OpenPose based Smoking Gesture Recognition System using Artificial Neural Network

Tae-Yeong Jeong, Il-Kyu Ha*

Abstract: Smoking is an extremely important health problem in modern society. This study focuses on a method for preventing smoking in non-smoking areas, such as public places, as well as the development of an artificial neural network based smoking motion recognition system for more accurately recognizing smokers in such areas. In particular, we attempted to increase the rate of recognition of smoking behaviors using an OpenPose based algorithm and the accuracy of such recognition by additionally applying a hardware device for recognizing cigarette smoke. In addition, a preprocessing method for inputting a dataset into the proposed system is proposed. To improve the recognition performance, four types of dataset models were created, and the most suitable dataset model was selected experimentally. Based on this dataset model, test data were created and input into the proposed neural network based smoking behavior recognition system. In addition, the nearest neighbor interpolation method was selected experimentally as an image interpolation approach and applied to the image preprocessing. When applying experimental data based on learned data, the developed system showed a recognition rate of 70-75%, and the smoking recognition accuracy was increased through the addition of the hardware device.

Keywords: artificial neural network; OpenPose algorithm; smoking gesture recognition system; smoking recognition sensor

1 INTRODUCTION

Smoking is an important issue in modern society and is directly related to health and the environment. In the past, people smoked everywhere; however, as the awareness of secondhand smoke has increased, the perception of smoking and its health effects has changed, and negative views on smoking have increased, resulting in an increase non-smoking areas. However, smoking in public places, such as stations, schools, and downtown areas where many people gather, is still common despite laws prohibiting such activity, causing various problems including environmental pollution and second-hand smoke [1]. In the Republic of Korea, according to Article 9 of the National Health Promotion Act, a non-smoking area within 10 m of the outer wall of a public building is designated as a non-smoking area [2]. However, indiscriminate smoking frequently occurs near buildings that are not designated as smoking areas. In addition, although smoking areas are used in many locations to prevent smoking in non-smoking areas, smokers often do not use these areas for various reasons. For example, in many cases, the smoking areas are not fully functional [3-5]. Therefore, this study focused on a method for preventing smoking in non-smoking areas. We developed an artificial neural network-based smoking behavior recognition system. In particular, we used the OpenPose-based algorithm to design our recognition model. Several characteristic tasks, such as preprocessing of training data, selection of efficient training datasets, and interpolation of test data, were performed to improve the performance of the recognition module. A hardware device was developed for judging smoking behavior by recognizing cigarette smoke, to guarantee the recognition of smoking behavior and improve its accuracy.

2 RELATED WORK
2.1 Analysis of Existing Studies

Tab. 1 list the characteristics of existing studies on human motion recognition and major studies on the OpenPose based skeleton model. In [6], a skeleton model is used with the part affinity fields (PAF) method, which connects points placed on the joints of the body, and captures the posture of an individual with a relatively high accuracy. In [7], although a partitioning method is proposed for recognizing multiple people in an image using a deepcut structure; it has a limitation in that it takes a long time to recognize a single image. In [8], an AI-based recognition technique is proposed for recognizing human hands as objects; however, it has difficulty estimating various human postures. The approach in [9] has the advantage of directly learning the inference process, unlike the general object recognition methods from previous studies, but has difficulty detecting complex or rare postures and obtaining datasets for rare postures for artificial intelligence learning. In [10], the spatio-temporal affinity fields (STAF) method is used to strengthen the existing PAF skeleton model method. However, when a large number of objects appear, the computational speed increases linearly, and the objects are not properly recognized when the camera framerate is high. In [11], a regional multi-person pose estimation (RMPE) framework method is proposed for estimating multiple human objects. Although RMPE more accurately identifies the existence of multiple objects, it has difficulty estimating detailed postures. In [12], a method is proposed for generating a 3D human object skeleton to estimate the motion of human objects in sporting events. The method focuses on image processing using pipeline structures, rather than machine learning.

In [13-17], major studies on OpenPose based skeleton models are described. In [13], an OpenPose based skeleton model is used to determine the key point, i.e., the center of movement in the human body, and the characteristics of the falling behavior are investigated. This is a study on the detection of unusual human behaviors. In [14], the authors pointed out the problem of using markers as a method for estimating human motion and proposed a method for estimating the positions and postures in 3D without markers when applying OpenPose. In [15], also using OpenPose, a method is proposed for estimating a falling posture by detecting the movement of the center of the joint of a person in 2D in real time. The motion posture of a person is

determined based on 25 joint points. In [16], an OpenPose skeleton based basketball free-throw posture analysis model is proposed and the accuracy of free-throw posture prediction is analyzed. In [17], a method for predicting the 3D skeleton of a human hand using OpenPose is proposed, allowing the robot to recognize the hand movements of the human operator. As described above, there are various methods and models for estimating the posture of human objects. Based on the OpenPose based skeleton model [6], recently developed at Carnegie Mellon University, we focused on the development of a smoking behavior prediction model.

