Investigation of Hybrid Deep Learning Networks in Forecasting Energy Supply

Yuriy Zaychenko, Helen Zaichenko and Oleksii Kuzmenko

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Institute for Applied System Analysis, Prospect Beresteiskyi, 37, Kyiv, 03056, Ukraine

Abstract

In this paper the intelligent methods for solving the problem of short- and middle-term forecasting of electricity sales to ultimate customers are considered. A hybrid GMDH-neo-fuzzy network, GMDH and ARIMA were studied. Neo-fuzzy neurons are chosen as nodes of a hybrid DL network. The optimal parameters of the hybrid DL network and GMDH were found. The synthesis of the optimal structure of hybrid GMDH-neo-fuzzy network for short- and middle-term forecasting was performed. Experimental studies of hybrid GMDH-neo-fuzzy network, GMDH and ARIMA for short- and middle-term forecasting have been conducted. The accuracy of the obtained forecasts was compared. The expediency of applying the researched methods of artificial intelligence for the considered intervals is substantiated.

Keywords

Hybrid DL network, GMDH, ARIMA, short-term, middle-term forecasting

1. Introduction

Problems of forecasting non-stationary time series and market indexes at stock exchanges pay great attention of managers of enterprises and various scientific researchers. For its solution were developed and applied for a long time powerful statistical methods, first of all ARIMA [1, 2]. In recent years, various intelligent methods, and technologies, such as fuzzy logic systems and neural networks, have also been proposed and widely used for forecasting in economics and technology.

The Group Method of Data Handling (GMDH), proposed and developed by acad. Alexey Ivakhnenko [3, 4], is an effective tool for forecasting and modeling non-stationary time series. This self-organization method allows the algorithm to build the optimal model structure for forecasting as it goes along, so it happens automatically. Besides GMDH has one more very important property: it may work with short samples. Method GMDH and fuzzy GMDH were successfully applied for forecasting in economy and financial sphere for long time.

Popular neural networks are used for forecasting in the financial sector, so MLP [5], fuzzy neural networks [6, 7], neo-fuzzy networks [8] and deep learning (DL) networks [9] can be alternative approaches. A new class of neural networks - hybrid DL-networks based on the GMDH method [10] initiated the emergence of a new trend in DL-networks. Their feature is the synthesis of the optimal network structure in the learning process, in addition to the learning of the neuron weights. This happens due to applied self-organization. The learning method allows you to adjust the weights in these networks layer by layer, rather than simultaneously. This solves the problem of gradient explosion or decay phenomenon. For networks with many layers, this fact is very important.

Initially, neurons with two inputs were used as nodes of a Wang-Mendel hybrid network in this field [10]. After the development of the DL neo-fuzzy network, Yamakawa neo-fuzzy neurons were used as nodes [8, 11]. Their important difference is that you don't need to train fuzzy sets, only neuron weights. Compared to Wang-Mendel neurons, this reduces the computational cost and minimizes the

International Scientific Symposium “Intelligent Solutions” (IntSol-2023), September 27-28, 2023, Uzhhorod, Ukraine
EMAIL: zaychenkoyuri@ukr.net (Yu. Zaychenko); syncmaster@bigmir.net (H. Zaichenko); oleksii.kuzmenko@ukr.net (O. Kuzmenko)
ORCID: 0000-0001-9662-3269 (Yu. Zaychenko); 0000-0002-4630-5155 (H. Zaichenko); 0000-0003-1581-6224 (O. Kuzmenko)

overall training time. Based on the studied two classes of hybrid DL networks, their effectiveness for forecasting in the financial sphere was compared.

In the previous papers these methods were applied and investigated in the problem of forecasting in financial sphere for market indices Dow Jones industrial average and NASDAQ. That is why it is interesting to investigate the effectiveness of ARIMA, GMDH and hybrid DL networks in forecasting in other areas, such as technology and production, specifically in short- and middle-term forecasting tasks. The goal of this paper is to investigate the accuracy of intelligent methods – hybrid DL networks, GMDH and ARIMA at the problem of forecasting Electricity Sales to Ultimate Customers, Residential (USA), June 7, 2023 – at the different forecasting intervals (short-term and middle-term), compare their efficiency and to determine which computational intelligence methods are the most perspective for forecasting in the economy and technology.

2. A review of the evolving hybrid GMDH-neo-fuzzy network

The architecture of the evolving hybrid DL-network is shown in Fig. 1. The input of the system accepts an $(n \times 1)$-dimensional vector of signals that are considered input. Then the first hidden layer receives this signal. At this level there are $n_1 = c_n^2$ nodes, each of which has strictly two inputs.

Outputs $N^{[1]}$ of the first hidden layer form the output signals to be further transmitted to the selection block located after the first hidden layer.

