A new hybrid-system method of Machine Learning using a new method of fractal geometry and a new method of graph theory

Matej Babič[1]

[1]*Jožef Stefan Institute, Slovenia,*
E-Mail: babicster@gmail.com

Abstract. In the paper we present a hybrid system method to predict the volume of the robot-laser-hardened specimens when one of the parameters in the existing model cannot be measured or calculated the intelligent-system. Also, we have a model of the intelligent system to predict the volume of hardened specimens developed by someone, but we can not calculate one parameter in it. Thus, we develop a new method of the hybrid intelligent system to solve this problem. We develop a hybrid of genetic programming and multiple regression. To predict the volume of hardened specimens, we use teh neural network, genetic algorithm and multiple regression. The genetic programming modelling results show a good agreement with the measured volume of hardened specimens. We analyse the SEM picture of the microstructure of robot-laser-hardened specimens with a mathematical method. In this open problem we use the graph theory and fractal geometry. Fractal dimensions are calculated using image processing of a SEM micrographs in combination with a box-counting algorithm using ImageJ software.

Keywords: image processing, intelligent system, visibility graphs, fractal dimension

Novi hibridni sistem strojnega učenja z uporabo novem etode fracktalne geometrije in nove metode teorije grafov

V prispevku bomo predstavili metodo hibridnih sistemov za za napovedovanje volumna robotsko lasersko kaljenih vzorcev, če ne moremo izmeriti ali izračunati enega od parametrov v obstoječem modelu inteligentnega sistema. Torej, imamo model inteligentnega sistema za napoved volumna kaljenih vzorcev, ki ga je nekdo naredil, vendar ne moremo izračunati enega od parametrov v tem modelu. Tako smo razvili novo metodo hibridnih inteligentnih sistemov za reševanje teh problemov. Razvili smo hibrid genetskega programiranja in multiple regresije. Za napovedovanje volumna kaljenih vzorcev smo uporabili nevronske mreže, genetsko progamiranje in multiplo regresijo. Rezultati modeliranja genetskega programiranja pokažejo dobre rezultate. Analizirali smo mikrostrukturo SEM slik robotsko lasersko kaljenih vzorcev z matematično metodo. Uporabili smo metodo teorije grafov in fraktalno geometrijo. Fraktalne dimenzije so bile izračunane z obdelavo SEM posnetkov z algoritmob box-counting s programom ImageJ.

1 INTRODUCTION

To investigate the possibility of application of fractal analysis and graph theory to a heat-treated surface, we have examined the relation between teh volume, fractal dimensions and density of a visibility graph in a 3D space depending on various parameters of the robot laser cell. The robot laser surface hardening [1] heat treatment is complementary to the conventional flame or inductive hardening. Fractals [2] are a new branch of mathematics and art. Perhaps this is the reason why most people recognize fractals only as pretty pictures useful as backgrounds on the computer screen or original postcard patterns. Many people are fascinated by the beautiful images termed fractals. Extending beyond the typical perception of mathematics as a body of complicated, boring formulas, fractal geometry mixes art with mathematics to demonstrate that equations are more than just a collection of numbers. In mathematics and computer science, graph theory [3] is the study of graphs which are mathematical structures used to model pairwise relations between objects. A "graph" in this context is made up of "vertices" or "nodes" and lines called edges that connect them. In the most common sense of the term, a graph is an ordered pair G=(V. E) comprising a set Vvof vertices or nodes together with a set E of edges or lines which are 2-element subsets of V (i.e., an edge is related with two vertices, and the relation is represented as an unordered pair of the vertices with respect to the particular edge). To avoid ambiguity, this type of the graph may be described precisely as undirected and simple. Machine learning [4] applications at EPFL range from the natural language and image processing to scientific imaging as well as computational neuroscience. In this paper we use the method of intelligent system multiple regression, neural network and genetic programming to images processing [5] to calculate the fractal dimension in a 2D

space, 3D space and calculate the density of visibility graphs [6] in a 3D space. At the end we use a new method of the hybrid intelligent systems to predict porosity of hardened specimens.

