Intelligent Identification System of the Process Liquid Solutions Composition

Oleksandr Bezsonov [1], Oleg Ilyunin [1], Alisher Khusanov [2], Oleg Rudenko [1] and Oleg Sotnikov [1]

[1] *Kharkiv National University of Radioelectronics, Nauky ave., 14, Kharkiv, 61166, Ukraine*
[2] *M.Auezov South Kazakhstan State University, Shymkent, Tauke Khan St., KZ-160012, Republic of Kazakhstan*

Abstract

Changing the composition of process liquid solutions is inevitable and affects directly on process sustainability, safety and reliability, resource efficiency, solutions' life span and harmful waste reduction. The authors share their experience gained in the development method of intelligent identification of the pickling solution composition for the metallurgical industry. The method of solution opaque salts concentration definition based on a fuzzy estimate of the concentration of salt components in technological liquid solutions by the number of pixels with certain digital color codes from digital images of control doses of the solution processed by the computer vision method is proposed. The values of the RGB pixel color estimates lie in fuzzy intervals for each of the salt components, which made it possible to build a fuzzy classifier. The presented procedure for obtaining a fuzzy classifier can be used to develop IT tools for operational monitoring of the technological liquid solutions state.

Keywords

Computer vision, classifier, c-means fuzzy method, membership function, solution's composition, concentration

1. Introduction

Standard task of different processes is to maintain an optimal level of concentration of an individual component of working solutions. The lack of timely and objective information about the concentration changing leads to excess consumption of solutions, reducing the efficiency of the equipment and even worse, the emergency shutdown of the process.

For example, these problems are in next processes: calcination of circulation water in district heating and the formation of deposits on the heat exchanger plates and the loss of sedimentary ferrous sulfate monohydrate in the bath and on the heat exchanger's plates of continuous pickling lines in metallurgy. Sulfuric acid is one of the most important industrial acids with over 190 million tones production in 2008 year with annual growth about 1.8% [1]. One of its most diverse area is steel pickling and descaling. The large manufacturers of steel production generally use hydrochloric acid for pickling processes, but although the amount of steel pickled by hydrochloric acid is more than pickled by sulfuric acid, the number of facilities using sulfuric acid exceeds those using hydrochloric acid [2].

Sulfuric acid pickling results pickling sludge, the waste product includes acidic water, iron sulfates, metallic salts and waste acid. Pickling sludge is the most hazardous waste as reported by EPA [3]. Continuous carbon steel pickling lines use the much water solution of sulfuric acid as pickling liquor. In addition, the solutions have to be regenerated in large volumes or dropped into the process water for neutralization with further water treatment and preparation of new fresh solution. All these measures

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: oleksandr.bezsonov@nure.ua (O. Bezsonov); oleh.iliunin@nure.ua (O. Ilyunin); alisher.khusanov@auezov.edu.kz (A. Khusanov); oleh.rudenko@nure.ua (O. Rudenko); olegsotnikov52@gmail.com (O. Sotnikov)
ORCID: 0000-0001-6104-4275 (O. Bezsonov); 0000-0002-7751-4814 (O. Ilyunin); 0000-0002-1563-6437 (A. Khusanov); 0000-0003-0859-2015 (O. Rudenko); 0000-0002-7848-4581 (O. Sotnikov)

increase the operation, resource and energy costs of the process and also pollute the production areas (acidifies and salts soil).

The carbon steel continuous sulfuric acid pickling process is carried out with pickling line that consists of several pickling baths. The process medium is the pickling solution (liquor) which is the diluted sulfuric acid. While diluted acid removes the oxides and other impurities it also attacks the steel surface and partially dissolves it. So it is very important to control such process parameters as dwell time, temperature and composition of pickling solution in baths. The only way to make the pickling process economically viable and ensure proper environmental protection is to develop a perfect control system. The pickling solution heating in special plate heat exchangers with temperature controller outside of pickling baths is progressive tendency in sulfur acid pickling [4, 5].

The analysis of pickling solution components in the pickling baths is significantly more complicated problem that directly connected with pickled carbon steel strip quality and pickling line capacity. Now widespread pickling line containing the regeneration node of the spent solution from, which meets the technical requirements for energy efficiency and control of emissions into the environment [6].

Fuzzy systems (FSs) originated from the fuzzy set theory proposed by Zadeh in 1965[7], which are based on fuzzy rules. As soft computing techniques, FSs have achieved great success in dealing with numerous problems of uncertainty, such as the prediction of the irrigation water infiltration rate [8], the seismic vulnerability assessment of buildings [9–11], the automatic classification of crop disease images [12], etc. FSs are more interpretable and intuitive methods that can match any set of input–output data by each fuzzy rule [13].

2. Problem analysis

Numerous mathematical models of salts crystallization in solutions often do not allow with sufficient accuracy to determine the development process. Sometimes the models include additional factors, taking into account the buildup of deposits with increasing number of crystallization centers of sediments [14]. The main disadvantage of these models is the absence in each case of any information about values of the constants. Existing methods and equipment of the solution composition control require additional laboratory analysis with mandatory operational personal participation or not sufficiently versatile in regards to the solution composition. Now for measurement of salts` ultimate permissible concentration (PC) in the pickling solution the direct chemical titration and non-destructive testing (NDT) or non-contact methods are used. Electric NDT method with high sensitivity and relative simplicity of the instrument implementation is one of the most informative. Opportunities for operational monitoring of the concentration of electrolytes in solutions (which are the solution of ferrous sulfate in sulfuric acid) open a direct conductivity measurement. But existing analyzers are designed for measuring the pickling solution of hydrochloric acid, which is quite stable ternary salt system. Analyzers provide a way to measure the overall pH in the temperature range of the solution $T=5\div80^{o}C$, but do not allow to determine the concentration of some components in aggressive sulfate environment.