Table 1 Comparison of existing studies on motion recognition

Studies	Characteristic
[6]	Introduces filter learning that can encode not only the location of body parts but also the connection relationships between body parts, and encodes expressions (part affinity fields) including location and direction information into 2D vectors.
[7]	Devised a method to conduct calculations more efficiently by changing the structure of Deepcut; however, several to several hundred seconds may be required depending on the image.
[8]	A technique is presented for separating the hands by mixing color and depth images, and by learning the skin model in real time, the skin model is adaptively updated for the lighting environment, thereby resulting in a stronger recognition performance.
[9]	To solve the difficulty of taking complex poses with the existing graphic model, a method of directly learning the inference process is used over a graphical pose estimation method.
[10]	Based on the PAF(Part Affinity Fields) method, the TAF(Temporal Affinity Fields) method is applied in the existing online tracking approach. PAF shows only those parts constituting the skeleton while connecting key points, whereas TAF strengthens the connectivity by connecting the surrounding key points, similar to a graph.
[11]	A regional multi-person pose estimation (RMPE) framework is proposed as a method of estimating multiple objects of a person. Although it achieves a more accurate identification of multiple objects, there is a lack of detail in their posture.
[12]	A method for generating a 3D human object skeleton in estimating the motion of human objects in sports events is proposed. It focuses on image processing using pipelines rather than learning.
[13]	A study is conducted on determining the key points in the human body and investigating the characteristics of falling behavior using an OpenPose based skeleton model.
[16]	An OpenPose skeleton based basketball free throw posture analysis model is proposed and the accuracy of the free throw posture prediction is analyzed.

2.2 OpenPose based Skeleton Model

OpenPose [6] is a method of human pose estimation that predicts the body, face, and knuckles of an individual using only a single camera. Early research began with a method published by Carnegie Mellon University in 2017. Based on a deep learning convolutional neural network, it estimates the feature points of the body, hand, and face from only images and photographs. With a traditional method, it is necessary to attach related devices to the human body, which is expensive and limited in terms of space, and achieves varying results depending on the body type of the person. With the development of knowledge and computing performance, deep learning technology has been used in the field of

computer vision [21]. By combining these technologies, it has become possible to predict human postures using only images. Fig. 1 show the human posture prediction process based on OpenPose.
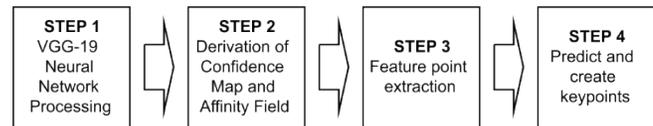

Figure 1 Prediction process of human posture based on OpenPose

The human posture prediction process based on OpenPose is as follows: The first stage is the VGG-19 [18] neural network processing stage. In this step, a batch-type input image is input into the VGG-19 neural network. After the input image passes through the neural network, it becomes image data with a feature size of $28 \times 28 \times 512$. These data become the input data for step 2. In step 2, a confidence map and affinity field are derived. A confidence map is used to determine the joint position of a person in an image, and the preference field is used to determine who owns the joint extracted from the image. Step 3 involves extracting the feature points. At the beginning of step 3, although meaningless features are created according to the input image, these features are compared with the posture of the individual, optimization is performed, and the features gradually converge in the direction pointing toward the joint position of the person. Subsequently, they continuously pass through the next branch and predict a 2D preference field based on the similarity between each part of the human body. In Step 4, a keypoint is predicted. In other words, keypoints of individuals in the image are predicted and generated using the confidence map and preference field calculated in the previous step. When each joint from the reliability map is combined, the greedy algorithm (greedy relaxation) determines the owner of each joint if there are several people in the image. By repeating this process, it becomes possible to predict the human posture.

3 DESIGN OF PROPOSED SYSTEM
3.1 Image Data Processing Procedure Used by the Proposed System

The proposed system uses the OpenPose based skeleton algorithm to identify and analyze the smoking behavior of an object from motion images captured by a CCTV camera. Fig. 2 shows the image data processing procedure used by the proposed system.

The image processing procedure of the proposed system is as follows: First, the image data to be used for learning are preprocessed to facilitate data processing, and the preprocessed smoking image is input into the OpenPose skeleton model. The images processed by the skeleton model undergo a learning process. As a result, a smoking behavior recognition model is created. CCTV footage is input into the generated smoking behavior recognition model to determine whether smoking is taking place. To increase the accuracy of the smoking behavior recognition, the result of the smoking behavior recognition model is transmitted to the hardware device and combined with the smoking behavior result of the hardware device, allowing the smoking behavior to be

