

Machine learning for two-phase gas-liquid flow regime evaluation based on raw 3D ECT measurement data

Radosław WAJMAN¹*, Jacek NOWAKOWSKI¹, Michał ŁUKIAŃSKI, and Robert BANASIAK¹

Institute of Applied Computer Science, Lodz University of Technology, Stefanowskiego 18, 90-537 Łódź, Poland

Abstract. This paper presents a study on applying machine learning algorithms for the classification of a two-phase flow regime and its internal structures. This research results may be used in adjusting optimal control of air pressure and liquid flow rate to pipeline and process vessels. To achieve this goal the model of an artificial neural network was built and trained using measurement data acquired from a 3D electrical capacitance tomography (ECT) measurement system. Because the set of measurement data collected to build the AI model was insufficient, a novel approach dedicated to data augmentation had to be developed. The main goal of the research was to examine the high adaptability of the artificial neural network (ANN) model in the case of emergency state and measurement system errors. Another goal was to test if it could resist unforeseen problems and correctly predict the flow type or detect these failures. It may help to avoid any pernicious damage and finally to compare its accuracy to the fuzzy classifier based on reconstructed tomography images – authors' previous work.

Keywords: neural networks; machine learning; data augmentation; electrical capacitance tomography; industrial process monitoring.

1. INTRODUCTION

The rapid expansion of computerization in the last decades opened new possibilities for industry. Thanks to it, most industrial processes can be automated for better efficiency and final product quality. As a result, companies can cut costs while maximizing profits without compromise. A good example can be various applications of computer science for process tomography. The non-invasive methods of observing various chemical and physical processes play a vast role in industry. For this purpose, different types of equipment can be used depending on specific needs, such as electrical, ultrasound, or optical tomography [1]. For instance, it can be used as a diagnostic tool to classify the characteristics of two-phase flow (TPF). This information may be crucial to adjust the parameters of the process controlling the supplying devices (e.g. air pressure, liquid flow rate) to make the flow stable with desired characteristics.

Various solutions within the Industry 4.0 concept involve machine learning (ML) [2] to: increase interconnectivity and smart automation, ensure higher production quality and energy usage optimization [3, 4], autonomous vehicles usage [5], personal assistant systems [6], security systems based on face recognition [7], etc. These innovations were possible thanks to developments made in the field of ML.

Current research was conducted on ML algorithms for identifying TPF types. For this purpose, a model of an artificial neural network was built. Training data was prepared based on raw measurement data collected using the 3D electrical capacitance tomography (ECT) diagnostic technique [8, 9]. Measurement data was acquired as part of a measurement campaign for two

research projects completed previously by the authors. Those studies aimed to identify types of TPF processes by using fuzzy logic algorithms.

The collected and well-described measurement data was stored since performed experiments. Currently, technologies and algorithms for effective classification based on ML have emerged. Hence, the motivation to undertake the current research was to restore and reuse data often referred to as trash data (refurbishing) [10]. In Sections 2 and 3, the state-of-the-art is presented showing that AI and tomographic techniques have not been used so far together to recognize the TPF regime. Moreover, the stored set of collected measurement data was insufficient to build a reliable artificial neural network (ANN) model. The known image data augmentation algorithms could not be used in this case due to the specificity of the diagnostic data format. Therefore, new dedicated algorithms for augmenting tomographic measurement data for building a training set were developed within the current research. The results of the model operation were evaluated and compared with those obtained with the fuzzy models. The effectiveness of identification and the workload related to the preparation of the fuzzy and neural network models were compared. Also, different experiments were conducted to demonstrate the high adaptability of the artificial neural network to malfunctions. Three scenarios of the breakdown of the measuring device were proposed. Unique errors simulating the sensor and tomograph measurement card damages were applied to the test data, which makes an additional scientific contribution to this article.

2. BACKGROUND OF THE STUDY

2.1. Two-phase gas-liquid flow processes

The two-phase gas-liquid flows (TPF) are essential to many industrial processes. Some of the numerous examples are the

*e-mail: radoslaw.wajman@p.lodz.pl

Manuscript submitted 2023-07-04, revised 2023-12-08, initially accepted for publication 2023-12-10, published in February 2024.

aeration processes [11] in chemical reactors [12], flotation processes [13], and water and sewage aeration systems [14]. The main task of aeration systems is producing a proper fraction of aerated liquid and oxygen. The TPF processes also occur in the bubble columns [15]. Their purpose is to implement various physical and chemical processes. Controlling the interfacial area often determines the progress intensity of these processes. There is also a separate group of industrial processes in which gas bubbles may be formed in the liquid due to chemical reactions. It can occur, for example, in chemical reactors or the process of electrolysis [16], where the gas phase is a product (often kind of a by-product) of a chemical reaction. The increasing needs of industry [17] for a simple, versatile, relatively inexpensive, non-invasive, and rapid method of process diagnosis and control for TPFs justify the importance of the research topic. One of the main challenges in TPF control is the possibility of a priori prediction of features and the type of mixture flow based on known apparent velocity values, the properties of individual phases, and, finally, the bed installation geometry (such as diameter and inclination angle of the pipeline). The observed structures of the gas-liquid mixture in pipes during vertical flow movement are consistent with those presented by Nicklin and Davidson [18]. These structures, influenced by physical variables such as superficial velocity, viscosity, density, and surface tension, have been used to classify these flow regimes. Knowledge of the characteristics and the gas-liquid flow regime is essential for designing and implementing industrial-scale research facilities and numerical modelling. The continuous monitoring and diagnosis of abnormalities can provide valuable information about the dynamic state of the process and allow for continuous and automatic control. That inspired the authors to develop a method and diagnostic system that can be applied to measurement data and classification of the flow type.

