IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2019, accepted June 11, 2019, date of publication June 14, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923002

Prediction of the Location of the Glottis in Laryngeal Images by Using a Novel Deep-Learning Algorithm

JONG SOO KIM[1], YONGIL CHO[2], AND TAE HO LIM[3]

[1]Institute for Software Convergence, Hanyang University, Seoul 04763, South Korea
[2]Hanyang University Hospital, Seoul 04763, South Korea
[3]Department of Emergency Medicine, College of Medicine, Hanyang University, Seoul 04763, South Korea

Corresponding author: Tae Ho Lim (erthim@gmail.com)

This work was supported by the Hanyang University, Seoul, South Korea, under Grant 201800000000277 and Grant 201800000003039.

ABSTRACT A novel deep-learning algorithm for artificial neural networks (ANNs) was developed and presented in this paper, which is intuitively understandable, simple, efficient, and completely different from the back-propagation method, i.e., randomly selecting weight factors and bias values of an ANN and adjusting their values by small random amounts during the training session where it does not need to calculate the gradients of the training error to adjust weight factors as does the back-propagation method. The algorithm was applied to predict the location of the glottis in airway images obtained using a video airway device. The glottic locations were marked in 1,200 airway images captured using GlideScope® and fiberoptic laryngoscopy. With the randomly selected 1,000 training set data, 84 ANN models were trained using the above algorithm. We sought an ANN model that minimized the average training error for all training set data by reducing the input image resolution. As the resolution was reduced, the average training error decreased to its lowest level at 30×30 pixels. Eventually, the 900-98-49 ANN model was selected as the prediction model for the location of the glottis; it was the model with the lowest training error, i.e., the highest learning rate. The selected prediction model was applied to the remaining 200 test set data to obtain the test accuracy, and we obtained that the accurate prediction and the adjacent prediction rates were 74.5% and 21.5%, respectively. Reducing the input image resolution to an appropriate level could yield better prediction of the glottic location in airway images. This ANN model can help clinicians perform intubation by presenting the predicted location of the glottis.

INDEX TERMS Artificial neural network, deep learning, glottis, location, video airway device.

I. INTRODUCTION

Endotracheal intubation is an important medical procedure for patients with difficulty in spontaneous breathing, airway maintenance problem due to unconsciousness, general anesthesia induction, and cardiopulmonary problem. The endotracheal intubation is a life-saving procedure that is performed in situations such as cardiac arrest, respiratory arrest, when there is a high risk of aspiration, inadequate oxygenation, inadequate ventilation, and airway obstruction [1]. Figure 1 shows the anatomy of the larynx, which is visible during tracheal intubation. The glottis is an opening between two vocal cords (Figure 1a). Plastic endotracheal tubes should be inserted into the trachea through the glottis while performing intubation (Figure 1b).

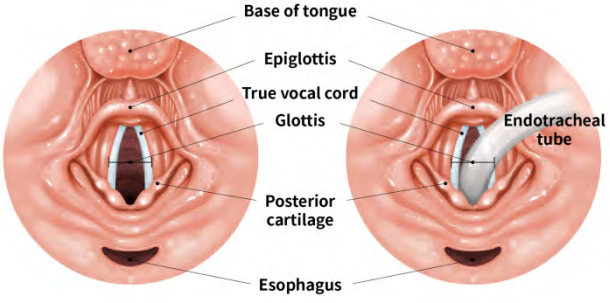

FIGURE 1. Anatomy of larynx visible during tracheal intubation.

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

However, endotracheal tubes are occasionally inserted into inappropriate structures, such as the esophagus, rather than the glottis. If the endotracheal tube is inserted into the wrong structure after intubation, it can lead to serious complications, such as hypoxemia and cardiac arrest [2], [3]. A study in a prehospital setting reported that 56% of patients with esophageal intubation died after emergency department arrival [4]. Therefore, it is very important to accurately identify the location of the glottis when intubating.

However, it is sometimes difficult to determine the location of the glottis during intubation. First, distinguishing the glottis from the esophagus is difficult when intubation is performed by less-experienced personnel who are unfamiliar with the anatomical structures of the airway. Second, the anxiety of the physician can be increased in cases of emergency airway management rather than prepared situations. This anxiety reduces the confidence of the physician, which may lead to the inability to accurately detect the glottis and to intubation failure. In addition, the glottis may be difficult to identify if it is only partially visible or if the anatomy of the larynx is deformed by edema or tumor.

Video laryngoscopy is a method in which the camera is inserted at the tip of the blade. Video laryngoscopy improves glottic visualization and reduces incidents of esophageal intubation [5], [6]. However, there are discrepancies in the findings of various studies regarding whether video laryngoscopy increases the first-attempt success rate [5], [6]. In one study, the success rate of prehospital intubation was found to be lower when video laryngoscopy was used than when it was not [7]. Furthermore, esophageal intubation can occur even when video laryngoscopy was used [5]–[7]. A single esophageal intubation increases the risk of desaturation, aspiration, and cardiac arrest [8].

Recently, computer-based image recognition technology has been rapidly developing. Imaging recognitions using artificial neural networks (ANNs) have recently been studied in various fields of medicine, and high diagnostic predictive powers were reported in the diagnosis of diabetic retinopathy [9], [10], pathologic diagnosis of lymph node metastasis in breast cancer [11], classification of tuberculosis in chest X-ray [12], and anatomical classification in esophagogastroduodenoscopy [13]. Although artificial intelligence technologies were applied to predict the glottic opening in airway images of manikins [14], studies to predict the glottic opening or the glottic location in laryngeal images of patients obtained during clinical practice are scarce.

Feedback while presenting the predicted location of the glottis on the monitor of devices using a microvideo camera, such as video laryngoscopes, flexible intubating endoscopes, and fiberoptic and video intubating stylets, could help clinicians perform intubation. In this study, we sought the prediction model for the glottic location in airway images acquired during clinical practice by using ANNs.

An ANN that imitates the neuronal structure of an animal [15] has been applied to implement functions similar to the brain [16]. We thought that the actual learning process of an animal would never be mathematical and unnatural, as in the back-propagation method [17]–[19], and would be similar to the biological evolution of animals based on a trial-and-error process. Therefore, a novel deep-learning algorithm for ANNs based on the Monte Carlo simulation was developed and applied in this study. An ANN includes hundreds of thousands or more unknown variables, i.e., weight factors and bias values. The novel deep-learning algorithm is an optimization process that applies the Monte Carlo simulation to determine the weight factors and the bias values which minimize the average training error for learning data.