determined. The hardware device developed in this study can be used to increase the accuracy of the software when recognizing smoking behavior. The device might mistake the presence of residual gas as the occurrence of smoking and is therefore used only as an auxiliary tool.
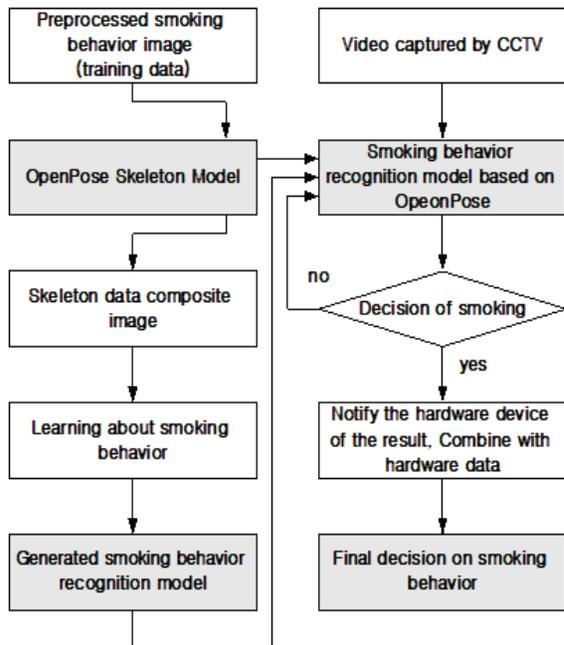

Figure 2 Image data processing procedure used by the proposed system

3.2 Image Data Preprocessing Procedure for Machine Learning

With the proposed system, preprocessing of the learning image data is required to create a smoking behavior recognition model. Fig. 3 shows the preprocessing of the training data. In the training image, the longer horizontal and vertical lengths are adjusted to a standard of 500 mm, and the missing part is filled in with black. The middle image in Fig. 3 illustrates this phenomenon. An OpenPose based skeleton model is applied to this image to create a dataset for learning the smoking behavior. The image on the right side of Fig. 3 shows this phenomenon. Using this image, an artificial intelligence neural network learns the smoking behavior for model creation. This model is combined with the skeleton model to input CCTV images in real time and detect whether a human object is smoking. If the individual is determined to be smoking, the result is sent to the hardware device.
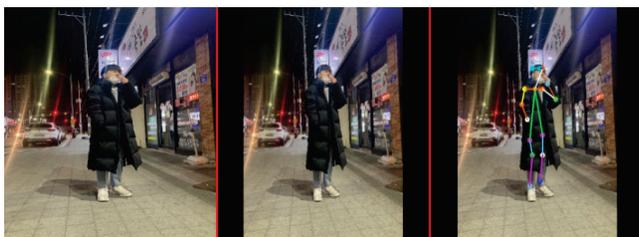
Figure 3 Preprocessing of the training data

3.3 Hardware Development to Increase the Accuracy of Smoking Behavior Recognition

A hardware device is used as an auxiliary tool to increase the accuracy of the smoking image recognition in the proposed system. Fig. 4 shows the systematic structure of the proposed hardware device. It consists of sensors that can detect a variety of smoke produced by cigarettes, a warning sound generation device, and a Raspberry Pi microcontroller. Fig. 5 shows the operational procedure of the hardware device. First, the CCTV footage is input into the above-mentioned smoking behavior recognition OpenPose based skeleton model to determine whether an individual is smoking. If it is determined that smoking behavior is taking place, the information is transmitted to a Raspberry Pi, which controls the hardware sensor. Second, when data are received from the Raspberry Pi, the sensors detecting gases and flames generated from smoking respond. Third, each sensor collects information on the smoking behavior and cigarette smoke. Fourth, when smoking occurs, the sensor values for collecting information on the presence of gases increase, and when the increased sensor values exceed a certain reference value, they are recognized as cigarette smoke. As a final step, when all sensors detect cigarette smoke, it is finally determined whether smoking is occurring, and a warning sound indicating a non-smoking area is output. By calculating an average value through repeated experiments, a standard value for cigarette smoke recognition is achieved.
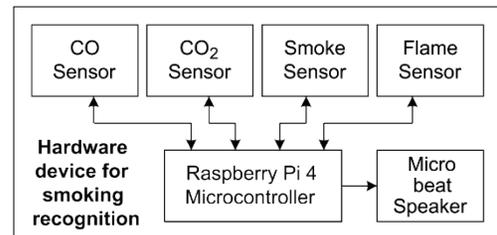