2.2. Three-dimensional electrical capacitance tomography (3D ECT)

As mentioned in the introduction, archive measurement data, acquired during the completion of previous projects, was used for this study. The diagnostic system was based on 3D ECT. The sensors used in this diagnostic method consist of a set of measuring electrodes located around the tested object without interfering with the characteristics of the ongoing process. In the 3D mode, the electrodes are mounted in a few rings instead of one. The measurement is performed similarly between all possible variations of pairs of electrodes without repetition, scanning the space in many cross-sections, not in one.

2.3. Experimental setup

The experimental setup is dedicated to the non-invasive study of TPFs. It is designed and built on a semi-industrial scale and ensures that measurements are conducted in conditions that take place during industrial processes. Each of the individual parts of the setup (horizontal and vertical) is equipped with measuring pipelines with three different internal diameters – 34, 53.6, and 81.4 mm, respectively, which facilitates testing TPF processes on a different scale. The length of the measuring section of

horizontal pipelines is approximately 7.5 m and the vertical one is approximately 4 m high.

2.4. Measurement campaign

The data for training the models was obtained from 32 electrodes of a 3D ECT system and validated with dedicated flow maps. Note that for supervised learning the measurement data needs to be described in a manner adequate to the expected result of the model prediction. Next, the liquid and air flow rates were changed respectively to set different experimental conditions and achieve various flow patterns. The experiment procedure was summarized and marked in the flow map. Simultaneously, ECT measurement data was recorded with measured liquid and air mass flow rate values, air pressure at different rig points, time stamps, and recognized flow patterns, which were verified by the process engineering expert.

3. MACHINE LEARNING

3.1. Classification

The process of categorizing objects into groups based on specific characteristics is called classification. This broad set of algorithms and methods is used to fulfil this task. ML uses various algorithms to build so-called models based on sample-categorized data, commonly known as the training dataset. Regarding how the model should be trained, two main approaches are distinguished: supervised and unsupervised learning [19]. In supervised learning, the used dataset must consist of two components – the input component, which contains expected characteristics, and the component with desired outputs. During the process of training, some algorithms try to adjust the model (internal weights, etc.) to match the described vectors. Neural network algorithms tend to be the default choice regarding categorization. Extensive modularity while creating a model provides a variety of possibilities in terms of usage, from a simple grouping of text to classifying objects based on an image. However, unsupervised learning takes only the input vector as its dataset without desired output values. The purpose of these types of algorithms dictates such a change. The task of unsupervised learning is not to match the input with the output but to discover patterns and commonalities in each set of information. Items with similar characteristics are grouped into clusters.

The data provided for the project was categorized. That is why the authors decided on supervised learning. The neural network was chosen from the described prediction algorithms due to its common utilization.

3.2. Neural network

The popularity of ANNs [19] has significantly increased over the last decade. This can be attributed to supervised ML and clustering. The process of training the neural network is divided into epochs. Epoch is a cycle of training the model with all available data. It can be described as follows:

- Feeding data to model – the training set is put into the input layer and forwarded through all hidden layers until it reaches

the output. In this step, neuron values are calculated to find a proper answer.

- Model validation – after getting the results from a model, they are compared with expected values. That is done by calculating the loss function for each layer. Information about errors calculated by this function is later used to adjust the internal weights of the model — a value closer to 0 is better. What is more, the dataset used for training can contain some extreme values called outliers that are outside the range of what is expected and unlike the other data.

Accuracy is another criterion to estimate processing quality. While loss describes an error made by ANN, accuracy measures the exactness of the prediction.

- Model adjustment – depending on the calculated loss, the model tries to adjust the internal weights of neuron connections between layers. Its main goal is to reduce network error.

The training process tries to optimize both accuracy and loss. It can stop when a certain loss threshold is reached, or it starts to increase. But the most often used criteria are epoch lengths. Even though the loss value decreases with more epochs, finishing training earlier is sometimes more efficient to counteract overfitting (perfect fit with the input data, which inaccurately predicts untrained data).

3.3. Machine learning in ECT and flow processes applications

In the world literature, some studies may be found that deal with TPF classification. They tend to use data from ECT sensors. Measurement datasets can be processed in numerous ways, i.e. by creating a statistical model such as the hidden Markov model (HMM). It can provide satisfactory results in terms of identification while being fast enough [20]. An example of another mathematical approach is fuzzy clustering. Generally, algorithms operating on fuzzy logic provide reasonable accuracy at around 90% [21]. The verification showed the system weakness when lowering the flow rate to a very small value [22]. While described approaches work in most cases, they overlook current trends in computer science.

However, some applications of ANN dedicated to this purpose can also be found in world literature. The authors of [23] built and tested several ANN models to make predictions of TPF patterns based on recorded readings of measurements of power devices, installation design, and the nature of the liquid. This does not determine the method as universal but rather specific to a particular case. In turn, the next two research cases described in studies [24, 25] solve this problem using a convolutional neural network (CNN) and camera observation. While the results in the above-mentioned studies indicate high prediction accuracy and a universal approach, the applied visualization technique is limited to only one 2D surface and does not contain information about the spatial distribution of flow fractions, which is provided with the 3D ECT technique (as discussed in Section 2.2).

In the field of industrial tomography, there are some applications of machine learning. In [26], the source data used for classification was in the form of images. It was then necessary to reconstruct the image that can limit the system ability to work

in real-time industrial applications, which is important in dynamic flow processes. Moreover, researchers did not use a neural network but were focused on comparing fuzzy logic, support vector machine (SVM), and SVM with PCA (principal component analysis). They tested it for four types of vertical TPFs: annular, bubbly, churn, and slug. The identification rates varied depending on the method used. Other work for categorization employed the CNN, achieving 94% accuracy [27]. However, the authors did not utilize electrodes that measure electrical field, but the ultrasound Doppler velocimetry. In contrast, in the literature review research can be found that focuses on vertical flow classification by applying electrical field measurements. Nevertheless, the data used to train the network was numerically simulated [28]. The authors did not perform any test regarding incorrect data acquisition on the pre-trained model.