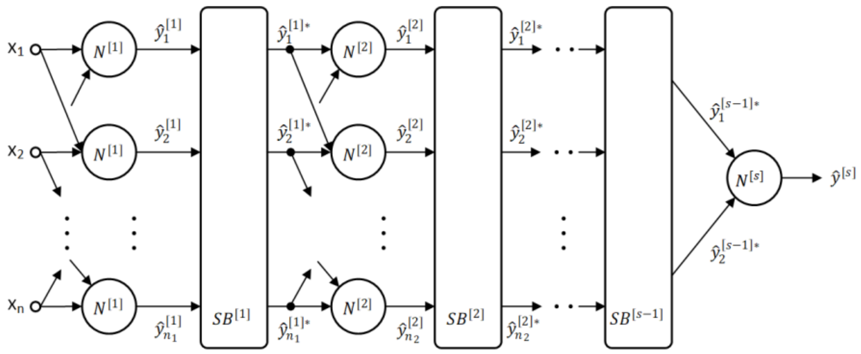

Figure 1: Evolving GMDH-network

It selects among the output signals $\hat{y}_l^{[1]}$ $n_1 *$ (where $n_1 *= F$ is so-called freedom of choice) most precise signals by some chosen criterion (mostly by the mean squared error $\sigma^2_{y_l^{[1]}}$). Among these $n_1 *$ best outputs of the first hidden layer $\hat{y}_l^{[1]} *$ n_2 pairwise combinations $\hat{y}_l^{[1]} *, \hat{y}_p^{[1]} *$ are formed. These signals are fed to the second hidden layer, that is formed by neurons $N^{[2]}$. After training these neurons output signals of this layer $\hat{y}_l^{[2]}$ are transferred to the selection block $SB^{[2]}$ which choses F best neurons by accuracy (e.g. by the value of $\sigma^2_{y_l^{[2]}}$) if the best signal of the second layer is better than the best signal of the first hidden layer $\hat{y}_1^{[1]} *$. Other hidden layers work in a similar way. The evolution of the system continues until the best signal of the selection block $SB^{[s+1]}$ is worse than the best signal received on the previous s-h layer. After that, you need to return to the previous layer to select the best node neuron $N^{[s]}$, which will have some output signal $\hat{y}^{[s]}$. The sequential movement from this neuron (node) back takes place along its connections and passes through all previous layers, which makes it possible to build the resulting structure of the GMDH-neo- fuzzy network.

As a result, due to the GMDH algorithm, it is possible to obtain a well-trained network with an optimal structure that was synthesized automatically. High-dimensionality problems, as well as vanishing or exploding gradients, are avoided because the learning is sequential layer-by-layer.

2.1. The role of the Neo-fuzzy neuron in the hybrid GMDH system

In Fig. 2 shows the architecture of the node selected as the quality for the proposed GMDH system. This is a neo-fuzzy neuron (NFN) proposed by Takeshi Yamakawa et al. in [9]. It is a non-linear system that has one output and several inputs. In the proposed GMDH system, neo-fuzzy-neurons with only two inputs are used, which implements the following mapping:

$$\hat{y} = \sum_{i=1}^{2} f_i(x_i) \tag{1}$$

where $\hat{y}$ is the output of the system, x_i is the input signal i $(i = 1,2,\dots,n)$. The nonlinear synapses NS_i are the building blocks of a neo-fuzzy neuron. Their task is to convert the input signal in the form of

$$f_i(x_i) = \sum_{j=1}^{h} w_{ji}\mu_{ji}(x_i) \tag{2}$$

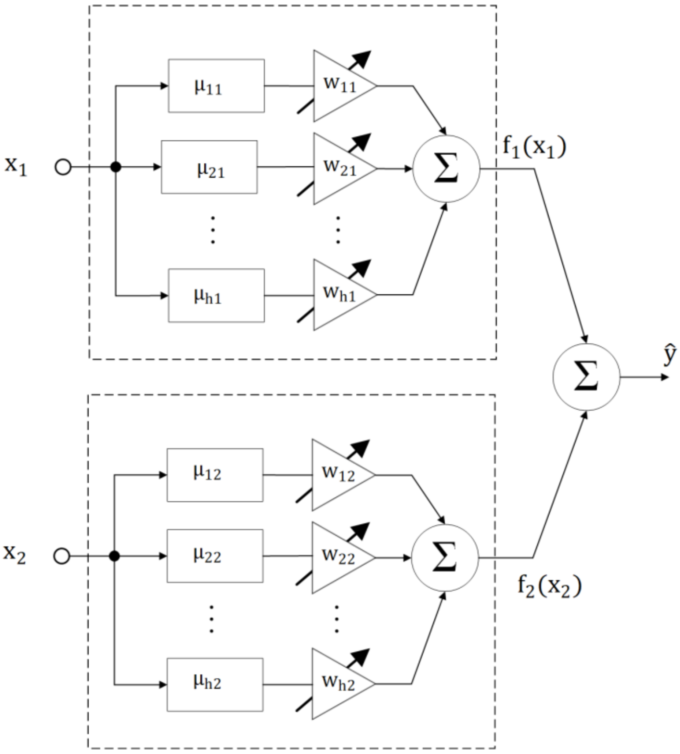

Figure 2: Architecture of neo-fuzzy neuron with two inputs

and fuzzy inference is realized in the form: if x_i is x_{ji} *then the output is* w_{ji},where w_{ji} is the synaptic weight in the consequent, x_{ji} is a fuzzy set whose membership function is μ_{ji} [11].