Figure 4 Structure of the proposed hardware device
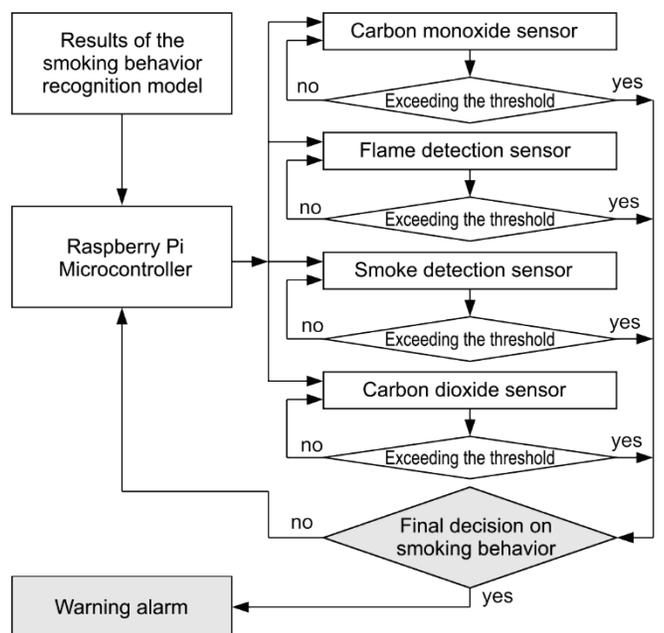
Figure 5 Operating procedure of hardware device

```
1   aug = ImageDataGenerator(
2       rotation_range = 20,
3       zoom_range = 0.5
4       width_shift_range = 0.2,
5       height_shit_Range = 0.2,
6       shear_range = 0.15,
7       horizontal_flip = True,
8       fil_mode = "nearest")
9
10  baseModel = MobileNetVe(weight = "imagenet",
11      include_top = false, input_tensor = Input(shape=(224, 225, 3)))
12
13  headModel = baseModel.output
14  headModel = AveragePooing2D(pool_size=(7, 7))(headModel)
15  headModel = Flatten(name = "flatten")(headModel)
16  headModel = Dense(128, activation = "relu")(headModel)
17  headModel = Dropout(0.5)(headModel)
18  headModel = Dense(2, activation = "softmax")(headModel)
19  model = Model(inputs = baseModel.input, outputs = headModel)
20
21  for layer in baseModel.layer:
22      layer.trainable = False
23  opt = Adam(lr=INIT_LR, decay = INIT_LT / EPOCHS)
24  model.compile(loss="binary_crossentropy", optimizer=opt,
25      metics=["accuracy"])
26
27  H=model.fit(aug.flow(trainX, trainY, batch_size = BS),
28      steps_per_epoch = len(trainX) // BS,
29      validation_data = (testX, testY),
30      validation_steps = len(testX) // BS,
            epochs = EPOCHS)
```

Figure 6 Neural network algorithm for smoking behavior recognition

4 IMPLEMENTATION OF PROPOSED SYSTEM
4.1 Creation of Smoking Recognition Neural Network

In this paper, we describe the process of generating a smoking recognition neural network using the proposed system. Fig. 6 shows the main parts of the proposed smoking recognition neural network algorithm. The implemented smoking recognition artificial intelligence neural network uses the TensorFlow module (ImageDataGenerator) [19], by which the number of training data is increased through a slight transformation when learning the images. To create a large number of smoking data, an image is created and used by rotating, zooming, and moving based on the original image. The detailed parameters of this module are presented in Tab. 2. MobileNetV2 is used as the base model, and is a neural network light enough to be applied in a mobile environment. Because of the small number of parameters applied, the burden of manipulating a neural network can be reduced and the computational cost can be decreased.

Table 2 Image generator parameters

Parameters	Contents
Rotation_range	Random rotation angle within 20 degrees
Zoom_range	15% random zoom range
Width_shift_range	15% left and right movements
Height_shift_range	15% vertical movement
Shear_range	15% floor push strength
Horizontal_flip	Randomly flip horizontally
Fill_mode	When there is a blank space in the image from a reduction in rotation.

Average pooling is applied on the spatial data using the AveragePooling2D module to construct a neural network.

With this module, the parameter (7, 7) indicates that the image is scaled down to 1/7 for a two-dimensional space, and thus an image size of $14 \times 14 \times 3$ is created. The generated image is input into the Relu activation function, negative values are discarded, and positive values remain. Subsequently, it is normalized using the Dropout function to solve the overfitting problem, which is a phenomenon in which a large number of data are learned for only a specific model, and the accuracy of the analysis is significantly lowered for the untrained data. Finally, using the activation function Softmax, the input value is normalized to a value between zero and 1 as an output, and the sum of the output values is always 1. Softmax can be used for binary classification by allowing only the largest value of the result of this function to have a true value and the rest to have false values. Because a compilation is a binary classification of smoking and non-smoking, the model was constructed using the typical loss function NBinary_crossentropy.

The main software packages used for the smoking recognition model are listed in Tab. 3. Python 3.8 is used, and the latest versions of other artificial intelligence libraries such as TensorFlow and Keras were applied.

Table 3 Open software package versions

Libraries	Version	Libraries	Version
Python	3.8	Matplotlib	3.4.3
TensorFlow	2.5.0	TensorFlow-estimator	2.5.0
Keras	2.4.3	Keras-Preprocessing	1.1.2
scikit-learn	0.24.2	scicy	1.6.2
pandas	1.3.3	pip	21.2.2
OpenCv2	4.0.1	TensorBoard	2.5.0

4.2 Training Process of Dataset for Smoking Recognition Model

The training process of the dataset used for creating the smoking recognition model is as follows: First, a dataset was prepared. In this step, smoking behavior images were collected for data mining using Kaggle. Initially, 3,000 images, including 1,500 smoking and 1,500 non-smoking images, were prepared. Second, an image-resizing step was applied. The sizes of the 3,000 dataset images prepared were all different, and we therefore resized them to $500 \times 500 \times 3$ for image learning. Resizing was conducted based on the longer horizontal and vertical lengths, and bilinear interpolation with high efficiency was used. The third step is image padding. When applying the second step, because resizing was applied based on the longer distance between the horizontal and vertical lengths, space occurred on one side. Empty spaces were filled with black. Resizing not only allows the images to be uniform, it also increases the recognition rate of the skeleton model. Fig. 7 shows an example of a resized image. When applying a skeleton to the original image, the arm part of the original data in the left image is not normally covered with the skeleton model; however, in the image on the right, it can be seen that the arm is normally covered with a skeleton model. Fourth, this stage involves the creation of a skeleton neural network. A skeleton neural network model was created by applying the skeleton model to the padded image in the previous step. It took approximately 15 h to process approximately 3,000 sheets.