4. MODEL DEVELOPMENT

4.1. Measurement data structure and augmentation

When the training dataset is small, ANNs will struggle to create robust generalizations, hindering their ability to make satisfactory predictions. The data augmentation process aims to enhance the model generalization by generating synthetic data and applying appropriate random transformations to existing data. In theory, having more data for training, validation, and testing is beneficial. This holds partly true because a higher quantity of good-quality data improves the trained model. However, the sheer amount of data alone does not independently determine the effectiveness or accuracy of a given model. Therefore, the augmentation process must be moderate and tailored to the specific problem. Outliers in the training dataset can lead to instability or a failure to converge during the training process. Incomplete, inconsistent, and missing data can significantly degrade prediction results, potentially rendering them inaccurate and recommendations misleading. In essence, a model will be rendered useless and will certainly not fulfil its intended purpose if not properly handled.

The world state-of-the-art provides numerous solutions for the problem of data augmentation [29]. However, they focus on 2D and 3D images, audio, or even text data. The tomography measurement data have many limitations and a unique structure that differs from the format of traditional signals and additionally should be analysed as a function of time in the case of TPF diagnostics. This means that an innovative approach to this problem was needed.

The single 3D ECT data frame consists of 496 measured electrical capacitances between all possible electrode pairs in the 32-electrode sensor. Those measurements inside the frame are variations without repetition. This means that the first value in a frame is a measurement between the first and second electrodes. The second value represents the pair of the first and third electrodes, etc. The system used within experiments can collect 12 frames per second.

It is necessary to have a sequence of frames to get the characteristics of a flow. Usually, it takes up to a few seconds to determine what type of TPF occurs. The provided data consists

of measurements from different runs and flow types handled in separate files. Depending on the flow, a file could consist of 360 up to 720 frames. The data was grouped into files containing 100 frames to unify sets. Even though the number of frames seems to be large, the dataset size is insufficient to train the network. This would negatively impact the model prediction. The sample set was augmented to improve accuracy.

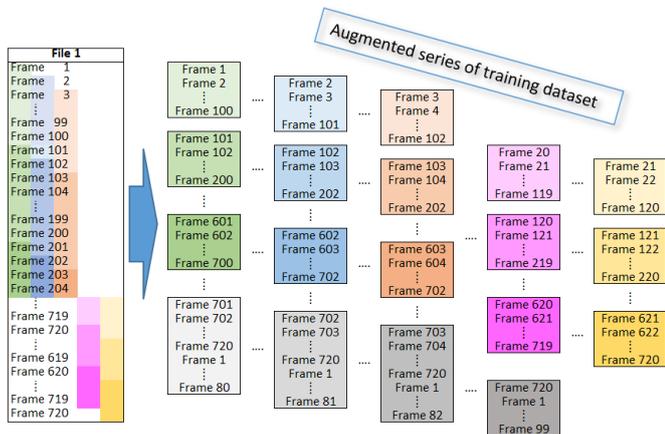


Fig. 1. Process of training dataset augmentation for one of the original files with 720 measurement frames

The algorithm for this is as follows. Take the first 100 frames for the given file and save them to a new one (see Fig. 1). Do the same with the next 100 frames and continue this until the set is not precious, then complete it with an adequate frame amount from the beginning of the set. In the next step, do the same tasks, but the start frame number is set to the next one (i.e. from 2 to 101 frame). Stop the algorithm if all possible variations without repetitions are determined. The next step for dataset augmentation was to take the last ten frames from a new augmented file of the same run and flow and then append them at the beginning of the next one. The results of this operation were saved in separate files. It enlarged the dataset and added historical information and continuity to each sample. The sample polls before categorizing into training and test sets are bubbly flow – 81 samples and 12 samples for horizontal and vertical orientation respectively, disturbed flow – 200 samples for horizontal orientation, stratified flow – 27 samples for horizontal orientation, and annular flow – 28 samples for vertical orientation.

After those operations, data was considered ready for split into training and test sets. The test set consisted of two files of each category. That, in total, gave six dataset files for horizontal flow and four for vertical. The test dataset used in training the neural network does not contain extended files.

4.2. Structure of models and training

The models used in this work were written in Python programming language with the help of the TensorFlow library. Neural networks for vertical and horizontal flows include one input layer, one output layer, and five hidden layers. That gives seven layers in total. The difference between these two networks is the number of neurons used in the output layer. The horizontal flow

contains three neurons on its output, while the vertical has only two. These result from the number of classes.

The softmax activation function was used in the output layer instead of the hyperbolic tangent. The critical difference between those two functions is that the softmax converts values from the last hidden layer to a probability distribution.

The model was configured to use cross-entropy to compute losses, with the implementation provided by the library in the sparse categorical cross-entropy (SCCE) class. The learning rate was set to 0.00001. The value was empirically chosen after multiple attempts of training the network.

5. EXPERIMENTS AND RESULTS

5.1. Training and accuracy

The training was performed using an Adam optimizer. Models were trained on AMD Ryzen 7 5700G CPU with 32GB of RAM, running on GNU/Linux operating system. The training process of the horizontal flow model took 242.1 seconds. The vertical model was trained in 253.5 seconds. The results are shown in plots: change of loss over the epoch, and confusion matrix of predictions (Figs. 2–4).

