Optical and Quantum Electronics (2022) 54:501
https://doi.org/10.1007/s11082-022-03799-1

Influence of data scaling and normalization on overall neural network performances in photoacoustics

K. Lj. Djordjević[1] · M. I. Jordović-Pavlović[2] · Ž. M. Ćojbašić[3] · S. P. Galović[1] · M. N. Popović[1] · M. V. Nešić[1] · D. D. Markushev[4]

Received: 26 September 2021 / Accepted: 26 April 2022

Abstract
In this paper, the influence of the input and output data scaling and normalization on the neural network overall performances is investigated aimed at inverse problem-solving in photoacoustics of semiconductors. The logarithmic scaling of the photoacoustic signal amplitudes as input data and numerical scaling of the sample thermal parameters as output data are presented as useful tools trying to reach maximal network precision. Max and min–max normalizations to the input data are presented to change their numerical values in the dataset to common scales, without distorting differences. It was demonstrated in theory that the largest network prediction error of all targeted parameters is obtained by a network with non-scaled output data. Also, it was found out that the best network prediction was achieved with min–max normalization of the input data and network predicted output data scale within the range of [1–10]. Network training and prediction performances analyzed with experimental input data show that the benefits and improvements of input and output scaling and normalization are not guaranteed but are strongly dependent on a specific problem to be solved.

Keywords Photoacoustic · Semiconductors · Artificial neural networks · Thermal diffusion · Thermal expansion · Photothermal · Inverse problem

This article is part of the Topical Collection on Photonics: Current Challenges and Emerging Applications, Guest edited by Jelena Radovanovic, Dragan Indjin, Maja Nesic, Nikola Vukovic and Milena Milosevic.

✉ M. I. Jordović-Pavlović
miroslava.jordovic-pavlovic@vpts.edu.rs

[1] Vinca Institute of Nuclear Sciences-National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia

[2] Western Serbia Academy of Applied Studies, Trg Svetog Save 34, Uzice, Serbia

[3] Faculty of Mechanical Engineering, University of Niš, Niš, Serbia

[4] Institute of Physics Belgrade - National Institute of the Republic of Serbia, University of Belgrade, Belgrade-Zemun, Serbia

Published online: 12 July 2022

1 Introduction

It is a well-known fact that data scaling and normalization in neural networks are the techniques usually applied as part of the data preparation process. Their ultimate goal is to change the values of data in the dataset to a common scale without distorting input differences. The main difference, followed in this article, between scaling and normalization is that, in scaling, the data range is changing, helping one to compare different variables in the same range. In normalization, the data range is changed from the original so that all values are within the range of 0 and 1, without distorting differences in the ranges of values or losing information. In general, scaling helps one to compare different variables on equal footing and normalization is used trying to improve the network model due to its numerical stability and quality of the training process. Both scaling and normalization can be applied in neural networks not only on input (for example, arbitrary signals) but on output data (predicted parameters) as well (Balderas-Lopez 2006; Balderas-López et al. 2002; Govorkov et al. 1997; Melo and Faria 1995).

In photoacoustics (PA), scaling and normalization process is usually used in experimental data analysis for accurate and reliable characterization of the investigated materials, preferably in two cases: (1) to eliminate the influence of the measuring system and (2) to find the differences in photoacoustic signal behavior by detecting small changes that correspond to possible thermal, mechanic and/or electronic parameters variations. In both cases, photoacoustics uses the data set containing two features: amplitudes and phases in the modulation frequency range from (20–20 k) Hz. In the given frequency range the phase varies from (0 to 360) deg, while the amplitude variations are within the range of (10^{-3}–1^{-8}) a.u. and lower. It is obvious that these two features are in different ranges (Balderas-Lopez 2006; Balderas-López et al. 2002; Calderón et al. 1998; Dramićanin et al. 1995; Govorkov et al. 1997; Melo and Faria 1995; Todorović et al. 2014, 1995). Within the framework of artificial neural networks (ANNs), defining the main feature as an input is a very important issue (Djordjevic et al. 2020, 2019; Jin et al. 2015; Kim 1999). Many authors define the amplitudes as the main feature and more important predictor in photoacoustic signal analysis than phases (El-Brolossy and Ibrahim 2010; Velasco et al. 2011). But phases are the constitutive part of the signals, too, also very sensitive to the sample parameters changes (Pichardo-Molina and Alvarado-Gil 2004). It means that both amplitudes and phases have to be considered as predictors of equal importance (Ordóñez-Miranda and Alvarado-Gil 2009).Based on our experience and the application of ANNs in photoacoustics, the equal importance of these predictors allows us to use, for simplicity, only one of them to characterize the sample, without significant changes in the quality of sample parameters prediction (Arridge et al. 2019; Yahyaoui's et al. 2018). This is the reason why, in this article, we will use only amplitudes of the photoacoustic signal as an ANNs input.

In our previous articles we have shown that the application of ANNs in photoacoustics could improve experimental procedures in many ways: better accuracy and precision in investigated sample parameters prediction, better control of the experimental conditions together with approaching to the real-time characterization of the investigated sample, etc. (Djordjevic et al. 2020, 2019). Here, in this article, we will try to show why the different types of scaling and normalization procedures of input and output data could be beneficial to the accuracy, precision and numerical stability of the network predicted parameters, and to the process of network training acceleration. To do that logarithmic scaling and min–max and max normalizations will be applied on input data used in the ANNs training process. At the same time, simple numerical scaling will be used for network output data

(predicted sample thermal and geometric parameters such as: thermal diffusivity, linear coefficient of thermal expansion, thickness…) to find possible benefits to ANNs performances. Our analysis of training, stability and accuracy of network prediction will rely on the ANNs that are trained with or without scaling and/or normalization of input and output data aiming to find their influence to overall network performances. All analyzes will be done on a theoretical model first. Then special attention will be given to the benefits of scaling and normalization within the application of ANNs on experimental data.

