

Article

Driver Facial Expression Analysis Using LFA-CRNN-Based Feature Extraction for Health-Risk Decisions

Chang-Min Kim ¹, Ellen J. Hong ², Kyungyong Chung ³ and Roy C. Park ^{4,*}¹ Division of Computer Information and Engineering, Sangji University, Wonju 26339, Korea; changingstart@gmail.com² Department of Computer and Telecommunications Engineering, Yonsei University, Wonju 26493, Korea; jeonghee.hong@gmail.com³ Division of Computer Science and Engineering, Kyonggi University, Suwon 16227, Korea; dragonhci@gmail.com⁴ Department of Information Communication Software Engineering, Sangji University, Wonju 26339, Korea

* Correspondence: roypark1984@gmail.com

Received: 26 February 2020; Accepted: 22 April 2020; Published: 24 April 2020



Abstract: As people communicate with each other, they use gestures and facial expressions as a means to convey and understand emotional state. Non-verbal means of communication are essential to understanding, based on external clues to a person's emotional state. Recently, active studies have been conducted on the lifecare service of analyzing users' facial expressions. Yet, rather than a service necessary for everyday life, the service is currently provided only for health care centers or certain medical institutions. It is necessary to conduct studies to prevent accidents that suddenly occur in everyday life and to cope with emergencies. Thus, we propose facial expression analysis using line-segment feature analysis-convolutional recurrent neural network (LFA-CRNN) feature extraction for health-risk assessments of drivers. The purpose of such an analysis is to manage and monitor patients with chronic diseases who are rapidly increasing in number. To prevent automobile accidents and to respond to emergency situations due to acute diseases, we propose a service that monitors a driver's facial expressions to assess health risks and alert the driver to risk-related matters while driving. To identify health risks, deep learning technology is used to recognize expressions of pain and to determine if a person is in pain while driving. Since the amount of input-image data is large, analyzing facial expressions accurately is difficult for a process with limited resources while providing the service on a real-time basis. Accordingly, a line-segment feature analysis algorithm is proposed to reduce the amount of data, and the LFA-CRNN model was designed for this purpose. Through this model, the severity of a driver's pain is classified into one of nine types. The LFA-CRNN model consists of one convolution layer that is reshaped and delivered into two bidirectional gated recurrent unit layers. Finally, biometric data are classified through softmax. In addition, to evaluate the performance of LFA-CRNN, the performance was compared through the CRNN and AlexNet Models based on the University of Northern British Columbia and McMaster University (UNBC-McMaster) database.

Keywords: facial expression analysis; line segment feature analysis; dimensionality reduction; convolutional recurrent neural network; driver health risk

1. Introduction

In our lives: emotion is an essential means to deliver information among people. Emotional expressions can be classified in one of two ways: verbal (the spoken and written word) and non-verbal