For the case of sulfuric acid pickling one of the main challenges is the presence in pickling solution *FeSO_4* salts in three forms: $FeSO_4 \cdot 7H_2O$, $FeSO_4 \cdot 4H_2O$, $FeSO_4 \cdot H_2O$. Ferrous sulfate salts are resulted because of reactions of carbon steel strip scale and its iron part with sulfuric acid. The use of exact analytical relationships for the prediction of this process is difficult and not always correct. Mainly ferrous sulfate heptahydrate $FeSO_4 \cdot 7H_2O$ and ferrous sulfate monohydrate $FeSO_4 \cdot H_2O$ are considered for sulfur acid pickling. The ferrous sulfate tetrahydrate $FeSO_4 \cdot 4H_2O$ was not investigated so much. As the acid content of the pickling solution is decreased the ferrous sulfates content appropriately is increased. With adding sulfuric acid the pickling process is continued but ferrous sulfates content increases. Under certain temperature and concentration conditions the pickling solution will only hold a certain amount of dissolved salts and the excess will crystallize. For sulfuric acid pickling such conditions may be occurred when ferrous sulfate crystals will change from one type to another. It is called as phase change [15]. Phase change of ferrous sulfate heptahydrate (clear green coarse crystals soluble in hot water) to ferrous sulfate monohydrate (very fine white cement-like precipitate, non-soluble) begins when the temperature of the solution reaches $T=80^{o}C$ and leads to the

precipitation of monohydrate on heat transfer surfaces of heat exchangers or heating elements of pickling baths, baths walls etc., that does the pickling process difficult. At the metal plants of Ukraine, producing roll strip made of carbon steel, as a pickling solution traditionally used a solution of sulfuric acid with concentration c up to 20÷24% and temperature range of the solution $T=85\div98^{o}C$. In this range of parameters of the total presence of $FeSO\_4$ salts in pickling solution is capped by the process order at 15%, and the presence of $FeSO\_4 \bullet H\_2O$ is regulated by the limit of 4%. So the effective control of PC of ferrous sulfate salts is very important for proper pickling process operation suggestion.

Between modern non-contact composition identification system the ultrasonic one seemed perfect. But for case of sulfuric acid pickling this method has a drawback. Ultrasound can provoke the crystallization of salts in pickling solution about hundreds times more intensive than for usual conditions and it may to lead to unpredictable formation of $FeSO\_4$ *n*-hydrates that significantly decreases the quality of pickling liquor.

Effective control of continuous technological processes is possible only on the basis of the creation of more accurate models and methods, which must be insensitive to significant noise and measurement errors. Such requirements are met by intelligent methods for monitoring and identifying dynamic systems that are operated with significant uncertainty about the characteristics of the object and the environment, based on the combination of the principles of artificial neural networks and fuzzy control theory. Various aspects of the intellectual identification of objects and their parameters were studied.

However, there are three challenges at least in developing optimal FSs through the analysis of the research status:

- Optimizing FSs to achieve higher accuracy and faster convergence is now worthy of in-depth study.
- The number of fuzzy rules increases exponentially with the dimensionality of the input, which leads to the computation not being able to be completed within a reasonable time. Hence, it is difficult for FSs to deal with high-dimensional problems.
- As the number of fuzzy rules increases, the interpretability of FSs will be affected. Therefore, how to solve the high-dimensional problem on the basis of ensuring better interpretability is one of the current bottlenecks. In recent years, great efforts have been made to improve FSs [16].

The adaptive neuro-fuzzy inference system (ANFIS) formed by the gradient descent (GD) algorithm plus least squares estimation (LSE) has been widely applied in national energy demand forecasting [17], geographic temperature forecasting [18], and so on.

However, the convergence speed of GD or GD plus LSE is very slow, and they easily fall into local optimal solutions.In view of the above problems, new techniques based on the ANFIS optimized by EAs have been emerging. For example, a method of optimizing the ANFIS based on an improved firefly algorithm and the differential evolution optimization algorithm was proposed by Balasubramanian [19], which has been proven to be effective in the application of medical disease prediction. Ehteram et al. used three optimization algorithms (sine– cosine algorithm, particle swarm optimization algorithm, and firefly algorithm) to optimize the ANFIS to improve the prediction accuracy [20]. However, all the above methods only focus on the optimization of FSs and still have challenges in solving high-dimensional data. The ANFIS with PCA were combined to predict the Iranian ionospheric time series, which can shorten the convergence time and obtain the optimal solution [21]. It was solved [22] the problem of predicting the fracture pressure of defective pipelines more effectively through the combination of PCA and the ANFIS.

The structure of the ANFIS is composed of an adaptive neural network and an FS. It not only inherits the adaptive learning ability of neural networks, but also keeps the interpretability of FSs. The ANFIS can adjust the parameters according to prior knowledge so that the predicted values are closer to the target values, which has achieved great success in many applications [23]. The SC-ANFIS is proposed in this paper, which applies subtractive clustering (SC) to construct a Sugeno fuzzy inference system in the ANFIS. The SC-ANFIS can effectively avoid the combinatorial explosion of fuzzy rules when the dimensionality of the input is very high. In addition, the fuzzy rules generated by SC are more consistent with the data than those obtained without clustering. The input space can be divided appropriately, and the number of membership functions (MFs) and the parameters for each input domain can be reasonably determined [24].

Two methods to construct fuzzy inference systems are known well:

- fuzzy c-means clustering;
- grid partitioning [25].

It has been proven that SC is better than other algorithms [26], and it was adopted as the method to generate the fuzzy inference system in the ANFIS.

Furthermore, relevant studies have demonstrated that the performance of the Gaussian MF is better than others in many nonlinear complex problems [27,28].

Taking into account considered above, the present paper is devoted to development of low-cost intelligent identification system of the process liquid solutions composition based on fuzzy c-means clustering. The system is considered as one of the ways to control the concentration of *FeSO_4* in the pickling solution and the consumption of sulfuric acid for the continuous process of pickling carbon steel.

3. Fuzzy estimation procedure

3.1. Description of process, model and identification system

In the task of selective color segmentation, there are no reliable and fully automated methods yet. To solve them in a natural way, fuzzy soft approximations can be used [29]. Ferrous sulfate salts which affect the quality of the pickling solution, are presented as three compounds [30] (see Table 1).

Table 1

Color classifier of FeSO_4•n-hydrates

k	Chemical formula	Crystal	Formation temperature, ºC	Color cluster	LR-interval estimation of RGB - code of color		
					R_{L-R}	G_{L-R}	B_{L-R}
A	FeSO_4•7H_2O	Melanterite	21<	clearly green	81÷145	147÷183	81÷111
B	FeSO_4•4H_2O	Rosenite	>= 21	green	0÷2	128÷125	64÷82
C	FeSO_4•1H_2O	Szomolnokite	>= 80	white	245÷255	245÷255	245÷255

For definite values of bath temperature and pickling liquor flow rate through heat exchangers the concentration change Δc of *FeSO_4* influences on pickling line processing efficiency. These changes estimated with fuzzy intervals are significant disturbance factors of uncertainty for the process.

This dynamic process may be characterized with monotonic function $C_{FeSO\_4}(t)$ non-decreasing on interval $(0; t_0)$ and limited with the value $sup(c(t))= c$.

