## ORIGINAL RESEARCH ARTICLE

# Characterization of ink-based phantoms with deep networks and photoacoustic method

**Hui Ling Chua[1], Audrey Huong[1,*], Xavier Ngu[2]**

[1] Department of Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja 86400, Johor, Malaysia

[2] Institute for Integrated Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja 86400, Johor, Malaysia

**\* Corresponding author:** Audrey Huong, audrey@uthm.edu.my

## ABSTRACT

This study aims to explore the feasibility of using an in-house developed photoacoustic (PA) system for predicting blood phantom concentrations using a pretrained Alexnet and a Long Short-Term Memory (LSTM) network. In two separate experiments, we investigate the performance of our strategy using a point laser source and a color-tunable Light-Emitting Diode (LED) as the illumination source. A single-point transducer is employed to measure signal change by adding ten different black ink concentrations into a tube. These PA signals are used for training and testing the employed deep networks. We found that the LED system with light wavelength of 450 nm gives the best characterization performance. The classification accuracy of the Alexnet and LSTM models tested on this dataset shows an average value of 94% and 96%, respectively, making this a preferred light wavelength for future operation. Our system may be used for the noninvasive assessment of microcirculatory changes in humans.

*Keywords:* photoacoustic; concentration; phantom; Alexnet; LSTM

## 1. Introduction

Ink is a pigmented solution or liquid material that marks or colors a surface to create an image, text, or pattern. A significant application of ink is in the preparation of medical phantoms. In most practical cases, prior knowledge of ground truth is unavailable, so phantoms are often prepared for evaluating and validating the performance of a proposed approach[1] or instruments under development, particularly those intended for clinical use before the in-vivo study. Phantom for clinical applications (or medical model) can reproduce optical properties (e.g., absorption and scattering properties) of tissues[2,3], mimicking blood flow conditions[4,5] and even the organs within the human body[6,7] so that the anatomical structures or underlying physics[8] of human can be sufficiently comprehended. The key benefits of inks are that they are chemically and spectroscopically stable and harmless, making them a suitable candidate for preparing tissue-simulating phantoms. In a study by Cui et al.[9], the anisotropic cardiac patch with myocardial fibre orientation was created using a gelatin-based printable ink composed of Gelatin Methacrylate (GelMA) and Polyethylene Glycol Diacrylate (PEGDA) to study the treatment of myocardial infarction. Others used gelatin ink staining to evaluate microvascular perfusion and reported serum MiR-98-5p as a feasible marker for microvascular perfusion[10]. Another work by Bachir and Dargham[11] used Indian ink to prepare a phantom simulating subcutaneous veins in their study of blood vessel

1

localization in dark skins. Other applications include the study of color shift on the surface of a cardiac phantom created from Polyvinyl Alcohol (PVA) hydrogel with temperature-responsive pigment inks to visually identify the catheter ablation zone by Imanishi et al.[12]. These innovative tissue and organ models and phantoms are beneficial in allowing validation studies of a system or method without risking harm to human subjects.

Among the studies of phantom models that are worth mentioning include the work by Li et al.[13]. The authors demonstrated a Fibre Optic Spectrometer (FOS) system for characterizing skin pigmentation. This system is used to characterize an agar phantom dyed with red, green, and blue inks, with an accuracy of 96%, and was shown to discriminate the ink injection sites from normal tissue based on light scattering. In the study of Lam et al.[14], Multimodal Diffuse Optical Spectroscopy Imaging (MM-DOSI) technique was used for layered compositional analysis and metabolic tracking of tissue phantoms based on the absorptivity of the underlying chromophores (i.e., light absorbers). It was reported by Anugrah et al.[15] that the absorber's (i.e., ink) concentration inside a phantom can be exploited to distinguish between different kinds of tissue. The hydrogen (O–H) bonds absorption peak increased with gelatin content, allowing localization of human soft tissue based on the categorization of Computed Tomography (CT)-Number values as a function of gelatin content.

Photoacoustic imaging (PA) is another technology that has gained popularity as a noninvasive tool for studying optical absorption[10,16]. This approach combining optics and acoustics improves imaging sensitivity compared to absorption spectroscopy techniques. This technique begins with the transfer of light energy to the target samples. Acoustic waves are produced following the thermal expansion of tissues caused by light absorption in the medium. An ultrasonic transducer can be used to detect the produced waves, and their waveforms may be studied on an oscilloscope for further investigation and evaluation. Since chromophores' light absorption is proportional to their concentration in the medium, PA imaging can be used to quantify the concentrations of chromophores present in the imaged region[17,18]. Roy et al.[19] investigated the influence of microfluidic channel diameters on the emitted acoustic signals using the methylene blue solution as its marker. The acoustic signal intensity was shown to vary linearly with increasing dye concentration, implying that this technique is feasible to characterize a dye concentration. In the work by Dolet et al.[20] multispectral PA imaging was used to estimate the fractional absorbers concentration (which is taken as oxygen saturation ($SO_2$)) in the mice tumors model.

The growth of Artificial Intelligence (AI) has revolutionized data science[21,22], industry practice[23], and particularly the medical sector[24,25]. Deep learning and machine learning are the subsets of AI. Machine learning techniques include Support Vector Machine (SVM), K-neighbors (KNN), Decision Tree, and Linear Regression. SVM is less computationally intensive than KNN and easier to read, but its application may be limited to linearly separable datasets. The Decision Tree is the most commonly used classifier due to its ease of implementation. However, slight changes in the data may result in significant changes in the tree's structure. Linear Regression performs extraordinarily well for linearly separable data, while an Artificial Neural Network (ANN) is suitable for tasks that a linear method cannot handle. Deep learning is a representation-learning method that has multiple levels of representation. They are created by composing simple but nonlinear modules; each transforms the representation from a relatively simple level (begins with the raw input) to a higher level with complex and more abstract information. A pretrained deep model is a stored network previously trained on a large dataset for image-classification application. Since the representation it has learned is adequate (or readily transferable) for a given job, such as image classification or object detection, the pretrained model saves time by avoiding the need to train the network from scratch. Convolutional Neural Network (CNN) pretrained models include Alexnet, GoogleNet, and Visual Geometry Group Network (VGGNet). Alexnet is one of the most popular used networks using convolutional operation and max pooling to extract spatial-dependent features in the data. It evolved from the need to better the ImageNet challenge results. It has shallow architecture, which places smaller demand on computational resources. GoogleNet is an inception module that

performs several convolutions with varying filter sizes and pooling in a single layer. Instead of requiring users to select the layer that produces the best results, the network determines this after training[26]. Pretrained VGGNet is a deeper network with much smaller filters. On the contrary, a self-developed deep learning model requires human expertise in its architecture design to categorize fresh, unseen data. It can learn more precise characteristics, perform well even when a pretrained network fails, and work better with additional data.

