1. Grzegorz KŁOSOWSKI[1], 2. Michał OLESZEK[2,3], 3. Dariusz WÓJCIK[2,3], 4. Konrad NIDERLA[2]

Lublin University of Technology (1), WSEI University (2), Research & Development Center Netrix S.A. (3)
ORCID: 1. 0000-0002-4678-4512; 2. 0000-0002-0114-2794; 3. 0000-0002-4200-3432, 4. 0000-0003-1280-0622

doi:10.15199/48.2025.02.16

The use of LSTM with self-attention layer for monitoring industrial reactors using impedance and capacitance tomography

Abstract. *The study aimed to develop a method to improve imaging resolution in industrial electrical tomography. For this purpose, an algorithmic model of a neural network based on LSTM layers was developed, with two input sequences and a self-attention layer as characteristic features. Electrical impedance tomography measures voltage, and electrical capacitance measures capacitance. In the proposed solution, the LSTM network turns measurements into images using both types of measurements. The hybrid approach reduces the underdetermination of the inverse problem, improving imaging quality. To objectively evaluate the proposed solution, the new model was compared with the classical model, using only EIT measurements. Four quantitative criteria were used, in the form of popular indicators used to measure image quality. The results confirmed the advantage of the proposed approach over the classical model.*

Streszczenie. *Celem badań było opracowanie metody poprawiającej rozdzielczość obrazowania w przemysłowej tomografii elektrycznej. W tym celu opracowano model algorytmiczny sieci neuronowej opartej o warstwy LSTM, której cechą charakterystyczną były dwie sekwencje wejściowe oraz zastosowanie warstwy samouwagi. W proponowanym rozwiązaniu, sieć LSTM transformująca pomiary na obrazy wykorzystuje zarówno pomiary napięć, charakterystyczne dla elektrycznej tomografii impedancyjnej, jak również pomiary pojemności elektrycznej. Podejście hybrydowe redukuje niedookreśloność problemu odwrotnego, co poprawia jakość obrazowania. W celu obiektywnej oceny proponowanego rozwiązania porównano nowy model z modelem klasycznym, wykorzystującym jedynie pomiary EIT. Zastosowano cztery kryteria ilościowe, w postaci popularnych wskaźników wykorzystanych jako miary jakości obrazów. Rezultaty potwierdziły przewagę proponowanego podejścia nad modelem klasycznym. (**Zastosowanie LSTM z warstwą samouwagi do monitorowania reaktorów przemysłowych za pomocą tomografii impedancyjnej i pojemnościowej**).*

Keywords: electrical tomography; machine learning; industrial tomography; LSTM networks
Słowa kluczowe: tomografia elektryczna; uczenie maszynowe; tomografia przemysłowa; sieci LSTM

Introduction

Industrial tank reactors are key elements of almost all process manufacturing systems. They are used in various industrial sectors, such as chemical, petrochemical, pharmaceutical, food and energy. Reactors are also used in biogas plants and sewage treatment plants. The main areas of reactor applications include chemical synthesis, raw material processing, drug and food production, and energy generation.

Continuous monitoring of the processes inside the reactor directly affects the quality of the product and the economics of industrial processes. Precise control of reaction parameters allows for optimizing efficiency, minimizing raw material consumption and reducing waste. However, due to the complex physical conditions prevailing inside the reactor—high temperature, pressure, aggressive chemical environment and lack of light—observing chemical reactions and dynamic changes using traditional methods is impossible.

Classical methods of monitoring reactor operation rely on sensors that provide indirect point information about the current state of the process. This limits the ability to fully understand the phenomena and delays the activities related to process control. This is important, especially when observing phase changes, such as identifying gas bubbles and crystals formed in a liquid during dynamic processes.

Tomography is the only one known non-invasive method for imaging cross-sections of the reactor interior. Images provide valuable information on the phase distribution and dynamic changes occurring in real time. The use of impedance and capacitance tomography allows monitoring processes without the need to interfere with the reactor structure [1,2].