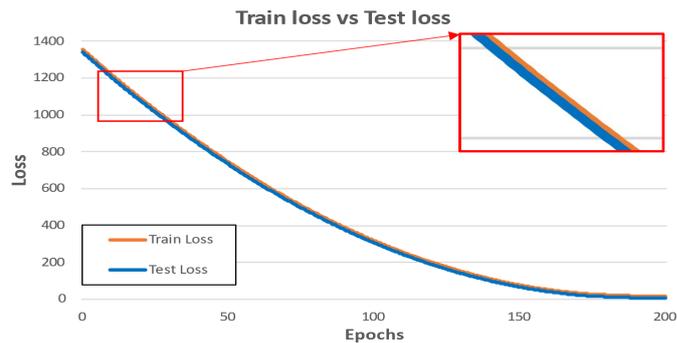


Fig. 2. Change of loss function over epochs for horizontal flow

Actual Class	(a) Training Set			(b) Test Set		
	disturbed	bubbly	stratified	disturbed	bubbly	stratified
disturbed	197	1	0	2	0	0
bubbly	0	79	0	0	2	0
stratified	4	0	21	0	0	2

Fig. 3. Confusion matrices of horizontal flow model prediction: (a) training set; (b) test set

Loss over epochs shows the impact of network changes on the error in time. The start value of the loss is positive. Its initial value depends on the network structure and the problem which the network tries to solve. However, the loss should decrease as a function of the epochs. Besides getting possibly the highest accuracy, the goal of the network is to obtain a loss close to 0. In the current study, the applied algorithm of SCCE sets high initial values for both loss functions (Figs. 2 and 4). This situation

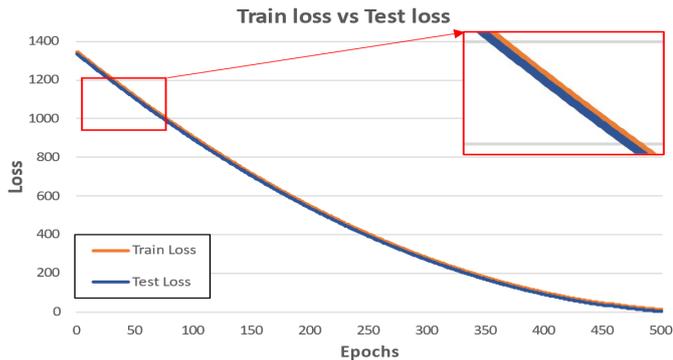


Fig. 4. Change of loss function over epochs for vertical flow

means that the model does not fit the problem correctly. Please note that the data reflects the nature of the flow process, which is dynamic and stochastic. In addition, the measurement data is noisy. The ECT technique measures small values of capacitance changes in the range of fF (femtofarads). Such a measurement is on the verge of measuring the resolution of ECT devices. Therefore, it is difficult to determine the correct structure of the model experimentally.

Nevertheless, the training algorithm successfully trained the model so that the loss functions for both the training and test data tended to zero, and the trace of both functions coincided.

Simultaneously, the prediction accuracy was logged over epochs. The accuracy range was between 0 and 1, where 1 determined that the network prediction is 100% correct. At the beginning of training, the value was low and started to increase. Accuracy fluctuated because the network tried to optimize itself. If the values of the prediction accuracy and the loss function reach the predetermined threshold levels, the training may be interrupted.

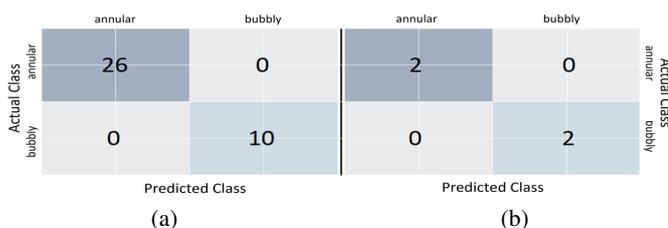


Fig. 5. Confusion matrices of vertical flow model prediction: (a) training set; (b) test set

A confusion matrix of predictions (Figs. 3 and 5) is a type of matrix that illustrates how the network classified the data. The matrix row represents the classes to which the data sample belongs. The columns correspond to predicted classes produced by the network. If the confusion matrix is diagonal, the network has correctly classified all data samples. Otherwise, the matrix shows what the correct class should be, e.g. in Fig. 3a data sample was classified as bubbly flow, but its real class is disturbed.

As expected, the loss functions presented in Figs. 2 and 4 show a decline over the epoch. Also, lines on the plots for test and training sets are aligned. The trained models do not show

signs of being overfitted. The accuracy of the test set in both cases is 100%. Such high values could be attributed to a low pool of samples (even including augmentation). The same can be said about the precision of the prediction of the training set of the vertical model. However, the prediction accuracy of the horizontal model for the training set is 98.3%. The confusion matrix shows (Fig. 3a) that this model classified one disturbed flow as bubbly and four stratified flows as disturbed. This non-perfect fit visible in confusion matrices is due to the imperfect model. The training dataset could contain some frames problematic for the network. The sequences of values in those frames could be close enough to each other. However, they described the different flows or transition patterns that can lie near the border lines between the flow types on a flow map. Please note that training the model was not based on flow maps and interphase border lines. If specific TPF patterns are defined during the model training, it attempts to match them with new measurements received online from ECT. If the model lacks information about potential transition patterns, accurate predictions cannot be anticipated. Moreover, exposing the network to more data during training would benefit the model, increasing the prediction correctness. This was proved by the experiments done as part of this research. The models were tested, adding various distortions to the measurement data. It could be observable, especially when the added noise deformed a part of the frame so that it would appear closer to another flow type.

6. THE ADAPTABILITY OF THE ANN MODEL TO EXCEPTIONAL SITUATIONS – DISCUSSION

In industrial conditions, the flow facility and the diagnostic or control devices are often exposed to exceptional situations or failures. Despite such significant obstacles, the neural network model should resist unforeseen problems, correctly predict the flow type, or detect these failures to avoid any pernicious damage. Therefore, a part of the research described in the article is the simulation of several emergencies. Each experiment examines models under different circumstances. Data had to be adapted for each test individually. Therefore, both the test and training sets were extended to simulate the following scenarios:

- Single electrode noise– failure simulation of one electrode from the whole measurement ECT system.
- Noise within pairs of electrodes of a single measuring card of the ECT system.
- Noise within a single ring of electrodes – simulation of electromagnetic interference, caused by, for example, other electromagnetic field sources placed near the electrodes.

Any extended datasets created for training purposes were excluded from simulations. The purpose of the experiments was to see how the models would react to originally gathered data from electrodes. Trials focused either on the single ring of electrodes, the separate measuring card, or the particular electrode.