AI has been extended in the field of characterizing tissue phantoms. Linear Regression is used in Bachir and Dargham[11] for differentiating vein depths using an 830 nm diode laser in illuminating Indian ink within an embedded polymer tube, which mimics subcutaneous veins, and reported a regression coefficient of 0.88. In the report by Youn et al.[27], an aortic phantom connecting a closed-circuit pulsatile flow system was developed to simulate the bolus geometry of the human abdominal aorta. The ANN with Multilayer Perceptron (MLP) architecture was used to predict the ideal bolus geometry with the corrected discrimination coefficient of 0.919. Zhang et al.[28] proposed a CNN named ring-array deep learning network (RADL-net) designed to reduce limited-view and under-sampling artifacts in PA data reconstructed from vascular images on a transparency film and in-vivo experiment using a three-quarter ring transducer array. It was shown that the net algorithm resulted in significant improvements in the image quality. RADL-net helped to restore the pixel intensity degraded by the limited-view issues while removing the background noise. In addition, the results showed that hemoglobin absorption was lower (in comparison to black ink), rendering a poorer image quality than in the phantom studies. The use of this CNN and PA was shown to be able to localize a target located at a depth of 40–60 mm below the surface. OxyGAN, a Conditional Generative Adversarial Network (CGAN) made up of two CNNs, i.e., a generator and a discriminator, was introduced by Chen and Durr[29] to estimate absorbers-related concentration (i.e., oxygen level) using single structured-light pictures. The authors reported a considerably good accuracy of 93%, suggesting the technique's robustness to various tissue types. An encoder-decoder CNN architecture incorporating customized modules was proposed by Johnstonbaugh et al.[30]. This system was trained to detect the origin of PA wavefronts inside an optically scattering deep-tissue medium[30]. Even though CNN has been widely used in the field of photoacoustics, this method is well-known for the issue of vanishing gradients when the number of network layers increases[31]. Considering there might be delays of undetermined duration between critical occurrences in a time series, Long Short-Term Memory (LSTM) networks are suitable for classification, processing, and data related to time series[32]. In the phantom and in-vivo experiments of Siami-Namini et al.[31], the authors investigated the feasibility of an LSTM architecture for estimating blood flow index using Diffuse Correlation Spectroscopy (DCS). The results showed a linear relationship between blood flow changes and the flow velocity with a Pearson's correlation of 0.95 and 0.99, respectively.

There has yet to be any documented work on the study of phantom characterization using LSTM and based on PA signals. Since a pretrained network can be used as an alternative for the classification problem, Alexnet is employed herein for comparison work. The training and testing of the pretrained Alexnet and LSTM models were performed on a DELL laptop with 16 GB RAM, 64-bit Windows 10, Intel® Xeon™ i7-1700M CPU @3.20 GHz. All the simulations have been carried out using MATLAB (2020b). It is also our intention to identify the best wavelength for phantom characterization. For this purpose, we used a 633 nm laser and color-tunable LED, producing six visible light bands as the system light source.

## 2. Materials and methods

### 2.1. PA detection system and phantom preparation

Since the PA has been reported to vary with the wavelength-dependent medium's absorptivity, we investigated the performance of PA-based phantom characterization using different light wavelengths. We considered two different light sources, i.e., a 633 nm laser source and a color-tunable LED source. **Figure 1** shows an in-house assembled PA system using a 633 nm continuous wave laser source (R-30993 New-port

Corp.), while **Figure 2** shows the system using the LED light source. This multi-wavelength illumination system consists of two 5 mm ultra-bright transparent white LEDs (model. 5LED-UL-W) used to consecutively illuminate a set of color filters (model. FKB-VIS-10 Thorlabs). In this study, we used six visible light wavelengths, namely 450 nm, 500 nm, 550 nm, 600 nm, 650 nm, and 700 nm, for demonstration. A modulating light signal is produced after passing the light through an Acousto Optic Modulator (AOM), which is controlled by a radiofrequency (RF) driver with a carrier frequency of 15 MHz. Meanwhile, an ultrasonic flaw detector (EPOCH 650, Olympus Corp, Japan) is used for detecting the produced acoustic energy. A wideband bandpass filter with fixed cut-off frequencies (0.5–4 MHz) inside the EPOCH 650 was used to eliminate baseline drift and suppress high-frequency noise.
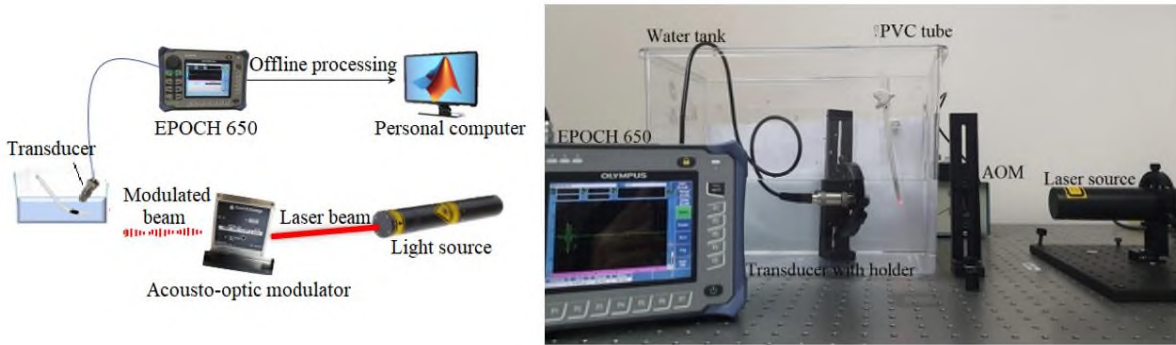


**Figure 1.** Laser-based PA system for ink concentration characterization (left) and its experiment setup (right).
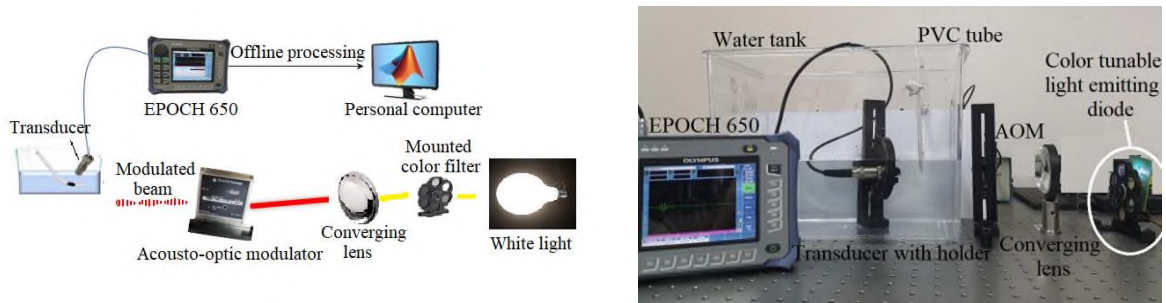


**Figure 2.** PA system using color-varying LED as its illumination system (left) and its experiment setup (right).

A single transparent Poly Vinyl Chloride (PVC) hollow tube (model: PRO 797–388) with an inner diameter of 4 mm installed inside a water tank (dimensions of 36 cm × 18 cm × 20 cm) is used to represent the blood vessels. The water tank was filled with tap water up to two-thirds of its capacity to establish an effective acoustic coupling between the phantom and transducer for the measurement. The illuminated tube containing the dye was completely immersed in the water throughout the whole experiment. In our study, commercial black ink was used as the dye. The peak wavelength of its light absorption is measured to be 450 nm using a USB spectrometer (model Ocean Optics USB 4000). Ten liquid samples were prepared and allowed to flow through the tube one at a time. They are undiluted ink (as the control sample) and water mixed with nine ink concentrations. The ink volumes considered are 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2, and 3 ml, respectively, added into 1000 mL of tap water. We considered a small concentration increment of 0.2 mL at the low ink concentrations to evaluate the sensitivity of the PA system in detecting small changes in signals and the performance of the classification system. These samples were stirred for about 15 seconds to ensure homogeneous mixing. The light absorption-induced pressure changes are measured with a flat acoustic transducer (V323-SU/2.25 MHz, Olympus NDR) immersed in the water.