2.2. Neo-fuzzy neuron training algorithm

The standard local quadratic error function is used as the goal function (i.e., the learning criterion):

$$E(k) = \frac{1}{2}\left(y(k) - \hat{y}(k)\right)^2 = \frac{1}{2}e(k)^2 = \frac{1}{2}\left(y(k) - \sum_{i=1}^{2}\sum_{j=1}^{h} w_{ji}\mu_{ji}(x_i(k))\right)^2 \tag{3}$$

Using a conventional stochastic gradient descent algorithm, it can be minimized.

In the case of a predefined dataset, the training process can be performed in a single epoch in batch mode. For this purpose, the conventional least squares method is used [11]

$$w^{[1]}(N) = \left(\sum_{k=1}^{N} \mu^{[1]}(k)\mu^{[1]T}(k)\right)^{+} \sum_{k=1}^{N} \mu^{[1]}(k)y(k) = P^{[1]}(N)\sum_{k=1}^{N} \mu^{[1]}(k)y(k) \tag{4}$$

where $(\bullet)^{+}$ denotes the pseudo-inverse of the Moore-Penrose (y(k) is assumed to be the real value of the external reference signal).

With the sequential receipt of training observations, i.e., in the online mode, the recurrent form of the ANN can be represented as

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + \dfrac{P^{ij}(k-1)\left(y(k) - \left(w_l^{ij}(k-1)\right)^T \varphi^{ij}(x(k))\right)\varphi^{ij}(x(k))}{1 + \left(\varphi^{ij}(x(k))\right)^T P^{ij}(k-1)\varphi^{ij}(x(k))}, \\ P^{ij}(k) = P^{ij}(k-1) - \dfrac{P^{ij}(k-1)\varphi^{ij}(x(k))\left(\varphi^{ij}(x(k))\right)^T P^{ij}(k-1)}{1 + \left(\varphi^{ij}(x(k))\right)^T P^{ij}(k-1)\varphi^{ij}(x(k))}. \end{cases} \quad (5)$$

3. Data set

As the data set for forecasting were taken monthly Electricity Sales to Ultimate Customers, Residential (USA) since 01-2002 till 01-2023 taken. The whole sample consisted of 251 instances. The sample was divided into training and test subsamples. The dynamics of monthly energy power supply to Ultimate Customers, Residential (USA) is shown in the Fig. 3.