2 Photoacoustic theory and experimental response

The photoacoustic signals are generated as a consequence of the thermal state changes in semiconductors due to the absorption of monochromatic modulated light source having the intensity of $I = I_0\left(1 + e^{i\omega t}\right)$ (I_0 is the incident light amplitude and $\omega = 2\pi f$ where f is the modulation frequency). Using theoretical model of composite piston the photoacoustic signals in semiconductors, so called total PA signals $\delta\tilde{p}_{\text{total}}(f)$ generated by the investigated sample can be presented in the form (Calderón et al. 1998; Dramićanin et al. 1995; Mandelis 1999; Markushev et al. 2018, 2015, 2019; Todorović et al. 2014, 1995):

$$\delta\tilde{p}_{\text{total}}(f) = K_1\tilde{T}_{\text{s}}(l,f)e^{i\left(\omega t-\frac{\pi}{4}\right)} + K_2\alpha_{\text{T}}\frac{R_{\text{s}}^4}{l^3}\int_{-l/2}^{l/2} z\,\tilde{T}_{\text{s}}(x,f)dx + K_3 d_{\text{n}}\frac{R_{\text{s}}^4}{l^3}\int_{-l/2}^{l/2} z\,\delta\tilde{n}_{\text{p}}(x,f)dx \tag{1}$$

where $K_i, i = 1, 2, 3$ are the constants dependent on the thermodynamic state of the gas in the photoacoustic cell and its geometrical properties, R_s and l are the radius and thickness of the sample, respectively, α_{T} is the coefficient of linear expansion of the sample, d_{n} is the coefficient of electronic deformation, $\delta\tilde{n}_{\text{p}}(x,f)$ is the concentration of minority excess charge carriers and $\tilde{T}_{\text{s}}(x,f)$ is the temperature distribution along the sample. Knowing that the solutions of $\tilde{T}_{\text{s}}(x,f)$ and $\delta\tilde{n}_{\text{p}}(x,f)$ can be found using complex analysis, Eq. (1) can be presented, also, in the simplified complex form:

$$\delta\tilde{p}_{\text{total}}(f) = A_{\text{total}}(f)\exp\left[i\phi_{\text{total}}(f)\right] \tag{2}$$

where $A_{\text{total}}(f)$ is the total signal amplitude and $\phi_{\text{total}}(f)$ is the total signal phase. Usually, measured experimental signals $\tilde{S}^{\text{exp}}_{\text{total}}(f)$ are the combinations of $\delta\tilde{p}_{\text{total}}(f)$ and influence of the measuring system $\tilde{H}_{\text{total}}(f)$ (the deviations caused by the microphone and accompanied electronics), written as:

$$\tilde{S}^{\text{exp}}_{\text{total}}(f) = \tilde{p}_{\text{total}}(f)\cdot\tilde{H}_{\text{total}}(f) \tag{3}$$

To eliminate the influence $\tilde{H}_{\text{total}}(f)$ of the measuring system, correction procedure of $\tilde{S}^{\text{exp}}_{\text{total}}(f)$ was established aiming to find characteristic parameters of the used instruments (microphone and accompanied electronics) and remove $\tilde{H}_{\text{total}}(f)$ from Eq. (3). In such way only $\delta\tilde{p}_{\text{total}}(f)$ remains in the Eq. (3), suitable to fit with Eq. (1). Mentioned fitting procedure will allows one to extract all thermal (for example D_{T},α_{T},$\tilde{T}_{\text{s}}(x,f)$) and electronic (for example $\delta\tilde{n}_{\text{p}}(x,f)$,$d_{\text{n}}$) parameters needed for sample material characterization.

In this article, theoretically obtained $\delta\tilde{p}_{\text{total}}(f)$ signal amplitudes (Eqs. (1, 2)) are used to form the network training database. Measured experimental signals $\tilde{S}^{\text{exp}}_{\text{total}}(f)$ are corrected (Eq. (3)), to obtain $\delta\tilde{p}^{\text{exp}}_{\text{total}}(f)$ amplitudes which are presented to the networks and

used for "intelligent characterization" of semiconductors (ANNs prediction of the sample parameters).

3 Neural networks model design

As it was explained in (Djordjevic et al. 2020, 2019), we used the simplest model in creating our network, the so-called "feed-forward network", having an input, one hidden and the output layer. The input layer consists of 72 neurons (amplitudes) following the standard number of modulation frequencies used in our experiments. The hidden layer consists of 50 neurons, following the criteria that the number of neurons is between the size of the input and output layers, ~ 2/3 the size of the input + output layer. The output layer consists of 3 neurons, each corresponding to the number of predictions we want to make (typical investigated sample thermal—D_T, α_T and geometric—l parameters). The sigmoid activation function (Logistic Sigmoid) is used to convert the input into a more useful output with values between 0 and 1.

Network database consists of 5491 amplitudes (Fig. 1a) of the PA theoretical signals $\delta\tilde{p}_{\text{total}}(f)$ obtained with Eqs. (1, 2), changing the sample parameters in the following ranges: thermal diffusivity $D_T = (8.1 - 9.9) \cdot 10^{-5}\text{m}^2\text{s}^{-1}$, linear expansion $\alpha_T = (2.34 - 2.86) \cdot 10^{-6}\text{K}^{-1}$ with a step of 1.25%, and the sample thickness $l = (1 - 10) \cdot 10^{-8}\text{m}$, with a step of 50 μm. From this database of 5491 signals, every 50th signal (110 in total) was taken for the independent test, and 5381 amplitudes was obtained for training base. In such way satisfying variety, density and volume of relevant data is achieved. Sometimes, simple amplitude scaling on unity (scaled on 1) is performed (Fig. 1b) to emphasize differences in frequency domain (Djordjevic et al. 2020, 2019).