The head of this sensor was placed next to the illumination area and at 3 cm from the region under study. This device was connected to the flaw detector for measuring ultrasonic waves as a function of time. A total

of 100 signals were collected under each ink concentration. The screenshot of these signals was saved on a microSD memory card before they were processed and analyzed offline.

## 2.2. Signal restoration

Since EPOCH 650 is not equipped to save raw data, signal restoration has been performed for the image frame to convert the image into numeric or matrix form for use with LSTM. Shown in **Figure 3** is the screenshot image during one of the experiments (left) and the restored signal (right). This is feasible because of the color change in the image (i.e., green color is the measured signal against the dark background). The screenshot image (left) was first converted into a black-and-white binary image using the im2bw function (i.e., pixel "0" represents the background, and pixel "1" represents the object pixel). Next, the binarized 2D image matrix is converted into a 1D matrix form of size $1 \times 494$ (i.e., $X_1$, ..., $X_{494}$), fed into the network input layer for further classification and analysis.
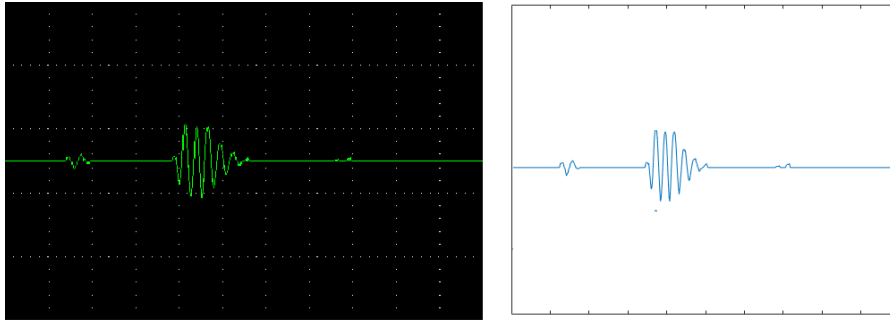


**Figure 3.** A screenshot of the PA signal measured using the V323-SU transducer (left) and the restored signal obtained from the binarization process and data transformation (right).

## 2.3. Data distribution

The obtained PA signals were used to train, validate, and test the employed model. A total of 100 PA signals have been measured for each ink concentration and illumination wavelength. These signals were randomly divided into 70%/15%/15% splits for training, validation, and testing purposes in **Table 1**. The validation set was used to provide evidence of over- or under-fitting of the model during the network training. Each model was trained three times using these data, and the average is reported in the results section.

**Table 1.** The distribution of wavelength-dependent PA data used for training, validation, and testing of the Alexnet and LSTM model.

| Class label | Ink concentration (ml)/L | Dataset* | | |
|---|---|---|---|---|
| | | **Training data** | **Validation data** | **Testing data** |
| 1 | 0.2 | 70 | 15 | 15 |
| 2 | 0.4 | 70 | 15 | 15 |
| 3 | 0.6 | 70 | 15 | 15 |
| 4 | 0.8 | 70 | 15 | 15 |
| 5 | 1.0 | 70 | 15 | 15 |
| 6 | 1.2 | 70 | 15 | 15 |
| 7 | 1.4 | 70 | 15 | 15 |
| 8 | 2 | 70 | 15 | 15 |
| 9 | 3 | 70 | 15 | 15 |
| 10 | Undilute | 70 | 15 | 15 |
| Total | | 700 | 150 | 150 |

*Numbers of data shown are for each wavelength.

## 2.4. Transfer learning Alexnet

Alexnet is one of the most popular CNN models, which architecture consists of five convolutional layers, the max-pooling layers, three fully connected layers, and normalization layers. This model adopts non-saturating Rectified Linear Unit (ReLU) activation function, which outperformed sigmoid in training. It is the process of applying filters over an image to output feature maps for the given input data. This network in **Figure 4** has an input layer of size $494 \times 329 \times 3$, consistent with the original image size. No image preprocessing has been performed on the input images. The input is fed into the network composed of a series of convolutional, ReLU layers, and Max Pooling layers. A dropout layer of 0.50 is placed after layers 19th and 22nd to prevent overfitting of the model. While the earlier layers were frozen during the training, the last fully connected layer (i.e., layer 23) was set as 10 according to the target class size in **Table 1**.
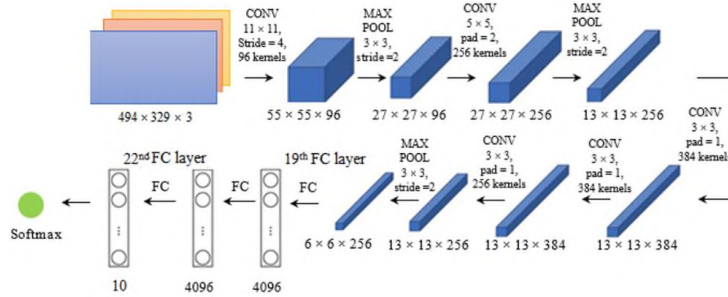


**Figure 4.** Architecture of the Alexnet model used in this study.

## 2.5. LSTM model

The LSTM model is another classification model used to classify different ink concentrations because of its outstanding prediction results compared to the other RNN model[33]. It is a Recurrent Neural Network (RNN) that deals with the vanishing gradient problem faced by CNN. The forget gate, input gate, and output gate are the three basic components of an LSTM. The first section determines whether the information from the preceding timestamp is important to remember or may be ignored. The cell attempts to learn new information from the input in the second phase. Finally, the cell sends updated information from the current timestamp to the next timestamp in the third component. It can deal with sequential data and learn time series information, and it is easy in its implementation. **Figure 5** shows the LSTM network that is used for the PA signal classification task. We used 150 hidden layers in the LSTM model to extract the important time features of the 1D matrix, as it was found that a large hidden size value greater than 150 increases the training time without further enhancing the classification accuracy. This is followed by three fully connected layers, FC, of sizes 100, 35, and 10. A dropout layer with a value of 0.3 was set right before the last FC layer to prevent the overfitting of the model. The output of the FCs is fed to a SoftMax classifier to classify the output into ten classes.
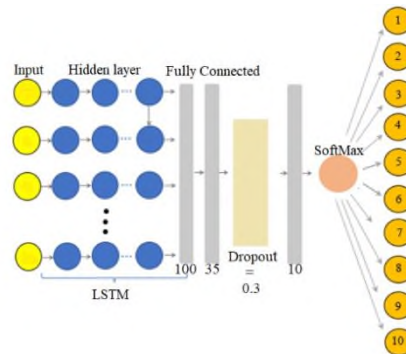


