Introduction

The respiratory system is a network of tissues and organs and is one of the most intricate systems in the human body [1]. The therapeutic effectiveness of inhaled aerosols depends on their spatial distribution within the respiratory tract and on the amount of the Active Pharmaceutical Ingredients (APIs) present in them [2-4]. However, the process of concentrating decomposed ambient particulate matter locally within the respiratory tract may lead to both lower and upper respiratory tract diseases [5-7]. Hence, detailed characterization of aerosol particle transportation and deposition is essential to quantitatively analyze their therapeutic and deleterious effects upon inhalation. Additionally, deposition of aerosol particles in airways plays a crucial role in the delivery of aerosol drugs [4,8].

The study of aerosol transport and deposition due to bronchial tube airflow can improve our understanding of the damaging or beneficial effects of their inhalation. The suspended particles in the aerosol are of numerous shapes and sizes ranging from nano-sized particles (diameter less than 1 μm) to large-sized pollens (diameter greater than 100 μm) including therapeutic aerosols, ultrafine dust, microbial aerosols, asbestos, pollen, and fumes [9-11].

Recognizing the size distribution of inhaled therapeutic/non-therapeutic aerosols is beneficial for assessing the harmful or useful effect of using aerosol. In this study, we aim to detect and segment the aerosol particle sizes into three parts of a respiratory system automatically using Machine Learning (ML) algorithms [8,11,12]. ML techniques employ a range of strate-
gies and data to reproduce specific outputs from complex engineering/biological systems. Image segmentation algorithms are subsets of ML techniques and are vital to many computers vision and image processing applications [13-15]. Segmentation is employed in many image processing fields such as medical imaging [16,17], object tracking [18,19], and satellite imaging [20-22]. The wide use of the segmentation algorithms can be attributed to the segmentation outcomes that directly affect the performance of the whole system [20,23]. Image segmentation strategies are broadly classified into four categories: edge-based, region-based, threshold-based, and deep learning methods [24-26].

Region-based approaches primarily search for some seed points inside the image and appropriate region growing methods are employed subsequently to reach the boundaries of the objects [27,28]. Edge-based algorithms try to identify the edges or contours inside the input image. Therefore, segmentation is affected by a determination of the region margins inside the image. The threshold-based techniques generally employ the histogram of the input image for identifying single or multiple thresholds [29-31].

In the last few years, Deep Learning (DL) strategies have yielded remarkable success in better segmentation results compared to other hand-crafted feature extraction methods in the different fields of computer vision tasks [17,32-36]. Convolutional Neural Networks (CNNs) are considered a type of DL model with high abilities for extracting and learning crucial features. Moreover, CNN models are able to obtain the best possible features needed for feeding to other models (classic models) [37,38].

In this study, we suggest a CNN-based strategy to recognize and segment particles inside the respiratory system. To investigate the decomposition of particles, the Weibel Airway (WA) model [39] was adopted in this study as indicated in Figure 1.

The remaining parts of this paper are organized as follows: Initially, a texture descriptor approach is described in section 2.1. The characteristics and architecture of the suggested CNN model are presented in section 2.2. In section 2.3, we propose a matching strategy to find all particles inside the image. Section 3 describes the implementation details of the suggested model. Section 4 provides the conclusions.

Materials and methods

This section is divided into two sub-sections. Firstly, we describe the textural analysis that is useful to identify some significant textural information. We subsequently describe the procedure for finding more informative features to identify the border of objects by employing a CNN model.

Texture analysis (2.1)

Textural information as image features is very valuable in many computer vision and image processing applications [40,41]. There is a broad literature on textural analysis in the machine vision literature where the principal emphasis has been on synthesis, segmentation, and classification. Textural information is used as input features and has been employed in different applications such as medical image analysis, text analysis, and aerial and satellite image analysis as a descriptor [30,42,43].

In texture segmentation and classification, the aim is to divide the input image into a set of similar textured regions (homogeneous). Some of these similarities are size, orientation, shape, texture, pattern, color, etc [44,45].