2 MATERIALS PREPARATION

Firstly, we harden tool steel with a robot laser cell. We change two parameters, i.e. speed v ∈ [2, 5] mm/s and temperature T ∈ [1000, 1400] °C. A detailed characterization of the microstructure before and after surface modifications is made using a field emission scanning electron microscope (SEM), JEOL JSM-7600F. In a (8-bit) grey scale image, each picture element has an assigned intensity that ranges from 0 to 255. The grey scale image is what people normally call a black and white image, but the name emphasizes that such an image will also include many shades of grey.
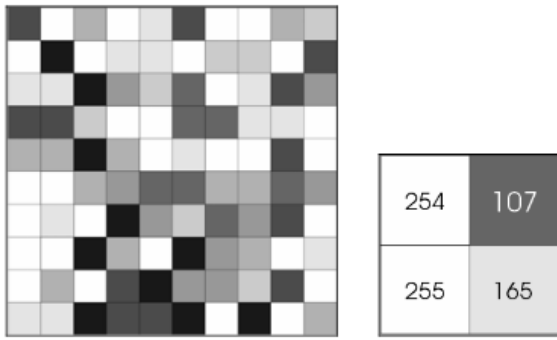

Figure 1: Each pixel has a value from 0 (black) to 255 (white). The possible range of the pixel values depends on the colour depth of the image, here 8 bit = 256 tones of the grey scales.

A normal grey scale image has an 8 bit colour depth = 256 grey scale. A true colour image has 24 bit colour depth $8 \times 8 \times 8$ bits = $256 \times 256 \times 256 = \sim 16$ million colours. Also, the SEM pictures were converted into binary images, from which we calculate the fractal dimension and into a 3D graph, from which we calculate volume. For each (x,y,z) we use only the z coordinate. The maximal value of coordinate z is 255. Also, we calculate volume of the robot-laser-hardened specimens with (1)

$$V=(z_1+z_2+...z_n)/255 \times n$$

for all n.

A new method to calculate the fractal dimension in a 3D space is presented in [7]. Also, we firstlx find all coordinates (x,y,z) of the SEM picture. Here we use the ImageJ program. Then we use only z coordinates to estimate Hurst exponent H. All the z coordinates are presented in a *2D* space component graph, which is continuous. Also, all points (x_i, y_0, z_i) present the first space component in *2D* graph for all points (x_i, z_i). All points (x_i, y_1, z_i) present the second graph of space component in a *2D* graph for all points (x_i, z_i). The graph of the space component for all y_i, $\forall i$. Then we combine all these graphs of the space component into one graph of the space component. For this long graph of space component, we can estimate hurst exponent H. We calculate fractal dimension $D=3-H$.

The visibility graph which is a fundamental structure studied in the field of computational geometry poses some special challenges. Some of the early applications include computing the Euclidean shortest paths in the presence of obstacles and decomposing the two-dimensional shapes into clusters. The visibility graph is the simplest form of the visibility maps. Commonly, all visibility maps consist of the nodes which share an edge if two nodes are within the line of sight of each other. The new method to construct the visibility graph is presented in [8].

We use an intelligent system method, i.e. genetic programming, neural network and multiple regression to modelling results.

Genetic Programming (GP) [9] is a technique that utilizes the principles of evolution and computer programming to solve problems. The basics of the GP algorithm operate as follows. A large number of the potential solutions – the Population – are tested against the problem. The results of this test are used to select high-performance solutions, which are then modified slightly to create the next Population. As the algorithm progresses, new modifications that improve the solutions will slowly percolate into the Population until a complete solution to the problem is found. The algorithms used to conduct the search based optimisation are also known to perform well in the presence of partial, noisy and missing data. This makes them attractive tools with which to approach the complex, noisy and incomplete world in which the software engineer has to find engineering solutions.

Neural networks [10], with their remarkable ability to derive the meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either human or other computer techniques. Neural networks are not only different in their learning processes but also different in their structures or topology. Neural networks have a broad applicability to the real world business problems. In fact, they have already been successfully applied in many industries. Since neural networks are best at identifying patterns or trends in data, they are well suited for prediction or forecasting the needs including: sales forecasting, industrial process control, customer research, Neural Networks and its Application in Engineering, data Validation and Risk Management.

