1. Krzysztof KRÓL[1], 2. Tomasz RYMARCZYK[1,2], 3. Michał OLESZEK[1], 4. Piotr BOŻEK[1], 5. Paweł TCHÓRZEWSKI[1] 6. Edward KOZŁOWSKI[3]

Research and Development Center, Netrix S.A (1), University of Economics and Innovation, Lublin (2), Lublin University of Technology (3)
ORCID. 1. 0000-0002-0114-2794, 2. 0000-0002-3524-9151, 3. 0000-0002-8979-9650, 4. 0000-0003-4929-1251, 5. 0000-0001-8344-8828, 6. 0000-0002-7147-4903

doi:10.15199/48.2022.01.21

Using a classification tree to identify seepage in flood embankments

***Abstract**. The article presents a method of controlling infiltration in flood embankments by means of impedance tomography with the use of classification tree prediction. The analysis was performed using electrical impedance tomography and image reconstruction using machine learning methods, the results of the reconstruction were compared and various numerical models were used. The main advantage of the presented solution is the possibility of analyzing spatial data and high processing speed. The key parameters in electrical tomography are the speed of analysis and the accuracy of the reconstructed objects. The reconstruction algorithm is obtained by solving the inverse problem. Classification trees were used to obtain feedback on the degree of water permeability of the embankment.*

***Streszczenie.** Artykuł przedstawia metodę kontroli przesiąków w wałach przeciwpowodziowych za pomocą tomografii impedancyjnej z wykorzystaniem predykcji drzewa klasyfikacyjnego. Analizę przeprowadzono z użyciem elektrycznej tomografii impedancyjnej i rekonstrukcji obrazu z wykorzystaniem metod uczenia maszynowego, porównano wyniki rekonstrukcji i zastosowano różne modele numeryczne. Główną zaletą prezentowanego rozwiązania jest możliwość analizy danych przestrzennych oraz duża szybkość przetwarzania. Kluczowymi parametrami w tomografii elektrycznej są szybkość analizy i dokładność rekonstruowanych obiektów. Algorytm rekonstrukcji uzyskuje się poprzez rozwiązanie problemu odwrotnego. Drzewa klasyfikacyjne zostały wykorzystane do uzyskania informacji zwrotnej o stopniu przesiąkliwości nasypu (**Wykorzystanie drzewa klasyfikacyjnego do identyfikacji przesiąkania wałów przeciwpowodziowych**).*

Keywords: Classification tree, tomography, seepage in flood embankments
Słowa kluczowe: Drzewo klasyfikacyjne, tomografia, przesiąkanie wałów przeciwpowodziowych

Introduction

Many different methods are used to solve optimization problems [1-11]. The presented solution is based on electrical impedance tomography [12-18]. Ordinary destructive methods cannot research identifying the problem of seepage or moisture inside flood embankments. Non-destructive methods must be used for this purpose. One of such methods is the impedance tomography method, which works very well in this type of research. The task of tomography is to reconstruct what is happening inside the examined area. It can be said that the task is to define a model in which conductivity should be calculated based on $x\epsilon R^m$ signals from individual electrodes. Finite elements are used for the calculations in this paper. Thus, the reconstruction model can be defined as a set of classification trees corresponding to the finite elements in the analysed visual field [19-20].

Classification tree

In tomography, the task consists of the reconstruction of the viewing area. In other words, the task is to define a model, where based on signals $x\epsilon R^m$ from electrodes, we should estimate a conductivity for the finite elements. The viewing area is created as a set of finite elements. Thus the reconstruction model can be defined as a set of classification trees corresponded to finite elements in the analysed viewing area. Let $D=\{(x_{(i)},y_i):x_{(i)}\epsilon R^m, y_i\epsilon\{0,1\}, 1\le i\le n\}$ be the learning set for established finite element. For i-th observation, $1\le i\le n$ the vector $x_{(i)}\epsilon R^m$ denotes the realisations of independent (input) variables, in tomography usually there are the signals obtained from the electrodes, whereas $y_i\epsilon\{0,1\}$ denotes the belonging to appropriate class concerning the conductivity. The tree-based method consists of partition (splitting) of the feature space into a set of separable regions and fitting constant values to appropriate regions. This method is straightforward and powerful. Below we consider a classification problem for response variable Y. Entire space of features Rm we split into $S_1,S_2,\ldots,S_k$ regions, where $S_i\cap S_j=0$ for $1\le i\ne j\le k$ and $S_1\cup S_2\cup\ldots\cup S_k=R^m$. Based on input vector $x\epsilon R$ we predict the output variable Y as follows