**Figure 5.** Architecture of the proposed LSTM model.

## 2.6. Network training and hyperparameters tuning

Both the abovementioned classification models were trained using an Adaptive Moment (ADAM) optimizer because of its rapid computation time and fast converging speed, as reported by Gupta[34], while some of the important hyperparameters, namely Epochs number, mini batch-size, initial learning rate, and gradient descent threshold, were manually tuned. Since all parameters should be nonzero and real positive, lower bounds are set close to the origin to prevent exploring in the negative direction, as shown in **Table 2**. Properly selecting hyperparameters would ensure better extraction of useful features for classification.

In our experiments, the epoch number was adjusted from 1 to 3000, while the initial learning rate ranged from 0.0001 to 1. In addition, the mini batch-size was allowed to vary from 1 to 256. Gradient descent is an efficient optimization process that aims to find the local or global minimum of the cost function. It is not feasible to reduce the cost function's value in infinitesimal steps; thus, the gradient threshold was arbitrarily adjusted in the range of 0.001 to 1 to avoid the non-convergence results. The upper limit of these parameters shown in **Table 2** is chosen as we noticed a negligible change in the performance metrics beyond these values. There is inconsistency in the upper limit of the epoch number and mini-batch-size used for training the models because our experimental results showed that the LSTM model required a longer time in the training to achieve acceptable classification accuracy; hence it is necessary to use larger numbers of epoch and mini-batch sizes for model convergence.

A total of 100 sets of these parameters combined based on the values in **Table 2** were attempted in our simulations. The prediction accuracy was shown to vary from 10% to 99%, while the training time varied between 20 minutes and 1400 minutes for both models. At the end of the process, the best combination is identified as the set that produced the best training accuracy (i.e., 100%).

**Table 2.** The upper and lower limits of the considered hyperparameters used in training the Alexnet and LSTM model.

| Parameter | Models | Limit | | Step of change | Best hyperparameter |
|---|---|---|---|---|---|
| | | **Lower** | **Upper** | | |
| Epoch number | Alexnet | 1 | 20 | 5 | 15 |
| | LSTM | 1 | 3000 | 100 | 3000 |
| Mini batch-size | Alexnet | 2 | 32 | $2^n$, where n = 2, 4, 6, …, 8 | 8 |
| | LSTM | 2 | 256 | $2^n$, where n = 2, 4, …, 6 | 128 |
| Initial learning rate | Alexnet and | $5 \times 10^{-4}$ | 1 | $5 \times 10^{-4}$ | $5 \times 10^{-4}$ |
| | LSTM | | | | - |
| Gradient descent threshold | Alexnet and | $1 \times 10^{-3}$ | 1 | $1 \times e^{-n}$, n = 3, 2, 1, 0 | $1 \times 10^{-3}$ |
| | LSTM | | | | - |

## 2.7. Traditional multi-class classification methods

Since most prior works[35,36] dealt with ink concentration prediction using hyperspectral imaging techniques, we used three traditional machine learning methods, namely SVM, KNN, and Decision Tree (DT), for comparison with the deep networks using the same PA dataset. The training and classification are done using MATLAB's multisvm, fitcknn, and fitctree functions, respectively, for SVM, KNN, and DT-based multi-class concentration classification. A KNN of $K$ value = 7 is used in the classifier, while the maximum split number is set to be 10 in the DT as they produced the best performance among other values from the pre-experiment runs.

## 2.8. Performance metrics

Mean classification accuracy, precision, specificity, and sensitivity in Equations (1)–(4) are used as the performance metrics to compare and evaluate the classification performance of the pretrained Alexnet and

LSTM models and traditional models used in this study. While accuracy refers to how near a forecast is to the true value, specificity and sensitivity measure how well a model can distinguish between different ink concentrations. Precision indicates the ability of the classification model to identify only the relevant data points in the prediction of ink concentration.

$$\text{Accuracy} = \frac{\Sigma_{i=1}^{T} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{T} \tag{1}$$

$$\text{Precision} = \frac{\Sigma_{i=1}^{T} \frac{TP_i}{TP_i + FP_i}}{T} \tag{2}$$

$$\text{Specificity} = \frac{\Sigma_{i=1}^{T} \frac{TN_i}{TN_i + FP_i}}{T} \tag{3}$$

$$\text{Sensitivity} = \frac{\Sigma_{i=1}^{T} \frac{TP_i}{TP_i + FN_i}}{T} \tag{4}$$

$T$ is the total number of class labels ($T = 10$). A true positive ($TP_i$) is a concentration class that is correctly predicted to be a specific class label, $i$. A false positive ($FP$) is a non-concentration class that is incorrectly predicted to be a class member, and a false negative ($FN$) is a concentration class member that is misclassified as a non-concentration class member. A true negative ($TN$) is a non-concentration class member correctly classified as a non-concentration class member.

## 3. Results

**Figure 6** shows the measured PA signal for the ink concentrations experiment using the light wavelength of 450 nm. The signals for this wavelength are chosen for demonstration because of their higher signal strengths compared to others, allowing better visualization of signal characteristics with sample concentration. The figure shows the produced values decrease with the dilution of the ink, consistent with the findings by Rajian et al.[37] that the signal amplitudes are directly proportional to the media absorption coefficient.
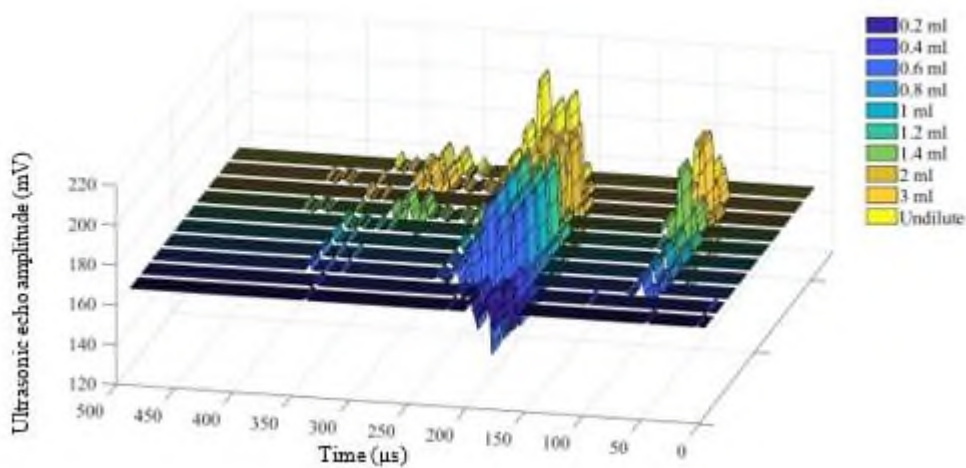