For the back-propagation method [17]–[19], the gradient descent method is employed to determine each weight factor by calculating the delta value using all or a part of the learning data repeatedly until the training error reaches the minimum according to a given learning rate, where the bias value of a node is treated as an additional weight factor of the node given its input value of 1.0 regardless of whether the bias value can be negative, i.e., a threshold rather than a bias. The training errors of the current ANN for all the learning data are calculated in the forward direction every time that all the hundreds of thousands or more weight factors of the ANN are adjusted by the gradient descent in the backward direction. The adjusting process of each weight factor is calculating the gradient of the training error using all or a part of the learning data, and is carried out individually for all the hundreds of thousands or more weight factors. Adjusting all the weight factors by the gradient descent are repeated until the training session is finished. Accordingly, there frequently needs massively parallel processing. Therefore, the back-propagation method requires considerably much computing resources, and sometimes the long-term dependencies problem occurs, caused by gradient vanishing or gradient exploding [20], leading to training failure. Consequently, to correctly extract an acceptable level of knowledge from the learning data, it is known that appropriate parameters should be entered empirically and that an approximate approach should be selected and applied in accordance with the nature of the learning data [21].

The new deep-learning algorithm in this study has a structure that is completely different from that of the back-propagation method. In the case of the new deep-learning algorithm, the average training error for all the learning data of the current ANN is calculated repeatedly in the forward direction only until the training error reaches the minimum, while randomly selected weight factors and bias values of the ANN are being adjusted by the amounts of randomly picked delta-values within a given range. Surely, the algorithm does not need to calculate the gradients of the training error or the adjusting amounts of weight factors and bias values using all or a part of the learning data during the training session, as does the back-propagation method. Namely, adjusting weight factors and bias values of an ANN by small random amounts, not by the computing-intensive gradient descent method applied individually to all the weight factors, during the training session is the main difference of the algorithm

from the back-propagation method. Therefore, the algorithm can be implemented with affordable computing resources because it is simple and efficient. The algorithm is detailed in the next section.

II. METHODS

A. BASIC SETTING

An ANN is a connected set of simple computing elements called ''nodes.'' There are usually organized into several layers; an input layer, an output layer, and several hidden layers in-between. Thus, it is called ''deep-learning'' method. Some ANNs even include hundreds or more hidden layers. Each node in the input layer, i.e., an input node, receives an input from outside of the ANN, and supplies it to all the nodes in the first hidden layer. Each node in a hidden layer, i.e., an intermediate node, receives inputs from all the nodes in the previous layer, calculates single output, and supplies it to all the nodes in the next layer. The signals are then propagated forward through the hidden layers to the output layer. The abstract representation of input data is extracted in the feed-forward process. The optimal number of hidden layers of an ANN depends on the character of input data.

Since the size and aspect ratio of the airway images were different from each other, the images were converted to squares (Figure 2a). Thereafter, their resolutions were reduced to 100×100, 70×70, 50×50, 45×45, 40×40, 35×35, 30×30, and 25×25 pixels (Figure 2b), and each resolution was applied to the input structure of an ANN model.

The number of input nodes of an ANN should be the total number of pixels of the original airway image of a reduced resolution (Figure 2b). The input value of each input node is the corresponding pixel value of the airway image of a reduced resolution; the pixel values were obtained via a black-and-white color-converting process using the following formula:

$$\text{value} = \text{red} \times 0.299 + \text{green} \times 0.587 + \text{blue} \times 0.114 \quad (1)$$

The value was then divided by the maximum value, 255, to convert it to a value between 0.0 and 1.0 (Figure 2c).

The same process was applied to the airway-marked images (Figure 2a′) to obtain the pixel values (Figure 2c′), and the images were then divided into seven horizontal sections and seven vertical sections. This division yielded 49 cells of an airway-marked image as shown in Figure 2c′. The images (Figure 2c and c′) were divided into 49 cells to predict the glottic location among 49 locations in an airway image.

The target value of the cell with the maximum average pixel value (Figure 2c′, red box) was set to 1, and the target values of the other 48 cells were set to 0. Actually, the target values of 49 cells to prepare training data were obtained via the converted pixel values from the airway-marked image of the original resolution (Figure 2a′), not from the airway-marked image of a reduced resolution (Figure 2b′) since there was no difference in the glottic location in the airway-marked image caused by changing the resolution of the image.

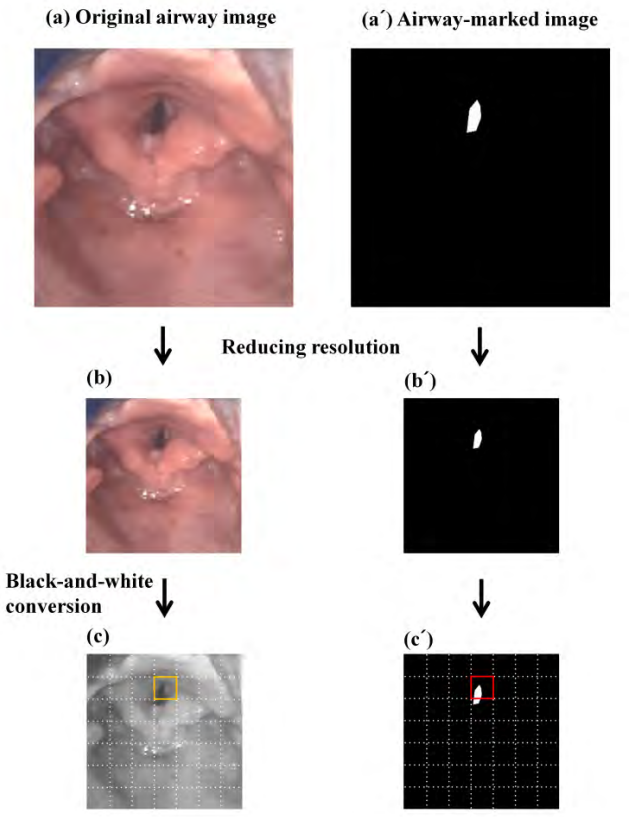

FIGURE 2. Example of the airway image processing for application to the artificial neural network models. (a) Original airway image. (a') Airway-marked image after marking the location of the glottis in white and the remaining areas in black. (b) Original airway image of a reduced resolution. (b') Airway-marked image of a reduced resolution. (c) Original airway image of a reduced resolution converted to black-and-white. The orange box indicates the position of the output node with the maximum output value among all 49 output nodes. (c') In this airway-marked image of a reduced resolution converted to black-and-white, the red box indicates the position of the cell with the maximum target value among all 49 cells.