(gestures, facial expressions, etc.) [1,2]. People communicate with others every day, and in such a process, facial expressions account for a significantly high proportion of meaning. Expressions can be used to accurately understand another person's emotional state, and since the perception of the emotion in facial expressions is an essential factor of social cognition, it could be seen as playing an essential role in diverse areas of life [3,4]. As described, facial expressions can be used to understand and empathize with other people's emotions. A number of services and prediction models for analyzing such expressions and understanding users' emotional states have been, and are being, studied [5,6]. Recently studied are lifecare services using gesture recognition or expression analysis to detect the risks to which the elderly and patients with chronic diseases are exposed. Currently, the whole world is entering an aging society, and accordingly, the number of patients with chronic diseases (hypertension, cardiovascular diseases, and coronary artery disease, etc.) is increasing. In addition, even in the low age group, the prevalence of chronic diseases increases due to changes in dietary habits (food with high calories and high sugar content, etc.), lack of exercise, and smoking, etc. [7,8]. From one generation to the next, humankind will require services to continuously monitor and manage chronic diseases. Unless medical service technology makes innovative progress, the demand for such services will continue to increase. There are many cases where the elderly or patients with diseases experience emergencies, major accidents, or death from disease-related reasons such as acute shock. In particular, since car accidents frequently occur due to acute diseases while someone is driving, it is necessary to take urgent action to prevent them [9–11]. When such accidents occur in the absence of a fellow passenger, the driver is unable to take prompt action. Moreover, as autonomous vehicles become more popular, they can operate on cruise control regardless of the status of the driver. In such cases, even after a driver's abnormal health status is detected, he or she might not be able to do something within the so-called "golden time" to address the problem. Although the mortality of patients with chronic diseases has decreased due to medical progress, it is still necessary to continuously manage and prepare for emergency situations [12,13]. To that end, services are being studied that alert friends, hospitals, police stations, etc., after detecting a driver at risk. However, since such studies involved prediction models based on external factor analysis, a driver's potential risk factors have not been applied, and it is difficult to predict accidents that occur due to internal factors. For prediction services in which internal factors apply, accurate prediction requires a particular device to be installed in the vehicle, and a given code of conduct is to be followed. As for the possibility of checking the driver's internal risk factors intuitively, it is possible to judge a dangerous situation through the driver's facial expressions. Thus, this study was conducted for the prediction of risk through recognition of the driver's facial expressions. Concerning current recognition of facial expressions, various studies are in progress [14]. As for the traditional face recognition techniques, classification models using the extraction of handcrafted features (local binary pattern (LBP), histogram of gradients (HOG), gabor, scale-invariant feature transform (SIFT)) for the sensible extraction of the characteristics of face images have usually been used [15–17]. These methods have a problem, however, in which the performance deteriorates when there are various changes in the face in an actual environment. It is difficult to choose the appropriate feature parameters according to the field of application. Transformation can be made in various shapes, but it is necessary to determine the optimal feature parameters via experiential elements and various experiments, which is a problem. Recently, the deep-learning technique has widely been used. As the face recognition technique based on deep learning itself learns high-level characteristics, using a large amount of data built up in various environments, it shows a high recognition performance even in a wild environment. Accordingly, DeepFace (based on AlexNet) uses the locally connected convolution layer to effectively extract local characteristics from the face region [18]. DeepID proposed a lightweight convolutional neural network (CNN)-based face recognizer, using an input resolution with smaller pixels than DeepFace [19]. VGGFace, which appeared later, learned a deep network structure consisting of 15 convolution layers using a data set for high-capacity face recognition made by itself through an Internet search [20]. In addition, various studies were conducted to improve the performance of face recognition models such as DeepID2, DeepID3, and GoogLeNet [21–23]. As yet,

if real-time video data are processed by a face recognition technique based on deep learning, it is necessary to learn many classes. Thus, they mostly show the structure in which a fully connected layer becomes larger, which accordingly decreases the batch size and acts as a factor disturbing convergence in the learning by the neural network. Accordingly, in this paper, to resolve such problems, facial expression analysis of drivers by using line-segment feature analysis-convolutional recurrent neural network (LFA-CRNN) feature extraction for health-risk assessment is proposed. A service using facial expression information to analyze drivers' health risks and alert them to risk-related matters is proposed. Drivers' real-time streaming images, along with deep learning-based pain expression recognition, were utilized to determine whether or not drivers are suffering from pain. When analyzing real-time streaming images, it may be difficult to extract accurate facial expression features if the image is shaking, and it may be difficult or impossible to run the analysis process on a real-time basis due to limited resources. Accordingly, a line-segment feature analysis (LFA) algorithm reduces learning and assessment time by reducing data dimensionality (the number of pixels). Also proposed is increasing the processing speed to handle large-capacity original data and high resolutions. Drivers' facial expressions are recognized through the CRNN model, which is designed to reduce input data dimensionality and to learn the LFA data. The driver's condition is understood based on the University of Northern British Columbia and McMaster University (UNBC-McMaster) database to understand the driver's abnormal condition. A service is proposed for coping with risks, spreading the dangerous conditions concerning health risks that may occur while driving through the notice by understanding the driver's conditions as suffering and non-suffering conditions.

This study is organized as follows. Section 2 presents the trends in face analysis research and also describes the current risk-prediction systems and services using deep learning. Section 3 describes how the dimensionality-reducing LFA technique proposed in this paper is applied to the data generation process, and also presents the CRNN model designed for LFA data learning. Section 4 describes how the UNBC-McMaster database was used to conduct a performance test.