The fifth stage is the data training stage. Machine learning was conducted based on the image data created in the previous step.
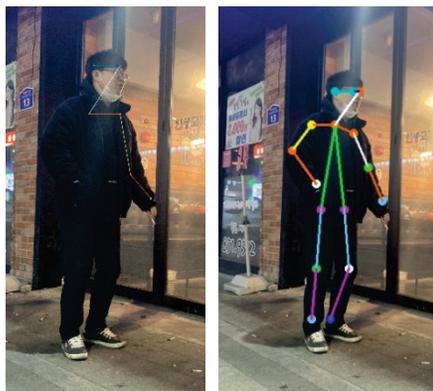

(a)   (b)
Figure 7 Resized data image (a) (b)

The following four test dataset models were created based on the neural network created in the previous step.

For test model 1, 1500 smoking images and 1500 non-smoking images to which the skeleton model was not applied were trained.

For test model 2, 1500 smoking images and 1500 non-smoking images to which the skeleton model was applied were trained.

For test model 3, 1,000 smoking images and 1,000 non-smoking images to which the skeleton model was applied were added, for a total training of 5,000 images.

For test model 4, 1,100 smoking images and 1,400 non-smoking images, i.e., 2,500 images, were trained by deleting garbage data to which the skeleton model was not properly applied.

A final test model was selected through a performance analysis of the generated models, and the selected model was used as the final model for smoking behavior recognition.

4.3 Hardware Device Implementation

For the proposed system, a smoking behavior detection device for determining tobacco smoke was developed as an auxiliary tool to increase the accuracy of detection. The system was developed using the sensors shown in Figure 8, and is located at the top of the surveillance camera placed in a non-smoking area, as shown in Fig. 9. A Raspberry Pi 4 is applied as the hardware device controller for smoke gas recognition and data transmission, and MQ7 and MH-Z14A sensors are used to measure CO and CO_2 in the air. In addition, a GSAS61-P110 smoke sensor is used to detect cigarette smoke and an NS-FDSM flame sensor is used to detect the flame used to light the cigarettes as well as cigarette cinder.

A socket program that can be accessed from an external IP is used to transmit the resulting software value to the Raspberry Pi. A PC equipped with a smoking-aware skeleton model is used as the client, a Raspberry Pi is placed as the server, and data transmission is smoothly applied during socket programming. For the transmitted data, a value of 1 is used for smoking, and a value of 0 is applied for non-smoking. When a value of 1 is received, the Raspberry Pi device starts measuring the sensor value. Hardware sensors collect information on smoking, and when the collected sensor values exceed a set threshold value, smoking behavior is determined. Finally, the system recognizes that smoking has occurred and outputs a warning sound.


Figure 8 Sensors and hardware devices
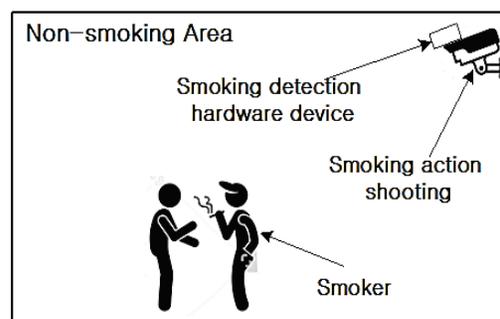
Figure 9 Smoking environment for experimentation

Table 4 Determination of sensor threshold values

Sensor	Threshold	Initial value	End value
CO	600	470	736
CO_2	1170	1182	1164
SMOKE	600	105	688
Flame	True	False	False

The threshold value of the sensor for cigarette smoke detection was determined through an environmental experiment on cigarette smoke using an acrylic box (40 × 40 × 40 cm). Tab. 4 shows the value of each gas measured for a period of approximately 60 s after the cigarette was lit (in order, the initial value when the sensor started measuring, the standard value indicating that smoking is occurring, and the value when the experiment was finished). The sensor threshold value is calculated by averaging the values collected from several environmental experiments. The CO, CO_2, SMOKE, and FIRE sensor threshold values were used

as the criteria for determining whether smoking is taking place.

4.4 Performance Analysis of Test Dataset Model

The most suitable dataset model for smoking recognition was selected through performance analysis experiments conducted on the four neural network test dataset models described in Section 4.2. The performance analysis proceeded as follows. First, a classification experiment was conducted for each model. This is an experiment on how well the classification (Classification_report) function was applied to the model (dataset). Tab. 5 shows the experimental results for models (dataset) 1 and 2, and Tab. 6 shows the experimental results for models 3 and 4.