$$f(x)=\sum_{j=1}^{k} c_j I_j(x), \tag{1}$$

$$I_j(x)=\begin{cases}1, for & x\in S_j\\ 0, for & x\notin S_j\end{cases} \tag{2}$$

and constant $c_j\epsilon R$ for $1\le j\le k$ denotes the mean value of response variable when the features belong to S_j region, i.e

$$c_j=\frac{1}{N_j}\sum_{x_{(i)}\in S_j} y_i \tag{3}$$

Nj is a length of subsequence for which the features belong to Sj region. From (1) we see, that the main task of building the classification tree consists in splitting the entire space of features into separated regions[12-26]. An objective function is defined as the sum of squares error, and the solution of the presented problem consists of the following task

$$\min_{\substack{S_1,S_2,\ldots S_k\\ S_j\bigcap S_2=0\ for j\ne l}} \sum_{i=1}^{n}\left(y_i-f\left(x_{(i)}\right)\right)^2 \tag{4}$$

To apply the regression tree in EIT, we utilise CART (Classification And Regression Tree) algorithm [25].

Receiver Operating Characteristic (ROC)

In the following, we will analyse the embankment model where the viewing area consists of $k\epsilon N$ finite elements. For each finite element, we determine the classification tree (1). Based on the signal $x\epsilon R^m$ obtained from electrodes, we estimate $P_j(Y=1|X)$ probability that the j-finite element belong to the seepage area, $1\le j\le k$. The reconstruction of viewing area is defined as a sequence $\{\hat{z}_j\}_{1\le j\le k}$

$$\hat{z}_j=\begin{cases}seepage & P_j(Y=1\mid X)>l\\ no\_seepage & P_j(Y=0\mid X)\le l\end{cases} \tag{5}$$

for the level $l\epsilon(0,1)$.

Some assumptions have to be taken into account, such as the finite element does not belong to the drainage zone and we take as a negative case (N) (elements belong to the

ground), while the finite element is contained in the drainage zone, we take as a positive case (P). To assess the quality of the reconstruction, the values of the confusion matrix are determined as follows: TP (True Positive) is the number of finite elements that correctly belong to the seepage zone, TN (True Negative) is the number of finite elements that were correctly identified as belonging to the background, FP (False Positive) is the number of finite elements that belong to the background but were identified as belonging to the seepage zone (false alarm), FN (False Negative) is the number of finite elements that belong to the seepage zone but were identified as background.

Table 1. The confusion matrix

	Positive	Negative
Positive Prediction	TP	FP
Negative Prediction	FN	TN

To describe the quality of seepage detection of the viewing area based on classification tree (1), we calculate the classical characteristics (ratios) [21-22]. Accuracy represents the part of the visual area that the model correctly recognised. On the other hand, it is only one possible measure that presents the correctness of recognition. In the EIT, the possibilities of seepage finding should also be described during the recognition of the embankment state. To determine the ability of the classifier based on the application of the classification tree [21-22], we determine the Receiver Operating Characteristic (ROC curve) curve. This curve shows the relationship between sensitivity and specificity during the reconstruction. The diagonal of the square on the ROC curve describes a strategy based on guessing the seepages during the recognition. If the ROC is above the diagonal, the recognition technology is much better than guess.