Thus, the number of output nodes should be 49. The number of hidden layers of an ANN was set to 1, 2, or 3. The number of intermediate nodes in a hidden layer was changed from 49 to 196, which corresponded to one to four times the number of output nodes. Accordingly, various ANN models were constructed.

B. DEEP-LEARNING ALGORITHM

A novel deep-learning algorithm for ANNs that differed from the back-propagation method [17]–[19] was developed and applied. The deep-learning algorithm imitates biological evolution, repeating a substantial number of generations to adapt to a given environment according to the principle of the survival of the fittest. The algorithm consists of (1) randomly selecting weight factors and bias values based on a given selecting ratio of the variables and adjusting their values by small random amounts within the range from -0.1 and +0.1, (2) accepting or rejecting the adjustment depending on whether or not the new values decrease the average training

error for all the learning data, and (3) repeating above two steps. The annealing was obtained not by changing the "temperature" but by gradually reducing the selecting ratio of the variables, i.e., randomly selecting weight factors and bias values among hundreds of thousands or more variables of an ANN, as training proceeded. Initial weight factors and bias values were randomly chosen in the range from -0.2 and +0.2. A training session was terminated after 10 repetitions of a training cycle, where the selecting ratio of the variables of a training cycle was decreased steadily from 15% to 1.5%. The algorithm is presented below.

For the activation function of each node of an ANN, sigmoid functions were employed as follows:

$$y_j = \frac{2}{1+e^{-x_j}} - 1 \tag{2}$$

$$y_j = \frac{1}{1+e^{-x_j}} \tag{3}$$

$$x_j = \sum_i w_{ij} y_i + b_j \tag{4}$$

where (4) for (2) and (3) denotes the summation over all nodes in the previous layer, i.e., each node j in a given layer receives an input y_i from a node i in the previous layer; w_{ij} indicates the weight factor between nodes i and j; and b_j indicates the bias value of node j. The activation functions (2) and (3) were applied to an intermediate node in a hidden layer and an output node, respectively.

Figure 3 shows a schematic representation of the algorithm structure. Initially, for all the variables (total number = N), including all weight factors and bias values of an ANN even comprising up to hundreds or more hidden layers, the values are set independently of each other by randomly selected values up to $+X$ (given 0.2). According to the selecting ratio of the variables (P; given 15% initially), a part of the variables is randomly selected. For the selected part of the variables, the values are changed independently of each other by the amounts of randomly picked delta-values up to $+Y$ (given 0.1). The maximum repeating number (R; given 30) is defined as the number of changing values for a given selected part of the variables. The total selecting range of the variables (S; given 900%) determines the frequency of randomly selecting a part of the variables. The total number of loops (L; given 10) indicates the number of repetitions of a training cycle. The detailed procedure of the algorithm is described below.

(a) The algorithm initializes all the variables of an ANN, yielding values randomly from $-X$ to $+X$ for a weight factor and from 0 to $+X$ for a bias value. The current loop number (l) is set to 1.

(b) P divided by l yields the current selecting ratio of the variables (p), i.e., $p = P/l$. The currently selected number of the variables (s) is set to 0.

(c) Based on p, the algorithm randomly selects a part of the variables (= N *p). s is increased by p, and the current repeating number (r) is set to 0.

(d) For the selected part of the variables, i.e., weight factors and bias values, the algorithm changes their values by the amounts of randomly picked delta-values from $-Y$ to $+Y$. If a bias value is changed to be negative, i.e., less than 0, it changes the value to be positive. r is then increased by 1.

(e) The algorithm calculates the output values of the current ANN for all training set data. The average training error $< E >$ for all training set data is calculated as follows:

$$< E >= \frac{1}{T}\sum_{k=1}^{T} E^k, \quad E^k = \sum_{j=1}^{m} \left(\hat{y}_j^k - y_j^k\right)^2 \tag{5}$$

where T denotes the number of training set data; m denotes the number of output nodes of an ANN; and y_j^k and $\hat{y}_j^k$ indicate the output value and the corresponding target value of an output node j for given training set data k, respectively. If the average training error $< E >$ is decreased compared with that in the previously accepted ANN $< E >_{\min}$, the current values of all the variables of the current ANN are accepted, i.e., $< E >_{\min} =< E >$, and step (f) is then followed. Otherwise, if r is less than the maximum repeating number (R), then step (d) is followed. If r reaches R, then step (f) is followed.

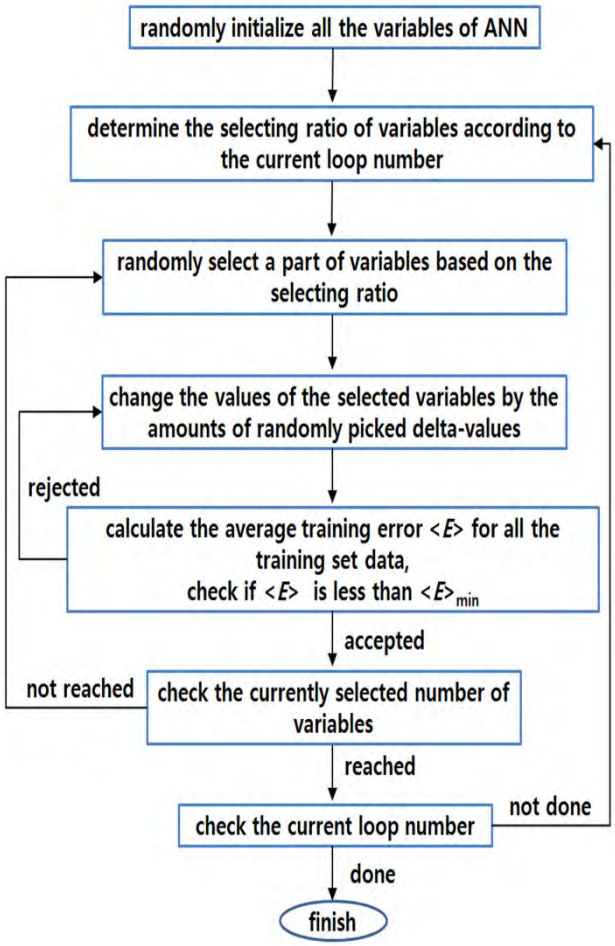

FIGURE 3. Architecture of the algorithm.