If summary PC of *FeSO_4 •nH_2O* in liquor reaches $C_{FeSO\_4}(t_0) = c$ the system becomes ineffective or stops abnormally. The control action for such situation is the liquor refreshment (Fig.1).

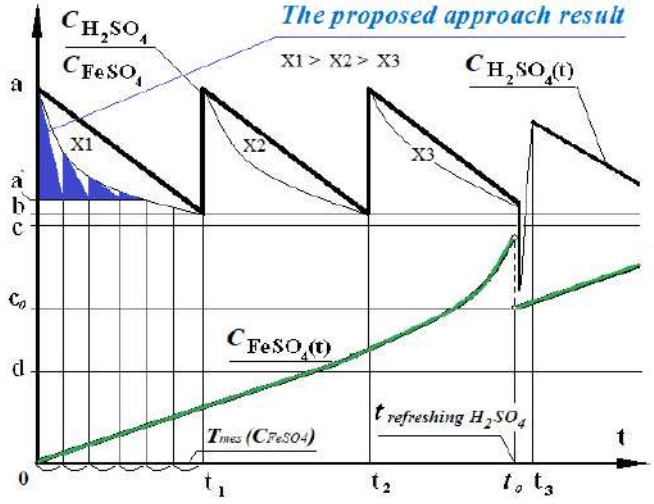

Figure 1: Pickling solution refreshment

The part of liquor is directed to regeneration unit and fresh regenerated pickling liquor is supplied to bath till reaching scheduled $C_{FeSO\_4}(t_0+\Delta t_r)= c_0$. The proposed approach saves the solution in process up to 25%, and, respectively reduces the harmful emission and working solution consumption. The system still is effective in case PC of *H_2SO_4* belongs to interval $C_{H\_2SO\_4} = (a`; a)$ and becomes ineffective or stops in case PC reaches value $C_{H\_2SO\_4} = b$. Segments X1, X2, X3 in Fig.1 represent an integrated assessment of savings flow solution.

Control actions are generated on the base of monitored *c(t)* values (partial or complete refreshment or full update of pickling bath solution, heat exchanger cleaning with turning on the back-up one). As a Mamdani model, the modeled object is represented as a black box, characterized by a lack of information about what physical phenomena is happening inside itself [31]. The development of the Mamdami model is the formation of the production rule set with a minimum number *k* that describe a mapping of its inputs (the vector X_n) to the output *Y*, which will provide the most accurate approximation of the real system in the sense of average absolute error [32]. Each rule specifies some fuzzy point in the display space defined Cartesian product $X1 \times X2 \times \dots X_n \times Y$ in the next form:

$$\begin{gathered} R_1: IF\ (x_1 \approx a_1)\ AND\ (x_2 \approx a_2) \dots\ AND\ (x_n \approx a_n)\ THEN\ (y_1 \approx b_1) \\ \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\ R_k: IF\ (x_1 \approx z_1)\ AND\ (x_2 \approx z_2) \dots\ AND\ (x_n \approx z_n)\ THEN\ (y_k \approx b_k) \end{gathered} \tag{1}$$

A fuzzy graph in the display space is formed from a set of fuzzy points. The mechanism of interpolation between points depends on the used tools of fuzzy logic.

The approach of active salts concentration *c(t)* measurements in the pickling bath based on fuzzy estimation of this value with using the classifier is proposed.

In general, the problem of classification is to determine the size and location of the membership functions of the individual classes in the feature space $X1 \times X2 \times \dots X_n$. The structure of classifier is presented below:

$$\langle CL, X, ST, R(Y, x)\rangle, \tag{2}$$

where *CL* – class label < color of the salts in the test solution volume >, *X* – < valid interval characteristic values in the Red-Green-Blue (RGB) color coordinate space >, *ST* – < area of the salts in the test solution volume >, *R(Y, x)* – is a set (defined on the set *X*) of fuzzy interval estimations of *Y* – <values of parameters measured in series of *N* concentration measurements>.

The process of knowledge extracting about the object [33] is to obtain statistics of the measurement inputs – (*X*, *ST* and *R*) and output (*c*). The parametrization of the exponential membership functions is given below (paragraph 3.2).

The proposed intelligent identification system is implemented in the production site of control of solution parameters in the following composition:

- light source of constant calibrated spectrum, which depends on the optical properties of the components of the solution placed under the cuvette;
- a cuvette for a solution that is automatically washed before each measurement with a given period of T_{mes} = 45 min;
- dispensers;
- digital photo-recorder-analyzer FESTO SBOC-Q for 6 million pixels (with a matrix of 1280 x 1024 pixels, shooting speed 150 frames/s) with a data channel (USB, RS-485, HDMI), which forms graphic image files of the pickling solution in the cuvette in graphic format.

After the time interval equal to $0{,}1 \cdot T_{mes}(C_{FeSO\_4})$ (Fig.1), the dosed volume of the solution, circulating through the technological circuit "bath-heat exchanger-bath", is taken and placed in a cuvette.

The solution dose is analyzed using an optical microscope and a digital camera that generates 2D graphic files in .JPG and .PNG format. These files are processed by the computer vision methods image identification system, developed in Open Source Computer Vision Library [34]. After recording the obtained measurements in the database to build a classifier, the same dose of the solution is analyzed by the dry residue method to obtain reference salt concentration values. The method of optical segmented color identification used in the system [35] was partly applied to instantaneously measure mineral concentrations in flotation froths and other stages of beneficiation processes [36].

The system of ferrous sulfates *n*- hydrates image identification according to color and occupied area has to analyze automatically the images and classify each pixel of the image according four objects

(*FeSO_4 •7H_2O, FeSO_4 •4H_2O, FeSO_4 •H_2O* and pickling solution background color). It calculates the volumetric parts of analyzed objects (%), their weight parts (%), and their surface area per unit of volume (sm^2 /sm^3). The system is able to analyze up to 256 images step by step, keep them in memory as table and present these data as graphical plots for dispersion. The presentation of Graphic User Interface (GUI) of intelligent identification system of the process liquid solutions composition is illustrated in Figure 2.

Each measurement from *m* measurements, that is equivalent to fixed state of pickling line, is a quantitative estimation V_k of the *k*-th hydrate presence: the surface occupied by appropriate color is calculated according to pixels with color code number $V_k(RGB)$ taking into account:

$$V_k = S_k \cdot V_k(RGB) \tag{3}$$

where $V_k(RGB)$ describes the membership of a pixel in the *k*-th color cluster according to the RGB code LR-interval (see Table 1) and takes the values 0 or 1 in the case of clear visibility of the color class, S_k – the number of pixels allocated to the *k*-th color cluster, defined by images analysis results.