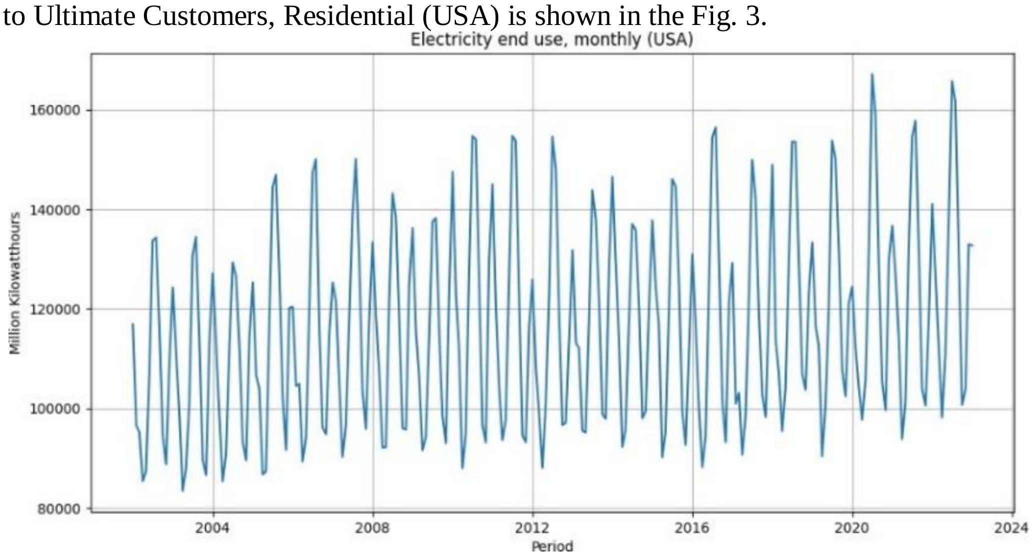

Figure 3: Dynamics of the index monthly energy supply

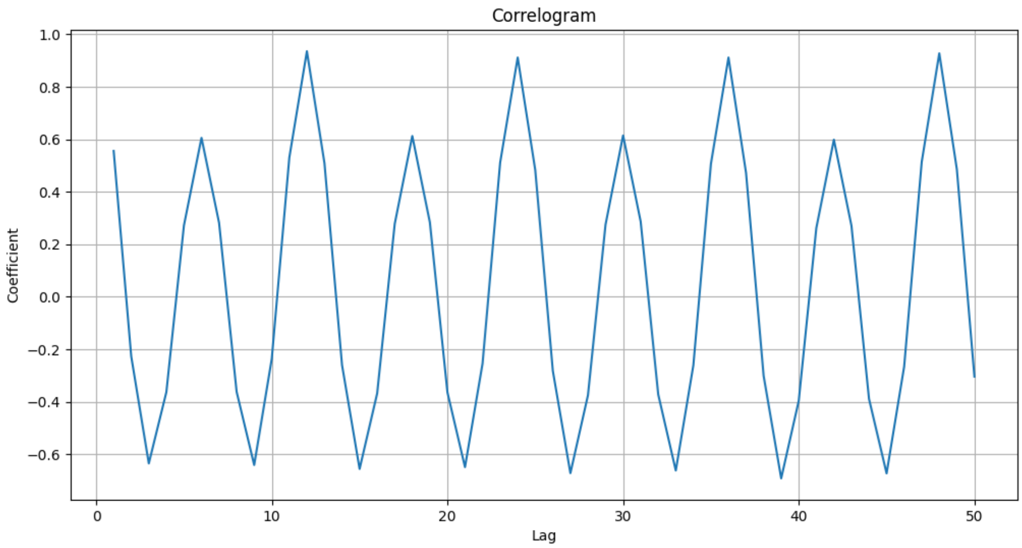

Figure 4: Correlogram

Analyzing the presented curve, one may conclude that there is strong correlation between preceding and conceding values and the process is periodical with period 6 months. Autocorrelation function (ACF) was determined for this process of power supply which is shown in the Fig. 5.

The check for stationarity of this process was performed using Dickey-Fuller test.

P-value: 0.5117527467140699 > 0.05. As it follows from this test the initial time series is not stationary. Using differencing this time series was transformed to the stationary one that's confirmed by Dickey-Fuller test with P-value: 1.3594288749888985e-14 < 0.05.

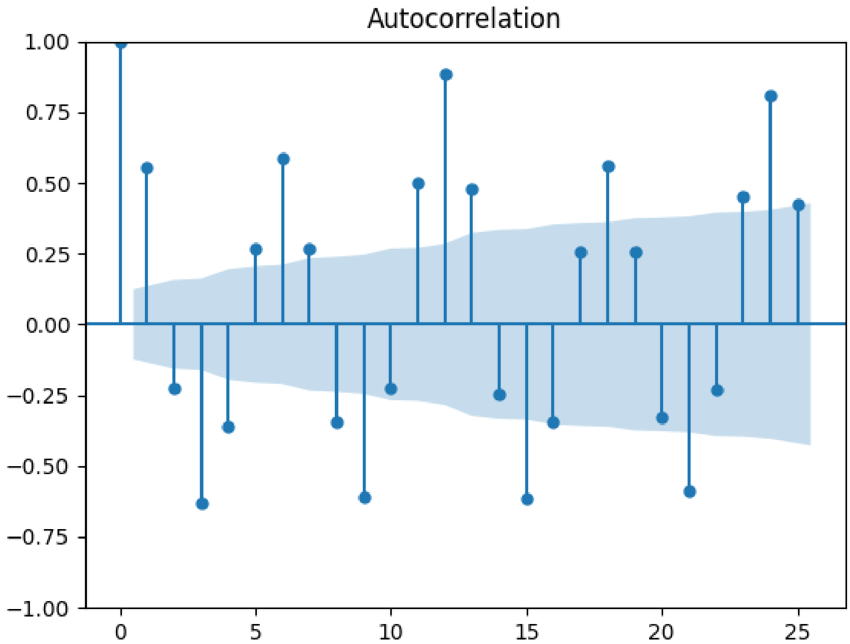

Figure 5: ACF of time series power supply

4. Experimental investigations

In the investigations was explored the forecasting accuracy of hybrid DL neo-fuzzy networks at various forecasting intervals: short-term forecasting with intervals 1, 3 5 and 7 days and middle term forecasting with intervals 10 and 20 days. At the first step the variable experimental parameters of hybrid network were chosen which are presented in the Table 1.

Table 1

Experimental parameters

Parameter	Value
Membership functions	Gaussian
Number of inputs	3; 4; 5
Number of linguistic variables	3; 4; 5
Ratio (percentage of the training sample)	0,6 (60%); 0,7 (70%); 0,8 (80%)
Criterion	MSE; MAPE
Forecast interval	1; 3; 5; 7; 20; 30

The optimization of these parameters was performed, in result the following optimal values were determined inputs: 3; linguistic variables: 3; ratio: 0,7.

After that the structure optimization of hybrid DL neo-fuzzy network was constructed using GMDH method. The process of structure generation is presented in the Table 2.

In result the optimal structure of three layers was determined: at the first layer 3 inputs, second layer – two neurons, third layer – one output neuron.

Further the training of the best hybrid network was carried out using method SGD (stochastic gradient descent with variable step. Flow chart of forecasting results for interval 1day in presented in the Fig. 6. The values of MSE and MAPE for this experiment are shown in the Table 3.

In the Fig. 6. flow chart of MAPE values for the best model of hybrid network is shown.

Further the similar experiments of hybrid network were performed with forecasting interval 3, 5, 10 and 20 days. After optimization the parameters and structure of hybrid network it was trained using training subsample. The forecasting accuracy at the test sample for interval 3 days is presented at the Table 4. In the succeeding experiments forecasting accuracy of Hybrid neo-fuzzy network was investigated with forecasting intervals 5, 10 and 20 days.