Typical supervised learning is used here in network training, applying regression technique for output data prediction together with back-propagation algorithm that assures

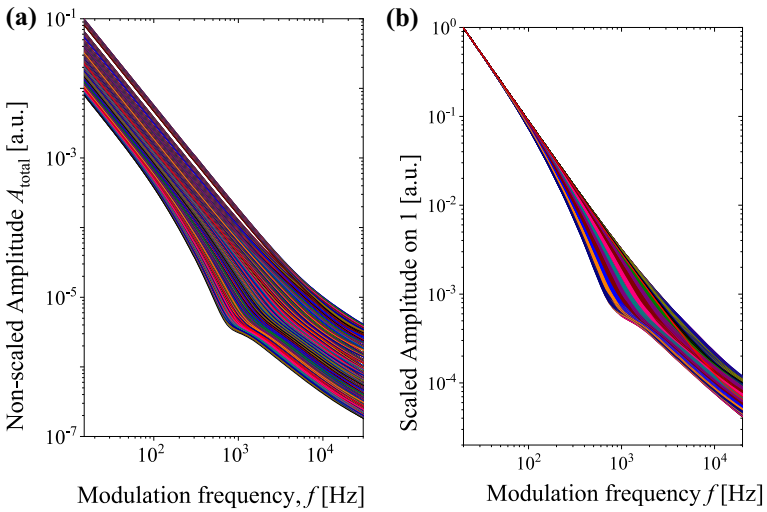

Fig. 1 **a** Non-scaled amplitude characteristics of the theoretical photoacoustic signals (Djordjevic et al. 2020, 2019), that forms the network training database; **b** simple amplitude scaling on unity to emphasize differences in frequency domain. The arrows indicate an increase in the thickness l

satisfying network prediction accuracy finding optimal values of weights. The training, testing and validation procedure was performed for input signal amplitudes and output parameters normalization or scaling (Figs. 4 and 5).

To evaluate our network model the test of its prediction accuracy on data from the training set which was not used at all during training was performed. Networks trained in mentioned way are used to predict thermal parameters of experimentally obtained PA signals from three n-type silicon samples with different thicknesses, because training set of photoacoustic signals amplitudes correspond to the measured signals amplitudes $\delta\tilde{p}^{\exp}_{\text{total}}(f)$ of all possible experimental settings of an open photoacoustic cell (Djordjevic et al. 2019, 2020, Jordovic-Pavlovic et al. 2020a, b).

4 Input and output data scaling and normalization

It is obvious from Fig. 1 that scaling is needed due to the large changes in amplitude values (few orders of magnitude). Scaling of PA amplitudes $A(f)$ from the training base (Fig. 1a) applied here is performed using the logarithmic function (log scaled) based on the Bode plots, having the form

$$A_{\log}(f_i) = 20 \cdot \log A(f_i), \quad i = 1, 2, ...72 \tag{4}$$

and the results based on the Eq. (4) are presented in Fig. 2a. This type of scaling is usually performed to change the amplitude values to a common scale, within the range of corresponding phase values. At the same time, it is expected that scaling will map amplitude-frequency variances from log–log (Fig. 1a) to lin-log scale (Fig. 2a).

Generally speaking, max and min–max normalizations are used to rescale amplitudes to the ranges of (0,1] and [0,1], respectively, and to equalize amplitude-frequency variances as much as it is possible.

First type of normalization applied on $A(f)$ is the normalization to the maximum absolute value (Max norm) of the base frequency vectors defined as (Jordovic-Pavlovic et al. 2020a, b):

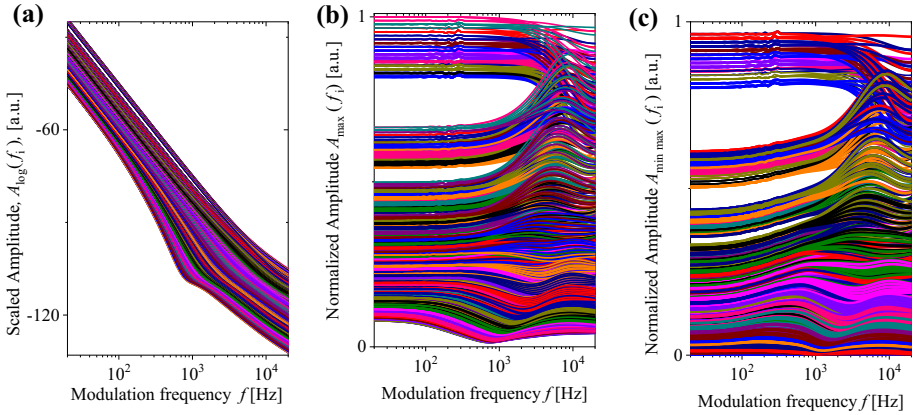

Fig. 2 Databases for neural network training: amplitudes **a**. scaled by logarithmic function (Djordjevic et al. 2020, 2019), and **b**. rescaled by max and **c**. min–max normalization

$$A_{\max}(f_i) = \frac{A(f_i)}{\left|\max(A_j(f_i))\right|}, \quad j = 1, 2, ..., 5381, \quad i = 1, 2, ...72. \tag{5}$$

This type of normalization gives amplitude values within the range of $0 < A_{\max}(f_i) \leq 1$ and changes (equalizes) the amplitude-frequency variances to some extent in the given modulation frequency range of 20 Hz–20 kHz (see Fig. 2b).