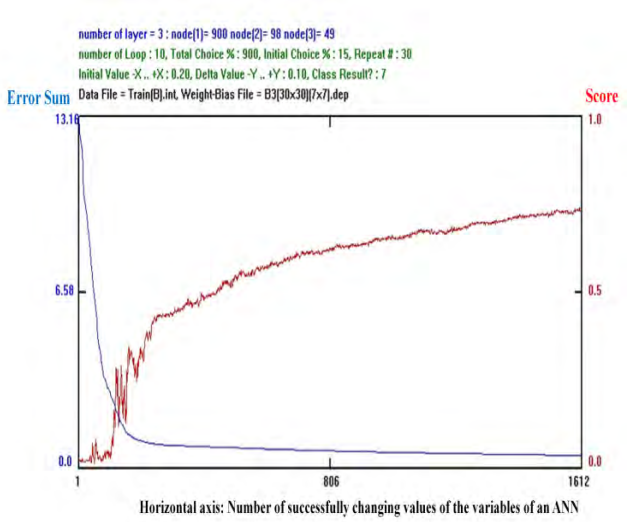

FIGURE 4. Computer screen for the training progress of an artificial neural network (ANN). The blue curve denotes the "error sum," which is the average value of training errors for all the training set data. The red curve denotes the "score," which is the average value of the arbitrary scores for all the training set data.

(f) If s is less than the total selecting range of the variables (S), then step (c) is followed. Otherwise, l is increased by 1. If l is not larger than the total number of loops (L), then step (b) is followed. If l is greater than L, then the algorithm finishes the training.

Figure 4 shows a computer screen for an ANN training progress. The blue curve denotes the "error sum," which is the average value of training errors for all the training set data $< E >$ in (5). Thus, the "error sum" $< E >_{\min}$ is the absolute criterion of training in progress. The horizontal value in Figure 4 indicates the number of successfully changing values of the variables since the average training error $< E >$ is less than that in the previously accepted ANN $< E >_{\min}$.

The red curve denotes the "score," which is the average value of the arbitrary scores for all the training set data. For convenience, x was defined as the position of the output node with the maximum output value (Figure 2c, orange box), and y was described as the position of the cell with the maximum target value (Figure 2c′, red box). The arbitrary score for given training data is 1 if x corresponds with y, 0.33 if x falls on one of the positions of the eight neighboring cells of y, and 0 for other cases. As shown in Figure 2c′, the position of the cell with the maximum target value among 49 cells is considered to be the accurate position of the glottis and given an arbitrary score of 1.0, and one of the positions of the eight neighboring cells of the accurate position is considered to be the adjacent position of the glottis and given an arbitrary score of 0.33. The "score" is a subsidiary reference value for ANN model training and testing.

The abovementioned deep-learning algorithm, tentatively named the "Kim-Monte Carlo algorithm," was developed by the first author (J. S. Kim). All computer software programs for image processing and ANN were programmed in Visual C++ language and executed on Windows PCs.

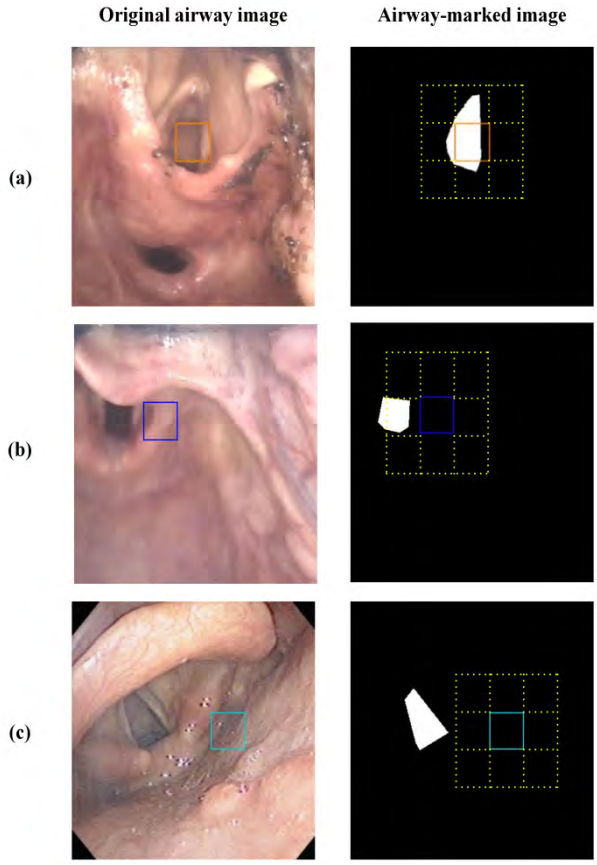

FIGURE 5. Pairs of original airway image and airway-marked image classified by the location of the square predicted by the artificial neural network. (a) Accurate prediction: The predicted location (orange square) overlapped with the glottis, marked in white. (b) Adjacent prediction: One of the eight adjacent equal-sized squares (yellow dotted) of the predicted location (blue square) overlapped with the glottis, marked in white. (c) Inaccurate prediction: The predicted location (cyan square) was not near the glottis.

C. EVALUATION CRITERIA

The training accuracy was obtained for the 1,000 training set data, and the test accuracy was obtained for the 200 test set data. The position of the output node with the maximum output value, i.e., the predicted location of the glottis is indicated by a colored square in Figure 5. The predicted location overlapping with the glottis, marked in white, was considered to be "accurate prediction" (Figure 5a). "Adjacent prediction" was defined as one of the eight adjacent equal-sized squares of the predicted location overlapping with the glottis, marked in white (Figure 5b). The size of the square used to predict the glottic location was arbitrary; therefore, "adjacent prediction" could be helpful in intubation as well as "accurate prediction." "Inaccurate prediction" was defined in cases aside from those of "accurate prediction" and "adjacent prediction"; it was considered to occur when the predicted location by the ANN is not near the glottis, marked in white (Figure 5c).

III. RESULTS

A. DATA SET

This study was performed using laryngeal images retrospectively obtained at a teaching university hospital in Seoul, Korea. The images were captured using GlideScope®(a video laryngoscope) and fiberoptic laryngoscopy. By the GlideScope®, images were captured during intubation in an emergency room from September 2017 to May 2018. By the fiberoptic laryngoscopy, images were obtained by an otolaryngologist from January 2017 to January 2018. This study was approved by the institutional review board (IRB) of Hanyang University Hospital (Seoul, Republic of Korea) with the need for informed consent waived (IRB No. HYUH 2018-08-018-002).