Table 2

Structure generation (inputs: 3; variables: 3; ratio: 0,7)

NFN	SB 1	SB 2	SB 3
(0, 1)	0.138		
(0, 2)	0.23		
(1, 2)	0.116		
((0, 1), (0, 2))		0.062	
((0, 1), (1, 2))		0.06	
((0, 2), (1, 2))		0.067	
(((0, 1), (0, 2)), ((0, 1), (1, 2)))			0.078
(((0, 1), (0, 2)), ((0, 2), (1, 2)))			0.081
(((0, 1), (1, 2)), ((0, 2), (1, 2)))			0.068

Table 3

Forecasting accuracy of hybrid neo-fuzzy network at forecasting interval 1 day

Criterion	MSE	MAPE
Min:	23088.985	0.137
Avg:	564640071.554	13.99
Max:	3671767454.527	36.888

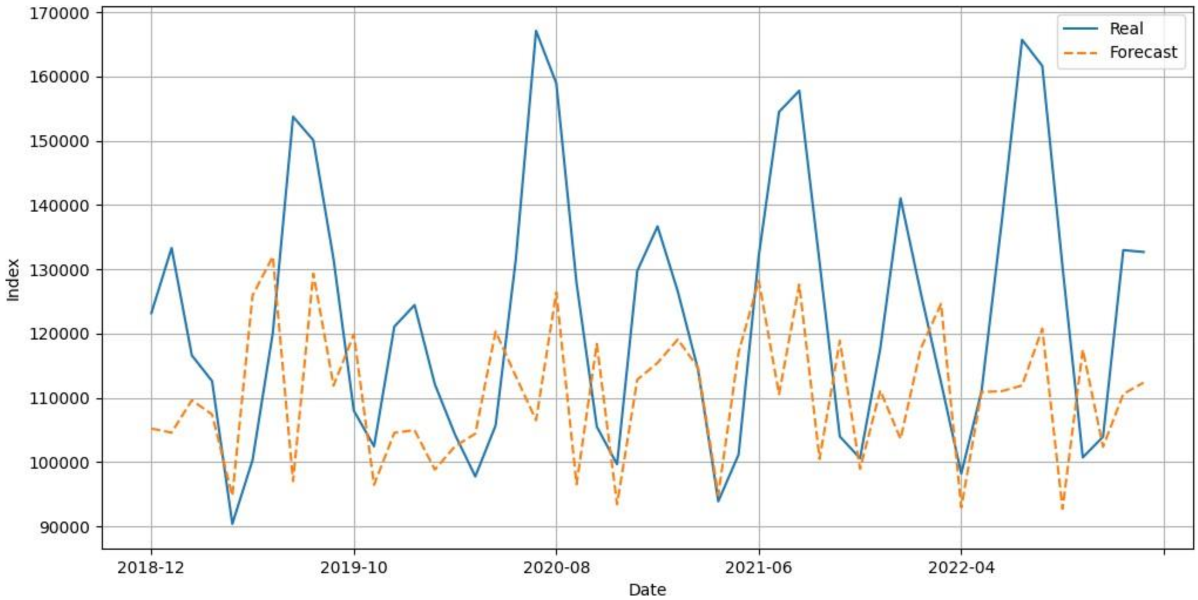

Figure 6: The most accurate forecast (inputs: 3; variables: 3; ratio: 0,7)

In the Table 5 accuracy of forecasting of the hybrid NFN optimal structure is presented with forecasting interval 10 days and in the table 6 with forecasting interval 20 days.

Table 4

Forecasting accuracy of hybrid network at forecasting interval 3 days

Criterion	MSE	MAPE
Min:	381106.04	0.592
Avg:	548163425.114	15.067
Max:	2486030587.063	38.021

Table 5
Forecasting accuracy of hybrid network at forecasting interval 10 days

Criterion	MSE	MAPE
Min:	19477.142	0.132
Avg:	538246827.982	14.074
Max:	2748009578.558	33.78

For estimating forecasting accuracy of hybrid DL network, it was compared with alternative methods: ARIMA and GMDH. The forecasting accuracy of GMDH for interval 1 day is shown in the Table 7 and for 5 days in the Table 8. The flowchart of forecasting results for the interval 5 days is shown in the Fig. 7 and for 20 days in the Fig. 8.

Table 6
Forecasting accuracy of hybrid network at forecasting interval 20 days

Criterion	MSE	MAPE
Min:	337367.726	0.472
Avg:	449715331.739	13.236
Max:	2692214636.887	33.064

Table 7
Forecasting accuracy of GMDH at interval 1 day

Criterion	MSE	MAPE
Min:	32949.209	0.121
Avg:	52857867.97	4.767
Max:	276493216.406	10.805

Table 8
Forecasting accuracy of GMDH at interval 5 day

Criterion	MSE	MAPE
Min:	1159357.608	1.096
Avg:	398583396.613	12.615
Max:	1617737584.052	25.174