6.1. Simulation of one electrode failure

First was the simulation of the first electrode disconnection. This can correspond to mechanical problems like the loose connection of the electrode. That required disfiguring the first 31 values

from a frame. To the original input signal (Fig. 6), the Gaussian noise

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(where μ is the mean and σ is the standard deviation) was added with μ equal to 0 and standard deviation: 0.1, 0.5, and 0.9. The distorted measurement values in a frame oscillated between -1 and 3 . See one of the results in Fig. 8. Figures 7 and 9 present the 3D images reconstructed for original and destroyed measurement data for horizontal and vertical flow. Compared to the original data, certain extra artefacts can be seen in images. The colours in the images reflect the normalized values for liquid (1 – red) and air (0 – blue).

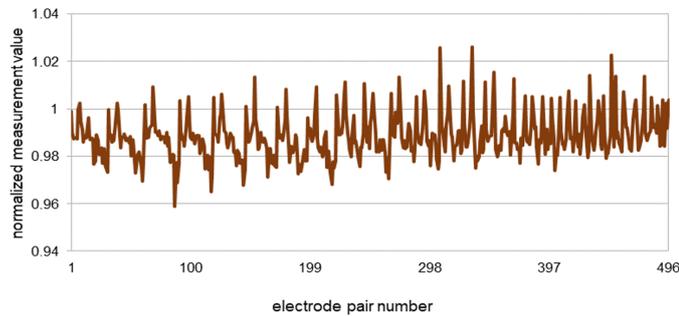


Fig. 6. An example frame with original measurement data

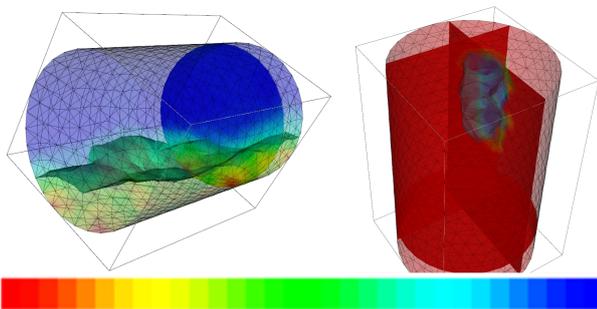


Fig. 7. Images reconstructed for some original data of horizontal and vertical flow

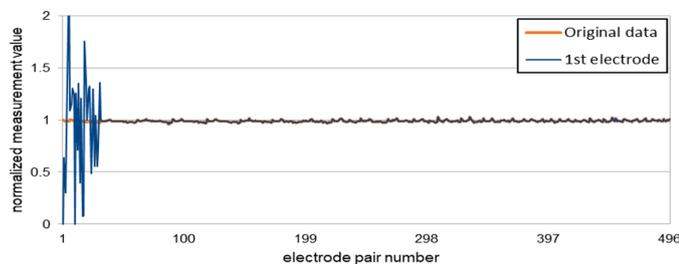


Fig. 8. An example frame with added noise ($\sigma = 0.5$) to measured values for all electrode pairs with the 1st electrode

The distortion of test data does not affect prediction for both horizontal and vertical flow compared to the original data (Fig. 6). In the case of the training set, the situation seems almost identical. The vertical model yields the same results as it would

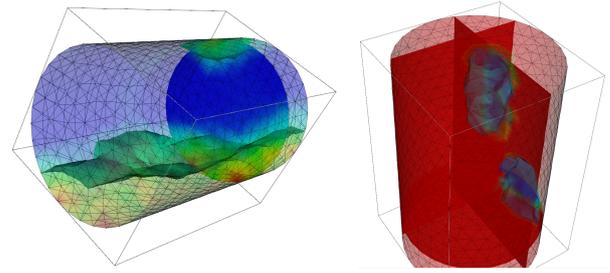


Fig. 9. Images reconstructed for some destroyed data of horizontal and vertical flow

without any distortion. Some misclassifications occur only for the horizontal model within the same classes as previously, just after training (Fig. 3a).

The reason for reliable results could be attributed to two aspects. Models used all values of a frame to make a prediction. That means, that the neural network can neglect or reduce parts of the frame where data values are higher than the average level. Secondly, it concerns how the data was prepared. Please note that some historical information is added when shifting parts of frames to the next frame. That gave the network a context of how the flow should look while going from one series of frames to another.

6.2. Simulation of ECT measuring card failure

The second experiment revolved around a simulation of one measuring card failure. To properly understand the concept, it is worth mentioning that in the research laboratory, 3D ECT measurement data was collected using the device named ET3 [30]. This device can connect 32 electrode sensors because it consists of eight measuring cards slugged to the main board, and each can control four electrodes. The construction of the main board slots in ET3 suffers from accidental disconnection of one card.

When one measuring card fails, the measurement values from attached electrodes are usually significantly out of the expected measurement range. In this scenario, electrodes from 1 to 4 were determined, as attached to the first card. The values representing the measurement pair, including these electrodes, were distorted (Fig. 10). Each frame of the set was modified. The Gaussian noise with μ equal to 1 was added to the original measurements. Figure 11 presents 3D images reconstructed for some measurement data destroyed in this scenario. Once more, horizontal and vertical models kept the same accuracy for modified

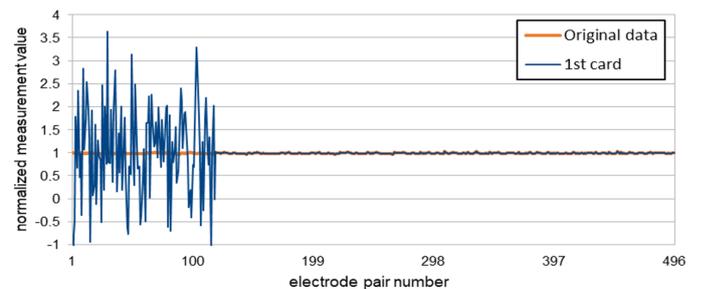


Fig. 10. An example of a frame where the 1st measurement card of an ECT device is broken

test data. In the case of an altered training set for both models, the neural network produced the same results as in the previous experiment (Table 1).