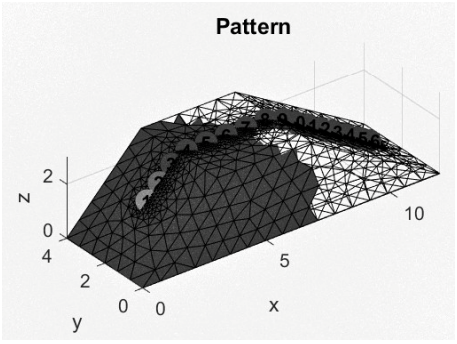

Fig.1. Model 1 - pattern

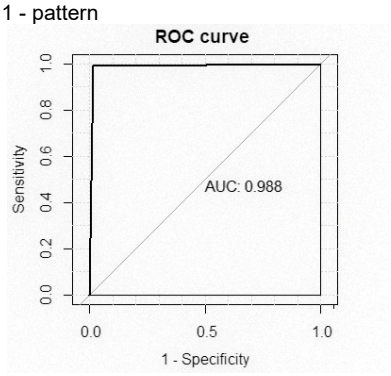

Fig.3. Classifier evaluation, for model, 1

Table 2. Confusion Matrix for example 1

	Positive	Negative
Positive Prediction	5269	36
Negative Prediction	90	4458

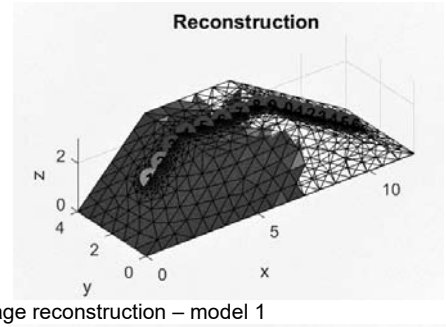

Fig.2. Image reconstruction – model 1

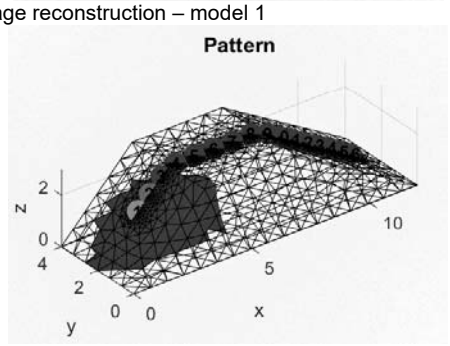

Fig.4. Model 2 - pattern

Table 3. Confusion Matrix for model 2

	Positive	Negative
Positive Prediction	8333	139
Negative Prediction	229	1152

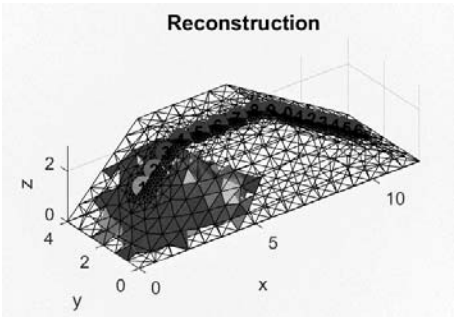

Fig.5. Image reconstruction – model 2

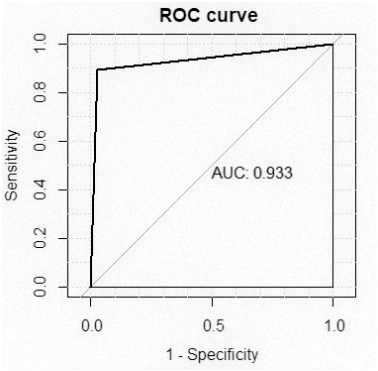

Fig.6. Classifier evaluation for model 2

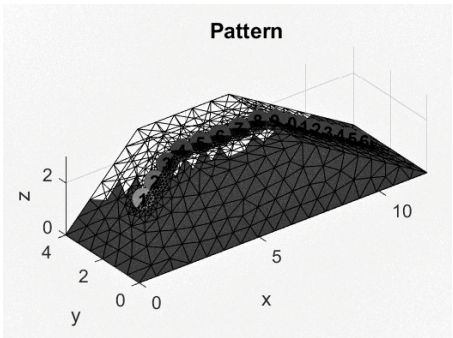

Fig.7. Model 3 - pattern

Table 4. Confusion Matrix for model 3

	Positive	Negative
Positive Prediction	8539	0
Negative Prediction	3	1311

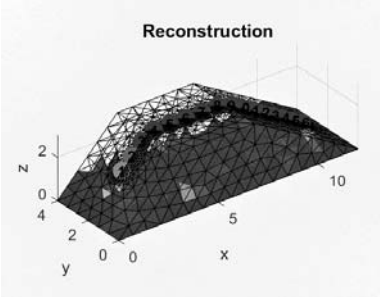

Fig.8. Image reconstruction – model 3

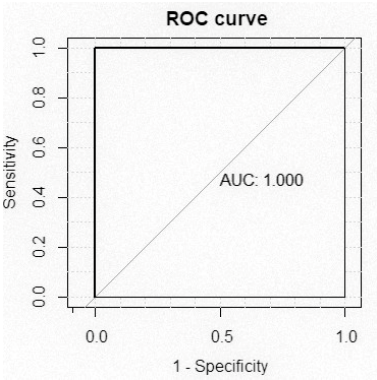

Fig.9. Classifier evaluation for model 3

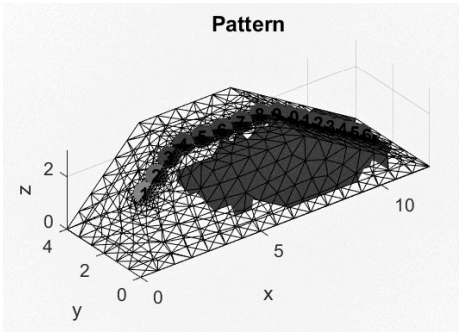

Fig.10. Model 4 - pattern

Table 5. Confusion Matrix for model 4

	Positive	Negative
Positive Prediction	8717	136
Negative Prediction	156	844

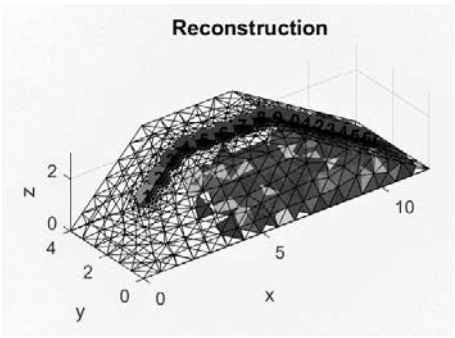

Fig.11. Image reconstruction – model 4

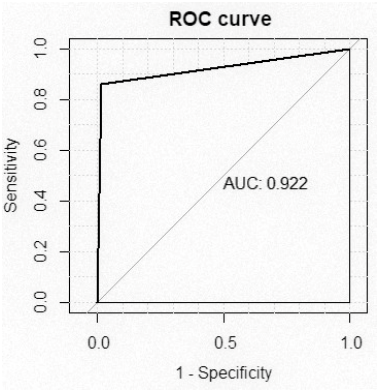

Fig.12. Classifier evaluation for model 4

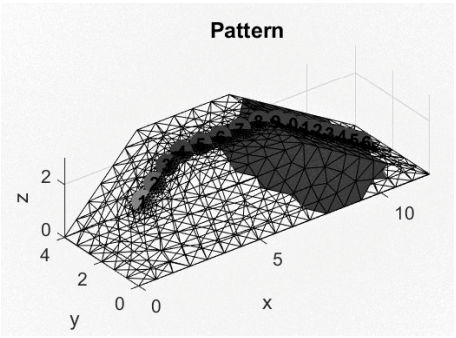

Fig.13. Model 5 - pattern

Table 6. Confusion Matrix for model 5

	Positive	Negative
Positive Prediction	6555	189
Negative Prediction	358	2751

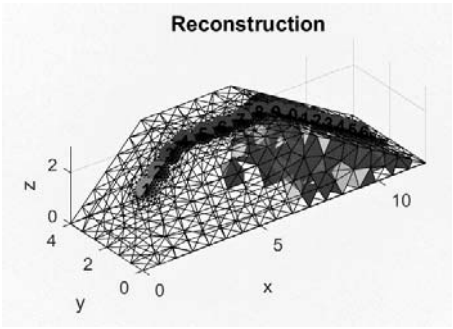