In general, during the setting of classifier the boundaries of the classes definition region can change in the space X_n. This leads to a situation where two classes are almost identical and cannot be distinguished. The classical evaluation of an object x_j belonging to a specific *i*-cluster is the smallest Euclidean distance between estimated object and centers of classes c_i selection:

$$d_{ij} = \|x_j - c_i\| = \sqrt{\sum_{l=1}^{n} \left(c_{l,i} - x_{l,j}\right)^2} \tag{4}$$

where c_i – is the coordinate of the cluster center relative to the axis OX_l in space X_n, x_j – is the coordinate of the classified object relative to the OX_l axis.

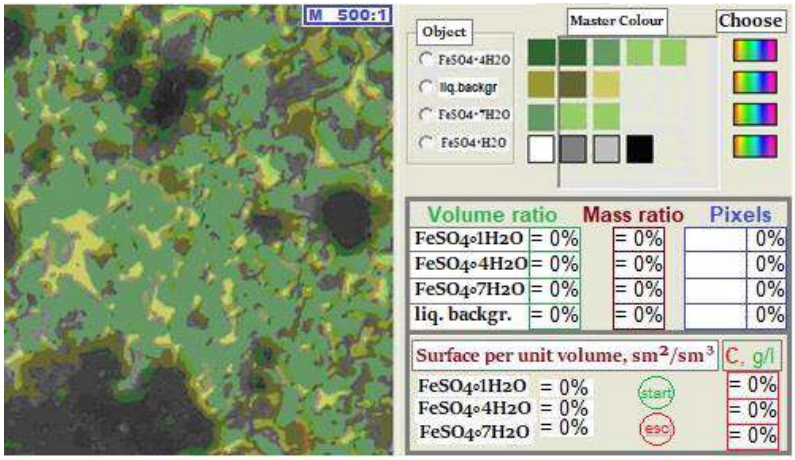

Figure 2: The presentation of system GUI

On the basis of statistics estimates of distances it is possible to construct radial membership functions $\mu(V_k(RGB))$ [37], although it is not always effectively in case of large dimension X_n. The proposed method uses a fuzzy classification with two-dimensional projections of multidimensional clusters [32]. Classes are clearly separated from each other and holding the classification is not difficult in our case.

Simultaneously the test titration to measure $C_{FeSO\_4}(t_i)$ and measurement of $V_k(t_i)$ are carried out at the t_i moment. Also *l* series of N_l reference measurements of concentration of the *k*-th salt $c_k(t_i)$ is made by gravimetric analysis of the dry residue with satisfactory accuracy δ_t.

On the base of statistical data on the state of salts in solution, obtained by the method of fuzzy classification, the membership $\mu(V_k(t))$ functions are formed.

The membership function $\mu(x)$ establishes the accordance between actually measured values: $V_k(t)$ – pixels number of and $C_k(t)$ – concentration of the *k*-th salt, establishing next relations:

$$V_k \rightarrow C_k \tag{5}$$

3.2. Model configuration

The formation and configuration of the classifier is made under the following assumptions: the *k*-th salt is clearly related to the class on the basis of color, are known its footprint with a relative accuracy of 1 pixel to 10^6 pixels (feature digital camera FESTO SBOC-Q) and mass, the specific volume resistance of mist in the test samples of dimensions. Given that the measurement accuracy of the input parameter of 0.005%, is not difficult to split samples of measurements between clusters. Function $C_{FeSO\_4}(t) = f(V_k(t))$ is considered to be one-dimensional, monotone and non-decreasing. Measurements are taken at equal time intervals. The configuring of classifier is based on a common clustering *k*-means method for one-dimensional case [32], the algorithm of the method is given in the article.

3.2.1. Definition of clusters' centers

The number of clusters in each of the *l* series, consisting of *N* measurements, is determined empirically – according to the frequency of manifestations and by the modal number of input and output values in a series of measurements. Each element of the sample of measurements is included in the closest cluster. The inclusion of elements into clusters can be done on the basis of formula (4).

The median (center) of the *i*-th cluster in the input coordinate of *v* is determined by the cumulative [32], taking into account all of the included elements according to the formula:

$$m_{V_i}(t) = \frac{1}{N_i}\sum_{j=1}^{N_i} v_j\,(t), i = 1, \dots, l \tag{6}$$

3.2.2. The method of weighted averages.

Three experimental series of a total elements number of 48 measurements (v_j, c_j) for $FeSO\_4 \bullet H\_2O$ at intervals of input values ε_i, the medians of the input and output values are represented on Fig.3.

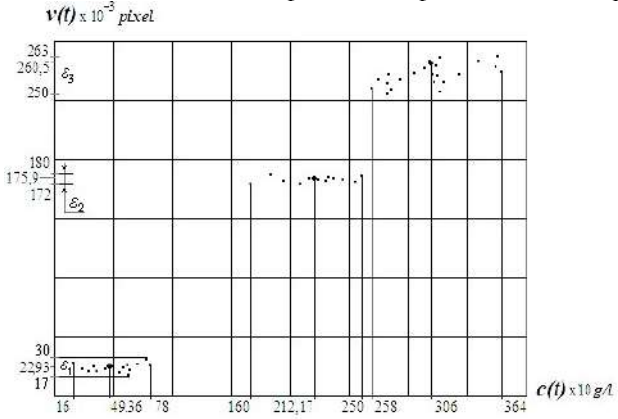

Figure 3: Data of three measurements series with input and output values medians

A series of *n* measurements forms a single cluster. As maximal close to v_i the *i*-th cluster is defined according to formula (4):

$$MIN(d_{ij}) = \left\| m_{V_j} - v_i \right\|, (j = 1, \dots, l) \tag{7}$$

Representation of c_i* in a neighborhood of a point $v_i = v_i \pm \delta$ can be calculated by using Gaussian MFs according to the formula:

$$c_i^*(v_i) = \frac{\sum_{j=1}^{n} c_j \cdot exp\left[\frac{-(v_i - v_j)^2}{2\delta^2}\right]}{\sum_{j=1}^{n} exp\left[\frac{-(v_i - v_j)^2}{2\delta^2}\right]} \tag{8}$$

where (v_j, c_j) – are the points of the real n measurements, δ – configurable value of the variation of the cluster membership function, which typically assigns a value:

$$\delta = \frac{1}{3} \cdot MIN \left\| m_{V_j} - v_i \right\|, \forall v_j \in \left(m_{V_j} - 0{,}5\varepsilon_j; m_{V_j} + 0{,}5\varepsilon_j \right) \quad (9)$$

where ε – is the size of the interval in which the cluster lies.