This paper describes the use of an LSTM network with a self-attention layer to monitor industrial reactors in the context of impedance and capacitance tomography [3]. By simultaneously using two types of measurements and a self-attention layer, the trained model better transforms measurements into tomographic images. The proposed method aims to improve the accuracy and efficiency of monitoring processes occurring inside reactors, which can help to optimize industrial processes and improve the quality of final products.

Measurement Hardware

The hybrid tomograph is a prototype sophisticated measurement apparatus developed in Netrix S.A. laboratories. It uses two techniques, Electrical Impedance Tomography (EIT) and Electrical Capacitance Tomography (ECT), to examine items [4–6]. This dual-method approach facilitates a thorough investigation by acquiring impedance and capacitance data for the subjects under examination. The Howland circuit is a crucial element of the hybrid tomograph, producing a consistent current signal for accurate impedance measurements. The Direct Digital Synthesis (DDS) module is fundamental to this circuit, facilitating accurate modulation and stability of the current signal. DDS technology facilitates the swift and precise creation of current signals with specified characteristics, including frequency, amplitude, and phase, improving the system's efficiency and accuracy.

Analog-to-digital converters (ADCs) are essential for transforming analog signals from several modules into digital data for subsequent analysis and image reconstruction. The tomograph utilizes high-resolution 16-bit ADC modules with rapid processing speeds of 500 kbps per channel. High resolution and speed provide accurate and instantaneous data capture, which is especially critical in applications where the condition of the measured item fluctuates rapidly. Figure 1 shows the physical model of an industrial reactor connected to a hybrid EIT/ECT device.

The apparatus features a digitally regulated gain module to provide excellent signal integrity before digitization. This module modulates the measurement signal's amplitude according to varying operating conditions through software control, providing flexibility and efficiency without needing

physical modifications. Digital gain control enables automatic calibration and adaptation to particular experimental requirements. The data controller manages data processing, analysis, and administration, converting raw input data into valuable information for analysis, diagnostics, or monitoring. The controller processes signals from the ADCs, performing functions such as filtering, decoding, and translating signals into suitable data formats. It also oversees memory resources and user interface components, including a touchscreen and graphical control panel. The data controller offers alternatives for data storage and transmission, enabling data to be stored on external media such as USB drives and delivered through interfaces like USB and Ethernet. This functionality expedites research and diagnostic procedures, allowing the tomograph to assimilate into extensive research networks for remote data analysis and processing. The control and data processing unit employs integrated FPGA and ARM Cyclone V circuitry. The Field-Programmable Gate Array (FPGA) handles specific tasks related to signal processing and real-time data processes, making it possible to make changes that fit different application needs. The ARM Cortex-A9 microprocessor facilitates comprehensive system control and manages advanced functions, such as user interface administration and external device connectivity. Dedicated ROM and RAM modules, communication interfaces like USB Host and 1000BASE-T Ethernet, and external DDR RAM enhance data management and system performance. The complete system is interconnected with a bus that enables communication and data transfer among all components.

Fig. 1. The measurement setup – a physical model of an industrial reactor connected to a hybrid EIT/ECT device

Methods

A refined Long Short-Term Memory (LSTM) network architecture is proposed, using a dual-channel "cell" structural variable at the input stage to accommodate distinct EIT and ECT measurement sequences. The first channel pertains to the EIT measurement sequence, whereas the second channel is allocated to the ECT measurement sequence [4]. This distinctive arrangement facilitates the independent and detailed examination of each tomographic measurement type, thus maintaining their unique attributes and any interrelations. The upgraded LSTM architecture incorporates a self-attention layer as a significant innovation. Figure 2 shows the arrangement of individual layers of the LSTM model.

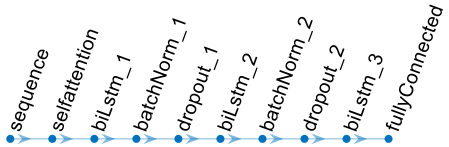

Fig. 2. Training performance of the two-sequence

Table 1 describes the structure of the LSTM network used in the research.