ImageJ software (from National Institutes of Health, Maryland, USA) was used to mark the glottis in the airway images where tracheal intubation should be performed in white and the remaining areas in black (Figure 2a′). Two emergency physicians (Y. Cho and T. H. Lim) with experience of performing more than 100 endotracheal intubation procedures marked the glottis location after agreement for each case. 1,200 pairs of images (see Figure 2a and a′) were obtained, of which 1,000 pairs were randomly sampled as the training set, and the remaining 200 pairs were used as the test set.

B. SUBGROUP ANALYSIS

In the subgroup analysis, the airway images were divided into the ''good view'' and the ''bad view'' groups. The good view images included those in which the glottis was not covered by foreign substances such as secretion or by the epiglottis (Cormack-Lehane grade 1 or 2a) (Figure 6a-e). The bad view images included those in which (1) the vocal cord could not be distinguished because it was covered by secretion or fog (Figure 6f and g), (2) the glottis was covered by the epiglottis (Cormack-Lehane grade 2b, 3, or 4) (Figure 6h), and (3) the vocal cord was difficult to distinguish because it was dark or out of focus (Figure 6i and j). The location where the glottis was thought to be was marked in the bad view images by referring to the structures such as the epiglottis and arytenoid cartilage, by consensus of the two emergency physicians (Y. Cho and T. H. Lim). Thereafter, the training and test accuracies were also obtained for the above two groups.

C. TRAINING RESULTS FOR THE ANN MODELS

Table 1 presents the ''score'' and ''error sum'' of the ANN models with 100×100, 70×70, 50×50, 45×45, 40×40, 35×35, 30×30, and 25×25 pixels; the last row indicates the average values of the ''score'' and ''error sum'' for a given resolution. For instance, 10000-196-98-98-49 ANN model in Table 1 denotes the ANN structure including 10,000 input nodes (for the resolution of 100×100 pixels), 196 intermediate nodes in the first hidden layer, 98 intermediate nodes in the second hidden layer, 98 intermediate nodes in the third hidden layer, and 49 output nodes.

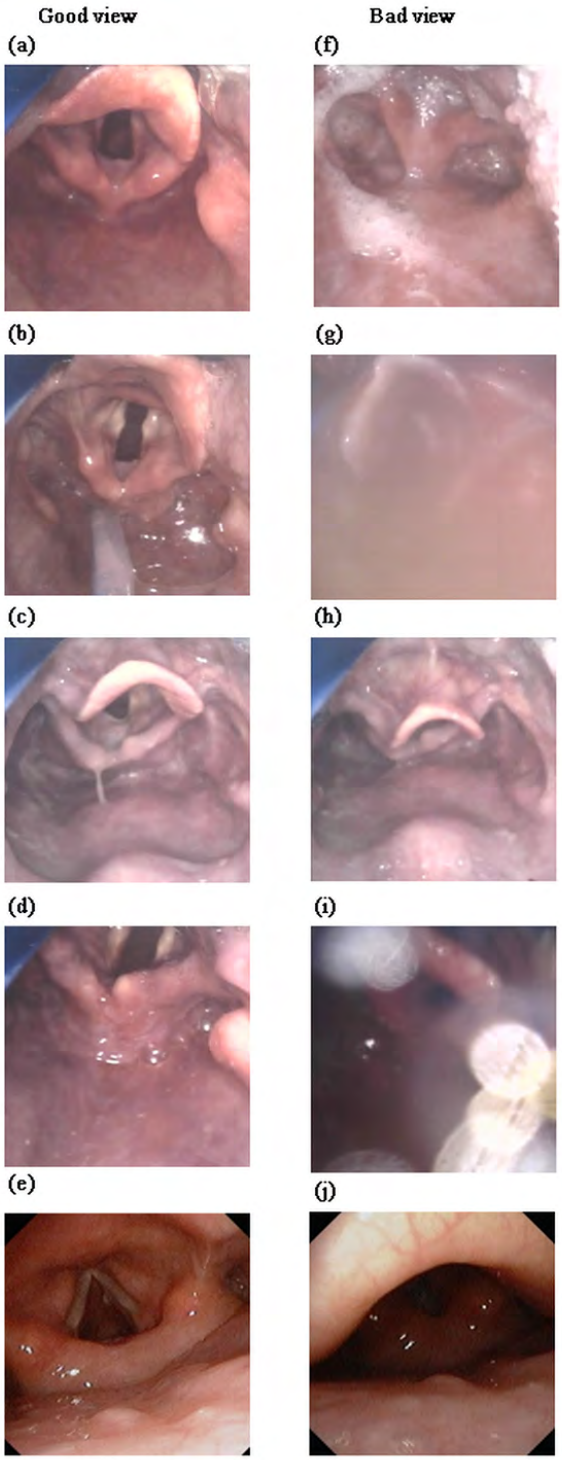

FIGURE 6. Two types of laryngeal images by set data by visualization of the glottis: Good view a (a-e) and bad view b (f-j). Good view: The glottis is not covered by foreign substances such as secretion or by the epiglottis (Cormack-Lehane grade 1 or 2a). Bad view: (1) the vocal cord cannot be distinguished because it is dark or covered by secretion; (2) the glottis is covered by the epiglottis (Cormack-Lehane 2b, 3, or 4); and (3) the vocal cord is difficult to distinguish because it is dark or out of focus.

As indicated in Table 1, the average ''error sum'' decreases as the resolution is reduced from 100×100 pixels to 30×30 pixels. At 25×25 pixels, the average ''error sum''

TABLE 1. Training results of the artificial neural network (ANN) models according to the resolution of the original airway images. (A) 100 x 100, (B) 70 x 70, (C) 50 x 50, (D) 45 x 45, (E) 40 x 40, (F) 35 x 35, (G) 30 x 30, (H) 25 x 25 Pixels, which correspond to the number of input nodes of the ANN models.