The Multiple Regression [11] operation studies the relationship between several predictor variables and a response variable. Many applications of regression analysis involve situations in which there are more than one regressor variable. A regression model that contains more than one regressor variable is called a multiple regression model. In multiple regression analysis, the relationship between one dependent variable and several

independent variables (called predictors) is analyzed. The regression equation takes the form

$$Y = b_0 + b_1 \times X_1 + \ldots + b_6 \times X_6 + e$$

where Y is the dependent variable, the b's are the regression coefficients for the corresponding x (independent) terms, b_0 is a constant or intercept, and e is the error term reflected in the residuals.
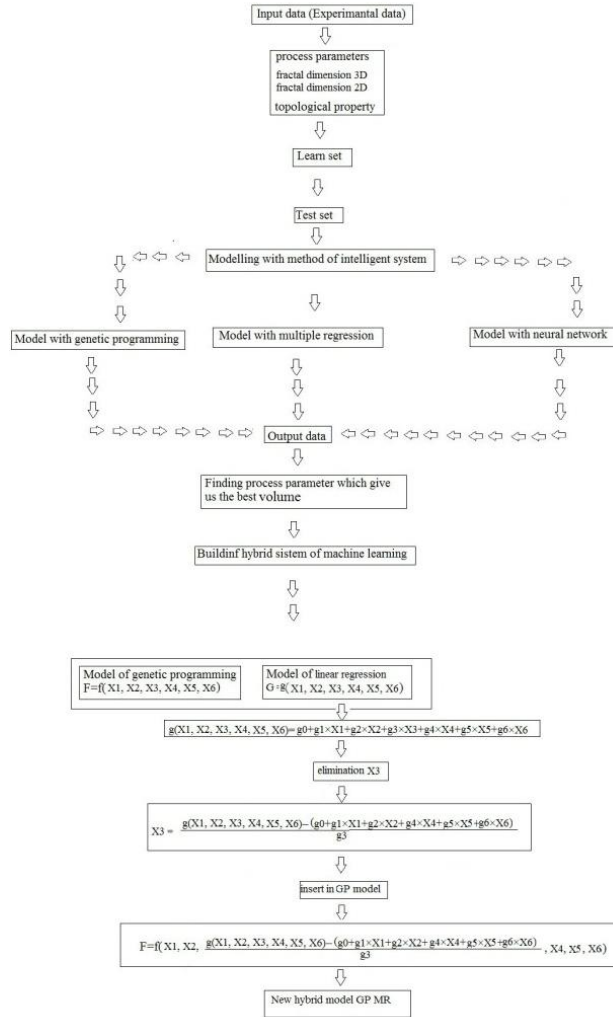

Figure 2: Intelligent hybrid systems model

3 RESULTS

In Table 1, the parameters of hardened specimens impacting the volume are presented. We mark specimens from P1 to P21. Parameter X1 presents the parameter of temperature in the degrees of Celsius [C], X2 presents the speed of hardening [mm/s], X3 presents the fractal dimension in a 2D space, X4 presents the fractal dimension in a 3D space, X5 presents the density of the visibility graphs in a 3D space and parameter X6 presents the basic volume of the specimens. The last parameter Y is the calculating volume of the robot-laser-hardened specimens. Table 2 presents the experimental and prediction data of the

volume of the robot-laser-hardened specimens. In Table 2 symbol S present the name of the specimens, E the experimental data, NM1 prediction with neural network with a 30% learn set, NM2 prediction with neural network with 50% learn set, NM3 prediction with neural network with the method one live out, R prediction with regression, GP prediction with genetic programming and H present prediction with the insert hybrid method of intelligent systems. As see from Table 1, specimen P17 has the highest density of the visibility graphs in a 3D space; i.e. 0.2832. Thus specimen P17 has the most complex graph. Specimen P16 has the maximal volume after hardening, that is 77.7%. The measured and predicted volume of the robot-laser-hardened specimens is shown in the graph in Fig. 3. The multiple-regression model:

$$Y = 222{,}7113 + 0{,}0004652 \times X1 - 4{,}8618 \times X2 - 23{,}6701 \times X3 - 18{,}7246 \times X4 - 433{,}491 \times X5 + 0{,}0355672 \times X6 \tag{1}$$

A regression model is presented in Eq. (1).. The genetic programming model presents a 11.90% deviation from the measured data, which is less than the regression model, with 31.40% deviation. The best neural network NN3 presents a 5.40% deviation from the measured data. Neural network NN1 presents a 27.73% deviation from the measured data and NN2 presents a 20.09% deviation from the measured data. Hybrid mathematical model of the intelligent system GP-MR is presented in Eq. (3) having a 18.00% deviation from the measured data.