Table 1. The two-sequence LSTM neural network architecture

No.	Layers
1	`sequenceInputLayer(2,"Name","sequence", "Normalization","zerocenter")` Sequence input with 2 dimensions (2×96)
2	`selfAttentionLayer(4,256,"Name", "selfAttention")` Self attention layer with 4 heads and 256 key and query channels
3	`bilstmLayer(512,"Name","biLstm_1", "OutputMode","sequence")` BiLSTM with 512 hidden units
4	`batchNormalizationLayer("Name", "batchNorm_1")` Batch normalization
5	`dropoutLayer(0.2,"Name","dropout_1")` 20% dropout
6	`bilstmLayer(1024,"Name","biLstm_2", 'OutputMode', 'sequence')` BiLSTM with 1024 hidden units
7	`batchNormalizationLayer("Name", "batchNorm_2")` Batch normalization
8	`dropoutLayer(0.2,"Name","dropout_2")` 20% dropout
9	`bilstmLayer(512,"Name","biLstm_3", 'OutputMode', 'last')` BiLSTM with 512 hidden units
10	`fullyConnectedLayer(numResponses, "Name","fullyConnected_1")` 20445-voxel fully connected layer

The model processes two measurement sequences as input, each consisting of 96-element vectors. The first sequence is derived from EIT, while the second is derived from ECT. The ECT vector originally contained 120 elements, but Lasso regularization was applied to reduce its dimensionality to 96, ensuring consistency with the EIT sequence. The model's output is a 3D reconstruction image consisting of 20,445 voxels.

The network architecture starts with a sequence input layer that processes the two input sequences. The data then passes through a self-attention layer with four heads and 256 key-query channels, which enhances the model's ability to focus on relevant parts of the sequence and model long-range dependencies. This is followed by a bidirectional LSTM (BiLSTM) layer with 512 hidden units, which allows the model to capture temporal relationships in both directions. Batch normalization and dropout are applied to regularize activations and prevent overfitting. A second BiLSTM layer of 1024 hidden units further processes the sequence data, using similar normalization and dropout techniques. The third BiLSTM layer of 512 hidden units uses only the final hidden state to distill information from the sequence, after which a fully connected layer outputs the final 3D reconstruction image. This layered design, integrating attention mechanisms, BiLSTM layers, and regularization techniques, is optimized for processing temporal and spatial relationships in the input sequences to produce a detailed and accurate 3D image.

The neural network architecture being examined is developed in MATLAB and is tailored for sequence-to-sequence mapping tasks. The network is engineered to process an input signal, which includes 2 channels and 96 time steps. Utilizing the Gauss-Newton method and the finite element approach within the Eidors toolbox (a Matlab extension), over 30,000 training cases were produced to address a straightforward problem. In tomographic data processing, raw EIT data consists of a vector of 96 voltage measurements, whereas ECT data comprises a vector of

120 capacitance measurements across 16 measurement channels. For the application of LSTM networks to these data streams, both channels must have an exact number of time steps [3]To synchronize the temporal aspects of these two modalities, the number of measures in the ECT data was decreased from 120 to 96. LASSO (Least Absolute Shrinkage and Selection Operator) regularization was utilized. Each series is depicted as a numChannels-by-numTimeSteps (2×96) numeric array, where numChannels denotes the number of channels and numTimeSteps represents the number of time steps in the sequence. Channel 1 comprises EIT measurements, while channel 2 encompasses the ECT measurement process [7-10].

Figure 3 shows the training graph of the neural network in question. In this case, Loss = mse (Mean Square Error). The shape resembling a hyperbola and the lack of fluctuations indicate the correct training process. This gives reason to believe that the trained network will be characterized by the ability to generalize.

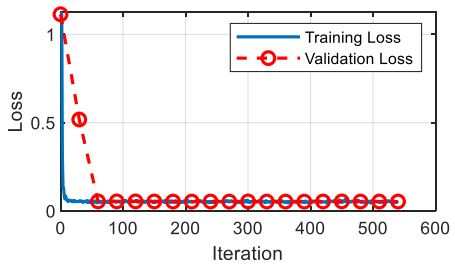

Fig. 3. Training performance of the two-sequence LSTM network

It is worth noting how the 120-element measurement sequence was shortened to its 96-element representation. For this purpose, LASSO (Least Absolute Shrinkage and Selection Operator) regularization was used.