**Figure 6.** The 3D visualization of the PA signals measured from different ink solution concentrations at 450 nm.

Based on this diagram, it can be noticed that signals from 0.6 mL to 1.4 mL suspensions of ink have similar amplitudes and signal trend characteristics. Therefore, in **Figure 7**, we compared the signals between 0.2 mL, 1 mL, 2 mL, 3 mL, and undiluted ink solution. In addition to the consistent increase in the signal strength, the delay in the arrival of the acoustic waves can be seen with increasing medium light absorption. The delay is consistent, particularly between 50 and 100 µs and from 150 to 300 µs, suggesting the likelihood of light scattering by light-absorbing particles within the medium. This increases the medium refractive index causing changes in the sound waves propagated in the medium[38]. Even though changes in the patterns and characteristics of the PA waves are pronounced for the sparsely-spaced concentration levels in **Figure 7**,

8

**Figure 6** shows that their trends can be relatively similar for closely-spaced concentration levels. To this end, the Alexnet and LSTM models are trained to classify the ink concentration. Using the best parameters setting in section 2.6 in the network training, the best and the worst performing models are presented in **Figures 8** and **9**. The mean and standard deviation of the considered metrics calculated from three consecutive runs are shown in **Table 3**. The average training time for Alexnet and LSTM is $37 \pm 4.71$ minutes and $1364 \pm 21.78$ minutes, respectively. **Table 4** shows the overall classification performance of these models averaged based on their results in **Table 3**. Also included in this table are the classification performance results of the considered traditional methods.
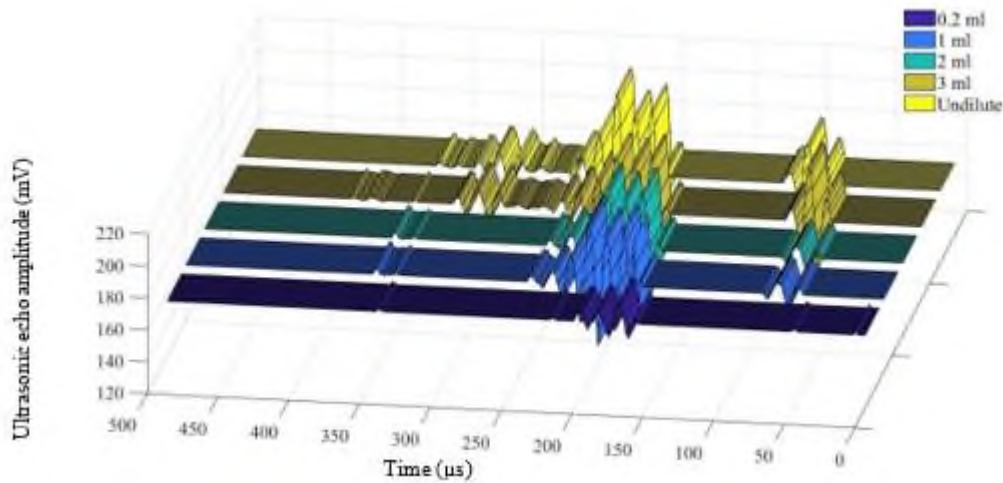


**Figure 7.** A comparison of the produced PA signals recorded for sparsely-spaced concentration levels at wavelength 450 nm.

**Table 3.** The average classification performance of Alexnet and LSTM tested on datasets of different wavelengths for ink concentration classification.

| Wavelength (nm) | Average model performance[I] (%) | |
| --- | --- | --- |
| | **Alexnet** | **LSTM** |
| 450 | Acc: $94.00 \pm 0.00$<br>Spec: $90.20 \pm 0.12$<br>Sen: $90.00 \pm 0.50$<br>Prec: $87.00$ | Acc: $96.00 \pm 0.97$<br>Spec: $93.80 \pm 0.17$<br>Sen: $99.30 \pm 0.59$<br>Prec: $93.00$ |
| 500 | Acc: $83.00 \pm 0.00$<br>Spec: $75.00 \pm 0.22$<br>Sen: $84.70 \pm 0.90$<br>Prec: $90.00$ | Acc: $87.30 \pm 0.84$<br>Spec: $84.30 \pm 2.55$<br>Sen: $87.90 \pm 2.63$<br>Prec: $91.00$ |
| 550 | Acc: $75.00 \pm 0.00$<br>Spec: $70.00 \pm 0.55$<br>Sen: $77.30 \pm 2.70$<br>Prec: $70.00$ | Acc: $94.00 \pm 0.41$<br>Spec: $94.90 \pm 0.52$<br>Sen: $94.70 \pm 2.03$<br>Prec: $93.00$ |
| 600 | Acc: $70.00 \pm 0.00$<br>Spec: $74.70 \pm 0.32$<br>Sen: $70.40 \pm 0.60$<br>Prec: $76.00$ | Acc: $70.00 \pm 0.30$<br>Spec: $76.80 \pm 0.26$<br>Sen: $72.00 \pm 0.34$<br>Prec: $88.00$ |
| 633* | Acc: $87.00 \pm 0.00$<br>Spec: $88.00 \pm 1.45$<br>Sen: $91.20 \pm 0.36$<br>Prec: $88.00$ | Acc: $76.00 \pm 0.64$<br>Spec: $81.81 \pm 0.21$<br>Sen: $76.70 \pm 0.92$<br>Prec: $85.00$ |
| 650 | Acc: $75.00 \pm 0.78$<br>Spec: $68.20 \pm 0.22$<br>Sen: $68.00 \pm 0.79$<br>Prec: $79.00$ | Acc: $72.70 \pm 0.60$<br>Spec: $67.40 \pm 0.15$<br>Sen: $70.21 \pm 2.13$<br>Prec: $82.00$ |

**Table 3.** (*Continued*).

| Wavelength (nm) | Average model performance[I] (%) | |
| --- | --- | --- |
| | **Alexnet** | **LSTM** |
| 700 | Acc: 66.00 ± 0.61<br>Spec: 63.30 ± 0.40<br>Sen: 65.00 ± 0.72<br>Prec: 70.00 | Acc: 58.92 ± 0.67<br>Spec: 65.80 ± 0.23<br>Sen: 64.37 ± 0.03<br>Prec: 58.00 |

[I]Acc = accuracy, Spec = specificity, Sen = sensitivity, Prec = precision.
* Laser illuminated.
Remarks: The best models are highlighted in blue, and the worst models are in orange.