Second normalization applied on $A(f)$ is the min–max normalization (Min–Max norm) which rescales the range of the amplitudes to [0,1] (Fig. 2c), using (Furundzic et al. 2017, 1998):

$$A_{\min\max}(f_i) = \frac{A(f_i) - A_{\min}}{A_{\max} - A_{\min}}, \; i = 1, 2, \ldots 72 \tag{6}$$

where $A_{\max} = \max\left(A_j(f_i)\right)$, $j = 1, \ldots, 5381$ and $A_{\min} = \min\left(A_j(f_i)\right)$, $j = 1, \ldots, 5381$.

This type of normalizations (Eq. (6)) is depicted in Fig. 2c, simultaneously changes (equalizes) the amplitude-frequency variances to some extent, slightly different from max one, in the investigated modulation frequency range.

It is assumed that both types of input data normalization could lead to better network predictions in critical areas (> 1 kHz) where some signal overlapping exists. Then, it can be said that databases in the range 0–1 (Fig. 2b and c) are more acceptable for machine learning because it is easier to form a weight matrix using numerical values of a certain smaller range. The slight difference that exists in max and min–max normalizations is due to the normalization of the maximum value of the signal at a certain frequency, while the normalization of min–max is the normalization to the range [0,1].

Two ANNs with different ways of scaling the output data are formed to analyze the influence of scaling on the network's training performances. As input parameters of both ANNs, the unchanged (non-scaled and non-normalized) PA signal amplitudes (Fig. 1a) are used. As we said earlier, the data of the output vectors are three parameters of the sample, given in the following ranges:$D_{\mathrm{T}} = (8.1 - 9.9) \cdot 10^{-5}\mathrm{m}^2\mathrm{s}^{-1}$, $\alpha_{\mathrm{T}} = (2.34 - 2.86) \cdot 10^{-6}\mathrm{K}^{-1}$, and $l = (1 - 10) \cdot 10^{-4}\mathrm{m}$. For the first neural network (NN1) with non-scaled output, we set the thermal diffusivity range as the smallest data in the output vector $(0.81 - 0.99)$; the linear expansion in the range $(2.34 - 2.86)$; the thickness in the range $(100 - 1000)$. For the second neural network (NN2), we scaled the output data in the range of values D_{T}, α_{T} and l, from 1 to 10: (8.10–9.90); (2.34–2.86); (1–10).

5 Results and discussion

5.1 Neural networks with non-scaled and scaled output

To test the network training performances (a plot of Mean Squared Error (MSE) vs Epochs), we train NN1 and NN2 networks. The results of obtained network training performances are given in Table 1, Fig. 4. It is obvious that NN2 networks shows much better results in network training performances, having larger number of epochs but same training time as in the case of NN1.

Network prediction accuracy is tested using maximal and average relative errors obtained comparing network prediction and parameters values used to form the theoretical amplitudes. The tests are performed on signals not presented to the network during

Table 1 Training performance of two neural networks with non-scaled input: the first with non-scaled output NN1, the second with scaled output NN2

Type of NN	Performance
NN1 non-scaled output	MSE 0.059084 at 809 epoch, training time 4 h 30 min
NN2 scaled output	MSE 0.000037822 at 1000 epoch, training time 4 h 30 min

the training: 110 randomly selected signals from the training database, and 24 randomly selected signals out of the training base (with random parameters within the range of parameter changes).

The results presented in Tables 1, 2 and 3 indicate that the scaling of output data is beneficial not only in the case of network training performance but also for network prediction performances, as well. Only the results obtained for the sample thickness prediction deviate from the expected ones (Tables 2 and 3). This could have been expected knowing that the theoretical model (Eq. (1)) considers thickness as a parameter obtained by fitting the larger error so the thickness prediction results can vary on a much larger scale than D_T and α_T.

5.2 Neural networks with non-scaled, scaled and normalized input

Based on the results presented in the previous paragraph NN2 network seems to be more acceptable than NN1 so, our choice for further theoretical analysis will be the NN2. Its training performances are analyzed using different PA signal amplitudes $A(f)$ data as an input in the form of non-scaled (Fig. 1a), scaled on 1 (Fig. 1b), logarithmically scaled (Fig. 2a), max (Fig. 2b) and min–max normalized (Fig. 2c). The results of such analysis are shown in Table 4, Fig. 5.

It is obvious that the worst result in Table 4 is network training with data scaled on unity. Logarithmic scaling significantly improves network performances, max normalization gives better results, while max–min normalization shows the best performances.

Table 2 Maximal and average (%) relative prediction error of two neural networks with differently scaled output layer on independent amplitude test

110	Max (%) relative error			Average (%) relative error		
Type of NN	D_T	α_T	l	D_T	α_T	l
NN1	11.8496	7.2159	0.4632	3.9795	1.7724	0.0632
NN2	0.4824	0.2173	0.4756	0.0661	0.0574	0.0720

Table 3 Maximal and average (%) relative error of two neural networks with differently scaled output layer data in amplitude prediction of randomly selected parameters in the range of changes of parameters

24	Max (%) relative error			Average (%) relative error		
Type of NN	D_T	α_T	l	D_T	α_T	l
NN1	9.2781	8.0963	3.3938	3.2013	1.8035	0.3448
NN2	5.7019	6.5365	4.5952	0.5167	0.3886	0.4285

Table 4 The NN2 performances with different input scaling and normalizations

Type of input normalization	Performance
Non-scaled	0.000037822
Scaled on 1	0.00015258
Log scaled	0.000045951
Max normalization	0.0000041492
Min–max normalization	0.00000041263

In general, the presented analysis shows that data normalization is beneficial in terms of network training performances.