In this study, we address the challenge of reducing a high-dimensional feature space to enhance computational efficiency and mitigate issues related to overfitting in predictive modeling. Specifically, we aim to minimize a 120-dimensional feature set to a 96-dimensional representation while retaining the most informative features. The original dataset, denoted as $\mathrm{X} \in \mathbb{R}^{n \times p}$, consists of $n = 30{,}000$ observations and $p = 120$ features.

We employ a feature selection technique based on the LASSO regression to achieve dimensionality reduction. LASSO regression is particularly suitable for feature selection due to its ability to impose L_1–norm regularization, which can shrink some coefficients exactly to zero, thereby performing variable selection implicitly.

The process involves iteratively treating each feature as the response variable and using the remaining features as predictors. For each feature $i \in \{1,2,\ldots,p\}$, we define the response vector $y = x_i \in \mathbb{R}^n$, where x_i is the i-th column of X and the predictor matrix $\mathrm{X}_{-i} \in \mathbb{R}^{n \times (p-1)}$, which consists of all columns of X except x_i. We then fit a LASSO regression model to estimate the relationship between y and X_{-i}:

$$\widehat{\boldsymbol{\beta}}^{(i)} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} |\boldsymbol{y} - \boldsymbol{X}_{-i}\boldsymbol{\beta}|_2^2 + \lambda |\boldsymbol{\beta}|_1 \right\} \quad (1)$$

where $\boldsymbol{\beta} \in \mathbb{R}^{p-1}$ is the vector of regression coefficients excluding the $i-th$ feature, $\|\cdot\|_2$ denotes the Euclidean norm, $\|\cdot\|_1$ represents the L_1–norm, and $\lambda > 0$ is the regularization parameter. In our analysis, we set $\lambda = 0.01$, which balances the trade-off between model complexity and feature selection.

For each regression, we obtain a set of coefficients $\widehat{\boldsymbol{\beta}}^{(i)}$. To quantify the importance of each feature across all regressions, we accumulate the absolute values of the coefficients. Specifically, the importance score s_j for feature j is computed as:

$$s_j = \sum_{\substack{i=1 \\ i \neq j}}^{p} \left| \hat{\beta}_j^{(i)} \right| \quad (2)$$

where $\hat{\beta}_j^{(i)}$ is the coefficient corresponding to feature j in the i-th regression. This summation excludes the case when $i = j$ since feature j is not included as a predictor when it is the response variable.

After calculating the importance scores for all features, we rank them in descending order. The top 96 features with the highest importance scores are selected for inclusion in the reduced feature set. Let $S \subseteq \{1,2,\ldots,p\}$ denote the indices of the selected features such that $|S| = 96$. The reduced dataset $X_{reduced} \in \mathbb{R}^{n \times 96}$ is then formed by extracting the columns of X corresponding to the indices in S:

$$X_{reduced} = X_{[:,S]} \quad (3)$$

This method efficiently detects and retains the most significant characteristics by utilizing the sparsity-inducing characteristic of LASSO regression. It identifies direct and indirect links within the data by examining the relationships among all features. The diminished feature set retains the critical information required for further studies while alleviating processing demands and minimizing potential multicollinearity. Using LASSO regression for feature selection offers a methodical and scalable strategy for dimensionality reduction in high-dimensional datasets. The 96-dimensional representation is the best compromise between making the computer work quickly and keeping important data from the original 120-dimensional feature space.

Results

Table 2 compares the quality of reconstructions obtained using the improved LSTM network with the standard network. The classic network used a single EIT measurement sequence of 96 electrode voltage values.