Table 5 Performance analysis results for neural network models 1 and 2

#model 1					#model 2				
	Precision	Recall	F1-score	Support		Precision	Recall	F1-score	Support
Not-smoking	0.79	0.82	0.81	255	Not-smoking	0.82	0.83	0.83	255
Smoking	0.88	0.86	0.87	396	Smoking	0.89	0.89	0.89	396
Accuracy			0.85	651	Accuracy			0.86	651
Macro avg.	0.84	0.84	0.84	651	Macro avg.	0.86	0.86	0.86	651
Weighted avg.	0.85	0.85	0.85	651	Weighted avg.	0.87	0.86	0.86	651

Table 6 Performance analysis results for neural network models 1 and 2

#model 1					#model 2				
	Precision	Recall	F1-score	Support		Precision	Recall	F1-score	Support
Not-smoking	0.83	0.84	0.84	439	Not-smoking	0.82	0.85	0.83	227
Smoking	0.88	0.87	0.87	571	Smoking	0.87	0.85	0.86	281
Accuracy			0.86	1010	Accuracy			0.85	508
Macro avg.	0.85	0.85	0.85	1010	Macro avg.	0.85	0.85	0.85	508
Weighted avg.	0.86	0.86	0.86	1010	Weighted avg.	0.85	0.85	0.85	508

The experimental results are as follows: First, precision is the ratio of the number of samples that actually belong to the positive class to the number of samples that are output as belonging to this class. The higher the precision is, the better the judgment of the model. A precision of 1.0 means that there are zero false positives (FPs). Tab. 7 shows a classifier matrix of the precision.

Table 7 Classifier matrix

		Real result	
		True	736
Classification result	True	Classification result	True
SMOKE	False	105	False

The meaning of each item is as follows, and the precision can be obtained using Eq. (1):
- True positive (*TP*): The model accurately predicts that the sample is higher than the threshold value.
- True Negative (*TN*): The model accurately predicts that the sample is lower than the threshold value.
- False positive (*FP*): The model inaccurately predicts that the sample is higher than the threshold value.
- False negative (*FN*): The model inaccurately predicts that the sample is lower than the threshold value.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Second, recall is the ratio of the number of samples that are output as belonging to the positive class to the number of samples belonging to that class. The higher the number is, the better the model. A recall of 1.0 means that the *FN* is zero. The recall can be calculated using Eq. (2).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Third, the F-score is a weighted harmonic average of the precision and recall. In particular, beta is the weight assigned to the precision, and when the value is 1, it is called the F1 score. The higher the number is, the better the model. The F1 score can be calculated through Eq. (3).

$$PF_\beta = \frac{(1 + \beta^2)(Precision \times Recall)}{(\beta^2)(Precision \times Recall)} \tag{3}$$

Fourth, accuracy refers to the ratio of the number of correctly predicted samples to all samples. The higher the number is, the better the model. In general, it is used as an optimization objective function in learning. The accuracy can be calculated through Eq. (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Fifth, the macro-average is the weight assigned to each class. The same weight is assigned to each class. In other words, an imbalance in the number of samples is not considered. Because it does not consider the imbalance in the number of samples, a larger penalty occurs when the performance of a minority class is low.

Finally, the weighted average is based on the number of samples belonging to each class. An imbalance in the number of samples is considered. Because the weighted average is applied, the influence of a class with a small number of samples is reduced.

The accuracy of the results for each dataset model were 0.85, 0.86, 0.86, and 0.85, respectively, indicating no significant differences. In addition, there were no significant differences in the precision, recall, and F1 scores. However, it can be determined that relatively uniform data can be output regardless of the input data because the difference in value between the non-smoking data and the smoking data for dataset model 4 is smaller than that of the other models. Therefore, we proceeded with the final smoking recognition experiment using dataset model 4.

To judge the smoking behavior, the image to which the skeleton model is applied must pass through a smoking recognition neural network. To input the image into the neural network, we need to resize it to 244 × 244 × 3, and in this case, interpolation is used. Interpolation is a method for estimating unknown values using known data. When changing the size of an image, if the ratio of the image is changed, new values must be assigned by mapping new pixel values to non-existent areas or by compressing the existing pixels. When an image is enlarged, an interpolation method for the pixels is applied, and when an image is reduced, a pixel merging method is used. There are five types of nearest-neighbor interpolation provided by OpenCV: neighbor, bilinear, bicubic, domain, and Lanczos interpolation [20].

Among these five interpolation methods, the nearest neighbor interpolation method was found to be the most suitable through experimentation. By applying each interpolation method to the test data consisting of 100 images, the proposed smoking recognition module is executed, and the interpolation method with the most correct answers is selected as the optimal interpolation approach. Finally, the selected interpolation method is applied to the image interpolation.