Table 1. Hardened specimen parameters

S	X1	X2	X3	X4	X5	X6	Y
P1	1000	2	1,9135	2,304	0,1936	8,3	28,0
P2	1000	3	1,9595	2,264	0,2208	8,3	20,6
P3	1000	4	1,9474	2,258	0,2144	8,3	25,4
P4	1000	5	1,9384	2,341	0,2256	8,3	22,5
P5	1400	2	1,9225	2,222	0,2445	8,3	22,9
P6	1400	3	1,9784	2,388	0,2221	8,3	19,3
P7	1400	4	1,954	2,25	0,2036	8,3	13,2
P8	1400	5	1,9776	2,286	0,2096	8,3	21,0
P9	1000	2	1,972	2,178	0,2352	28,0	24,0
P10	1000	3	1,858	2,183	0,2288	20,6	27,5
P11	1000	4	1,9784	2,408	0,2144	25,4	28,7
P12	1000	5	1,941	2,21	0,2352	22,5	23,7
P13	1400	2	1,9784	2,257	0,2208	22,9	26,3
P14	1400	3	1,581	2,265	0,232	19,3	22,1
P15	1400	4	1,965	2,433	0,1984	132	20,3
P16	1400	5	1,8113	2,289	0,1904	21,0	77,7
P17	800	0	1,9669	2,232	0,2832	8,3	15,3
P18	1400	0	1,9753	2,235	0,2688	8,3	31,8
P19	2000	0	1,9706	2,261	0,2416	8,3	36,9
P20	950	0	1,9631	2,282	0,2128	8,3	70,8
P21	0	0	1,9537	2,319	0,208	8,3	43,9

The genetic programming model:

$$Y = \frac{\frac{X1}{X4} + \frac{2 - X3 + X1 + \frac{X2}{3.516}}{-7071170 + 0.1740836 \times 0.11. X1 \times X5 + \frac{X4 - 5.85356}{X5}} - X5 \times (-5.85356 - X4 + X6)}{10 \times X4} - $$

$$\frac{6 \times X4 + X6 + \frac{X3 - 1.9381}{X2 - 0.1768} - \frac{X1 \times X1 \times X5}{X4 \times \left(2 \times X6 + \frac{X1 \times 5.85356}{2 \times X4 + 2 \times X6}\right)}}{10 \times ((-5.85356 + X4) \times X5 - \frac{X1}{X4 + X6} - \frac{X1}{-7.62513 \times (-5.85356 + X4) \times X4 + +\frac{X4 \times (X6 - X5)}{X5} + \frac{X1 \times X1 \times X5}{X4 \times (2 \times X6 - X4)}})} \tag{2}$$

Hybrid GP-R:

$$Y = \frac{\frac{-2227.113 - 0.004652 \times X1 + 48.618 \times X2 + 187.246 \times X4 + 4334.91 \times X5 - 0.355672 \times X6}{23.6701} - 0.19364}{(x4 - 5.85356) \times x5 - \frac{X1}{X4 + X6} - \frac{X1}{-2.68513 \times (-5.68513 + X4) \times X4 + \frac{X4 \times (X6 - X5)}{X5} - \frac{X1 \times X1 \times X5}{3.516 X4 \times (2 \times X6 - X4)}} + \frac{1}{236.701}}$$

$$+ \frac{\left(6 \times X4 + \frac{X1}{-17.5607 + \frac{X4 - 5.85356}{X5} + 0.170836 \times X1 \times X6} + X6 - \frac{X1 \times X1 \times X5}{X4 \times \left(2 \times X6 + \frac{5.85356 \times X1}{2 \times X4 + 2 \times X6}\right)}\right) \times \left(\frac{X2 - 1.1768}{10}\right)}{(x4 - 5.85356) \times x5 - \frac{X1}{X4 + X6} - \frac{X1}{-2.68513 \times (-5.68513 + X4) \times X4 + \frac{X4 \times (X6 - X5)}{X5} - \frac{X1 \times X1 \times X5}{3.516 X4 \times (2 \times X6 - X4)}} + \frac{1}{236.701}}$$