Table 2. Reconstruction quality assessment indicators

Method	#	MSE	PSNR	SSIM	ICC
Regular LSTM	1	0.0529	12.7574	0.6510	0.8240
	2	0.0371	14.3032	0.6337	0.8145
	3	0.0098	20.0652	0.6559	0.7938
	4	0.0940	10.2657	0.4926	0.8660
Improved LSTM	1	0.0184	17.3469	0.7182	0.8508
	2	0.0196	17.0573	0.6586	0.8449
	3	0.0064	21.8997	0.7501	0.8098
	4	0.0319	14.9609	0.5141	0.8665

The second difference was the lack of a self-attention layer. The values of MSE (Mean Squared Error), PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), and ICC (Intraclass Correlation Coefficient) indices were compared for both networks.

Table 2 displays the reconstruction quality indicators for regular and upgraded LSTM models in four measurement cases that match those in Figure 4.The typical LSTM model's MSE values vary from 0.0098 in case 3 to 0.0940 in case 4, demonstrating a wide reconstruction accuracy range. The PSNR values reveal that reconstruction quality inversely varies with MSE, with case 3 having the greatest PSNR of 20.0652 and case 4 the lowest at 10.2657. MSE and PSNR are better in most situations, especially in example 3, where the upgraded LSTM model has the lowest MSE of 0.0064 and the greatest PSNR of 21.8997, demonstrating its increased reconstruction capabilities. Compared to standard

models, improved LSTMs had higher SSIM values, with case 3 having the highest SSIM of 0.7501.Both models had good ICC values in all circumstances, indicating model output consistency and reproducibility. The upgraded LSTM had better ICCs, indicating greater reliability, especially in example 4, which achieved 0.8665. In most cases, the modified LSTM beats the normal model, with lower MSE, higher PSNR, SSIM, and ICC values, and superior reconstruction quality. A comparison of the reconstructions seen in Figure 4 also confirms this.