Fig.14. Image reconstruction – model 5

Table 7. Evaluation metrices for models

	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5
Accuracy	0.9872	0.9627	0.9997	0.9704	0.9445
Sensitivity	0.9920	0.8923	1.0000	0.8612	0.9357
Specificity	0.9832	0.9733	0.9996	0.9824	0.9482
Pos Pred Value	0.9802	0.8342	0.9977	0.8440	0.8849
Neg Pred Value	0.9932	0.9836	1.0000	0.9846	0.9720
Precision	0.9802	0.8342	0.9977	0.8440	0.8849
Recall	0.9920	0.8923	1.0000	0.8612	0.9357
Prevalence	0.4561	0.1310	0.1331	0.0995	0.2984
Detection Rate	0.4525	0.1169	0.1331	0.0857	0.2792
Detection Prevalence	0.4616	0.1402	0.1334	0.1015	0.3155

Conclusion

The article presents an analysis of leaks in flood embankments using a specially prepared numerical model. An electrical impedance tomography approach was used to measure and reconstruct the images. The results of the

work show that the presented method is effective. The image reconstruction was obtained thanks to the use of decision trees, which allowed for the analysis of spatial data and high processing speed.

Authors: *Ph.D. D.Sc. Eng. Tomasz Rymarczyk, prof., University of Economics and Innovation, Projektowa 4, Lublin,. E-mail: tomasz@rymarczyk.com; Krzysztof Król, Research&Development Centre Netrix S.A., Email: krzysztof.krol@netrix.com.pl; Michał Oleszek, Research&Development Centre Netrix S.A., E-mail:michal.oleszek@netrix.com.pl; Piotr Bożek, Research & Development Centre Netrix S.A., E-mail: piotr.bozek@netrix.com.pl; Paweł Tchórzewski, Research&Development Centre Netrix S.A., Email: pawel.tchorzewski@netrix.com.pl; Edward Kozłowski Ph.D.Eng.,Lublin University of Technology, Nadbystrzycka 38, Lublin, E-mail: e.kozlowski@pollub.pl;*