2. Related Research

2.1. Face Analysis Research Trends

In early facial expression analysis, various studies were conducted based on a local binary pattern (LBP). LBP is widely used in the field of image recognition thanks to its ability to recognize things, its strength against changes in lighting, and its ease of calculation. As LBP became widely used in face recognition, center-symmetric LBP (CS-LBP) [24] was used in a modified form that can show components in the diagonal direction, reducing the dimension of feature vectors. Also, some studies enhanced the accuracy of facial expression detection by using multi-scale LBP that multiplied the size of the radius and the angle [25,26]. However, the LBP technique is used with techniques for extracting feature vectors in order to increase accuracy. In this case, based on the field of the application, there is difficulty in choosing the appropriate feature vectors. Transformation in various forms is possible, but the optimal feature vector should be decided by experiential elements and from various experiments. If the LFA proposed in this study is used, the minimum necessary data are used when the face is analyzed, so data compression takes place autonomously. Also, since it can be performed through techniques for detecting the face and its outline, it can easily be used in various fields. Studies of face analysis based on point-based features utilizing landmarks are also in progress. Landmark-based face extraction has a very fast process of measuring and restoring the landmark, so it can immediately display changes in the face shape and facial expressions filmed in real time. The weight of the measured landmark can be lightened for uses and purposes, such as character and avatar. Jabon et al. (2010) [27] proposed a prediction model that could prevent traffic accidents by recognizing drivers' facial expressions and gestures. This prediction model generates 22 x and y coordinates on the face (eyes, nose, mouth, etc.) in order to extract facial characteristics and head movements, and it automatically detects movement. It synchronizes the extracted data with simulator data, uses them as input to the classifier, and calculates

a prediction for accidents. Also, Agbolade et al. (2019) [28] and Park (2017) [29] conducted studies to detect the face region based on multiple points, utilizing landmarks to increase the accuracy of face extraction. However, to prevent prediction of the landmark value from falling to the local minimum, it is necessary to pass through a process of correcting the result through plural networks based on the initial prediction value in cascade form. The difficulty in detection differs depending on the set value of the feature point of the face. The more subdivided the overall detected outline, the more difficult it gets. Also, if part of the face is covered, it becomes very hard to measure landmarks. If the LFA proposed in this study is used, it is somewhat possible to escape the impact of light, since only information about the segments is used. Also, there is no increase in the difficulty of detection.

Since the deep learning method shows high performance, studies based on CNNs and deep neural networks (DNNs) are actively conducted. Wang et al. (2019) [30] proposed a method for recognizing facial expressions by combining extracted characteristics with the C4.5 classifier. Since some problems still existed (e.g., overfitting of a single classifier, and a vulnerable generalization ability), ensemble learning was applied to the decision-making tree algorithm to increase classification accuracy. Jeong et al. (2018) [31] detected face landmarks through a facial expression recognition (FER) technique proposed for face analysis, and extracted geometric feature vectors considering the spatial position between landmarks. By implementing the feature vectors on a proposed hierarchical weighted random forest classifier in order to classify facial expressions, the accuracy of facial recognition increased. Ra et al. (2018) [32] proposed a deep learning structure in a block method to enhance the face recognition rate. Unlike the existing method, feature filter coefficients and the weighted values of the neural network (on the softmax layer and the convolution layer) are learned using a backpropagation algorithm. Performing recognition with the deep learning model that learned the selected block region, the result of face recognition is drawn from an efficient block with a high feature value. However, since the face recognition technique based on CNNs and DNNs should generally learn a large amount of classes, there is a structure in which the fully connected layer grows bigger. Accordingly, the structure acts as a factor reducing the batch size and disturbing convergence in the learning by a neural network. If the LFA proposed in this study is used, the input dimension is small. Thus, the disturbance in the convergence from learning (due to the decrease in the batch size that may be generated in the CNN and DNN) can be minimized.

2.2. Facial Expression Analysis and Emotion-Based Services

FaceReader automatically analyzes 500 features on a face from images, videos, and streaming videos that include facial expressions, and it analyzes seven basic emotions: neutrality, happiness, sadness, anger, amazement, fear, and disgust. It also analyzes the degree of the emotions, such as the arousal (active vs. passive) and the valence (positive vs. negative) online and offline. Research on emotions through analyzing facial expressions has been conducted in various research fields, including consumer behavior, educational methodology, psychology, consulting and counseling, and medicine for more than 10 years. It is widely used in more than 700 colleges, research institutes, and companies around the world [33]. The facial expression-based and bio-signal-based lifecare service provided by Neighbor System Co. Ltd. in Korea is an accident-prevention system dedicated to protecting the elderly who live alone and who have no close friends or family members. The services provided by this system include user location information, health information confirmation, and integrated situation monitoring [34]. Figure 1 shows the facial expression-based and bio-signal-based lifecare service, which consists of four main functions for safety, health, the home, and emergencies.