ANN models	Score [a]	Error sum [b]	ANN models	Score [a]	Error sum [b]
(A) 100×100 pixels			(B) 70×70 pixels		
10000-98-49	0.616	0.684	4900-98-49	0.647	0.641
10000-196-49	0.580	0.704	4900-196-49	0.646	0.634
10000-98-98-49	0.624	0.655	4900-98-98-49	0.645	0.641
10000-196-98-49	0.622	0.689	4900-196-98-49	0.642	0.635
10000-98-98-98-49	0.605	0.699	4900-98-98-98-49	0.645	0.668
10000-196-98-98-49	0.608	0.691	4900-196-98-98-49	0.638	0.667
Average	0.609	0.687	Average	0.644	0.648
(C) 50×50 pixels			(D) 45×45 pixels		
2500-98-49	0.658	0.625	2025-98-49	0.666	0.605
2500-196-49	0.635	0.636	2025-196-49	0.649	0.626
2500-49-49	0.663	0.610	2025-49-49	0.663	0.611
2500-63-49	0.675	0.599	2025-63-49	0.670	0.604
2500-98-98-49	0.661	0.635	2025-98-98-49	0.654	0.619
2500-196-98-49	0.647	0.642	2025-196-98-49	0.656	0.639
2500-49-49-49	0.661	0.623	2025-49-49-49	0.634	0.622
2500-98-49-49	0.649	0.629	2025-98-49-49	0.670	0.607
2500-98-98-98-49	0.652	0.648	2025-98-98-98-49	0.635	0.648
2500-196-98-98-49	0.644	0.654	2025-196-98-98-49	0.669	0.637
2500-49-49-49-49	0.644	0.632	2025-49-49-49-49	0.627	0.650
2500-98-49-49-49	0.645	0.639	2025-98-49-49-49	0.658	0.618
Average	0.653	0.631	Average	0.654	0.624
(E) 40×40 pixels			(F) 35×35 pixels		
1600-98-49	0.687	0.596	1225-98-49	0.684	0.579
1600-196-49	0.676	0.610	1225-196-49	0.666	0.615
1600-49-49	0.679	0.601	1225-49-49	0.664	0.610
1600-63-49	0.656	0.602	1225-63-49	0.679	0.593
1600-98-98-49	0.674	0.598	1225-98-98-49	0.678	0.609
1600-196-98-49	0.668	0.618	1225-196-98-49	0.684	0.617
1600-49-49-49	0.661	0.619	1225-49-49-49	0.672	0.607
1600-98-49-49	0.677	0.595	1225-98-49-49	0.660	0.609
1600-98-98-98-49	0.649	0.621	1225-98-98-98-49	0.653	0.629
1600-196-98-98-49	0.651	0.643	1225-196-98-98-49	0.647	0.638
1600-49-49-49-49	0.645	0.618	1225-49-49-49-49	0.638	0.647
1600-98-49-49-49	0.664	0.617	1225-98-49-49-49	0.658	0.606
Average	0.666	0.612	Average	0.665	0.613
(G) 30×30 pixels			(H) 25×25 pixels		
900-98-49	*0.684*	*0.579*	625-98-49	0.666	0.607
900-196-49	0.674	0.586	625-196-49	0.665	0.600
900-49-49	0.670	0.594	625-49-49	0.655	0.616
900-63-49	0.678	0.579	625-63-49	0.642	0.620
900-98-98-49	0.653	0.603	625-98-98-49	0.655	0.621
900-196-98-49	0.653	0.616	625-196-98-49	0.638	0.626
900-49-49-49	0.677	0.588	625-49-49-49	0.634	0.629
900-98-49-49	0.657	0.589	625-98-49-49	0.621	0.624
900-98-98-98-49	0.661	0.614	625-98-98-98-49	0.620	0.639
900-196-98-98-49	0.628	0.644	625-196-98-98-49	0.640	0.644
900-49-49-49-49	0.662	0.598	625-49-49-49-49	0.619	0.641
900-98-49-49-49	0.664	0.592	625-98-49-49-49	0.656	0.628
Average	0.663	0.599	Average	0.643	0.625

[a]The "score" is the average value of the arbitrary scores for all the training set data. The arbitrary score is 1 if the position of the output node with the maximum output value corresponds to the position of the cell with the maximum target value, 0.33 if the position of the output node with the maximum output value falls on one of the positions of the eight neighbouring cells of the position of the cell with the maximum target value, and 0 for other cases.

[b]The "error sum" is the average value of training errors for all the training set data. The training error is the summation over all the output nodes of the square of the difference between the output value of an output node and the corresponding target value for the given training data.

abruptly increases. Thus, we selected the models with 30×30 pixels because they had the lowest average "error sum" and thereby the highest learning rate. The 900-98-49 model and the 900-63-49 model had the lowest "error sum" (0.579) among the models with 30×30 pixels. We ultimately selected the 900-98-49 model as the prediction model for the location of the glottis because its "score" was greater than that of the 900-63-49 model.

D. PREDICTION ACCURACY OF THE SELECTED MODEL

Table 2 reports the training and test accuracies for predicting the glottic location in the airway images by the 900-98-49 model that was selected as the prediction model. For the 1,000 training set data, the accurate prediction rate was 78.1%, the adjacent prediction rate was 17.3%, and the inaccurate prediction rate was 4.6%. For the 200 test set data, the accurate prediction rate was 74.5%, the adjacent

TABLE 2. Prediction accuracy for the glottic location for the highest learning rate model (900-98-49 model).

Type of dataset (number of images)	Prediction (%), (number of images)		
	Accurate[a]	Adjacent[b]	Inaccurate[c]
Training set (1,000)	78.1% (781)	17.3% (173)	4.6% (46)
Test set (200)	74.5% (149)	21.5% (43)	4.0% (8)

[a]"Accurate prediction": the predicted location overlapped with the glottis.
[b]"Adjacent prediction": one of the eight adjacent equal-sized squares of the predicted location overlapped with the glottis.
[c]"Inaccurate prediction": the predicted location was not near the glottis.

TABLE 3. Prediction accuracy for the glottic location for the glottic location for the highest learning rate model (900-98-49 model) by the type of laryngeal images.

Type of images (No. of images)	Type of dataset (No. of images)	Prediction (%), (number of images)		
		Accurate[a]	Adjacent[b]	Inaccurate[c]
Good view (981)	Training set (813)	83.3% (677)	14.3% (116)	2.4% (20)
	Test set (168)	78.6% (132)	19.0% (32)	2.4% (4)
Bad view (219)	Training set (187)	55.6% (104)	30.5% (57)	13.9% (26)
	Test set (32)	53.1% (17)	34.4% (11)	12.5% (4)

[a]"Accurate prediction": the predicted location overlapped with the glottis.
[b]"Adjacent prediction": one of the eight adjacent equal-sized squares of the predicted location overlapped with the glottis.
[c]"Inaccurate prediction": the predicted location was not near the glottis.

prediction rate was 21.5%, and the inaccurate prediction rate was 4.0%.

E. PREDICTION ACCURACY FOR THE SELECTED MODEL BY THE TYPE OF VIEW

The training and test accuracies for predicting the glottic location in the airway images by the selected 900-98-49 model were obtained by dividing the training set and the test set into the two types of views: good views and bad views as reported in Table 3.