3.2.3. Calculation of displacements of clusters

The calculation of the distance of the displacement of the clusters centers relatively to their positions in the previous cycle clustering *(t-1)* is carried out in accordance with following expression:

$$\Delta m_i(t) = \| m_i(t) - m_i(t-1) \| \quad (10)$$

The compressed data is presented in table 2.

Table 2

The measurement sample, the distances between elements P_j and the cluster centers m_i.(1)

Cluster $_i$	*1*			*2*			*3*		
P_j	1	2	3	4	5	6	7	8	9
$V=x_1$	28	22	26	176	178	180	254	250	262
$C=x_2$	16	50	74	175	212	245	273	306	360
d_{1j}	***31.06***	***5.00***	***27.00***	197.19	224.34	250.84	321.03	342.43	392.00
d_{2j}	245.23	223.46	203.96	***35.06***	***2.00***	***35.06***	98.72	120.00	171.92
d_{3j}	373.82	351.37	331.00	159.01	127.00	101.24	***40.01***	***8.60***	***47.52***
$m_i(1)$ (v,c)	(25.33; 46.67)			(178; 210.67)			(255.33; 313)		

The result is $\Delta m_1 = 0.75$, $\Delta m_2 = 0.67$, $\Delta m_3 = 0.33$, each of the values exceeds the threshold value shift of the cluster center (δ_m =0.05). If the minimum offset Δm_i does not satisfy the specified condition:

$$\min(\Delta m_i(t)) \le \delta_m, \qquad i = 1, \dots, k \quad (11)$$

we will return to the allocation of elements to clusters. Otherwise, the adjustment of cluster centers m_i ends and moves to the next step. The condition (11) is not fulfilled and another iteration is carried out with results shown in table 3:

Table 3

The distances between elements P_j and clusters centers m_i.(2)

Cluster $_i$	1			2			3		
P_j	1	2	3	4	5	6	7	8	9
d_{1j}	***30.79***	***4.71***	***27.34***	197.91	225.04	251.51	321.74	343.12	392.67
d_{2j}	245.76	223.94	204.41	***35.73***	***1.33***	***34.39***	98.29	119.46	171.33
d_{3j}	374.02	351.58	331.23	159.18	127.20	101.48	***40,02***	***8.80***	***47.47***
$m_i(2)$ (v,c)	(25.33,46.67)			(178, 210.67)			(255.33, 313)		

$m_1(2)$ = (25.33, 46.67), $m_2(2)$ = (178, 210.67), $m_3(2)$ = (255.33, 313)

The offset of values of the cluster centers satisfies the accuracy requirements: Δm_1 (2) = 0.0047, $\Delta m_2(2)$ = 0.0033, Δm_3 (2) = 0.0033.

3.2.4. The definition and parameters of clusters' membership functions

Given the uneven distribution of cluster centers in the input space, as well as possible increasing the dimension of the model, it is appropriate to use Gaussian MFs [28] of the formula (9):

$$\mu_{V_i}(x) = exp\left[\frac{-(m_i - x)^2}{2{\delta_i}^2}\right], \quad i = 1, \dots, k \quad (12)$$

In the absence of measurements in a neighborhood of the coordinate system origin the dummy cluster, completely excluding any element, is declared (m_{V0} =0.00). The value of δ_i is chosen to be equal one-third the distance | m_i - m_{i+1}| in the input space *V* (in accordance to the 9-th formula). The values of δ_i are:

$$\delta_1' = \frac{|0 - 25.33|}{3} = 8.44, \qquad \delta_1'' = \frac{|178 - 25.33|}{3} = 50.89,$$
$$\delta_2 = \delta_3 = \frac{|178 - 255.33|}{3} = 25.78 \tag{13}$$

The exception (13) for the *1*-st cluster is transformed in the asymmetric Gaussian function:

$$\mu_{V_1}(v) = exp\left[\frac{-(m_1 - v)^2}{2(\omega \cdot \delta_1' + (1-\omega)\cdot \delta_1'')^2}\right] \tag{14}$$

where $m_1 = m_{V1}$ =25.33,

ω - auxiliary Boolean variable of the next form:

$$\omega = \begin{cases} 1 & for \quad 0 \le v \le m_1 \\ 0 & in \quad other \ cases \end{cases} \tag{15}$$

3.2.5. Construction of fuzzy models C(V)

As a result of the clusters projection in the input space *V* m_{V1} = 25.33, m_{V2} = 178, m_{V3} = 255.33 are obtained. Output single-point MFs are defined in points m_{C1} = 46.67, m_{C2} = 210.67, m_{C3} = 313. The membership functions of input and output parameters are presented in Fig.4.

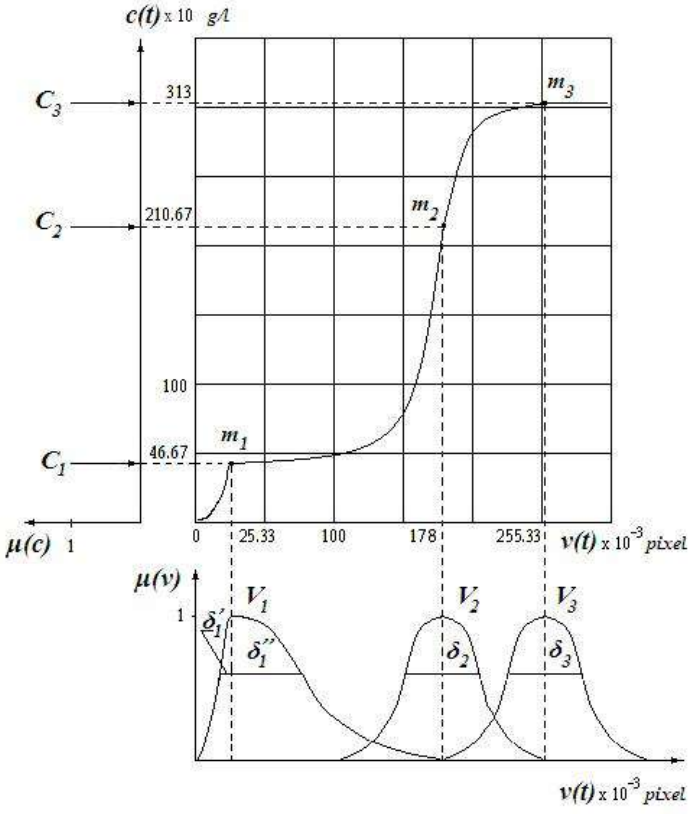

Figure 4: Resulting fuzzy dependence of *c(v)* the resulting clustering and membership functions

The value *c(v)* at the output of the fuzzy model is calculated by the formula:

$$c(v) = \frac{46.67 \cdot \mu_{V1}(v) + 210.67 \cdot exp\left\{\frac{-(v-178)^2}{664.44}\right\} + 313 \cdot exp\left\{\frac{-(v-255.33)^2}{664.44}\right\}}{\mu_{V1}(v) + exp\left\{\frac{-(v-178)^2}{664.44}\right\} + exp\left\{\frac{-(v-255.33)^2}{664.44}\right\}} \tag{16}$$

After definition of such dependencies for all kinds of salts from Table 1 and the dependence of the form *c(V)*, we have the possibility to carry out the operational control of salts' PC, and indirectly to determine the dynamic changes in the concentration of pickling solution C_s with the recalculation of the mass of the *SO_4* groups in *FeSO_4 n*-hydrates.