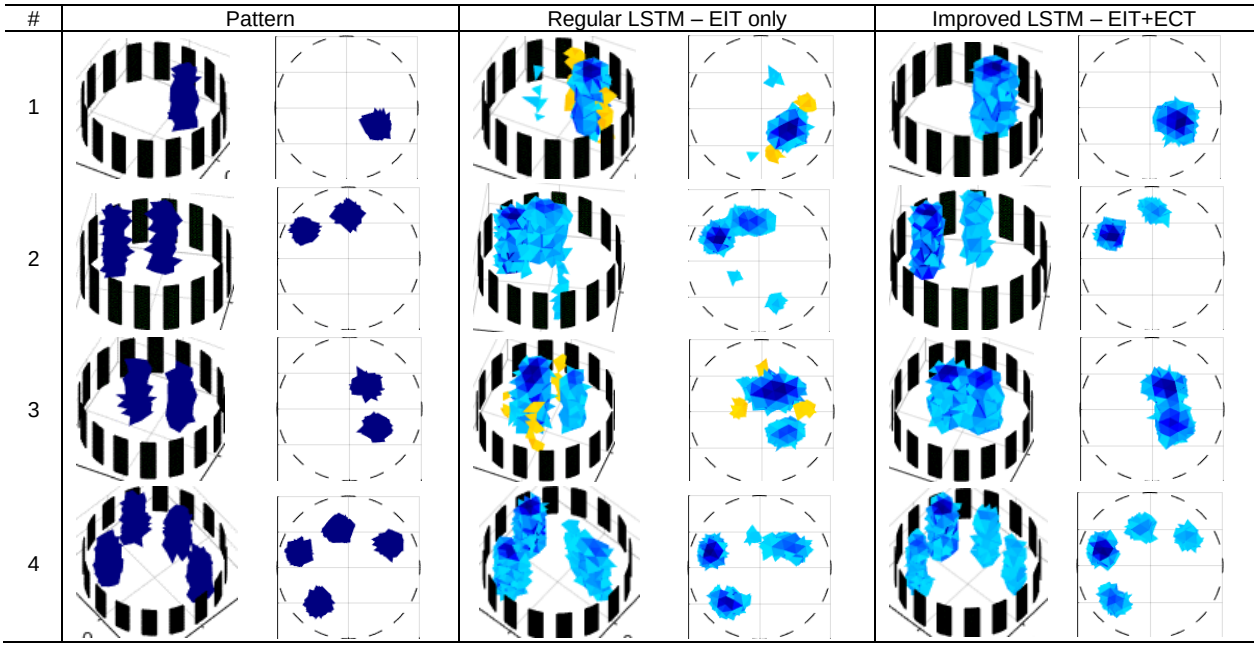

Fig. 4. Comparison of regular and improved LSTM networks

Conclusions

In this study, an improved algorithmic model was proposed that effectively transforms tomographic measurements into reconstruction images. For this purpose, a neural network of ten layers was trained, including three LSTM layers and a self-attention layer. The model input consists of two sequences - voltage and capacitive- and is worth noting. An original method using LASSO regularization was used to adapt the 120-element capacitive sequence to the 96-element voltage sequence. The results of the experiments revealed a clear advantage of the new approach over the classical approach. Further research will be conducted on preprocessing input measurement sequences using Fourier and Hilbert transforms.

***Authors**: Grzegorz Kłosowski, Ph.D. Eng., Lublin University of Technology, Lublin, Poland, E-mail: g.klosowski@pollub.pl; Michał Oleszek, MSc., WSEI University, Lublin, Poland; Research & Development Center Netrix S.A., Lublin, Poland, E-mail:: michal.oleszek@wsei.pl; Dariusz Wójcik, Ph.D., WSEI University, Lublin, Poland; Research & Development Center Netrix S.A., Lublin, Poland, E-mail:: dariusz.wojcik@wsei.pl; Konrad Niderla, MSc., WSEI University, Lublin, Poland; Research & Development Center Netrix S.A., Lublin, Poland, E-mail:: konrad.niderla@wsei.pl.*

REFERENCES

[1] Krawczyk, A.; Korzeniewska, E. Metaphysical and Medical Aspects of Electromagnetism. *Przeglad Elektrotechniczny*. 99 (2023), No.3, 124-127

[2] Pawłowski, S.; Plewako, J.; Korzeniewska, E. Influence of Structural Defects on the Resistivity and Current Flow Field in Conductive Thin Layers, *Electronics*, 9 (2020), 2164

[3] Kulisz M., Kłosowski G., Rymarczyk T., Hoła A., Niderla K., Sikora J., The use of the multi-sequential LSTM in electrical tomography for masonry wall moisture detection, *Measurement,* 234 (2024) 114860.

[4] Baran B, Kozłowski E, Majerek D, Rymarczyk T, Soleimani M, Wójcik D. Application of Machine Learning Algorithms to the Discretization Problem in Wearable Electrical Tomography Imaging for Bladder Tracking, *Sensors,* 23 (2023); No. 3,1553

[5] Przysucha B, Wójcik D, Rymarczyk T, Król K, Kozłowski E, Gąsior M. Analysis of Reconstruction Energy Efficiency in EIT and ECT 3D Tomography Based on Elastic Net. *Energies,* 16 (2023); No. 3, 1490

[6] Kłosowski G, Rymarczyk T, Niderla K, Kulisz M, Skowron Ł, Soleimani M. Using an LSTM network to monitor industrial reactors using electrical capacitance and impedance tomography – a hybrid approach, *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, 25 (2023). No.1

[7] Rybak G, Kozłowski E, Król K, Rymarczyk T, Sulimierska A, Dmowski A, Bednarczuk P. Algorithms for Optimizing Energy Consumption for Fermentation Processes in Biogas Production. *Energies,* 16 (2023); No. 24, 7972

[8] Kozłowski E., Borucka A., Oleszczuk P., Jałowiec T., Evaluation of the maintenance system readiness using the semi-Markov model taking into account hidden factors, *Eksploatacja i Niezawodność – Maintenance and Reliability,* 25 (2023) ;No. 4, 172857

[9] Soleimani M., Rymarczyk T., Kłosowski G., Ultrasound Brain Tomography: Comparison of Deep Learning and Deterministic Methods, *IEEE Transactions on Instrumentation and Measurement,* 73 (2024), 4500812

[10] Król, K., Rymarczyk, T., Niderla, K., & Kozłowski, E., Sensor platform of industrial tomography for diagnostics and control of technological processes. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska,* 13 (2023), No.1, 33–37