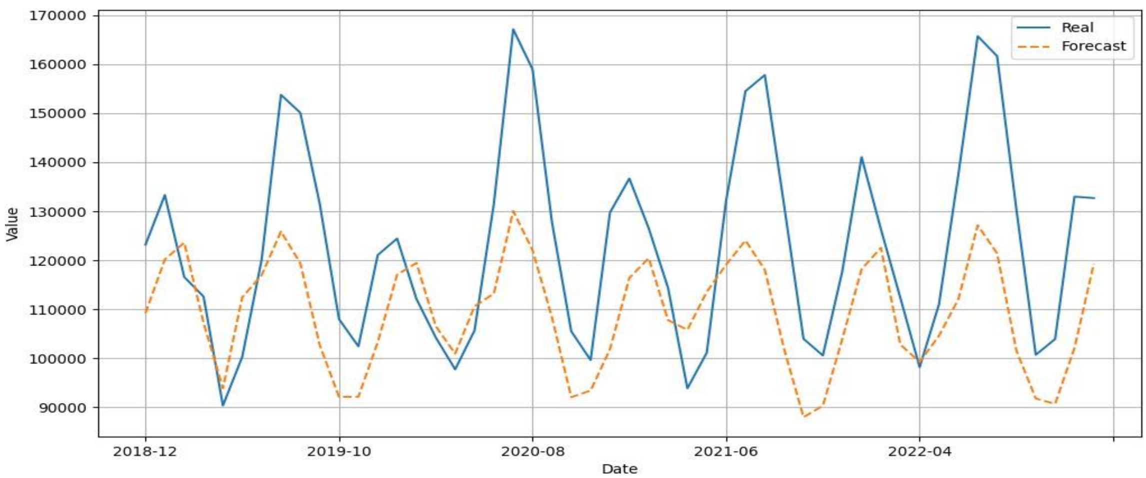

Figure 7: MAPE for the most accurate forecast by GMDH (inputs: 4; variables: 4; ratio: 0,8) for 5 days

In the next experiments forecasting efficiency of method ARIMA was investigated and analyzed. After the preliminary investigations the optimal parameters for ARIMA were found which were used in the following experiments. The forecasting accuracy of ARIMA for interval 1 day is presented in the Table 9 and for interval 5 days in the Table 10. The flowchart of real and forecasting results for ARIMA with interval 20 days is shown in the Fig. 9.

The comparative experiments were performed in which the accuracy of forecasting by hybrid DL network, GMDH and ARIMA at the different forecasting intervals was estimated and compared. The corresponding results are presented in the Tables 11, 12 and Fig. 10, 11.

Table 9

Forecasting accuracy of ARIMA at interval 1 day

Criterion	MSE	MAPE
Min:	2.291	0.001
Avg:	132427261.825	6.769
Max:	1197915373.458	20.712

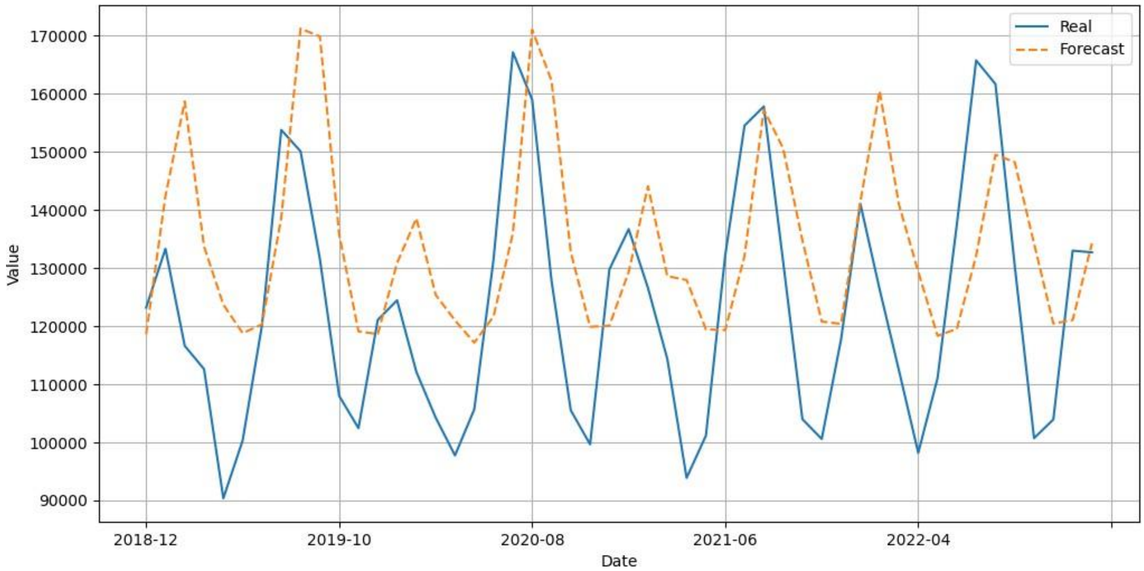

Figure 8: The most accurate forecast by GMDH (function: linear; inputs: 5; ratio: 0,6) 20 days