The safety function provides help/rescue services through tracing/managing the users' location information, tracing their travel routes, and detecting any deviations from them. The health function measures/records body temperature, heart rate, and physical activity level, and monitors health status. In addition, it determines whether or not an unexpected situation is actually an emergency by using facial expression analysis, and provides services applicable to the situation. The home function provides a service dedicated to detecting long-term non-movement and to preventing intrusions

by using closed-circuit television (CCTV) installed within the users' residential space. Lastly, the emergency function constructs a system with connections to various organizations that can respond to any situation promptly, as well as deliver users' health history records to the involved organizations.

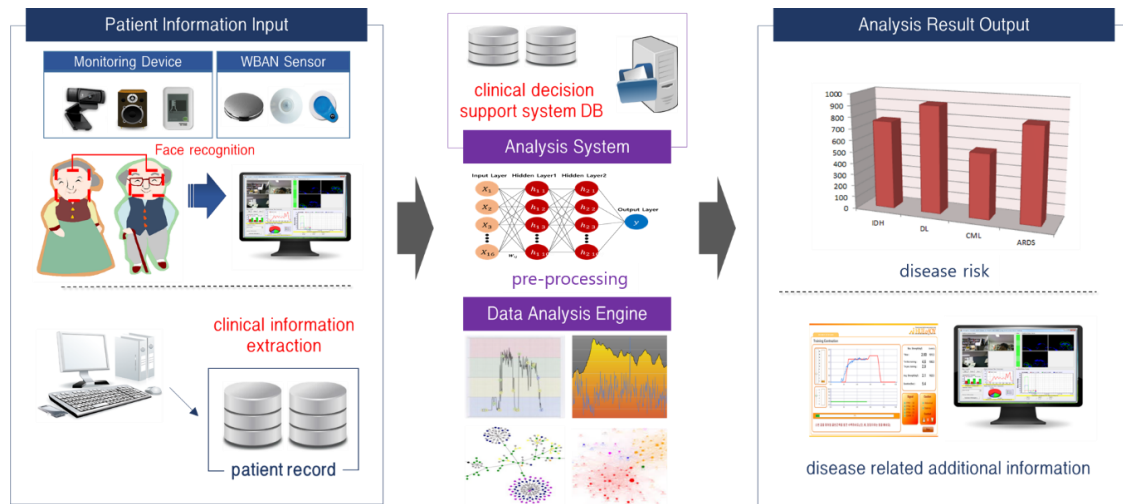


Figure 1. Facial expression and bio-signal-based lifecare service.

3. Driver Health-Risk Analysis Using Facial Expression Recognition-Based LFA-CRNN

It is necessary to compensate for senior drivers' weakened physical, perceptual, and decision-making abilities. It is also necessary to prevent secondary accidents, manage their health status, and take prompt action by predicting any potential traffic-accident risk, health risk, and risky behavior that might show up while driving. In cases where a senior driver's health status worsens due to a chronic disease, it becomes possible to recognize accident risks through facial expression changes. Accordingly, we propose resolving such issues with facial expression analysis using LFA-CRNN-based feature extraction for health-risk assessment of drivers. The LFA algorithm was performed to extract the characteristics of the driver's facial image in real time in the transportation support platform. An improved CRNN model is proposed, which can recognize the driver's face through the data calculated in this algorithm. Figure 2 shows the LFA-CRNN-based driving facial expression analysis for assessing driver health risks.