The standard NN2 network prediction accuracy test is performed with independent signals extracted from the amplitude base before training (Djordjevic et al. 2020, 2019). The results presented in Table 5 show small relative (%) errors for all types of normalization. This test confirms that the best prediction accuracy was achieved with min–max input normalization.

5.3 Application on experimental signals

All previous considerations and tests of network training performances and their prediction accuracy have been done on theoretical signals used for precise determination of changes in D_T, α_T and l. Interesting results are obtained by analyzing the network prediction accuracy of experimental samples with different thicknesses (Table 6). Unfortunately, the experimental signal $\tilde{S}^{exp}_{total}(f)$ amplitudes (Eq. (3)) measured by the standard experimental setup cannot be directly presented to the network as an input data because they are distorted under the influence of $\tilde{H}_{total}(f)$. As it was mentioned in Sect. 2, additional procedure is needed (Fig. 3), explained in details in (Aleksić et al. 2016; Markushev et al. 2015; Popovic et al 2016) to detect instrumental characteristics, calculate $\tilde{H}_{total}(f)$ and find undistorded signal $\delta\tilde{p}_{total}(f)$(Eq. (3)) which can be recognized by ANNs. Obtained $\delta\tilde{p}_{total}(f)$ for each sample are fitted using Eq. (1) to find D_T, α_T and l parameters (see Sect. 4). After these parameters estimation, amplitudes $A(f)$ (Eq. (2), non-scaled input) are presented to different ANNs (NN1, NN2) to find network predicted parameters D_T^{ANN}, α_T^{ANN} and l^{ANN}. Relative errors (%) are calculated comparing fitted and predicted parameters. The results of calculated experimental data are presented in Tables 6 and 7.

Table 5 Maximal and average (%) relative prediction error on the test of the independent signal amplitudes extracted before network training

Type of normalization	Max relative error %			Average relative % error		
Parameters	D_T	α_T	l	D_T	α_T	l
Non-scaled	0.4824	0.2173	0.4757	0.0661	0.0574	0.0720
Scaled on 1	1.1474	1.0572	1.0963	0.1417	0.1234	0.1700
Logarithmic normalization	0.3472	0.2289	0.3882	0.0661	0.0729	0.0679
Max normalization	0.1264	0.1998	0.1939	0.0254	0.0352	0.0265
Min–max normalization	0.0322	0.0770	0.0605	0.0068	0.0185	0.0084

Table 6 Parameter prediction D_T^{ANN}, α_T^{ANN} and l^{ANN} of amplitude neural networks NN1 and NN2.

Rel error (%)	Sample no.1 830 µm			Sample no.2 417 µm			Sample no. 3 128 µm		
Parameters	D_T^{ANN}	α_T^{ANN}	l^{ANN}	D_T^{ANN}	α_T^{ANN}	l^{ANN}	D_T^{ANN}	α_T^{ANN}	l^{ANN}
NN1	0.6556	0.3590	0.0133	0.6035	0.9553	0.0831	1.9162	1.6576	2.1096
NN2	0.0555	0.0421	0.0111	0.1290	0.0263	0.0592	11.7065	5.7570	11.5476

The relative (%) error of prediction of parameters of individual samples is given. Sample no.1 is 830 µm, sample no. 2 is 417 µm and 3 is 128 µm

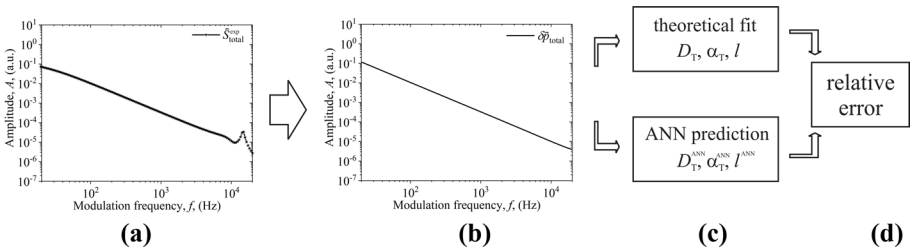

Fig. 3 Simplified procedure scheme for relative error calculation: **a** typical experimental signal (Eq. (3)); **b** signal corrected due to instrumental influence; **c** theoretical fit (Eq. (1)) and network prediction parameters; **d** relative error calculation

Table 7 The NN2 parameter prediction D_T^{ANN}, α_T^{ANN} and l^{ANN} with non-scaled, scaled and normalized inputs on experimental photoacoustic signals

Rel error (%)	Sample no.1			Sample no.2			Sample no. 3		
Parameters	D_T^{ANN}	α_T^{ANN}	l^{ANN}	D_T^{ANN}	α_T^{ANN}	l^{ANN}	D_T^{ANN}	α_T^{ANN}	l^{ANN}
Non-scaled	0.0555	0.0421	0.0111	0.1290	0.0263	0.0592	11.7065	5.7570	11.5476
Scaled on 1	0.0367	0.0812	0.0037	0.1120	0.0635	0.0011	2.8903	2.0515	2.9798
Log scaled	0.0450	0.0387	0.0204	0.0610	0.0339	0.0130	5.8210	2.3487	6.2749
Max norm	0.0672	0.0278	0.0005	0.0774	0.0316	0.0116	1.2337	0.1819	2.3816
Min–max norm	0.0366	0.0489	0.0188	0.0651	0.0124	0.0031	2.3220	0.3906	3.1880