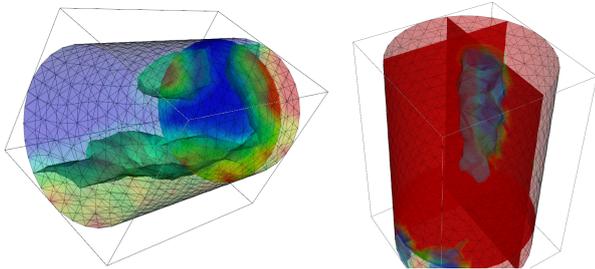


Fig. 11. Images reconstructed for some destroyed data of horizontal and vertical flow

Table 1

Prediction accuracy of models

σ	Horizontal model accuracy [%]		Vertical model accuracy [%]	
	Test dataset	Train dataset	Test dataset	Train dataset
0.1	100	98	100	100
0.5	100	98	100	100
0.9	100	98	100	1
1	100	98	75	100
1	100	98	100	96.4

- with added noise to the first electrode
- with added noise to the electrode pairs connected to the first measurement card of the ECT device
- with added noise to the measurement values for the values of electrode pairs from the first ring

However, the accuracy for the vertical model training set dropped from 100% to 75%. The 25% difference can be attributed to a small sample size – only 26 original data samples, especially when compared to the horizontal model, which used more original data for training (302 samples). Even though the first 122 values of a frame were modified, causing a loss of data at the level of 24.6%, the model for horizontal flow was able to upkeep a required prediction. This shows the model resilience to distortions of measurement data. As in the previous experiment, the condition of training the network impacts prediction. Also, the neural network might pay less attention to parts of data that are out of scale compared to the training dataset. In the case of the vertical model, the level of 100% accuracy on training data can confirm that. The decline in accuracy for the test set indicates that there are frames for which the network must consider the augmented parts before predicting. Moreover, the higher prediction accuracy in the model for the horizontal flow shows that providing a comprehensive training dataset allows neglecting that effect.

6.3. Simulation of electromagnetic interference

The last experiment simulated the impact of the outer electromagnetic field on the first electrode ring of the sensor. The 3D ECT device contains four electrode rings, each consisting of

eight electrodes. Therefore, the Gaussian noise was added to the measurement values correlated to the electrode pairs from the first ring (i.e. first eight electrodes) (Figs. 12 and 13). The noise intensity was at the same level as in the previous experiment.

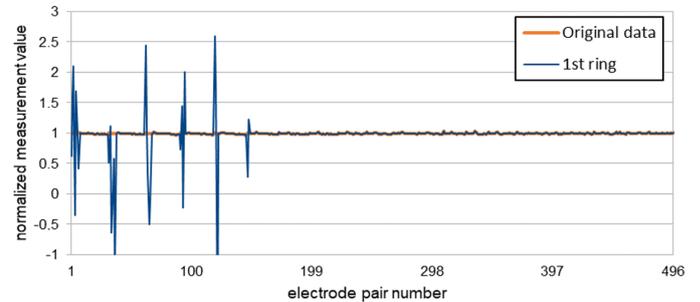


Fig. 12. An example of a measurement frame where an external electromagnetic field disrupted the measurements within the 1st electrode ring of the sensor

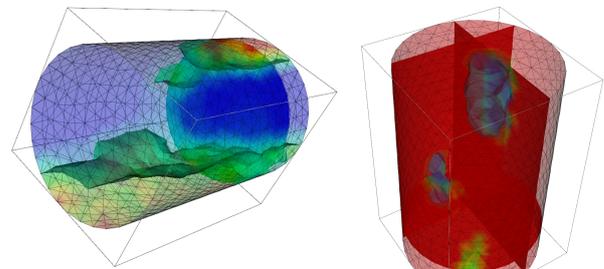


Fig. 13. Images reconstructed for some destroyed data of horizontal and vertical flow

Both horizontal and vertical models reached 100% accuracy for test datasets (Table 1). Horizontal model performance did not change for the training dataset, staying at 98%. The difference can be observed for the vertical model. Compared to the previous experiment, the prediction precision is lower and equal to 96.4% for the training dataset. Once more, the insufficient pool of training datasets impacts this.

In this case, the data was damaged aggregately at a lower level than in the previous attempt. In total, 28 values were modified, giving a loss of 5.6% of the information in a frame. The noisy values were cumulated into a series separated from the original data series. Hence, in this case, the model showed a much higher prediction accuracy and noticeable resistance to unexpected changes in some measurement values. A notable decrease of 3.6% was recorded only for the vertical flow model. Extending the training dataset should improve the model performance. The neural network can effectively filter damaged parts of frames, even when the noise is dispersed over the frame.

7. CONCLUSIONS

This research focuses on two aspects. Firstly, the ANN models were developed to classify the vertical and horizontal two-phase gas-liquid flows. In the article, the authors presented only part of the conducted experiments. It was decided to show results

only for two sections (one horizontal 53.6 mm and one vertical 81.4 mm) with the worse prediction level of NN models. For the described vertical case, the neural network had to distinguish two types: bubbly-type, disturbed-type, and three characteristic types for horizontal: bubbly-type, disturbed-type, and stratified-type. Training was done on data gathered from the 3D ECT system with 32 electrode sensors. The prediction was performed on raw measurement data without image reconstruction, ensuring the online mode if only the 3D ECT hardware could provide enough input measurement data. The archival measurement datasets used to train the models were insufficient to achieve high prediction accuracy. Therefore, some algorithms for data augmentation were developed and described.