In order to characterize textured images, various texture feature extraction techniques have been suggested. One can use traditional algorithms that employ a co-occurrence of matrix-based methods [33], fractal analysis [46], and filter-based approaches such as Gaussian Markov random fields [47], wavelet [48], and Gabor [23]. The Local Directional Pattern (LDP) technique is one of the most popular strategies that focus on the boundaries of objects in pre-defined directions [49]. Hence, LDP is able to recognize more prominent edges using an edge detection approach called “Kirsch filters”.

Kirsch filters (Kirsch kernels) are non-linear edge detectors and are utilized to explore the edge response values in eight directions in the vicinity. The results of applying Kirsch kernels to an input image are demonstrated in Figure 2. LDP features are achieved by calculating the obtained edge maps at each pixel position in all eight compass directions and creating a code from relative strength magnitude [49,50]. LDP converts the directional information of input images. Each bit of code is generated by considering local vicinity and obtaining robustness in noisy situations. The result of applying the LDP approach to an input image is demonstrated in Figure 3.

Our deep learning model (2.2)

In the previous step, we used a texture descriptor approach to extract significant features that are crucial in detecting the edges of all particles precisely. In this section, by employing a Convolutional Neural Network (CNN), we classify all the pixels...
inside the image into edge and non-edge type of pixels. By doing so, the exact borders of all circular objects (some of them are not visible completely) inside the image are detected.

The CNNs are popular and widespread deep learning (DL) pipelines that have become one of the most successful techniques in the field of machine learning (ML). Typically, a CNN structure consists of four layers: 1) convolutional, 2) pooling, 3) activation, and 4) fully-connected layers [37, 41].

The convolutional layers (conv layers) aim to learn hidden patterns and feature representations of the inputs [51]. Each neuron inside the feature map is connected to an area of neighbouring neurons in the prior layer. Such neighbourhoods are referred to as the neuron’s receptive fields in the last layer. The new feature map is generated by convolving the input with a learned filter. For generating each feature map, the filter is shared by all spatial locations of the input [38, 52, 53].

Similar to conv layers, pooling operators include a predefined patch (window) that is slid over all areas in the input based on its stride, calculating an output for each point traversed by the pooling window [54]. However, unlike kernels in the convolutional layer and the cross-correlation computation of the inputs, the pooling layer contains no kernel. Normally, pooling operators calculate either the average (mean-pooling) or the maximum (max-pooling) value of the elements in the pooling window [26, 55].

The Fully Connected Layers (FC) can end up with a Soft Max (SM) output layer to classify the input. The SM activation function is utilized in the output layer and is a multi-class version of the logistic regression [26, 56, 57].

This layer is able to convert all the negative values to zero. We utilized a 2×2 max-pooling layer after each conv layer to decrease the dimension of the obtained feature maps. Moreover, in order to avoid memorization, a dropout of 0.15 is used. For increasing the training samples, two augmentation strategies are utilized including random rotations and random Gaussian noise [26, 32, 58]. The utilized parameters for training the CNN model are shown in Table 1.

### Table 1: Parameters utilized for training the CNN model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch size</td>
<td>35x35</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Output number</td>
<td>2</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Batch size</td>
<td>7000</td>
</tr>
<tr>
<td>Learning Rate Drop Factor</td>
<td>0.15</td>
</tr>
<tr>
<td>Max Epochs</td>
<td>50</td>
</tr>
</tbody>
</table>

Finding circular objects (2.3)

In the last section, we extracted all edges of objects. In this part, we propose a searching approach that draws some circles around each object (local area) to find the best match. This process is demonstrated in detail in Figure 5. Inside a loop with 100 iterations, we generate different circles (red circles in Figure 5) with random radiuses (close to the target radius). Next, by comparing the border of generated circles with the target, we are able to find the best circle that represents the occluded target and can be added to a list for finding the whole objects. In other words, we calculate the over fitting pixels created by overlapping both borders of the object and the border of the generated circle to find the best fit.

The employed CNN model is shown in Figure 4 and has two similar feature exploration routes for extracting high-level and low-level features. Each feature extraction route has four convolutional layers. The first two and the last two conv layers explore low-level and high-level features, respectively. The number of utilized filters in each route increases with the depth: 8, 16, 32, and 64. A Rectified Linear Units (ReLU) layer is utilized for applying activation function in an element-wise manner.