The relative (%) error of prediction of parameters of individual samples is given. Sample no.1 is 830 µm, sample no. 2 is 417 µm and no.3 is 128 µm

The predictions of the thinnest sample are most accurate in the case of NN1 network, while in the case of thicker samples the better predictions are obtained with the NN2. Obviously, NN2 output scaling matches the scale of precise network prediction. This logic does not work with thinner samples where NN1 non-scaled output variables result in a more precise parameters prediction. As a conclusion based on Table 6, one can say that scaling the output data is useful in the case of PA signal amplitudes processing originating from thicker samples. Approaching the thinner ones, output scaling loses its importance. In our case (Fig. 1), since we work in most cases with thick samples, the choice of an NN2 network for further experimental amplitudes analysis proves to be a rational solution.

Neural networks NN2 with scaled outputs, but different rescaled and normalized inputs, were tested on adjusted (Eqs.4, 5 and 6) experimental signals as well. Prediction accuracy test and the results are presented in Table 7. The benefits of the input data scaling and normalization of the NN2 network can be seen, also, but not as obvious as with theoretical signals. The reason for that lies in the fact that the experimental signals are neither in frequency nor parameters changing steps as theoretical signals are. This is the reason why one can expect larger errors and more diversity of NN2 parameters prediction accuracy obtained with experimental signals. Largest errors are obtained mostly with non-scaled inputs.

6 Conclusion

In this article, the influence of different input and output data scaling and normalization on overall neural network performances in photoacoustics is presented.

In theory, simple numerical scaling is applied on network prediction parameters (diffusivity, coefficient of thermal expansion, thickness) as output data. This kind of scaling was found to be beneficial in the terms of network training and prediction and it was kept throughout the whole analysis. Various scaling and normalization methods (logarithmic scaling, scaling on unity, max and min–max normalization) were applied to the photoacoustic signal amplitudes as input predictors of equal importance compared to the signal phase. It was found that min–max amplitude normalization shows the best results in network training and prediction accuracy.

In the experiment, the results of network input scaling and/or normalization are not unambiguous. In general, each method has pros and cons based on the specific problem to be solved and one can decide based on the problem which scaling or normalization method is best suitable for the problem. The results obtained here with the silicon samples of different thicknesses suggesting that max normalization of input and non-scaled output data are the best choice to reach the highest quality of network overall performances in the case of thicker samples. In the case of thinner samples, various scaling and normalization methods can be only partially beneficial for network overall performances. As a principal conclusion, there is no universal input data normalization method that can be chosen in advance to improve network training performances and prediction accuracy.

Appendix

Appendix I Training, validation and test results of neural networks with non-scaled and scaled output

Performance and training time for training, validation and test of amplitude neural networks with non-scaled input but with: (a) non-scaled output—NN1, (b) scaled output—NN2, see Table 1. See Fig. 4.

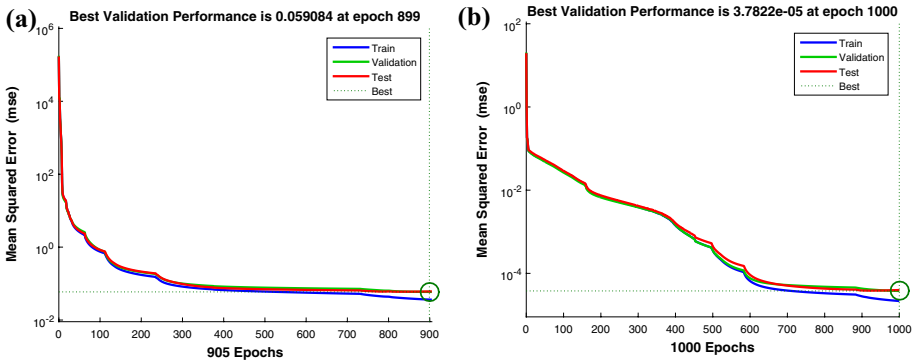

Fig. 4 Error function (Mean Squared Error) on the training set, validation set and test set depending on the number of epochs for **a** NN1, and **b** NN2 networks

Appendix II Training, validation and test results of neural networks with scaled output and scaled or normalized input

Training, validation and test results of amplitude neural network NN2 with (a) non-scaled input, (b) scaled input on 1, (c) log scaled input, (d) max normalization input, (e) min–max normalization input in order to achieve the best network performance, see Table 4 and Fig. 5:

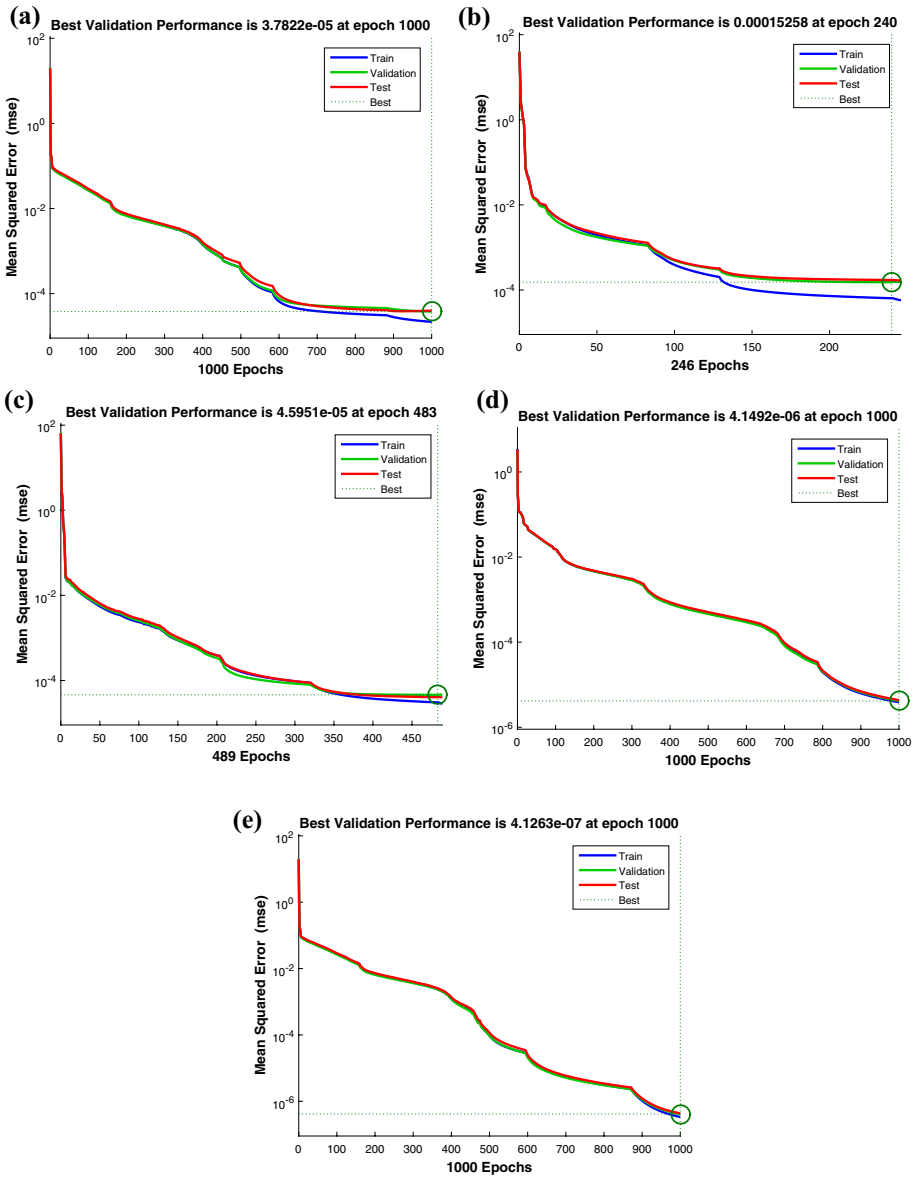

Fig. 5 Error function (Mean Squared Error) on the training set, validation set and test set depending on the number of epochs for: **a** non-scaled input, **b** scaled input on 1, **c** log scaled input, **d** max normalization input, **e** min–max normalization input

Acknowledgements This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, contract number 451-03-09/2021-14/200017.

Authors contributions Not applicable.

Funding Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they no conflict of interest.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethics approval Not applicable.

References

Aleksić, S.M., Markushev, D.K., Pantić, D.S., Rabasović, M.D., Markushev, D.D., Todorović, D.M.: Electro-acoustic influence of measuring system on the photoacoustic signal amplitude and phase in frequency domain. FACTA Univ. Ser. Phys. Chem. Technol. **14**(1), 9–20 (2016). https://doi.org/10.2298/FUPCT1601009A

Arridge, S., Maass, P., Öktem, O., Schönlieb, C.-B.: Solving inverse problems using data-driven models. Acta Numer. **28**, 1–174 (2019). https://doi.org/10.1017/S0962492919000059

Balderas-Lopez, J.A.: Photoacoustic signal normalization method and its application to the measurement of the thermal diffusivity for optically opaque materials. Rev. Sci. Instrum. **77**, 64902 (2006). https://doi.org/10.1063/1.2209951

Balderas-López, J.A., Mandelis, A., García, J.A.: Normalized photoacoustic techniques for thermal diffusivity measurements of buried layers in multilayered systems. J. Appl. Phys. **92**(6), 3047–3055 (2002). https://doi.org/10.1063/1.1500784

Calderón, A., Muñoz Hernández, R.A., Tomás, S.A., Cruz-Orea, A., Sánchez Sinencio, F.: Method for measurement of the thermal diffusivity in solids: Application to metals, semiconductors, and thin materials. J. Appl. Phys. **84**, 6327–6329 (1998). https://doi.org/10.1063/1.368957

Djordjevic, KL., Markushev, D.D., Ćojbašić, ŽM., Djordjevic, K.L.: Photoacoustic measurements of the thermal and elastic properties of n-type silicon using neural networks. SILICON **12**, 1289–1300 (2019). https://doi.org/10.1007/s12633-019-00213-6

Djordjevic, K.L., Markushev, D.D., Ćojbašić, ŽM., Galović, S.P.: Inverse problem solving in semiconductor photoacoustics by neural networks. Inverse Probl. Sci. Eng. **29**, 1–15 (2020). https://doi.org/10.1080/17415977.2020.1787405

Dramićanin, M.D., Ristovski, Z.D., Nikolic, P.M., Todorović, D.M.: Photoacoustic investigation of transport in semiconductors: Theoretical and experimental study of a Ge single crystal. Phys. Rev. B **51**, 226–233 (1995). https://doi.org/10.1103/PhysRevB.51.14226

El-Brolossy, T.A., Ibrahim, S.S.: Photoacoustic measurement of thermal properties of polystyrene metal oxide composites. Thermochim. Acta **509**, 46–49 (2010). https://doi.org/10.1016/j.tca.2010.05.020

Furundzic, D., Djordjevic, M., Jovicevic Bekic, A.: Neural networks approach to early breast cancer detection. J. Syst. Archit. **44**, 617–633 (1998). https://doi.org/10.1016/S1383-7621(97)00067-2