The neural network model accuracy is significantly higher than the authors' previous work. In the case of TPF-type identification based on reconstructed tomography images, the accuracy of fuzzy classification fluctuated mainly between 85% and 99% in the horizontal case or 65% and above in the vertical one. The more stable prediction was in the case of the study with fuzzy inference based on raw ECT data without images. The average correct identification rate was about 90%. The differences in flow recognition between the human expert opinion and the decision of the developed algorithms were only for one type of flow, which was in the area of transitional boundary. However, both techniques (i.e. fuzzy and ML) are highly efficient for this task. The fuzzy model requires adjusting the membership functions to each pipeline (horizontal and vertical) at different diameters. There is also a need to determine the membership degrees for different flow types and write all the rules. So, this workflow is more complex than preparing a training set for a neural network model, even with an augmentation procedure.

While developing the classification techniques applied under industrial conditions one cannot forget about exceptional situations and the possibility of getting incorrect measurements generated by system failure or interferences. Systems that rely on flow classification use that information to adjust their parameters, e.g. change gas or liquid pressure/flow. Incorrect categorization due to the mentioned problems will provide wrong fine-tuning, resulting in overall process degradation or equipment damage. That is why the second part of this research focused on how different measuring device failures and interferences will affect the trained model. The simulations of several emergencies were done. Each experiment examined models under different circumstances as the disconnection of one electrode in the sensor, measuring card failure, or even the interference from the external electromagnetic source. The analysis of the result proved the adaptability of the ANN model to such exceptional situations.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Henryk Fidos, Dr Tomasz Jaworski, and Paweł Fiderek for their technical help and Professor Dominik Sankowski and Professor Jacek Kucharski for their fruitful advice.

This work was financed by the Institute of Applied Computer Science (TUL) as a part of statutory project no. 501/2-24-1-2.

REFERENCES

- [1] M.R. Rzasa, "The measuring method for tests of horizontal two-phase gas-liquid flows, using optical and capacitance tomography," *Nucl. Eng. Des.*, vol. 239, no. 4, pp. 699–707, Apr. 2009, doi: [10.1016/J.NUCENGDES.2008.12.020](https://doi.org/10.1016/J.NUCENGDES.2008.12.020).
- [2] T. Rymarczyk, G. Kłosowski, A. Hoła, J. Sikora, P. Tchórzewski, and Ł. Skowron, "Optimising the use of Machine learning algorithms in electrical tomography of building Walls: Pixel oriented ensemble approach," *Measurement*, vol. 188, p. 110581, Jan. 2022, doi: [10.1016/J.MEASUREMENT.2021.110581](https://doi.org/10.1016/J.MEASUREMENT.2021.110581).
- [3] Z.M. Łabęda-Grudziak and M. Lipiński, "The identification method of the coal mill motor power model with the use of machine learning techniques," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 69, no. 1, p. e135842, Feb. 2021, doi: [10.24425/bpasts.2021.135842](https://doi.org/10.24425/bpasts.2021.135842).
- [4] T. Pajchrowski, P. Siwek, and A. Wójcik, "Adaptive controller design for electric drive with variable parameters by Reinforcement Learning method," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 68, no. 4, pp. 1019–1030, Oct. 2020, doi: [10.24425/bpasts.2020.134667](https://doi.org/10.24425/bpasts.2020.134667).
- [5] Q. Zhu, X. Ji, J. Wang, and C. Cai, "A machine learning-based mobile robot visual homing approach," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 66, no. 5, pp. 621–634, 2020, doi: [10.24425/bpasts.2018.124278](https://doi.org/10.24425/bpasts.2018.124278).
- [6] P. Mane, S. Sonone, N. Gaikwad, and J. Ramteke, "Smart personal assistant using machine learning," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017*, Jun. 2018, pp. 368–371, doi: [10.1109/ICECDS.2017.8390128](https://doi.org/10.1109/ICECDS.2017.8390128).
- [7] R. Szmurło and S. Osowski, "Ensemble of classifiers based on CNN for increasing generalization ability in face image recognition," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 70, no. 3, p. e141004, 2022, doi: [10.24425/bpasts.2022.141004](https://doi.org/10.24425/bpasts.2022.141004).
- [8] J. Kryszyn and W. Smolik, "2D modelling of a sensor for electrical capacitance tomography in ECTSIM toolbox," *Informatics Control Meas. Econ. Environ. Prot.*, vol. 7, no. 1, pp. 146–149, Mar. 2017, doi: [10.5604/01.3001.0010.4604](https://doi.org/10.5604/01.3001.0010.4604).
- [9] T. Rymarczyk, K. Król, E. Kozowski, T. Wołowicz, M. Cholewa-Wiktor, and P. Bednarczuk, "Application of Electrical Tomography Imaging Using Machine Learning Methods for the Monitoring of Flood Embankments Leaks," *Energies*, vol. 14, no. 23, p. 8081, Dec. 2021, doi: [10.3390/EN14238081](https://doi.org/10.3390/EN14238081).
- [10] J.F. Conway, "The Science Behind Trash Data," *20visioneers15*. [Online]. Available: <https://www.20visioneers15.com/post/trash-data>
- [11] P. Teng, J. Yang, and M. Pfister, "Studies of Two-Phase Flow at a Chute Aerator with Experiments and CFD Modelling," *Model. Simul. Eng.*, vol. 2016, pp. 1–11, Aug. 2016, doi: [10.1155/2016/4729128](https://doi.org/10.1155/2016/4729128).
- [12] S.L. Kiambi, H.K. Kiriamiti, and A. Kumar, "Characterization of two phase flows in chemical engineering reactors," *Flow Meas. Instrum.*, vol. 22, no. 4, pp. 265–271, Aug. 2011, doi: [10.1016/j.flowmeasinst.2011.03.006](https://doi.org/10.1016/j.flowmeasinst.2011.03.006).
- [13] B. Vadlakonda and N. Mangadoddy, "Hydrodynamic study of two phase flow of column flotation using electrical resistance tomography and pressure probe techniques," *Sep. Purif. Technol.*, vol. 184, pp. 168–187, 2017, doi: [10.1016/j.seppur.2017.04.029](https://doi.org/10.1016/j.seppur.2017.04.029).
- [14] M.R. Rzaşa, Z. Kabza, and R. Gasz, "Wastewater flow measurement with adjustable slurge flow meter," *Przegląd Elek-*