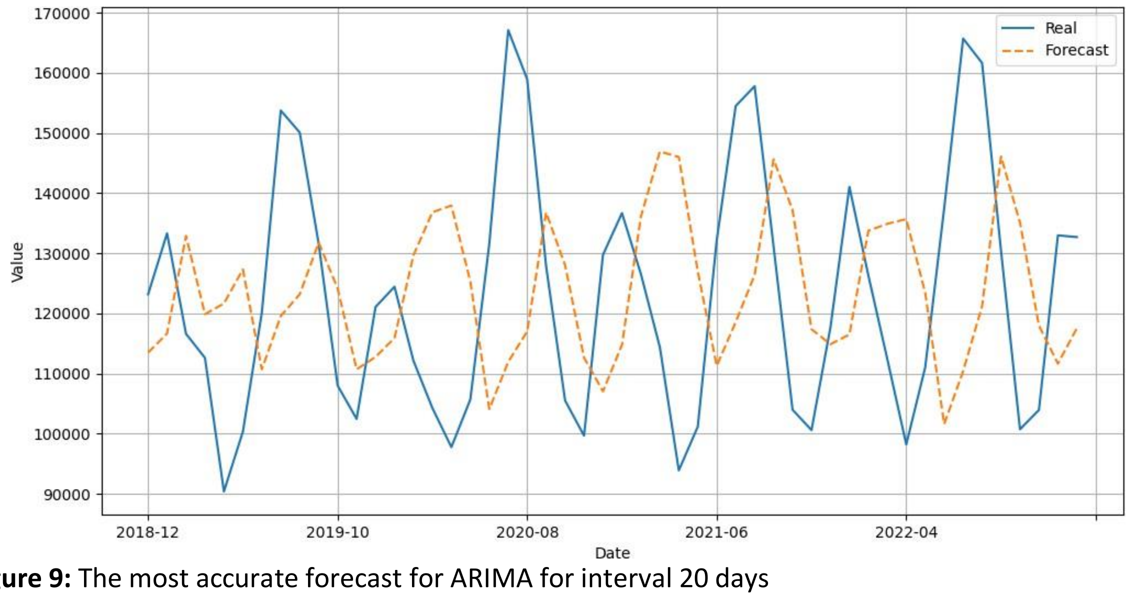

Figure 9: The most accurate forecast for ARIMA for interval 20 days

Table 20

Forecasting accuracy of ARIMA at interval 5 day

Criterion	MSE	MAPE
Min:	71329.525	0.238
Avg:	637786112.725	17.025
Max:	3294047244.213	49.151

Table 21
Average MSE values of the best models for different intervals

Interval	GMDH-Neo-fuzzy	ARIMA	GMDH
1	170490988.697	132430397.71	52857867.97
3	315557184.523	497392547.163	309777397.448
5	284852296.188	530711958.519	398583396.613
10	342855400.208	730109790.172	461671534.933
20	321175540.062	707721137.741	467578939.296

Analyzing the presented results in the Fig. 10 and 11 one may conclude that GMDH method appears to be the best at short term forecasting 1, 3 days which complies the theory.