REFERENCES

[1] Miłak, M., Leszczyńska, A., Grudzień, K., Romanowski, A., & Sankowski, D. (2019). Slug flow velocity estimation during pneumatic conveying of bulk solid materials based on image processing techniques. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 9(1), 11-14.

[2] Kryszyn, J., Wanta, D., Smolik, W. T. (2019). Evaluation of the electrical capacitance tomography system for measurement using 3d sensor. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 9(4), 52-59.

[3] Kłosowski G., Rymarczyk T., Wójcik D., Skowron S., Adamkiewicz P., The Use of Time-Frequency Moments as Inputs of LSTM Network for ECG Signal Classification, Electronics, 9(9), 1452, 2020.

[4] Korzeniewska E., Krawczyk A., Stando J., Torsion field - an example of pseudo-scientific concept in physics, Przeglad Elektrotechniczny, Volume97, Issue1, Page196-199, DOI10.15199/48.2021.01.41, 2021.

[5] Korzeniewska, E; Szczesny, A; Lipinski, P; Drozdz, T; Kielbasa, P; Miernik, A, Prototype of a Textronic Sensor Created with a Physical Vacuum Deposition Process for Staphylococcus aureus Detection, SENSORS Volume: 21 Issue: 1 Article Number: 183, 2021.

[6] Wajman, R; Banasiak, R; Babout, L, On the Use of a Rotatable ECT Sensor to Investigate Dense Phase Flow: A Feasibility Study, SENSORS Volume: 20 Issue: 17 Article Number: 4854, 2020.

[7] Banasiak, R.; Wajman, R.; Jaworski, T.; Fiderek, P.; Fidos, H.; Nowakowski, J.; Sankowski, D. Study on two-phase flow regime visualization and identification using 3D electrical capacitance tomography and fuzzy-logic classification. Int. J. Multiph. Flow 2014, 58, 1–14.