For the training set, the number of good view images was 813 (81.3%) and that of the bad view images was 187 (18.7%). For the test set, the number of good view images was 168 (84.0%) and that of the bad view images was 32 (16.0%). For the good view images, the accurate prediction rates were 83.3% for the training set and 78.6% for the test set as reported in Table 3. For the bad view images, the accurate prediction rates were 55.6% for the training set and 53.1% for the test set.

IV. DISCUSSION

The novel deep-learning algorithm for ANNs, tentatively named the "Kim-Monte Carlo algorithm," described in this paper is not fundamentally restricted to apply to an ANN including hundreds or more hidden layers as described in the detailed procedure of the algorithm in the "methods" section. However, to extract the abstract representation of the glottic location in airway images, ANN models including only 1, 2, or 3 hidden layers were more than sufficient as indicated in Table 1, where the average training error, 'error sum' of an ANN model does not significantly decrease as the number of hidden layers increases from 1 to 3 for all the resolutions.

The novel algorithm does not need to set heuristic parameters and apply an approximate approach in the training process, as does the back-propagation method [17]–[19], which has been the most commonly used method to date for training ANNs [21]. For the back-propagation method, each weight factor w_{ij} is adjusted by calculating the delta value Δw_{ij} using all or a part of the learning data, which is carried out individually for all the weight factors in the backward direction, repeatedly until the training session is finished as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \tag{6}$$

$$\Delta w_{ij}(t) = \alpha \sum_{k=1}^{T} \left(\hat{y}_j^k - y_j^k\right) \mathcal{A}'(x_j) y_i \tag{7}$$

where each node j in a given layer receives an input y_i from a node i in the previous layer; w_{ij} indicates the weight factor between nodes i and j; x_j denotes the summation over all nodes in the previous layer as indicated in (4); α denotes a given learning rate; $\mathcal{A}'$ indicates the first derivation of the activation function (2) or (3); T denotes the number of training set data; and y_j^k and $\hat{y}_j^k$ indicate the output value and the corresponding target value of a node j for given training set data k, respectively. Therefore, the back-propagation method using (6) and (7) takes massive computing resources for training an ANN even with not a large amount of learning data.

For the novel algorithm, randomly selected weight factors w_{ij} and bias values b_j are adjusted by the amounts of randomly picked delta-values within a given range, not by calculating the delta value Δw_{ij} using (7). Consequently, the algorithm is intuitively understandable, simple, and efficient. The advantage of such an efficient deep-learning algorithm is that it can consistently complete trainings of a much greater number of ANN models in a given time and with a given limit of computing resources. Accordingly, it is a favorable approach for finding a better predictive ANN model.

In this study, we sought the prediction model for the location of the glottis in airway images obtained using a video airway device during clinical practice. With 1,000 training set data randomly selected from 1,200 airway images, the training processes to predict the location of the glottis were conducted using 84 ANN models. Among them, the model with the highest learning rate yielded 78.1% accurate prediction and 17.3% adjacent prediction for the training set. As the ANN model was applied to 200 test set data, we obtained 74.5% accurate prediction and 21.5% adjacent prediction. The inaccurate prediction rates were similar for the training

set (4.6%) and for the test set (4.0%). Good training and test results were obtained without overfitting or overtraining. These results indicate that the sample size of 1,000 airway images was sufficient as the training set to predict the location of the glottis

We also compared the training results of the airway images of various resolutions. As the resolution was reduced, the learning rate increased and reached the highest learning rate at 30 × 30 pixels. Then, the learning rate decreased at a lower resolution, i.e., 25×25 pixels. Thus, rather than increasing the resolution of input images for ANNs, reducing the resolution to an appropriate level could yield better training and test results.

This ANN model was developed to help clinicians perform more accurate intubation by presenting the predicted location of the glottis. This approach will improve the quality of airway management, the accuracy and efficiency of intubation, and the self-confidence of clinicians to accurately identify the glottis. It is expected that this ANN model will reduce the incidence of complications and medical costs for malpractice.

Very few studies regarding automated detection of the glottic opening using artificial intelligence have been published. Carlson *et al.* [14] predicted the presence or absence of the glottic opening using k-nearest neighbor, support vector machines, decision trees, and neural networks. In their study, videos in which manikins were intubated were subdivided by 1-second unit images. The disadvantage is that manikins were used instead of actual patient airways. In their classification study, the test accuracy rate was 74–81% for predicting the presence or absence of the glottic opening.

There are some limitations of this study. First, the images used were obtained from a single center. Since all the subjects were Asian, validation is required to apply the same method to other sites and ethnic groups. Second, the accurate prediction rate for the good view images was 78.6% in the test set; however, that for the bad view images was lower (53.1%). The reason may be the low number of bad view images. Additional research involving more images is needed in the future. Third, this study included only laryngeal images of patients with normal anatomy, and there were no images of patients with cancer or congenital anomalies. In particular, no laryngeal image indicated a history of surgery or radiation therapy and an altered anatomical structure. The predictive power of the ANN in images of an abnormal larynx, in which tracheal intubation is difficult to perform, has not been studied; therefore, further studies are needed.

V. CONCLUSIONS

We sought the prediction model for the location of the glottis in airway images obtained during clinical practice by using 84 ANN models trained by the novel deep-learning algorithm described in this paper. As the highest learning rate model was applied to the test set, the accurate prediction rate was 74.5%, and the adjacent prediction rate was 21.5%. In addition, reducing the resolution of input images for ANNs to an appropriate level could yield better training and test results. This ANN model could help clinicians perform more accurate intubation by presenting the predicted location of the glottis.

ACKNOWLEDGMENT

(Jong Soo Kim and Yongil Cho are co-first authors.)

REFERENCES

[1] C. Kabrhel, T. W. Thomsen, G. S. Setnik, and R. M. Walls, "Videos in clinical medicine. Orotracheal intubation," *New England J. Med.*, vol. 356, no. 17, p. e15, Apr. 2007. doi: 10.1056/NEJMvcm063574.

[2] Y. Ono, T. Kakamu, H. Kikuchi, Y. Mori, Y. Watanabe, and K. Shinohara, "Expert-performed endotracheal intubation-related complications in trauma patients: Incidence, possible risk factors, and outcomes in the prehospital setting and emergency department," *Emergency Med. Int.*, vol. 2018, Jun. 2018, Art. no. 5649476. doi: 10.1155/2018/5649476.

[3] J. H. Jones, M. P. Murphy, R. L. Dickson, G. G. Somerville, and E. J. Brizendine, "Emergency physician–verified out-of-hospital intubation: Miss rates by paramedics," *Acad. Emergency Med.*, vol. 11, no. 6, pp. 707–709, Jun. 2004. doi: 10.1197/j.aem.2003.12.026.