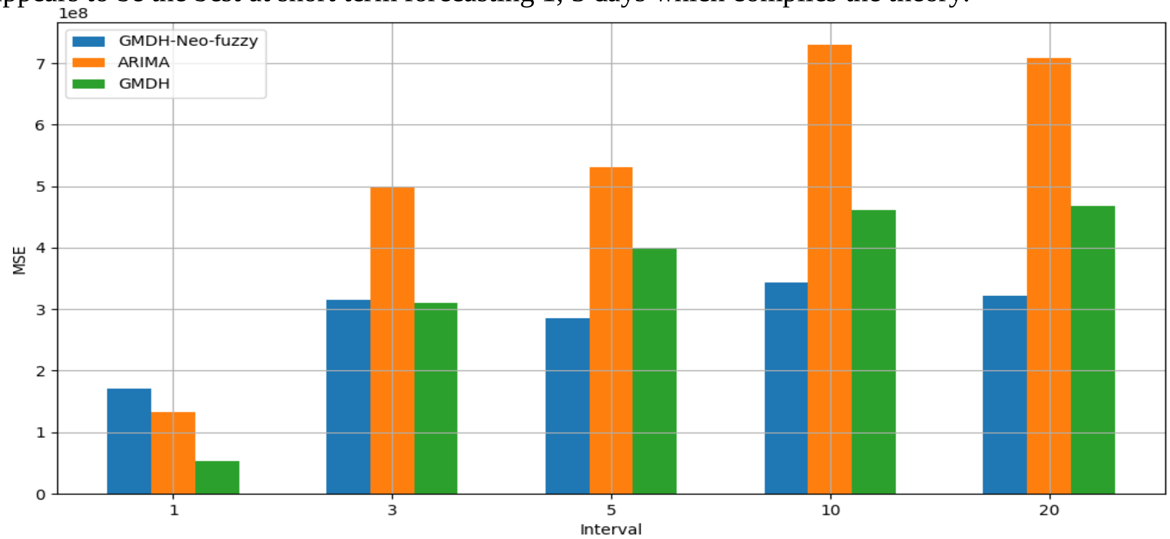

Figure 10: Average MSE values of the best models for different intervals

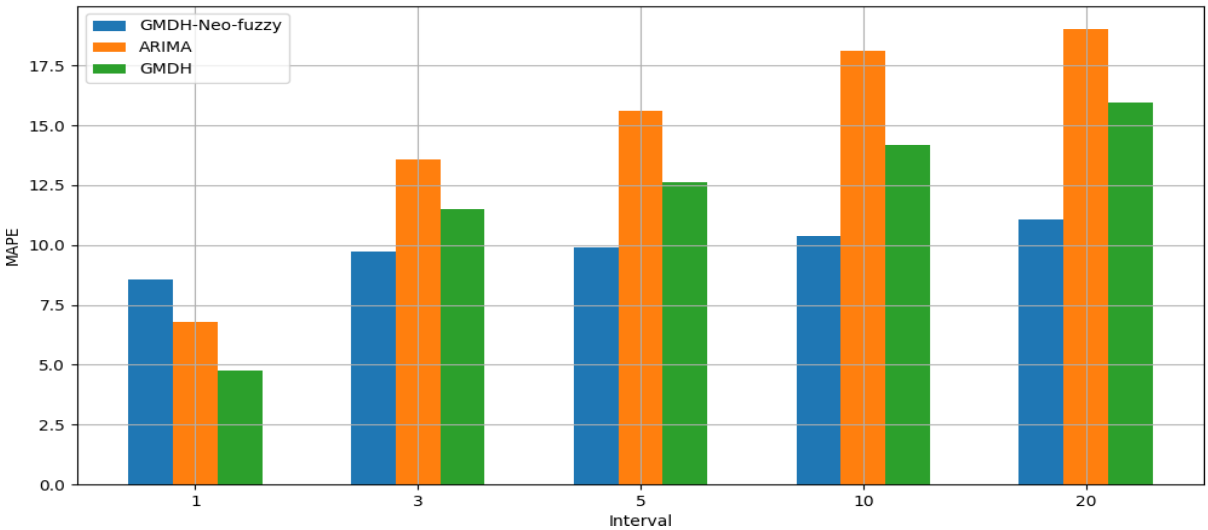

Figure 11: Average MAPE values of the best models for different intervals

Hybrid deep learning neo-fuzzy networks are the best at middle-term forecasting 5, 7, 10, 20 days. ARIMA appeared to be the worst by accuracy as compared with intelligent methods – hybrid DL networks and GMDH.

5. Conclusion

In this paper the investigations of artificial intelligence methods: hybrid Deep learning networks and GMDH and ARIMA were carried out in the problem of forecasting Electricity Sales to Ultimate Customers, Residential (USA) since 01-2002 till 01-2023.

Table 12
Average MAPE values of the best models for different intervals

Interval	GMDH-Neo-fuzzy	ARIMA	GMDH
1	8.571	6.768	4.767
3	9.723	13.581	11.518
5	9.914	15.609	12.615
10	10.359	18.132	14.202
20	11.064	19.024	15.95

During the experiments the optimal structure and optimal parameters: number of inputs, number of linguistic values, ratio training/test samples of hybrid neo-fuzzy networks were determined.

After optimization of hybrid neo-fuzzy networks and parameters of GMDH method the experiments on forecasting Electricity Sales to Ultimate Customers, were performed at different intervals: 1, 3, 5, 7 (short term forecast) and 10, 20 days (middle term forecast).

The accuracy of forecasting by Hybrid DL networks was compared with alternative methods – GMDH and ARIMA.

The analysis of obtained results have shown that GMDH method is the best at short term forecasting 1, 3 days while hybrid deep learning neo-fuzzy networks are the best at middle-term forecasting 7, 10, 20 days. Method ARIMA appeared to be the worst by accuracy as compared with intelligent method – hybrid DL networks and GMDH.

6. References

[1] Brockwell, Peter J. Introduction to time series and forecasting / Peter J. Brockwell and Richard A. Davis. – 2nd ed. Springer, 2002. – 429 p.

[2] Robert H. Shumway, David S. Stoffer. Time Series Analysis and Its Applications With R Examples. 4 th edition. Springer, 2017. – 562 p.

[3] A. G. Ivakhnenko, G. A. Ivakhnenko, J. A. Mueller, Self-organization of the neural networks with active neurons, Pattern Recognition and Image Analysis, 1994, 4, №2, pp. 177 188.

[4] A. G. Ivakhnenko, D. Wuensch, G. A. Ivakhnenko, Inductive sorting-out GMDH algorithms with polynomial complexity for active neurons of neural networks, Neural Networks, 1999, 2, pp. 1169 1173.

[5] S. S. Haykin, Neural networks: a comprehensive foundation, 2nd ed. Upper Saddle River, N.J: Prentice Hall, 1999.

[6] Osowski S. Neural networks for processing information. – Publishing House of the University of Technology Warszawska, Warszawa, (in Polish) – 2000.

[7] Wang F. Neural Networks Genetic Algorithms and Fuzzy Logic for Forecasting // Proc. Intern. Conf. Advanced Trading Technologies. – New York, 1992. – pp. 504 532.

[8] Yamakawa T., Uchino E., Miki T., Kusanagi H. A neo-fuzzy neuron and its applications to system identification and prediction of the system behavior // Proc. 2nd Intern. Conf. Fuzzy Logic and Neural Networks «LIZUKA-92». – Lizuka, 1992. – pp. 477 483.

[9] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT PRESS, 2016. URL: http://www.deeplearningbook.org.

[10] Yuriy Zaychenko, Yevgeniy Bodyanskiy, Oleksii Tyshchenko, Olena Boiko, Galib Hamidov. Hybrid GMDH-neuro-fuzzy system and its training scheme. Int. Journal Information theories and Applications, 2018. vol.24, Number 2. – pp. 156 172.

[11] Evgeniy Bodyanskiy, Yuriy Zaychenko, Olena Boiko, Galib Hamidov, Anna Zelikman. Structure Optimization and Investigations of Hybrid GMDH-Neo-fuzzy Neural Networks in Forecasting Problems. System Analysis & Intelligent Computing. Ed. Michael Zgurovsky, Natalia Pankratova. Book Studies in Computational Intelligence, SCI, vol.1022. Springer, 2022. – pp. 209 228.