$$+ \frac{\left((X4 - 5.85356) \times X5 + \frac{X1}{X4 + X6} + X6 - \frac{X1}{-2.68513 \times (-5.68513 + X4) \times X4 + \frac{X4 \times (X6 - X5)}{X5} - \frac{X1 \times X1 \times X5}{3.516 X4 \times (2 \times X6 - X4)}}\right) \times \left(\frac{X2 - 1.1768}{10}\right)}{(x4 - 5.85356) \times x5 - \frac{X1}{X4 + X6} - \frac{X1}{-2.68513 \times (-5.68513 + X4) \times X4 + \frac{X4 \times (X6 - X5)}{X5} - \frac{X1 \times X1 \times X5}{3.516 X4 \times (2 \times X6 - X4)}} + \frac{1}{236.701}}$$

$$\times$$

$$\frac{\left(\frac{\frac{X1}{x4} + \frac{X1}{-11.7071 + \frac{x4 - 5.85356}{x5} + 0.170836 \times x1 \times x5} - x5 \times (x \pm x4 - 5.85356)}{x4}\right)}{(x4 - 5.85356) \times x5 - \frac{X1}{X4 + X6} - \frac{X1}{-2.68513 \times (-5.68513 + X4) \times X4 + \frac{X4 \times (X6 - X5)}{X5} - \frac{X1 \times X1 \times X5}{3.516 X4 \times (2 \times X6 - X4)}} + \frac{1}{236.701}} \tag{3}$$

Table 2. Experimental and predicted data

| S	E	NN1	NN2	NN3	M	GP	H
P1	28	27,9	28,0	27,8	44,0	27,8	15,7
P2	20,6	21,5	20,6	22,7	27,1	22,5	19,7
P3	25,4	23,7	24,8	23,9	25,4	24,7	21,2
P4	22,5	24,1	23,4	23,8	14,3	20,0	17,9
P5	22,9	22,7	22,9	22,3	23,5	23,6	23,0
P6	19,3	19,0	19,0	18,5	23,9	18,0	17,2
P7	13,2	15,4	15,5	15,6	30,2	14,4	12,7
P8	21	18,7	19,2	19,1	21,5	19,1	17,8
P9	24	26,0	24,1	24,1	34,0	26,8	25,0
P10	27,5	32,3	27,2	27,7	31,9	25,5	23,2
P11	28,7	21,7	28,8	28,8	27,9	28,4	25,4
P12	23,7	27,3	23,9	23,8	17,6	24,2	22,4
P13	26,3	12,6	25,6	13,4	37,0	26,7	25,7
P14	22,1	52,8	52,9	22,2	35,2	25,1	24,4
P15	20,3	18,7	28,9	20,6	30,5	23,4	22,0
P16	77,7	27,6	25,3	77,8	38,3	29,1	27,0
P17	15,3	31,1	16,5	14,8	14,9	16,0	14,9
P18	31,8	43,2	40,7	32,1	21,2	22,5	22,1
P19	36,9	29,6	30,2	36,8	32,9	32,6	32,3
P20	70,8	40,7	28,6	70,7	44,7	57,3	52,4
P21	43,9	42,5	35,1	44,0	46,2	36,0	13,7
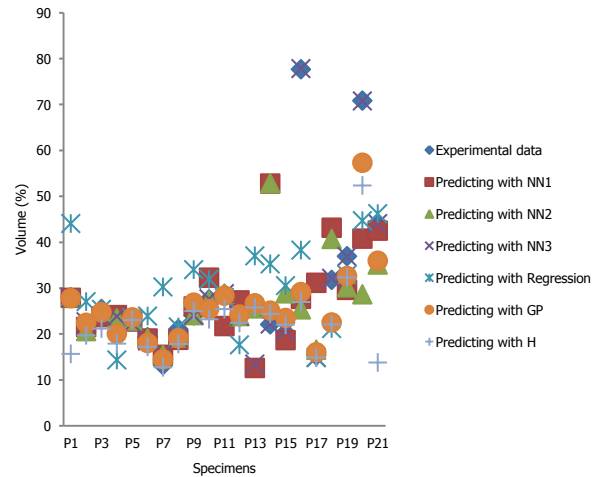
Figure 3. Measured and predicted volume of the laser-hardened robot specimens