**Experiments**

**Assessment metrics**

Three techniques are employed to evaluate the segmentation performance, including recall, precision, and F1-score. Sensitivity or recall is the True Positive Rate (TP). In other words, it is calculated by dividing the number of correct positive predictions by the summation of true positive and False Negative (FN). In addition, precision is calculated by dividing the true positive by the total of true positive and False Positive (FP).

\[
\text{Sensitivity or Recall} = \frac{TP}{TP + FN} \times 100 \quad (1)
\]

\[
\text{PPV or Precision} = \frac{TP}{TP + FP} \times 100\% \quad (2)
\]

\[
F_1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (3)
\]
**Experimental results and discussions**

We conduct experiments on a private dataset containing 10,000 images with the dimensions of 520×640. For evaluating the proposed method more accurately, we divide the respiratory system into three regions (upper, middle and lower) and evaluate each area separately. Figure 6 shows an example of dividing the respiratory system into three parts.

To have a clear understanding for comparison purposes, we use four other texture descriptors (Local Binary Pattern (LBP) [40], Local Directional Number Pattern (LDNP) [41], Local Ternary Pattern (LTP) [59], and Fuzzy Local Ternary Pattern (FLTP) [60]) to evaluate the segmentation performance. Quantitative results of different kinds of our structure are described in Table 2.

For each index in Table 2, the highest PPV, Sensitivity, and F1-score are highlighted in bold. The results in Table 2 clearly demonstrate that our technique is able to obtain the highest sensitivity values in regions 1 and 2 and the highest score for region 3 is obtained by LDNP. The structures based on LTP and FLTP have achieved good accuracy, but these approaches may or may not work if given more color similarity in the local areas inside the input images. Besides, there is a minimum difference between the values of PPV using LTP and FLTP. Another interesting point is that the worst scores for all measures are obtained using LBP in all areas. Additionally, the segmentation results in terms of PPV using LBP, LTP, and FLTP methods are generally under 90%. By employing the LDP strategy, all criteria are improved in comparison to other approaches, but the sensitivity value in region 3 employing LDNP is still higher.

Table 2: Comparison between the results of our strategy using the LDP model and four different texture descriptor techniques.

<table>
<thead>
<tr>
<th>Method</th>
<th>PPV (%)</th>
<th>Sensitivity (%)</th>
<th>F1-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Region 1</td>
<td>Region 2</td>
<td>Region 3</td>
</tr>
<tr>
<td>Local Binary Pattern (LBP) [40]</td>
<td>76</td>
<td>72</td>
<td>71</td>
</tr>
<tr>
<td>Local Directional Number Pattern (LDNP) [41]</td>
<td>89</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td>Local Ternary Pattern (LTP) [59]</td>
<td>86</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>Fuzzy Local Ternary Pattern (FLTP) [60]</td>
<td>87</td>
<td>82</td>
<td>83</td>
</tr>
<tr>
<td>Proposed method (LDP)</td>
<td>94</td>
<td>93</td>
<td>90</td>
</tr>
</tbody>
</table>

**Discussions and conclusions**

In this work, we presented a method for segmenting aerosol-based particles inside the respiratory system. We initially employed a texture descriptor technique to represent more unique features and for obtaining the border of each particle (object) more accurately. Then, by applying the original image and the encoded image to a CNN model, an edge map of the input image is created. The network only requires reasonable data for the training phase. Lastly, we suggested a fitting circle approach to compare each object with a lot of potential circles to find the best match and recognize the object. A comparison of the results obtained in this study with some texture descriptor approaches is given in Table 2. The comparison with these approaches demonstrates that the proposed method segment aerosol-based particles in the respiratory system with at least a 5% improvement in precision.

**Funding:** The funding sources had no involvement in the study design, collection, analysis or interpretation of data, writing of the manuscript or in the decision to submit the manuscript for publication.

**Declaration of interests:** We declare no conflict of interest.

**References**

958-964.


60. P Kavitha, K Vijaya. Fuzzy local ternary pattern and skin texture properties based counter measure against face spoofing in biometric systems. Computational Intelligence. 2021; 37: 559-577.