Within the cluster, the integrated function (16) is implemented using Gaussian MFs by the FCM method (Fuzzy Classifier Means), v_i and v_{i-1} are accepted as the centers of neighboring classes. The current value of $v(t) = x$, obtained during the execution of TP, belong to the *i*-th class by the criterion of the minimum Euclidean distance from the estimated object to the center of the *i*-th cluster - v_i.

4. Discussions

Simultaneous accounting of the number of pixels of certain colors using computer vision tools corresponding to different types of salts in a solution makes it possible to fuzzy estimate the concentrations of several components of the solution at the same time. The proposed approach is based on measuring changes in the physical properties of solutions caused by the change in the concentration of the individual components. The obtained dependencies can form the basis of software converter for fuzzy estimation of the concentration of salts in solution. Using equation (8) it is possible to find a representation of *c** at an arbitrary point in *v*, which is close to the sample measurement. The model applied in the system is static due to the low speed of the process. In high-speed multi-reagent processes, the problem of short-term prediction often arises. By introducing a time coordinate [22] and/or *m* additional parameters into the space of arguments, it is possible to obtain a predictive model that works in a short prediction period Δt according to the following rules:

$$\begin{gathered} R_1: IF\ (x_1(t) \approx a_1)\ AND (x_2(t) \approx a_2) \dots AND (x_{n+m}(t) \approx a_{n+m})\ THEN\ (y_1(t+\Delta t) \approx b_1) \\ \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\ R_k: IF\big(x_1(t) \approx z_1(t)\big)\ AND (x_2(t) \approx z_2) \dots AND (x_{n+m}(t)) \approx z_{n+m})\ THEN\ (y_k(t+\Delta t) \approx b_k) \end{gathered} \tag{17}$$

The surface purity coefficient – q_e, which generally takes into account the area and thickness of the pickled defects [38], and hence the mass of iron oxides that have reacted with the pickling solution, can be used as an additional parameter for constructing an ANN with the indicated properties. The study points in the neighborhood of PC (v, c^*_{PC}) will provide an opportunity to synthesize ANFIS [39].

In the case of a large number of measurements, the data are compacted by the method of weighted averages, replacing them with pairs (v^*_j, c^*_j) at 9 points with intervals ε_i, for the medians of the input and output values and the formation of 3 clusters for our example.

Also, to estimate the content of different salts of multihydrates in the solution, it is proposed to use the estimation of the brightness of individual points by *RGB* components, to calculate the number of pixels of color components in digital images of the analyzed surface as *Y* (formula 2) [40]:

$$Y = 0.299R + 0.587G + 0.114B \tag{18}$$

where *Y* – is the brightness in shades of gray, *RGB* estimates are in the range 0÷255.

It should be noted, that the greatest influence on the process has a change in the concentration of *FeSO_4 •H_2O*. The model for this type of monohydrate must take into account the errors associated with the presence of measurement noise. The following is proposed for this purpose.

The range of initial values of the monohydrate *FeSO_4 •H_2O* concentration *C* is divided into segments having the same width, due to the required accuracy of measurements δ_{C1}. The formation of the neurofuzzy model is carried out based on the results of experimental serial measurements in M reference points of the model – cluster centers (V_n, C_n) with uniform sampling step and segments of values ($(n-1)\delta_{C1}$, $n\delta_{C1}$), n= 1,2,… *M*, where *M* is estimated as:

$$M = [C_n/\delta_{C1}] \tag{19}$$

Fuzzy sets C_n, formed at the output of the model in the process of data accumulation are replaced by singletons (cluster medians), which coincide with the modal values $C_1 = \delta_{C1}$, $C_2 = 2\delta_{C1}$,…, $C_n = n\delta_{C1}$, to the maximum allowable value according to the process requirements:

$$C_n = \sup\big(c(t)\big) = c_0 \tag{20}$$

5. Conclusions

Procedures for the identification and control of the pickling solution components PC are formalized using elements of fuzzy logic. The identification of salt compounds in the pickling solution is solved by the method of selective optical segmenting by color and build of concentrations components classifiers using Gaussian membership functions. The expansion of the arguments space will allow to increase the sensitivity of the classifier and to conduct a more detailed clustering of states, characterizing quantitative regularities of changes in the component composition of the solution. Obtained plots may be used for further conversion of measured physical parameters into solution concentration, process efficiency and some other characteristics.

Finally, the obtained intelligent identification system of the process liquid solutions composition based on analytical calculation fuzzy model may be realized as an ARM microcontroller with high-speed graphic coprocessors of the Cortex series using the open software library OpenCV (Open Source Computer Vision Library) for image segmentation to enhance the automation procedure [41]. The new system has a direct influence on the improvement of the process sustainability safety and reliability and reduces the operation costs pollution and overall production expenses.

The implementation of the proposed system, which is built into the overall control system of the pickling line, will enable more accurate on-line automated control of the concentration of pickling solution components.

Approbation of the proposed intelligent identification system (in combination with the existing pickling solution temperature control system) showed the following results. Acid consumption for the pickling process is reduced from 1230 to 976 kg/h approximately while maintaining the same capacity of the pickling line. This is due to an increase in the accuracy and frequency of measurements ($T_{mes}(C_{FeSO\_4})$ – Fig.1). Accordingly, the load on process water purification from sludge and regeneration of sulfuric acid and the consumption of process water for the preparation of pickling solution decreased by 26% [38]. Figure 5 shows the graphs of sulfuric acid mass consumption rate for the work shift G_b – before the application of the intelligent identification system, and G_{rbn} – after.