[4] S. H. Katz and J. L. Falk, "Misplaced endotracheal tubes by paramedics in an urban emergency medical services system," *Ann. Emergency Med.*, vol. 37, no. 1, pp. 32–37, Jan. 2001. doi: 10.1067/mem.2001.112098.

[5] A. De Jong, N. Molinari, M. Conseil, Y. Coisel, Y. Pouzeratte, F. Belafia, B. Jung, G. Chanques, and S. Jaber, "Video laryngoscopy versus direct laryngoscopy for orotracheal intubation in the intensive care unit: A systematic review and meta-analysis," *Intensive Care Med.*, vol. 40, no. 5, pp. 629–639, May 2014. doi: 10.1007/s00134-014-3236-5.

[6] H.-B. Huang, J.-M. Peng, B. Xu, G.-Y. Liu, and B. Du, "Video laryngoscopy for endotracheal intubation of critically ill adults: A systemic review and meta-analysis," *Chest*, vol. 152, no. 3, pp. 510–517, Sep. 2017. doi: 10.1016/j.chest.2017.06.012.

[7] J. Jiang, D. Ma, B. Li, Y. Yue, and F. Xue, "Video laryngoscopy does not improve the intubation outcomes in emergency and critical patients—A systematic review and meta-analysis of randomized controlled trials," *Crit. Care*, vol. 21, no. 1, p. 288, Nov. 2017. doi: 10.1186/s13054-017-1885-9.

[8] T. C. Mort, "Esophageal intubation with indirect clinical tests during emergency tracheal intubation: A report on patient morbidity," *J. Clin. Anesthesia*, vol. 17, no. 4, pp. 255–262, Jun. 2005. doi: 10.1016/j.jclinane.2005.02.004.

[9] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016. doi: 10.1001/jama.2016.17216.

[10] D. S. W. Ting, C. Y. L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, and E. Y. M. Wong, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017. doi: 10.1001/jama.2017.18152.

[11] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, and O. Geessink, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017. doi: 10.1001/jama.2017.14585.

[12] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017. doi: 10.1148/radiol.2017162326.

[13] H. Takiyama, T. Ozawa, S. Ishihara, M. Fujishiro, S. Shichijo, S. Nomura, M. Miura, and T. Tada, "Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks," *Sci. Rep.*, vol. 8, no. 1, May 2018, Art. no. 7497. doi: 10.1038/s41598-018-25842-6.

[14] J. N. Carlson, S. Das, F. De la Torre, A. Frisch, F. X. Guyette, J. K. Hodgins, and D. M. Yealy, "A novel artificial intelligence system for endotracheal intubation," *Prehospital Emergency Care*, vol. 20, no. 5, pp. 667–671, Sep./Oct. 2016. doi: 10.3109/10903127.2016.1139220.

[15] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943. doi: 10.1007/BF02478259.

[16] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958. doi: 10.1037/h0042519.
[17] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard Univ., Cambridge, MA, USA, 1974.
[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA, USA: MIT Press, 1986.
[19] S. Sathyanarayana, "A gentle introduction to backpropagation," *Numeric Insight*, vol. 7, pp. 1–15, Jul. 2014.
[20] T. Lin, B. G. Horne, P. Tiňo, and C. L. Giles, "Learning long-term dependencies in NARX recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1329–1338, Nov. 1996. doi: 10.1109/72.548162.
[21] F.-F. Li, A. Karpathy, and J. Johnson, *Lecture: Software Packages Caffe/Torch/Theano/TensorFlow*. Stanford, CA, USA: Stanford Univ., 2016.

JONG SOO KIM received the B.S. degree in polymer engineering from Hanyang University, Seoul, South Korea, in 1980, and the M.S. and Ph.D. degrees in physical chemistry on theoretical study of polymer solution from the Korea Advanced Institute of Science and Technology, Seoul, in 1982 and 1985, respectively.

He was a Software Engineer with DACOM, Inc., Seoul, from 1985 to 1988, a Postdoctoral Fellow in theoretical biophysics with the University of Pittsburgh, Pennsylvania, USA, from 1988 to 1989, a Visiting Fellow with the Division of Computer Research and Technology, National Institutes of Health, MD, USA, from 1989 to 1992, and the President and a Software Developer with Real Image Corporation, Seoul, from 1999 to 2013. He is currently a Professor with the Institute for Software Convergence, Hanyang University, Seoul. His research interests include deep-learning algorithm for artificial neural networks, computer-aided clinical diagnosis, classification, and localization on medical images and signals, molecular graphics, computer graphics programming to convert 2D to 3D images, and statistical mechanics using molecular dynamics and Monte Carlo simulation.

YONGIL CHO received the B.S. (M.D.) degree in medicine from Hanyang University, Seoul, South Korea, in 2009, and the M.S. degree in medicine from Jeju National University, Jeju, South Korea, in 2017. He received the emergency medicine residency training from Hanyang University Hospital, Seoul, South Korea, from 2010 to 2013.

From 2017 to 2019, he was a Fellow of Emergency Medicine with Hanyang University Hospital, Seoul, where he is currently a Clinical Assistant Professor. His research interests include big data analysis, emergency medicine, airway management, and deep learning.

TAE HO LIM received the B.S. (M.D.), M.S., and Ph.D. degrees in medicine from Hanyang University, Seoul, South Korea, in 1991, 1999, and 2003, respectively. He received the general surgery residency training from Hanyang University Hospital, Seoul, from 1994 to 1997, and an Emergency Medicine Residency Training from the Gangnam Severance Hospital, Seoul, from 1999 to 2000.

He was a full-time Instructor with the Department of Emergency Medicine, Hanyang University Hospital, Seoul, from 2001 to 2002. From 2003 to 2007, he was an Assistant Professor of emergency medicine with the Hanyang University College of Medicine, a Visiting Scholar with the Department of Emergency Medicine, State University of New York-Stony Brook, New York, USA, from 2008 to 2009. He was an Associate Professor of emergency medicine, from 2008 to 2012 and has been a Professor of emergency medicine, since 2013. He has been the Head of the Department of Emergency Medicine, College of Medicine, Hanyang University, since 2004.

Since 2015, he has been the Director of the Convergence Technology Center for Disaster Preparedness, Hanyang University. His research interests include emergency and critical care, artificial intelligence in medicine and biomedical engineering, especially for emergency medical devices.