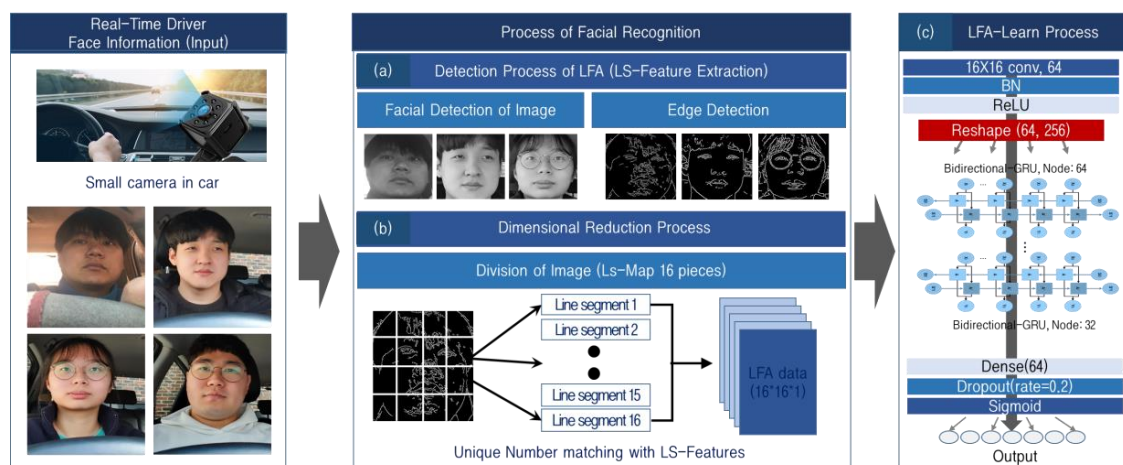


Figure 2. Line-segment feature analysis-convolutional recurrent neural network (LFA-CRNN)-based facial expression analysis for driver health risk assessment.

The procedures for recognizing and processing a driver's facial expressions can be divided into detection, dimensionality reduction, and learning. The detection process is a step of extracting the core areas (the eyes, nose, and mouth) to analyze the driver's suffering condition. In the step, there is a preconditioning process to solve the problem that the core areas are not accurately recognized. To extract features from the main areas of frame-type facial images segmented from real-time streaming images, multiple AdaBoost-based input images are divided into blocks. In the dimensionality reduction process, the LFA algorithm reduces the learning and reasoning time by reducing data dimensionality (the number of pixels) in order to increase processing speeds to handle large-capacity original data. High-resolution data and the dimensionality of the input data are reduced. Lastly, in the learning process, drivers' facial expressions are recognized through the CRNN model designed to learn the LFA data. In addition, to confirm a driver's abnormal status based on the UNBC-McMaster shoulder pain expression database, the service proposed determines if the driver is in pain, identifies the driver's health-related risks, and alerts the driver to such risks through alarms.

3.1. Real-Time Stream Image Data Pre-Processing for Facial Expression Recognition-Based Health Risk Extraction

Because pre-existing deep-learning models utilize the overall facial image for facial recognition, areas such as the eyes, nose, and lips serving as the main factors for analyzing drivers' emotions and pain status are not accurately recognized. Accordingly, through a detection process module, pre-processing is conducted for dimensionality reduction and learning. To analyze the original data transferred through real-time streaming, input images are segmented at 85 fps, and to increase the recognition rate, the particular facial image sections required for facial expression recognition are extracted using the multi-block method [35]. In particular, in cases where a multi-block is big or small during the blocking process, pre-existing models are unable to accurately extract features from the main areas, and this causes significant errors relating to recognition and learning. To resolve such issues, multiple AdaBoost is utilized to set optimized blocking, and then sampling is conducted. Figure 3 shows the process of detecting particular facial areas. A Haar-based cascade classifier is used to detect the face; Haar-like features are selected to accurately extract the user's facial features, and the AdaBoost algorithm is used for training. At this point, since features can be seen as a face/background-dividing characteristic and as a classifier, each feature is defined as a base classifier or a weak classifier candidate. During iterations, the training samples select one feature demonstrating the best classification performance, and the selected feature is used as the weak classifier in the iteration. The final weak classifiers are used in the weighted linear combination process to acquire the final strong classifiers.