4.5 Smoking Recognition Experiment

Experiments were conducted to analyze the performance of the proposed neural network model. Dataset model 4, described in Section 4.4, was applied during the experiment, and the nearest neighbor interpolation method was used as the image interpolation method. As the experimental results. A total of 198 images of smokers and 198 images of non-smokers were applied during the experiment. The correct answer rate was 0.75 for the smoking images (148 correct answers) and 0.7 for the non-smoking images (139 correct answers).

Fig. 10 shows example results when applying the neural network smoking behavior recognition module proposed in this study. The object number appears on the head of the person who is smoking. When the smoking behavior recognition module identifies a human object, a skeleton model is applied, and for each object to which the skeleton model is used, it is determined whether a smoking action has occurred. As shown in the figure, when the system judges that both the left and right objects are smoking, it outputs the messages "AI has detected that the Nth object is smoking" and "Send a value of 1 (smoking behavior) to the manager" in the output window below. The recognition results are then transmitted to the hardware device. Tab. 8 shows the result of the recognition module being transmitted to the hardware

device, where the final result is determined according to the sensor value. When a smoking action is determined, the message "The standard value has been exceed and output a warning sound" is displayed, as shown in the red box.

Figure 10 Results of smoking behavior recognition module

Table 8 Final results of the recognition module and hardware device

	Output Example
Algorithm execution and output example	CO: 580, CO$_2$: 1173, Smoke: 811, Fire: False
	CO: 610, CO$_2$: 1173, Smoke: 837, Fire: False
	The standard value has been exceed and output a warning sound.
	CO: 625, CO$_2$: 1173, Smoke: 840, Fire: False
	The standard value has been exceed and output a warning sound.
	CO: 637, CO$_2$: 1173, Smoke: 833, Fire: False
	The standard value has been exceed and output a warning sound.
	…

5   CONCLUSIONS

In this study, an artificial neural network based smoking behavior recognition model using an OpenPose based skeleton module for detecting smoking in a non-smoking area and providing a warning was proposed, along with a hardware device for improving the accuracy of smoking behavior recognition. To improve the recognition performance, the best dataset model among the four learning dataset models was selected experimentally, and the best performing method among the five interpolation methods was applied to generate the final recognition model. Based on the performance analysis of the proposed model, the smoking behavior recognition rate was 70% for *TP*s and 75% for *FN*s. The hardware device combines the transmitted results with the results recognized by the sensor and generates an accurate warning sound when the result of the smoking behavior recognition module is transmitted to the hardware device.

Compared with related studies [9], the image recognition rate of the proposed system is similar to the average recognition rate of 72% for each part of the body; however, this study is significant and differs in the following ways.

First, the subject of this study was the intricate posture of smoking, rather than easy-to-judge postures, such as walking and falling, addressed in other studies. This makes it difficult to judge only the numerical results of the motion recognition accuracy shown in each study. Second, several characteristic tasks were performed to increase the smoking behavior recognition rate. A preprocessing process was first performed for the convenience of processing learning data. Four dataset models were established, and the best training dataset was determined through experiments. Finally, various interpolation methods were tested, and the test data were processed using the nearest neighbor interpolation method to increase the recognition probability of the test data. These tasks are characteristic image recognition procedures of this study and are factors that can improve the performance of the proposed smoking behavior recognition module in the future. Third, a hardware device was developed to verify and supplement the accuracy of the smoking motion recognition software module. That is, a hardware device was developed and applied as an auxiliary tool to increase the accuracy of smoking behavior recognition. That is, the hardware device is notified of the result of the software recognition model, and thus the final smoking behavior is recognized and judged. In the final performance experiment fusing the software recognition module and hardware device, when the software result is transferred to the hardware device, the hardware device properly senses and reacts to the gas generated from smoking.

Despite the significance of this study, additional studies are required in the future. First, it is necessary to quantitatively expand the training image data for machine learning as well as select and train an image suitable for the OpenPose algorithm. In this study, 1500 smoking and 1500 non-smoking data points were used as training data. It is necessary to learn more data and select and apply various posture data to increase the recognition rate. Second, a better dataset model is needed to improve smoking recognition accuracy. In this study, the best-performing dataset model was selected based on precision and accuracy, but it is necessary to develop more diverse dataset models and use the best-performing dataset model. Finally, the hardware device that recognizes smoking motion needs to be improved. In other words, it is essential to add a sensor that can detect several gases produced by cigarette smoke, and it is equally important to select a sensor that can identify the same gas with excellent detection performance.