Furundzic, D., Stankovic, S., Jovicic, S., Punisic, S., Subotic, M.: Distance based resampling of imbalanced classes: With an application example of speech quality assessment. Eng. Appl. Artif. Intel. **64**, 440–461 (2017). https://doi.org/10.1016/j.engappai.2017.07.001

Govorkov, S., Ruderman, W., Horn, M.W., Goodman, R.B., Rothschild, M.: A new method for measuring thermal conductivity of thin films. Rev. Sci. Instrum. **68**, 3828–3834 (1997). https://doi.org/10.1063/1.1148035

Jin, J., Li, M., Jin, L.: Data normalization to accelerate training for linear neural net to predict tropical cyclone tracks. Math. Probl. Eng. **2015**, 1–8 (2015). https://doi.org/10.1155/2015/931629

Jordovic-Pavlovic, M., Kupusinac, A.D., Djordjevic, K.L., Galovic, S.P., Markushev, D.D., Nesic, M.V., Popovic, M.N.: Computationally intelligent description of a photoacoustic detector. Opt. Quantum Electron. **52**, 1–14 (2020a). https://doi.org/10.1007/s11082-020-02372-y

Jordovic-Pavlovic, M.I., Markushev, D.D., Kupusinac, A.D., Djordjevic, K.L., Nesic, M.V., Galovic, S.P., Popovic, M.N.: Deep neural network application in the phase-match calibration of gas-microphone photoacoustics. Int. J. Thermophys. **41**, 1–10 (2020b). https://doi.org/10.1007/s10765-020-02650-7

Kim, D.: Normalization methods for input and output vectors in backpropagation neural networks. Int. J. Comput. Math. **71**, 161–171 (1999). https://doi.org/10.1080/00207169908804800

Mandelis, A.: Progress in photoacoustic and photothermal science and technology (Vol. IV). SPIE Press, Bellingham, 98227-0010, USA (1999) ISBN: 9780819435064

Markushev, D.D., Ordonez-Miranda, J., Rabasovic, M.D., Galovic, S.P., Todorović, D.M., Bialkowski, S.E.: Effect of the absorption coefficient of aluminium plates on their thermoelastic bending in photoacoustic experiments. J. Appl. Phys. **117**, 245309 (2015). https://doi.org/10.1063/1.4922718

Markushev, D., Markushev, D., Galovic, S., Aleksic, S., Pantic, D., Todorovic, D.: The surface recombination velocity and bulk lifetime influences on photogenerated excess carrier density and temperature distributions in n-type silicon. Facta Univ. Ser. Electron. Energ. **31**, 313–328 (2018). https://doi.org/10.2298/fuee1802313m

Markushev, D.K., Markushev, D.D., Aleksić, S., Pantić, D.S., Galović, S., Todorović, D.M., Ordonez-Miranda, J.: Effects of the photogenerated excess carriers on the thermal and elastic properties of n-type silicon excited with a modulated light source: theoretical analysis. J. Appl. Phys. **126**, 185102 (2019). https://doi.org/10.1063/1.5100837

Melo, W.L.B., Faria, R.M.: Photoacoustic procedure for measuring thermal parameters of transparent solids. Appl. Phys. Lett. **67**, 3892–3894 (1995). https://doi.org/10.1063/1.115308

Ordóñez-Miranda, J., Alvarado-Gil, J.J.: Thermal wave oscillations and thermal relaxation time determination in a hyperbolic heat transport model. Int. J. Therm. Sci. **48**, 2053–2062 (2009). https://doi.org/10.1016/j.ijthermalsci.2009.03.008

Pichardo-Molina, J.L., Alvarado-Gil, J.J.: Heat diffusion and thermolastic vibration influence on the signal of an open photoacoustic cell for two layer systems. J. Appl. Phys. **95**, 6450–6456 (2004). https://doi.org/10.1063/1.1711182

Popovic, M.N., Nesic, M.V., Ciric-Kostic, S., Zivanov, M., Markushev, D.D., Rabasovic, M.D., Galovic, S.P.: Helmholtz resonances in photoacoustic experiment with laser-sintered polyamide including thermal memory of samples. Int J Thermophys **37**, 116 (2016). https://doi.org/10.1007/s10765-016-2124-3

Todorović, D.M., Nikolić, P.M., Dramićanin, M.D., Vasiljević, D.G., Ristovski, Z.D.: Photoacoustic frequency heat-transmission technique: Thermal and carrier transport parameters measurements in silicon. J. Appl. Phys. **78**, 5750–5755 (1995). https://doi.org/10.1063/1.359637

Todorović, D.M., Rabasović, M.D., Markushev, D.D., Sarajlić, M.: Photoacoustic elastic bending in thin film–substrate system: Experimental determination of the thin film parameters. J. Appl. Phys. **116**(5), 053506 (2014). https://doi.org/10.1063/1.4890346

Velasco, D.S., Baesso, M.L., Medina, A.N., Bicanic, D.D., Koehorst, R., van der Hooft, J.J.J., Bento, A.C.: Thermal diffusivity of periderm from tomatoes of different maturity stages as determined by the concept of the frequency-domain open photoacoustic cell. J. Appl. Phys. **109**, 034703 (2011). https://doi.org/10.1063/1.3530735

Yahyaoui's, A., Yahyaoui, I., Yumuşak, N.: Machine learning techniques for data classification. In: Advances in renewable energies and power technologies, pp. 441–450. Elsevier (2018). https://doi.org/10.1016/B978-0-12-813185-5.00009-7