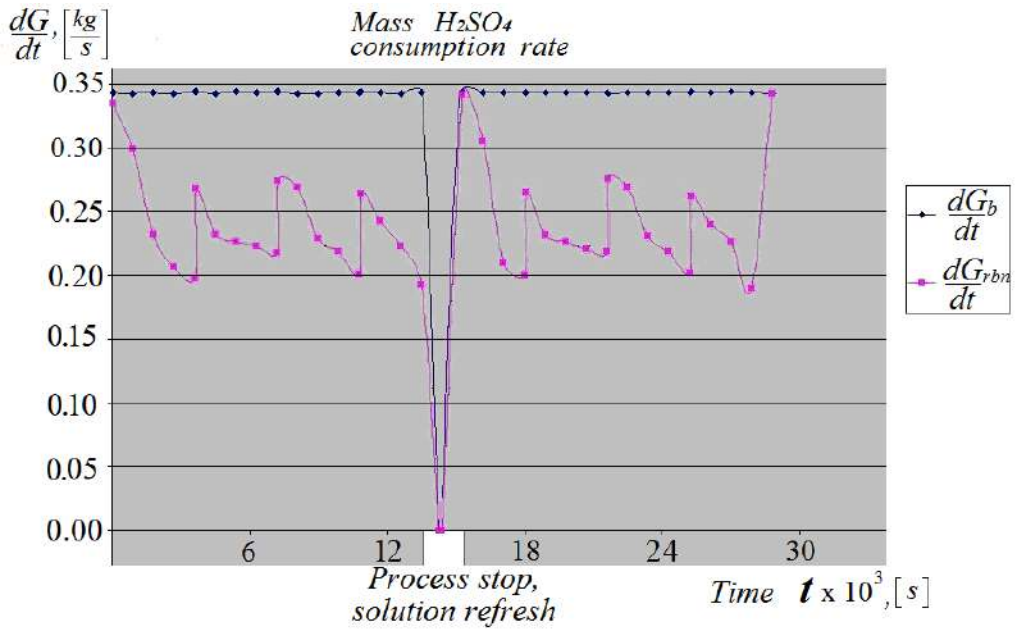

Figure 5: Graphs of sulfuric acid consumption

6. Nomenclature

ANN – artificial neural network.

ANFIS – Adaptive neuro-fuzzy inference system.
EPA – United States Environmental Protection Agency.
EAs – Evolutionary algorithms.
FSs – Fuzzy systems.
GD – Gradient descent.
LSE – Least squares estimation.
MFs – Membership functions.
NDT – non-destructive testing method of measurement.
PC – value of ultimate permissible concentration of salt in pickling solution, (g/l).
PCA – Principal component analysis.
SC – Subtractive clustering
T – temperature of the solution, ºC.
$C_{FeSO\_4(t)}$ – monotonic non-decreasing time function of FeSO_4 – salt concentration change, (g/l).
C_s – concentration of pickling solution, (g/l).
G – sulfuric acid mass consumption, (kg)
S_k – the number of pixels allocated to k-th color cluster.
V_k – quantitative estimation of the k-th hydrate presence in pickling solution.
$V_k(RGB)$ – boolean value, describes the membership of a pixel in the k-th color cluster according to the RGB code LR-interval.
d_{ij} – Euclidean distance between object x_i and the i-th cluster center c_i.
Δc – concentration change of a salt in pickling solution, (g/l).m_i – the i-th cluster center.
Δt – short prediction period of time (s).
q_e – surface purity coefficient (%).
δ_t – satisfactory accuracy of *FeSO_4* concentration measurement by traditional method.
δ_{C1} – required accuracy of *FeSO_4 •H_2O* concentration measurement.
δ_m – the set threshold value shift of the m_i in the arguments' space.
ε_i – interval of the i-th cluster's input values v.
$\mu(V_k(t))$ – membership function takes the values on interval [0, 1], establishes the accordance between really measured values of $V_k\,(t)$ and values of the k-th *FeSO_4* n-hydrate concentration $c_k(t)$ in pickling solution.

7. References

[1] Breuer, D., Heckmann, P., Gusbeth, K. et al. "Sulfuric acid at workplaces – appliciability of the new Indicative Occupational Limit Value to thoracic particles." Journal of Environmental Monitoring, 14 (2012):440-445.
[2] Recovery of spent sulfuric acid from steel pickling operation, *EPA Technology Transfer Capsule Report*, EPA-625/2-78-017, 30 p. URL: https://nepis.epa.gov/Exe/ZyNET.exe
[3] Sulfuric Acid and Ferrous Sulfate Recovery from Waste Pickling Liquor, *EPA Technology Transfer Capsule Report*, EPA-660/2-73-032, 82 p.
[4] European Commission, Reference Document on Best Available Techniques for the Ferrous Metals Processing Industry, December, 2001, 538 pp.
[5] Kapustenko, P., Perevertaylenko, A., Khavin, G., Arsenyeva, O. "Graphite Plate Heat Exchangers as Energy Saving Tool for Corrosive Media Duties." Chemical Engineering Transactions, Vol. 12, ICHEAP'08/PRES'07, Milano: Elioticinese Service Point Srl., 2007, pp.219-224. URL: https://folk.ntnu.no/skoge/prost/proceedings/icheap8pres07/pres07webpapers/114%20Kapustenko.pdf
[6] Martines, S., Marelli, S. "Modernisation of Pickling lines at the Magnitogorsk." Irons & Steel Works, 12 (2006): 233-236.
[7] Zadeh, L.A. Fuzzy sets. Inf. Control. 1965, 8, 338–353.
[8] Mattar, M.A.; El-Marazky, M.S.; Ahmed, K.A. Modeling sprinkler irrigation infiltration based on a fuzzy-logic approach. Span. J. Agric. Res. 2017, 15, 1–10.