**Table 4.** A comparison of the overall classification performance of Alexnet, LSTM, and traditional machine learning methods for ink concentration classification.

| Average model performance (in %) | | Accuracy | Specificity | Sensitivity | Precision |
| --- | --- | --- | --- | --- | --- |
| Model | Alexnet | 78.57 ± 9.15 | 75.67 ± 9.240 | 78.09 ± 9.93 | 80.00 ± 7.84 |
| | LSTM | 79.27 ± 12.62 | 80.69 ± 10.70 | 80.74 ± 12.32 | 84.29 ± 11.39 |
| | SVM | 59.81 ± 14.10 | 62.70 ± 12.18 | 52.84 ± 11.76 | 60.40 ± 13.21 |
| | KNN | 60.43 ± 14.10 | 60.22 ± 13.71 | 62.74 ± 11.09 | 64.73 ± 10.97 |
| | Decision Tree | 46.86 ± 20.50 | 51.07 ± 13.47 | 47.80 ± 9.21 | 51.33 ± 9.82 |

**Table 4** shows that even though LSTM produced a slightly better average performance in all metrics than Alexnet, the latter produces higher consistency with smaller standard deviation values due to the higher consistency in the results in **Table 3**. The layers in the Alexnet apply a set of learnable filters to the input image and each filter to detect spatial patterns in the dataset, whereas the LSTM network uses time series data and a control gate method to update the cells. Thus, the LSTM model may be sensitive to the time domain variation in the data, such as that noticeably seen near the end of the time series in **Figure 7**. An investigation into the signals at other wavelengths revealed larger data inconsistency and noise, especially for wavelength 700 nm. Training LSTM with corrupted signals made the network less capable of learning hidden patterns, leading to variations in the performance in **Figure 9** and **Table 4**, and it also demands a longer training time to achieve a reasonable performance. Meanwhile, the KNN model surpassed SVM and DT in terms of average accuracy, sensitivity, and precision. The DT method obtained the worst classification performance with the lowest values in all performance measures.

The confusion matrix of the best results from the Alexnet model is shown in **Figure 8** to achieve 95.3% testing accuracy with a training time of 30 minutes (using epochs: 20, mini batch-size: 15, gradient descent threshold: 0.001, and learning rate: 0.001). This is chosen from three simulation runs using the 450 nm dataset. The LSTM model achieved 99.3% testing accuracy using the same dataset at the price of a longer training time (i.e., 1404 minutes using epoch no.: 3000, mini batch-size: 256, gradient descent threshold: 0.001, and learning rate: 0.0005). **Figure 9** shows the classification performance of the worst-performing models. Both the Alexnet and LSTM have the lowest prediction accuracy using 700 nm data.

**Alexnet (top)**

| Output Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **15** 10.0% | 1 0.7% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 93.8% 6.3% |
| 2 | 0 0.0% | **14** 9.3% | 0 0.0% | 6 4.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 70.0% 30.0% |
| 3 | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 4 | 0 0.0% | 0 0.0% | 0 0.0% | **9** 6.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 5 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 6 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 7 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 8 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 9 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 100% 0.0% |
| 10 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 100% 0.0% |
| | 100% 0.0% | 93.3% 6.7% | 100% 0.0% | 60.0% 40.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | **95.3%** 4.7% |

Target Class

**LSTM (bottom)**

| Output Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **14** 9.3% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 2 | 1 0.7% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 93.8% 6.3% |
| 3 | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 4 | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 5 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 6 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 7 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 8 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 9 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 0 0.0% | 100% 0.0% |
| 10 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | **15** 10.0% | 100% 0.0% |
| | 93.3% 6.7% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | **99.3%** 0.7% |

Target Class

**Figure 8.** The confusion matrix of the best models. The diagram shows the class predictions for the 450 nm testing dataset using Alexnet (top) and LSTM (bottom).
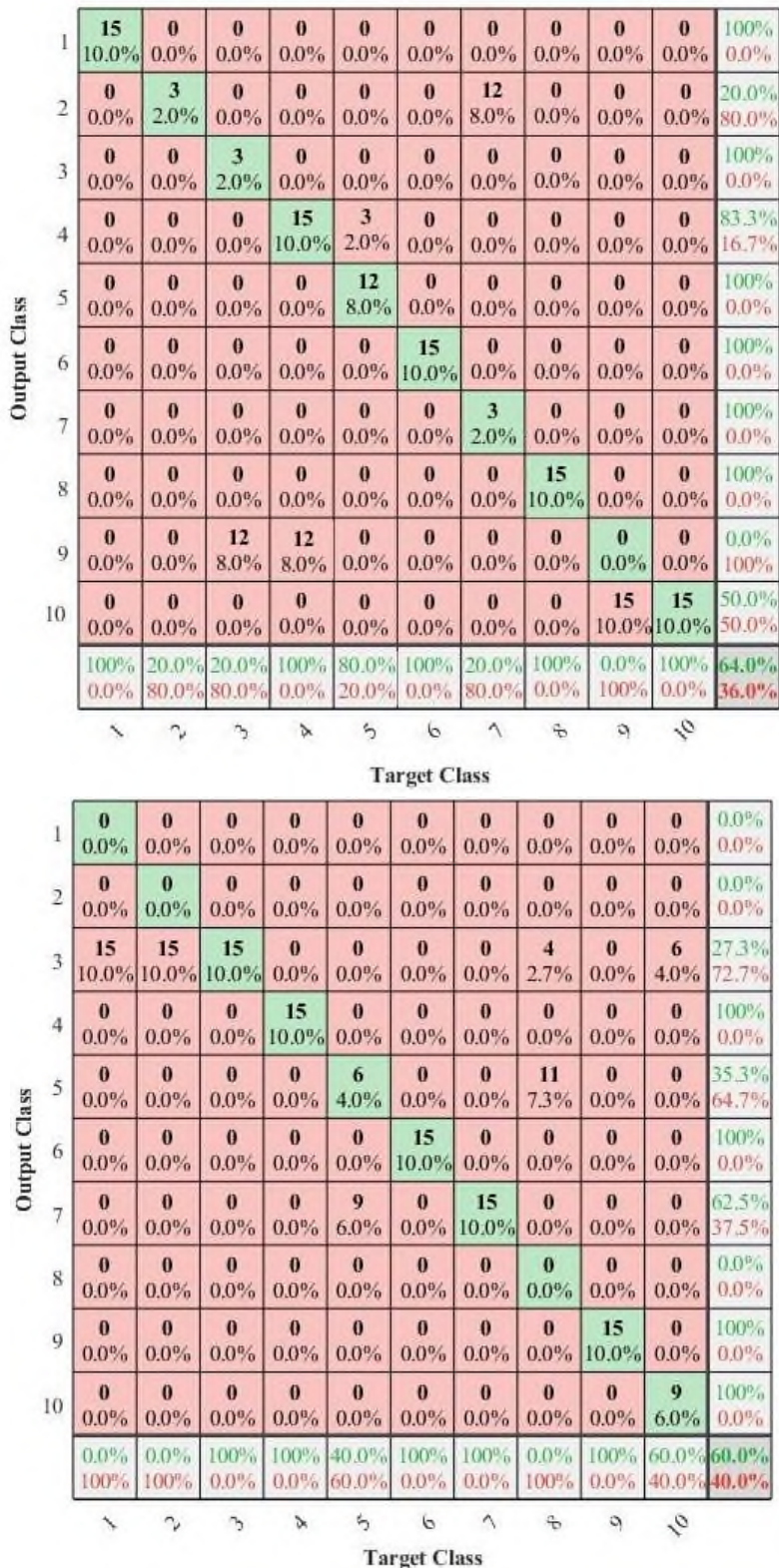
**Figure 9.** The confusion matrix of the worst models. The diagram shows the class predictions for the 700 nm testing dataset using Alexnet (top) and LSTM (bottom).

