of data fitting process
in fuzzy GMDH network model

In the simulation test, the structure of the fuzzy GMDH network model is $y = -9.815e - 1X12 + 5.7879e - 1$. The predicted value is 0.1785132, and the reduced value is 445031.1900003. The actual value is 45930.7300001, and the error is -1.95324%. The data-fitting process is shown in Fig. 6. The solid line represents the actual values, and the dotted line represents the training values. The data indicate a better fit.

5. Conclusion

The benefits of the Group Method of Data Handling (GMDH) include its ability to effectively achieve global optimization, its adaptability to work with small sample sizes, its robust self-organization capabilities, and its capacity to select and construct streamlined model structures autonomously. The rapid expansion of the sports sector introduces uncertainty about its future. Given the sports industry's relatively short history and limited data, assessing and advancing its value are challenging. Hence, this paper proposes a GMDH algorithm grounded in an optimized value chain to analyze the sports industry's value, aiming to offer a scientific foundation for forecasting its development. By examining the fundamental structure and processes of the GMDH model, the algorithm's strengths and weaknesses were evaluated. Fuzzy inference was incorporated to replace the basic processing unit in the GMDH algorithm, thereby enhancing its effectiveness. The fuzzy GMDH network was employed to model the sports industry in M Province, allowing for a comparison of the fit between the standard GMDH algorithm and the fuzzy GMDH network. The findings of this paper are as follows: the fuzzy GMDH algorithm, based on the value chain, provides a more precise prediction and evaluation framework for the sports industry's growth; it demonstrates high prediction accuracy, with minimal discrepancies between the goodness of fit and predicted values. This confirms the success of the study. However, the results also indicate that there is still room for improvement in the algorithm. Therefore, further development of the GMDH algorithm's model structure will be the primary focus of future research.

References

[1] Yi, H. E., Chen, J. (2015). Research on line loss data pretreatment in distribution network based on GMDH algorithm. *Dianli Xitong Baohu Yu Kongzhi/power System Protection Control*, 43 (9), 42-46.

[2] Kasaeian, A., Ghalamchi, M., Ahmadi, M. H., et al. (2017). GMDH algorithm for modeling the outlet temperatures of a solar chimney based on the ambient temperature. *Mechanics & Industry* 18 (2), 216

[3] Yin, W., Hu, W., Hui, F., et al. (2015). Inverse Determination of Material Parameters Based on Decoupled GMDH Algorithm. *China Mechanical Engineering* 26 (9), 1215-1221.

[4] Chang, F., Hwang, Y. (2015). A self organization algorithm for real time flood forecast. Hydrological Processes, 13 (2), 123-138.

[5] Antanasijevi, D., Antanasijevi, J., Pocajt, V., et al. (2016). A GMDH-type neural network with multi-filter feature selection for the prediction of transition temperatures of bent-core liquid crystals. Rsc Advances, 6 (102), 99676-99684.

[6] Osanaiye, O., Cai, H., Choo, K. K. R., et al. (2016). Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. *Eurasip Journal on Wireless Communications & Networking,* (1), 130.

[7] Inbarani, H. H., Bagyamathi, M., Azar, A. T. (2015). A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Computing Applications,* 26 (8), 1859-1880.

[8] Parsaie, A., Haghiabi, A. H. (2015). Predicting the longitudinal dispersion coefficient by radial basis function neural network. *Modeling Earth Systems Environment,* 1 (4), 1-8.

[9] Najafzadeh, M. (2015). Neuro-fuzzy GMDH systems based evolutionary algorithms to predict scour pile groups in clear water conditions. Ocean Engineering, 99, 85-94.

[10] Bodyanskiy, Y., Tyshchenko, O., Kopaliani, D. (2015). A hybrid cascade neural network with an optimized pool in each cascade. *Soft Computing*, 19 (12), 3445-3454.

[11] Heung Suk Hwang . (2006). Fuzzy GMDH-type neural network model and its application to forecasting of mobile communication, Computers & Industrial Engineering, Volume 50 (4), p. 450-457.

[12] Batkovskiy, A. M., Kalachikhin, P. A., Semenova, E. G., Telnov, Y. F., Fomina, A. V., Balashov, V. M. (2018). Configuration of enterprise networks, *Entrepreneurship and Sustainability Issues* 6(1), 311-328.

[13] Yuriy, P., Zaychenko, Yuriy, P., Zaychenko, Helen., Zaychenko, Helen Zaychenko. (March 2019). Fuzzy GMDH and its application to forecasting financial processes, System Research and Information Technologies. 10.20535/SRIT.2308-8893.2019.1.07

[14] Mikhail, Z., Zgurovsky., Yuriy, P., Zaychenko., Yuriy P., Zaychenko. (2016). Inductive Modeling Method (GMDH) in Problems of Intellectual Data Analysis and Forecasting. July 2016Studies in Computational Intelligence 652:221-260. DOI: 10.1007/978-3-319-35162-9_6. *In* The Fundamentals of Computational Intelligence: System Approach.

[15] Peykani, P., Gheidar, K. J. (2020). Performance appraisal of the research and development value chain for complex products and systems: a fuzzy three-stage DEA approach. *Journal of New Researches In Mathematics* 6 (25), 41-57

[16] Guo, Da-yong. (2019). A Multistage Fuzzy Comprehensive Evaluation Model based on AHP for Construction Enterprises through Value Chain Analysis, IOP Conference Series: *Materials Science and Engineering*.