[9] Harirchian, E.; Lahmer, T. Improved Rapid Assessment of Earthquake Hazard Safety of Structures via Artificial Neural Networks. IOP Conf. Ser. Mater. Sci. Eng. 2020, 897, 012014.
[10] Harirchian, E.; Jadhav, K.; Kumari, V.; Lahmer, T. ML-EHSAPP: a prototype for machine learning-based earthquake hazard safety assessment of structures by using a smartphone app. Eur. J. *Environ. Civ. Eng. 2021, 3,* 1–21
[11] Harirchian, E.; Hosseini, S.E.A.; Jadhav, K.; Kumari, V.; Rasulzade, S.; I̧sık, E.; Wasif, M.; Lahmer, T. A review on application of soft computing techniques for the rapid visual safety evaluation and damage classification of existing buildings. J. Build. Eng. 2021, 43, 102536.
[12] Fan, T.; Xu, J. Image Classification of Crop Diseases and Pests Based on Deep Learning and Fuzzy System. Int. J. Data Warehous. Min. 2020, 16, 34–47.
[13] Li, J.S.; Gong, Z.T. SISO intuitionistic fuzzy systems: IF-t-norm, IF-R-implication and universal approximators. IEEE Access. 2019, 7, 70265–70278.
[14] Bansal, B., Cheng, X.D., Muller-Steinhagen, H. "Analysis of "classical" deposition rate low for crystallization fouling." Chemical Engineering and Processing 47 (2008): 1201–1210.
[15] Hasler, F., The Whys and Hows of Sulfuric Acid Pickling and Recovery, *Esco Engineering*, 1992, revised February 1995, 61 pp.
[16] Cui, Y.Q.; Wu, D.R.; Huang, J. Optimize TSK Fuzzy Systems for Classification Problems: Minibatch Gradient Descent with Uniform Regularization and Batch Normalization. *IEEE Trans. FS. 2020, 28*, 3065–3075.
[17] Panapakidis, I.P.; Dagoumas, A.S. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. Energy. 2017, 118, 231–245.
[18] Kisi, O.; Demir, V.; Kim, S. Estimation of long-term monthly temperatures by three different adaptive neuro-fuzzy approaches using geographical inputs. J. Irrig. Drain. Eng. 2017, 143, 04017052. J. Intell. Eng. Informatics. 2020, 8, 374.
[19] Balasubramanian, K.; Ananthamoorthy, N.P. Improved adaptive neuro-fuzzy inference system based on modified glowworm swarm and differential evolution optimization algorithm for medical diagnosis. Neural Comput. Appl. 2021, 33, 1–12.
[20] Mohammad, A.; Fang, Y.T.; Ali, N.A.; Sarmad, D.L.; Yuk, F.H.; Osama, A.; Nadhir, A.A.; Ahmed, E.S. Performance improvement for infiltration rate prediction using hybridized Adaptive Neuro-Fuzzy Inferences System (ANFIS) with optimization algorithms. Ain Shams Eng. J. 2021, 12, 1665–1676.
[21] Razin, M.; Voosoghi, B. Ionosphere time series modeling using adaptive neuro-fuzzy inference system and principal component analysis. GPS Solut. 2020, 24, 1–13.
[22] Hcp, A.; Htd, B. Predicting burst pressure of defected pipeline with Principal Component Analysis and Adaptive Neuro Fuzzy Inference System. Int. J. Press. Vessel. Pip. 2021, 189, 104274.
[23] Karaboga, D.; Kaya, E. Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey. Artif. Intell. Rev. 2019, 52, 2263–2293.
[24] Abdolkarimi, E.S.; Mosavi, M.R. Wavelet-adaptive neural subtractive clustering fuzzy inference system to enhance low-cost and high-speed INS/GPS navigation system. GPS Solut. 2020, 24, 1–17.
[25] Mola, M.; Amiri-Ahouee, R. ANFIS model based on fuzzy C-mean, grid partitioning and subtractive clustering to detection of stator winding inter-turn fault for PM synchronous motor. Int. Trans. Electr. Energy Syst. 2021, 31, e12770.
[26] Tiwari, S.; Babbar, R.; Kaur, G. Performance Evaluation of Two ANFIS Models for Predicting Water Quality Index of River Satlu (India). Adv. Civ. Eng. 2018, 3, 1–10.
[27] Yilmaz, M.; Arslan, E. Effect of the Type of Membership Function on Geoid Height Modelling with Fuzzy Logic. Surv. Rev. 2008, 40, 379–391.
[28] Sambariya, D.K.; Prasad, R. Selection of Membership Functions Based on Fuzzy Rules to Design an Efficient Power System Stabilizer. Int. J. Fuzzy Syst. 2017, 19, 813–828.
[29] Jiří Močkoř, Petr Hurtík, Approximations of fuzzy soft sets by fuzzy soft relations with image processing application, May 2021, Soft Computing 25(13):1-11. doi:10.1007/s00500-021-05769-3
[30] de Busin, P.J.W., Viganico, E.M., de Almeida Silva R. et al. "Production of ferrous sulfate from steelmaking mill scale." International Journal of Scientific and Engineering Research 5(4) (2014): 353-359.

[31] Babuska, R, Verbruggen, H. B., A new identification method for linguistic fuzzy models, in: Proc. of the International Conference FUZZ-IEEE/IFES'95, Yokohama, Japan, 1995, pp.905-912.
[32] Piegat, A., Fuzzy Modeling and Control, Springer-Verlag Company, Heidelberg, 2011.
[33] Preuss, H.P., Tresp, V. "Neuro-Fuzzy Control.", Automatisierungstechnische Praxis 36 (5), (1994): 10-24.
[34] URL: https://opencv.org/
[35] Xiong, N., Shen, Y., Yang, K. *et al.* Color sensors and their applications based on real-time color image segmentation for cyber physical systems. *J Image Video Proc.* 2018, 23 (2018). doi.10.1186/s13640-018-0258-x
[36] JW Oestreich, WK Tolley, DA Rice, "The development of a color sensor system to measure mineral compositions." Minerals Engineering 8 (1995) 31–39. doi.10.1016/0892-6875(94)00100-Q
[37] Brown, M., Bossley, K.M., Mills, D.J., Harris C.J., High dimensional neuro-fuzzy systems: overcoming the curse of dimensionality, in: Proceedings of the International Conference FUZZ-IEEE/IFES'95, Yokohama, Japan, 1995, pp.2139-2146.
[38] Bezsonov O., Ilyunin O., Kaldybaeva B., Selyakov O., Perevertaylenko O., Khusanov A., Rudenko O., Udovenko S., Shamraev A., and Zorenko V. "Resource and Energy Saving Neural Network-Based Control Approach for Continuous Carbon Steel Pickling Process." Journal of Sustainable Development of Energy, Water and Environment Systems 7 (2019) 275–292. doi.10.13044/j.sdewes.d6.0249
[39] Kittisupakorn, P., Thitiyasook, P., Hussain, M. A. and Daosud, W. "Neural network based model predictive control for a steel pickling process.", J. Process Control 19 (2009): 579-590.
[40] Salomon D. Data compression – The Complete Reference. 4th Edition. London: Springer, 2007. PP. 17–46, 253–530. DOI 10.1007/978-1-84628-603-2.
[41] K Pulli, A Baksheev, K Kornyakov, V Eruhimov, Real-time computer vision with opencv, Communications of the ACM 55 (2012) 61–69 . doi.10.1145/2184319.2184337