Machine learning for two-phase gas-liquid flow regime evaluation based on raw 3D ECT measurement data

- trotechniczny*, vol. 98, no. 12, pp. 60–63, 2022, doi: [10.15199/48.2022.12.15](https://doi.org/10.15199/48.2022.12.15).
- [15] H. Abdulmouti, “Bubbly Two-Phase Flow: Part II- Characteristics and Parameters,” *Am. J. Fluid Dyn.*, vol. 4, no. 4, pp. 115–180, 2015, doi: [10.5923/j.ajfd.20140404.01](https://doi.org/10.5923/j.ajfd.20140404.01).
- [16] S.S. Lafmejani, A.C. Olesen, and S.K. K r, “VOF modelling of gas–liquid flow in PEM water electrolysis cell micro-channels,” *Int. J. Hydrogen Energy*, vol. 42, pp. 16333–16344, 2017, doi: [10.1016/j.ijhydene.2017.05.079](https://doi.org/10.1016/j.ijhydene.2017.05.079).
- [17] J. Niderla, T. Rymarczyk, and J. Sikora, “Manufacturing planning and control system using tomographic sensors,” *Inform. Control Meas. Econ. Environ. Prot.*, vol. 8, no. 3, pp. 29–34, Sep. 2018, doi: [10.5604/01.3001.0012.5280](https://doi.org/10.5604/01.3001.0012.5280).
- [18] C.E. Brennen, *Fundamentals of Multiphase Flows*, Cambridge. Cambridge University Press, 2005.
- [19] S. Osowski, B. Sawicki, and A. Cichocki, “Computational intelligence in engineering practice,” *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 69, no. 3, p. e137052, 2021, doi: [10.24425/bpasts.2021.137052](https://doi.org/10.24425/bpasts.2021.137052).
- [20] Y. Zhang and X. Ai, “The identification of two-phase flow regimes using HMM model based on ERT system and PCA feature extraction,” *Proc. – 2014 5th Int. Conf. Intell. Syst. Des. Eng. Appl. ISDEA 2014*, Dec. 2014, pp. 1053–1055, doi: [10.1109/ISDEA.2014.232](https://doi.org/10.1109/ISDEA.2014.232).
- [21] P. Fiderek, J. Kucharski, and R. Wajman, “Fuzzy inference for two-phase gas-liquid flow type evaluation based on raw 3D ECT measurement data,” *Flow Meas. Instrum.*, vol. 54, pp. 88–96, 2017, doi: [10.1016/j.flowmeasinst.2016.12.010](https://doi.org/10.1016/j.flowmeasinst.2016.12.010).
- [22] P. Wiedemann, A. D b, E. Schleicher, and U. Hampel, “Fuzzy flow pattern identification in horizontal air-water two-phase flow based on wire-mesh sensor data,” *Int. J. Multiph. Flow*, vol. 117, pp. 153–162, Aug. 2019, doi: [10.1016/j.ijmultiphaseflow.2019.05.004](https://doi.org/10.1016/j.ijmultiphaseflow.2019.05.004).
- [23] H.B. Arteaga-Arteaga *et al.*, “Machine learning applications to predict two-phase flow patterns,” *PeerJ. Comput. Sci.*, vol. 7, p. e798, 2021, doi: [10.7717/PEERJ-CS.798](https://doi.org/10.7717/PEERJ-CS.798).
- [24] F. Nie, H. Wang, Q. Song, Y. Zhao, J. Shen, and M. Gong, “Image identification for two-phase flow patterns based on CNN algorithms,” *Int. J. Multiph. Flow*, vol. 152, p. 104067, Jul. 2022, doi: [10.1016/j.ijmultiphaseflow.2022.104067](https://doi.org/10.1016/j.ijmultiphaseflow.2022.104067).
- [25] M. Ezzatabadipour, P. Singh, M. D. Robinson, P. Guillen-Rondon, and C. Torres, “Deep Learning as a Tool to Predict Flow Patterns in Two-Phase Flow,” *arXiv:1705.07117*, May 2017, Accessed: Sep. 11, 2018. [Online]. Available: <http://arxiv.org/abs/1705.07117>.
- [26] J. Xiao, X. Luo, Z. Feng, and J. Zhang, “Using artificial intelligence to improve identification of nanofluid gas–liquid two-phase flow pattern in mini-channel,” *AIP Adv.*, vol. 8, no. 1, p. 015123, Jan. 2018, doi: [10.1063/1.5008907](https://doi.org/10.1063/1.5008907).
- [27] Y. Zhang, A.N. Azman, K. W. Xu, C. Kang, and H.B. Kim, “Two-phase flow regime identification based on the liquid-phase velocity information and machine learning,” *Exp. Fluids*, vol. 61, no. 10, pp. 1–16, Oct. 2020, doi: [10.1007/S00348-020-03046-X/TABLES/7](https://doi.org/10.1007/S00348-020-03046-X/TABLES/7).
- [28] Y. Mi, M. Ishii, and L.H. Tsoukalas, “Flow regime identification methodology with neural networks and two-phase flow models,” *Nucl. Eng. Des.*, vol. 204, no. 1–3, pp. 87–100, Feb. 2001, doi: [10.1016/S0029-5493\(00\)00325-3](https://doi.org/10.1016/S0029-5493(00)00325-3).
- [29] C. Shorten and T.M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: [10.1186/S40537-019-0197-0/FIGURES/33](https://doi.org/10.1186/S40537-019-0197-0/FIGURES/33).
- [30] J. Kryszyn, W.T. Smolik, B. Radzik, T. Olszewski, and R. Szabatin, “Switchless charge-discharge circuit for electrical capacitance tomography,” *Meas. Sci. Technol.*, vol. 25, no. 11, p. 115009, Nov. 2014, doi: [10.1088/0957-0233/25/11/115009](https://doi.org/10.1088/0957-0233/25/11/115009).