## 4. Discussion

This work considered single-wavelength laser and multi-wavelength LED illumination systems to identify the best light wavelength for black ink concentration prediction. While our results in **Table 3** revealed a small difference in the performance of the two deep models using photoacoustic data, the prediction performance depends largely on the illuminating wavelength. The pretrained Alexnet slightly underperformed

compared to its time-recurrent counterpart, wherein the latter allowed data to travel through each layer, keeping only relevant information and discarding irrelevant information in each cell during the training phase. This characteristic of LSTM has slightly improved the overall accuracy, especially when there is a consistent trend in the concentration-dependent temporal changes in the acoustic wave signals, as shown in **Figures 6** and **7**. This pattern may be a significant feature for detecting concentration levels. Nonetheless, we observed larger fluctuations in their training and validation accuracies against the epoch count while training the model, which we attributed to the inefficiency of the hyperparameter tuning process. In addition, we do not rule out the possibility that the model is learning noise patterns in the data.

According to **Table 3**, the classification performance of both models generally decreases with an increase in wavelength. We attributed this observation to the peak light absorption at the wavelength of 450 nm, producing high signal intensity in **Figure 6**. The same (but weaker) effects have been produced by the highly intense and highly directional the laser system. **Table 4** shows superior classification results using the 10 mW 633 nm laser, compared to its neighbouring wavelengths at 600 and 650 nm (from the LED). The LED system used in this study has a higher total output intensity of 64 mW, but it has high-angle illumination beams of 120°, lowering the energy of light illuminating the selected region. Thus, producing a lower signal-to-noise performance than the laser beam. The worst results are observed in **Table 3** and **Figure 9** for models trained using the 700 nm dataset. The light absorption at this wavelength is the lowest, rendering high noise levels and a lack of distinctive features of the concentration-dependent photoacoustic signal. The inconsistency in the features that the networks used to recognize concentration levels at this wavelength implies the need for a more robust noise suppression system, for example, a lock-in amplifier, in the future.

Based on **Figure 8**, the Alexnet model failed to classify class 2 and class 4 signals, while the LSTM model produced near-perfect results for all classes except for class 1 (i.e., 1L of water added with 0.2 mL ink). The signals in class 1 were wrongly categorized as class 2 (i.e., 0.4 mL ink sample). These signals were investigated using the cross-correlation coefficient method (xcorr in MATLAB 2020b). The result shows a high correlation factor of 0.991, which explains the insufficiency of the models to differentiate between the two classes.

The results shown in **Table 4** comparing classification results from deep networks with a Decision Tree, SVM, and KNN showed the overall inferiority of the traditional methods, with an average performance of 46.86 ± 20.50%, 59.81 ± 14.10%, and 60.43 ± 14.10%, respectively. SVM, which uses a clear margin hyperplane to separate different classes, failed to establish a linear separation between signals of different ink concentrations. Meanwhile the Decision Tree that separates the data based on their signal behaviours cannot accurately identify signals with similar patterns. KNN, on the other hand, judged the outcome based on similarity in the neighbourhood. The considerable overlap in the signals' characteristics might have compromised the accuracy of this classifier. It is promising to note that using PA technology with deep learning produced an overall promising classification accuracy. This implies the feasibility of the PA technology and deep learning approach for predicting the optical properties of a scattering-absorbing medium. Such application includes characterizing blood components (such as blood carbon monoxide levels) for medical diagnosis, treatment and health research.

# 5. Conclusion

Our results show that the Alexnet and LSTM models can classify ink concentrations based on PA signals. They yielded considerably good results in terms of the high classification accuracy and precision of 95.3% and 87%, and 99.3% and 93% for Alexnet and LSTM, respectively. The LSTM model produced a better performance at the price of a longer training time. Through these efforts, we conclude that the PA system and time series deep learning models have a promising future in medical diagnosis and treatment due to their ability to give valuable insights into the optical properties of a medium (e.g., tissues) without the need for a contrast agent. A possible application of this strategy is in monitoring the changes in carboxyhemoglobin levels, in

which the differences in their PA signatures could be distinguished based on changes in light absorption. Future work deals with implementing the proposed system for animal studies before human experimentation in the study of drug delivery and disease progression and treatment. In addition, a noise filter may also be applied to remove the noise in the signals to achieve noise-free features and improve the signal's signal-to-noise performance.

## Author contributions

Conceptualization, AH; methodology, AH and XN; software, AH and HLC; validation, AH and HLC; formal analysis, AH and HLC; investigation, AH and HLC; resources, AH and HLC; writing—original draft preparation, HLC; writing—review and editing, AH; supervision, AH and XN; project administration, AH and XN; funding acquisition, AH and XN.

## Funding

## Availability of data and materials

The data used in this study will be made available upon formal request to the corresponding author.

## Conflicts of interest

The authors declare no conflict of interest.