[8] Jan Dusek, Jan Mikulka, Measurement-Based Domain Parameter Optimization in Electrical Impedance Tomography Imaging, Sensors 2021, 21(7), 2507

[9] Daniewski K., Kosicka E., Mazurkiewicz D., Analysis of the correctness of determination of the effectiveness of maintenance service actions. Management and Production Engineering Review 9 (2018); No. 2, 20-25.

[10] Romanowski, A. Contextual Processing of Electrical Capacitance Tomography Measurement Data for Temporal Modeling of Pneumatic Conveying Process. In Proceedings of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), Poznan, Poland, 9–12 September 2018; 283–286.

[11] Chen, B.; Abascal, J.; Soleimani, M. Extended Joint Sparsity Reconstruction for Spatial and Temporal ERT Imaging. Sensors 2018, 18, 4014.

[12] Rymarczyk T., Kłosowski G., Hoła A., Sikora J., Wołowiec T., Tchórzewski P., Skowron S., Comparison of Machine Learning Methods in Electrical Tomography for Detecting Moisture in Building Walls, Energies, 14(10), 2777; 2021.

[13] Rymarczyk T., Kozłowski E., Kłosowski G., Electrical impedance tomography in 3D flood embankments testing – elastic net approach, Transactions of the Institute of Measurement and Control, 42(4), 680-690, 2020.

[14] Rymarczyk T., Nita P., Vejar A., Woś M., Stefaniak B., Adamkiewicz P.: Wearable mobile measuring device based on electrical tomography, Przegląd Elektrotechniczny, 95(4), 211-214, 2019.

[15] Rymarczyk T., Kłosowski G., Tchórzewski P., Cieplak T., Kozłowski E.: Area monitoring using the ERT method with multisensor electrodes, Przegląd Elektrotechniczny, 95(1), 153-156, 2019.

[16] Koulountzios P., Rymarczyk T., Soleimani M., A quantitative ultrasonic travel-time tomography system for investigation of liquid compounds elaborations in industrial processes, Sensors, 19(23), 5117, 2019.

[17] Kłosowski G., Rymarczyk T., Kania K., Świć A., Cieplak T., Maintenance of industrial reactors based on deep learning driven ultrasound tomography, Eksploatacja i Niezawodnosc – Maintenance and Reliability; 22(1), 138–147, 2020.

[18] Kłosowski G., Rymarczyk T., Cieplak T., Niderla K., Skowron Ł., Quality Assessment of the Neural Algorithms on the Example of EIT-UST Hybrid Tomography, Sensors, 20(11), 3324, 2020.

[19] Liang, GH; Dong, F; Kolehmainen, V; Vauhkonen, M; Ren, SJ, Nonstationary Image Reconstruction in Ultrasonic Transmission Tomography Using Kalman Filter and Dimension Reduction, IEEE Transactions on instrumentation and measurement Volume: 70 Article Number: 4501012, 2021.T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer-Verlag New York Inc., 2009.

[20] James G., Witten D., Hastie T., Tibshirani R., An introduction to statistical learning, Springer-Verlag GmbH, 2013.

[21] Fawcett T., An introduction to ROC analysis, Pattern Recognition Letters. 27 (2006) 861–874.

[22] Hand D.J., Till ., R.J., A simple generalisation of the area under the ROC curve for multiple class classification problems, Machine Learning. 45 (2001) 171–186.

[23] Friedman J., Hastie T., Tibshirani R., Regularisation paths for generalised linear models via coordinate descent, Journal of Statistical Software. 33 (2010) 1.

[24] Breiman L., Friedman J., Stone C.J., Olshen R.A., Classification and regression trees, CRC press, 1984.

[25] Kozłowski E., Mazurkiewicz D., Sęp J., Żabiński T., The use of principal component analysis and logistic regression for cutter state identification, in: Innovations in Industrial Engineering, Springer International Publishing, 2021: pp. 396–405.

[26] Antosz K., Mazurkiewicz D., Kozłowski E., Sęp J., Żabiński T., Machining process time series data analysis with a decision support tool, in: Lecture Notes in Mechanical Engineering, Springer International Publishing, 2021: pp. 14–27.