4 DISCUSSION

The Volume of the SEM picture is a good predictor of the performance of a mechanical component, since the irregularities in the surface may form nucleation sites for cracks or corrosion. A statistically significant relationship is found between the volume, parameters of the robot laser cell, topological property density of the visibility graphs in a 3D space, fractal dimension in a 2D space and fractal dimension in a 3D space. The fractal analysis in a 2D and a 3D space and the topological property density of the 3D visibility graphs of a series of digitized surface microstructures from the robot laser surface-modified specimens indicate that useful correlations can be derived between the fractal dimensions and the surface microstructural features such as the volume. The best results of the volume are given by specimen P16. The fractal dimension in a 3D space of specimen P16 is 1.8113, in 3D space it is 2.289 and the density is 0.1904. It is hardened of a speed of 5 mm/s and the temperature of 1400 °C. In addition, the image analysis of the SEM images of the robot-laser-hardened specimens is an interesting approach. We use two methods of the intelligent system to predict the volume of the robot-laser-hardened specimens. We present a new hybrid system of intelligent system. In machine learning, the hybrid system is a very interesting method. We present a new method of the hybrid system. The hybrid system of machine learning presents a 18.00% deviation from the measured data.

5 CONCLUSSION

The paper presents a new hybrid system method of machine learning using a new method of fractal geometry and new method of graph theory and its application in mechanical engineering, especially in the robot-laser-hardened patterns. A relatively new method, fractal geometry and graph theory are used to analyse the SEM images of the laser-hardened specimens. The main findings can be summarized as follows:

1. The results demonstrate a viable potential in using the artificial neural networks, genetic programming and the regression approach in predicting the volume of the robot-laser-hardened specimens, even with limited data sets.

2. The fractal dimensions for pattern recognition are used.

3. The SEM pictures are converted into a 3D graph is calculated from which the volume is calculated.

4. The parameters of the robot laser cell giving the optimal volume are found.

5. A new method of Hybrid system is presented.

REFERENCES

[1] J. Grum, P. Žerovnik, R. Šturm: Measurement and Analysis of Residual Stresses after Laser Hardening and Laser Surface Melt Hardening on Flat Specimens; Proceedings of the Conference "Quenching '96", Ohio, Cleveland, 1996.

[2] Mandelbrot, B. B. The fractal geometry of nature. New York: W. H. Freeman, 1982:93.

[3] Ben-Moshe, B., Hall-Holt, O., Katz, M.J. and Mitchell, J.S.B. (2004). Computing the Visibility Graph of Points within a Polygon, Proc. 20th ACM Symp. on Computational Geometry, 27–35.

[4] TROST, Andrej, ŽEMVA, Andrej, ZAJC, Baldomir. A reconfigurable system for prototyping implementation of image processing algorithms. Elektrotehniški vestnik, ISSN 0013-5852.

[5] R. Kumar, M. Rattan. Analysis of various quality metrics for medical image processing. Int. J. Adv. Res. Comput. Sci. Software Eng., 2 (2012), pp. 137–144.

[6] Lacasa L., Luque B., Ballesteros F., Luque J., and Nuno J.C.(2008). From time series to complex networks: the visibilitygraph. Proceedings of the National Academy of Sciences USA105, 13, 4972-4975.

[7] Babič, Matej, Kokol, Peter, Guid, Nikola, Panjan, Peter. A new method for estimating the Hurst exponent H for 3D objects. Materias and technology, ISSN 1580-2949, 2014, letn. 48, 2, 203-208.

[8] Matej Babič. Doctoral dissertation. 2014.

[9] J. R. Koza. Course Notes for Genetic Algorithms and Genetic Programming. Spring, (2002).

[10] Graves, Alex; and Schmidhuber, Jürgen; Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, in Bengio, Yoshua; Schuurmans, Dale; Lafferty, John; Williams, Chris K. I.; and Culotta, Aron (eds.), Advances in Neural Information Processing Systems 22 (NIPS'22), December 7th–10th, 2009, Vancouver, BC, Neural Information Processing Systems (NIPS) Foundation, 2009, pp. 545–552.

[11] Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification". The Annals of Mathematical Statistics 25 (2): 290. doi:10.1214/aoms/1177728786.

Matej Babič received his Ph.D. degree in Computer Science from the Faculty of Electrical Engineering and Computer Science of the University of Maribor, Slovenia. He studied Mathematics at the Faculty of Education in Maribor. His research interest is in fractal geometry, graph theory, network, intelligent systems, hybrid machine learning and topography of materials after hardening.