Acknowledgements

6   REFERENCES

[1] Lee, G., Kim, J., & Jang, Y. (2020). A study on types of smoking location and space through analysis of smokers' behaviors. *Journal of the Korea Institute of the Spatial Design, 15*(8), 417-428. https://doi.org/10.35216/kisd.2020.15.8.417

[2] Kim, N. (2017). Change of University students' smoking behavior before and after non-smoking policy. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 7*(6), 419-426. https://doi.org/10.35873/ajmahs.2017.7.6.038

[3] Noh, J., Lee, Y., Yoo, K., & Yoon, J. (2018). Overseas trend of non-smoking area guidelines and outdoor smoking room. *Korean Public Health Research, 44*(3), 53-64. https://doi.org/10.22900/kphr.2018.44.3.005

[4] Jee, Y. & Kim, Y. (2021). Study on the factor influencing smoking cessation among adolescent smokers. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(8), 251-260. https://doi.org/10.47116/apjcri.2021.08.23

[5] Park, H. & Lee, C. (2021). Effects of anti-smoking advertising messages on smokers. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(9), 131-140. https://doi.org/10.47116/apjcri.2021.09.12

[6] Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172-186. https://doi.org/10.48550/arXiv.1812.08008

[7] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P., & Schiele, B. (2016). DeepCut: Joint subset partition and labeling for multi person pose estimation. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9. https://doi.org/10.48550/arXiv.1511.06645

[8] Urooj, A. & Borji, A. (2018). Analysis of hand segmentation in the wild. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4710-4719. https://doi.org/10.48550/arXiv.1803.03317

[9] Ramakrishna, V., Munoz, D., Hebert, M., Bagnell, J., & Sheikh, Y. (2014). Pose machines: Articulated pose estimation via inference machines. *Proceeding of the European Conference on Computer Vision, Lecture Notes in Computer Science*, 8690, 33-47. https://doi.org/10.1007/978-3-319-10605-2_3

[10] Raaj, Y., Idrees, H., Hidalgo, G., & Sheikh, Y. (2019). Efficient online multi-person 2D pose tracking withrecurrent spatio-temporal affinity fields. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 4620-4628. https://doi.org/10.48550/arXiv.1811.11975

[11] Fang, H., Xie, S., Tai, Y., & Lu, C. (2017). RMPE: Regional multi-person pose estimation. *Proceeding of the International Conference on Computer Vision ICCV 2017*, 2334-2343. https://doi.org/10.1109/ICCV.2017.256

[12] Bridgeman, L., Volino, M., Guillemaut, J., & Hilton, A. (2017). Multi-person 3D pose estimation and tracking in sports. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1-10. https://doi.org/10.1109/CVPRW.2019.00304

[13] Chen, W., Jiang, Z., Guo, H., & Ni, X. (2020). Fall detection based on key points of human-skeleton using OpenPose. *Symmetry, 12*(5), 1-17. https://doi.org/10.3390/sym12050744

[14] Slembrouck, M., Luong, H., Gerlo, J., Schutte, K., Cauwelaert, D., Clerdq, D., Vanwanseele, B., Veelaert, P., & Philips, V. (2020). Multiview 3D markerless human pose estimation from OpenPose skeletonsm. Proceeding of the International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS) 2020, Lecture Notes in Computer Science, 12002, 1-12. https://doi.org/10.1007/978-3-030-40605-9_15

[15] Lin, C., Dong, Z., Kuan, W., & Huang, Y. (2021). A framework for fall detection based on OpenPose skeleton and LSTM/GRU models. *Applied Sciences, 11*(1), 1-20. https://doi.org/10.3390/app11010329

[16] Nakai, M., Tsunoda, Y., Hayashi, H., & Murakoshi, H. (2018). Prediction of basketball free throw shooting by OpenPose. *Proceeding of the International Symposium on Artificial Intelligence (JSAI-isAI) 2018: New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, 11717, 435-446. https://doi.org/10.1007/978-3-030-31605-1_31

[17] Mazhar, O., Ramdani, S., Navarro, B., Passama, R., & Cherubini, A. (2018). Towards real-time physical human-robot interaction using skeleton information and hand gestures. *Proceeding of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1-7. https://doi.org/10.1109/IROS.2018.8594385

[18] Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceeding of the 3rd International Conference on Learning Representations, ICLR 2015*, 1-14. https://doi.org/10.48550/arXiv.1409.1556

[19] Urakawa, T., Tanaka, Y., Goto, S., Matsuzawa, H., Watanabe, K., & Endo, N. (2019). Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiology, 48(1), 239-244. https://doi.org/10.1007/s00256-018-3016-3

[20] Jose, S., Miguel, C., Rafael, A., & Marta, S. (2020). OpenCV basics: A mobile application to support the teaching of computer vision concepts. *IEEE Transactions on Education, 63*(4), 328-335. https://doi.org/10.1109/TE.2020.2993013

[21] Nesreen Samer E.-J., & Samy S. A.-N. (2018), Diabetes Prediction Using Artificial Neural Network. *International Journal of Advanced Scienc+e and Technology, NADIA, 121*, 55-64. https://doi.org/10.14257/ijast.2018.121.05

Authors' contacts:

Tae-Yeong Jeong, Graduate
School of Computer Science,
Kyungil University,
50, Gamasil-gil, Hayang-eup, Gyeongsan-si,
Gyeongbuk, 38428, South Korea
taeyeong.jeong419@gmail.com

Il-Kyu Ha, Professor
(Corresponding author)
School of Computer Science,
Kyungil University,
50, Gamasil-gil, Hayang-eup, Gyeongsan-si,
Gyeongbuk, 38428, South Korea
ikha@kiu.kr