## References

1. Urits I, Seifert D, Seats A, et al. Treatment strategies and effective management of phantom limb—Associated pain. *Current Pain and Headache Reports* 2019; 23(9): 1–7. doi: 10.1007/s11916-019-0802-0
2. Kleiser S, Ostojic D, Andresen B, et al. Comparison of tissue oximeters on a liquid phantom with adjustable optical properties: an extension. *Biomedical Optics Express* 2018; 9(1): 86–101. doi: 10.1364/BOE.9.000086
3. Ntombela L, Bamise A, Chetty N. Low-cost fabrication of optical tissue phantoms for use in biomedical imaging. *Heliyon* 2020; 6(3): e03602. doi: 10.1016/j.heliyon.2020.e03602
4. Jawad HJ, Sarimollaoglu M, Biris AS, Zharov VP. Dynamic blood flow phantom with negative and positive photoacoustic contrasts. *Biomedical Optics Express* 2018; 9(10): 4702–4713. doi: 10.1364/BOE.9.004702
5. Jonasson H, Fredriksson I, Larsson M, et al. Validation of speed-resolved laser doppler perfusion in a multimodal optical system using a blood-flow phantom. *Journal of Biomedical Optics* 2019; 24(9): 95002. doi: 10.1117/1.JBO.24.9.095002
6. Mumi ARO, Alias R, Abdullah J, et al. Assessment of electromagnetic absorption towards human head using specific absorption rate. *Bulletin of Electrical Engineering and Informatics* 2018; 7(4): 657–664. doi: 10.11591/eei.v7i4.1357
7. Wu Y, Cheng M, Wang W, et al. Development of chinese female computational phantom rad-human and its application in radiation dosimetry assessment. *Nuclear Technology* 2018; 201(2): 155–164
8. van Herten RLM, Chiribiri A, Breeuwer M, et al. Physics-informed neural networks for myocardial perfusion MRI quantification. *Medical Image Analysis* 2022; 78: 102399. doi: 10.1016/j.media.2022.102399
9. Cui H, Liu C, Esworthy T, et al. 4D physiologically adaptable cardiac patch: A 4-month in vivo study for the treatment of myocardial infarction. *Science Advances* 2020; 6(26): eabb5067. doi: 10.1126/sciadv.abb5067
10. Xie Z, Yang Y, He Y, et al. In vivo assessment of inflammation in carotid atherosclerosis by noninvasive photoacoustic imaging. *Theranostics* 2020; 10(10): 4694. doi: 10.7150/thno.41211
11. Bachir W, Dargham FA. Feasibility of 830 nm laser imaging for vein localization in dark skin tissue-mimicking phantoms. *Physical and Engineering Sciences in Medicine* 2022; 45(1): 135–142. doi: 10.1007/s13246-021-01096-x
12. Imanishi A, Kimura A, Miyamoto H, et al. Human organ phantoms for catheterization using the radiation crosslinking technique. *Journal of Applied Polymer Science* 2021; 138(33): 50818. doi: 10.1002/app.50818
13. Li T, Lu Z, Li Z, et al. In vivo detection of the margin of simulated melanoma based on a highly integrated and intelligent fiber optic spectrometer. *Acta Laser Biology Sinica* 2020; 29(6): 506–512. doi: 10.3969/j.issn.1007-

7146.2020.06.005

14. Lam JH, Hill BY, Quang T, et al. Multi-modal diffuse optical spectroscopy for high-speed monitoring and wide-area mapping of tissue optical properties and hemodynamics. *Journal of Biomedical Optics* 2021; 26(8): 85002. doi: 10.1117/1.JBO.26.8.085002

15. Anugrah MA, Suryani S, Ilyas S, et al. Composite gelatin/Rhizophora SPP particleboards/PVA for soft tissue phantom applications. *Radiation Physics and Chemistry* 2020; 173: 108878. doi: 10.1016/j.radphyschem.2020.108878

16. Manwar R, Mohsin Z, Xu Q. Signal and image processing in biomedical photoacoustic imaging: A review. *Optics* 2020; 2(1): 1–24. doi: 10.3390/opt2010001

17. Glatz J, Deliolanis NC, Razansky D, et al. Blind source unmixing in multi-spectral optoacoustic tomography. *Optics Express* 2011; 19(4): 3175–3184. doi: 10.1364/OE.19.003175

18. Razansky D, Vinegoni C, Ntziachristos V. Multispectral photoacoustic imaging of fluorochromes in small animals. *Optics Letters* 2007; 32(19): 2891–2893. doi: 10.1364/OL.32.002891

19. Roy K, Thomas A, Paul S, et al. An optofluidic dye concentration detector based on the pulsed photoacoustic effect. *Microfluidics, BioMEMS, and Medical Microsystems XIX* 2021; 11637: 89–95. doi: 10.1117/12.2582656

20. Dolet A, Ammanouil R, Petrilli V, et al. In vitro and in vivo multispectral photoacoustic imaging for the evaluation of chromophore concentration. *Sensors* 2021; 21(10): 3366. doi: 10.3390/s21103366

21. Raschka S, Patterson J, Nolet C. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information* 2020; 11(4): 193. doi: 10.3390/info11040193

22. Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. *Clinical Biochemistry* 2019; 69: 1–7. doi: 10.1016/j.clinbiochem.2019.04.013

23. Lee J, Davari H, Singh J, Pandhare V. Industrial artificial intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters* 2018; 18: 20–23. doi: 10.1016/j.mfglet.2018.09.002

24. Yu Kun-Hsing, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2018; 2(10): 719–731. doi: 10.1038/s41551-018-0305-z

25. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal* 2019; 6(2): 94. doi: 10.7861/futurehosp.6-2-94

26. Syarief M, Setiawan W. Convolutional Neural Network for maize leaf disease image classification. *Telkomnika (Telecommunication Computing Electronics and Control)* 2020; 18(3): 1376–1381. doi: 10.12928/TELKOMNIKA.v18i3.14840

27. Youn SW, Kwon J, Kim J, et al. Ideal bolus geometry predicted from in vitro pulsatile flow phantom and artificial neural networks for the optimization of image acquisition protocols for aortic contrast-enhanced computed tomography angiography. *Cardiovascular Imaging Asia* 2019; 3(2): 35–46. doi: 10.22468/cvia.2018.00248

28. Zhang H, LI H, Nyayapathi N, et al. A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography. *Computerized Medical Imaging and Graphics* 2020; 84: 101720. doi: 10.1016/j.compmedimag.2020.101720

29. Chen MT, Durr NJ. Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning. *Journal of Biomedical Optics* 2020; 25(11): 112907. doi: 10.1117/1.JBO.25.11.112907

30. Johnstonbaugh K, Agrawal S, Durairaj DA, et al. A deep learning approach to photoacoustic wavefront localization in deep-tissue medium. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2020; 67(12): 2649–2659. doi: 10.1109/TUFFC.2020.2964698

31. Siami-Namini S, Tavakoli N, Namin AS. The Performance of LSTM and BiLSTM in Forecasting Time Series. In: Proceedings of 2019 IEEE International Conference on Big Data (Big Data); 9–12 December 2019; Los Angeles, CA, USA. pp. 3285–3292.

32. Li Z, Ge Q, Feng J, et al. Quantification of blood flow index in Diffuse Correlation Spectroscopy using Long Short-Term Memory architecture. *Biomedical Optics Express* 2021; 12(7): 4131–4146. doi: 10.1364/BOE.423777

33. Alex S. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena* 2020; 404: 132306. doi: 10.1016/j.physd.2019.132306

34. Gupta A. A comprehensive guide on deep learning optimizers. Available online: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers (accessed on 7 October 2021).

35. Devassy BM, George S. Ink classification using Convolutional Neural Network. In: Proceedings of Conference NISKJ; 2019.

36. Wang S, He H, Lv R, et al. Classification modeling method for hyperspectral stamp-pad ink data based on one-dimensional Convolutional Neural Network. *Journal of Forensic Sciences* 2022; 67(2): 550–561. doi: 10.1111/1556-4029.14909

37. Rajian JR, Carson PL, Wang X. Quantitative photoacoustic measurement of tissue optical absorption spectrum aided by an optical contrast agent. *Optics Express* 2009; 17(6): 4879–4889. doi: 10.1364/OE.17.004879

38. Jakovljevic M, Hsieh S, Ali R, et al. Local speed of sound eatimation in tissue using pulse-echo ultrasound: Model-based approach. *The Journal of the Acoustical Society of America* 2018; 144(1): 254–256. doi: 10.1121/1.5043402