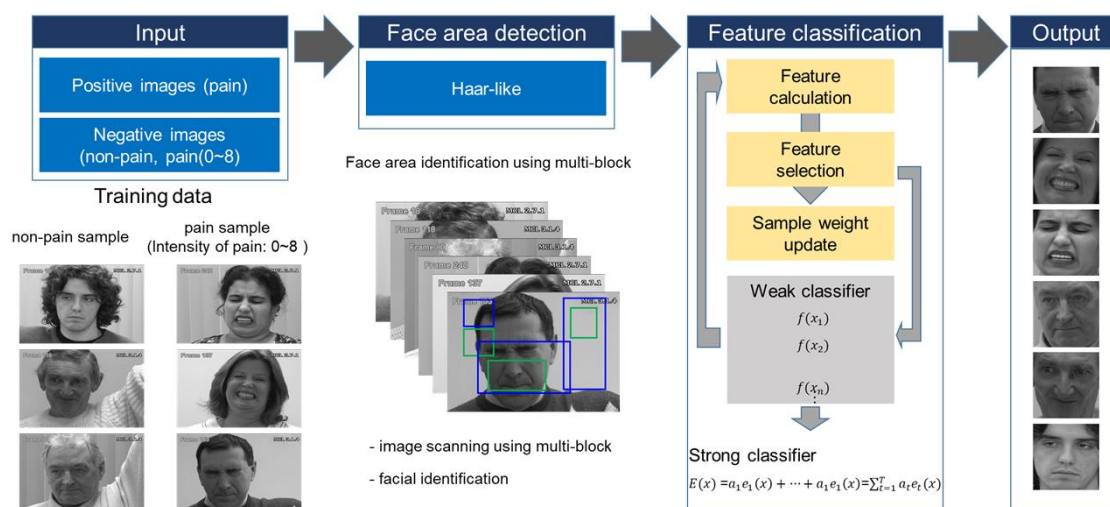


Figure 3. The multiple AdaBoost-based particular facial area detection process.

In the formula in Figure 3, $E(x)$ is the strong classifier finally found; e is the weak classifier drawn in the learning process, and a is the weighted value for the weak classifier. T is the number of repetitions. In this process, it is very hard to normalize the face if it is extracted without information such as the rotation and position of the face. Extracting the geometrical information of the face, it is necessary to normalize the face consistently. Faces can be classified according to their rotational positions, and if random images do not provide such information in advance, such rotational information must be detected during image retrieval. The detectors learned through multiple Adaboost are serialized, using the simple pattern of the face searcher. Using the serialized detectors, information can be found, such as the position, size, and rotation of the face. As for the simple pattern used in multiple Adaboost learning, the pattern in a basic form was used. 160 was chosen as the number of simple detectors to be found by Adaboost learning, and the processing speed of the learned detectors improved through serialization. The face region calculated through the above process detects the outline of the face through the Canny technique. This is the optimal technique option based on the experimental result. In the early stages, various outline detection techniques were used, but only the Canny method showed a high result.

3.2. Line-Segment Feature Analysis (LFA) Algorithm for Real-Time Stream Image Analysis Load Reduction

3.2.1. Pain Feature Extraction through LFA

Even after executing facial feature extraction through the procedures specified in Section 3.1, various constraint conditions may arise when extracting a driver's facial features from real-time driving images. In analyzing a real-time streaming image, it may be hard to extract accurate facial characteristics due to the motion of the image. Accordingly, since it is necessary to reduce the dimensionality of facial feature images extracted from real-time streaming images, the LFA algorithm is proposed. The proposed LFA algorithm is a dimensionality-reduction process that reduces learning and reasoning time by reducing data dimensionality (the number of pixels) to increase the processing speed in order to handle the original large-capacity, high-resolution data. To extract information from images, the line information on a 3×3 Laplacian mask's parameter-modified filter is extracted, a one-dimensional (1D) vector is created, and the created vector is utilized as the learning model's input data. Based on such a process, this algorithm creates new data through the line-segment features. LFA uses the driver's facial contour lines calculated through the detection process to examine and classify line-segment types. To examine the line-segment types, a filter, f , is used, and the elements {1, 2, 4, and 8} are acquired. Figure 4 shows the process where a driver's facial-contour line data are segmented, and the line-segment types are examined through the use of f .

Algorithm 1 Image Division Algorithm

Input: $[x_1, x_2, \dots, x_n]$

def Division and Max-pooling of image

$Y = \text{List}()$

for x_i in $[x_1, x_2, \dots, x_n]$ **do**

$\text{sub} = \text{List}()$

for w from 0 to D_w **do** // D_w, D_h denote the size of the image to be divided.

for h from 0 to D_h **do**

 // f_w, f_h denote the size of the filter.

$\text{sub.append}(x_i[w*fw: (w + 1)*fw, h*fh: (h + 1)*fh])$

$Y.append(\text{sub})$

$Y = \text{Max-pooling}(Y, \text{stride} = (2,2), \text{padding} = \text{'same'})$

Output: $Y[Y_1, Y_2, \dots, Y_n]$

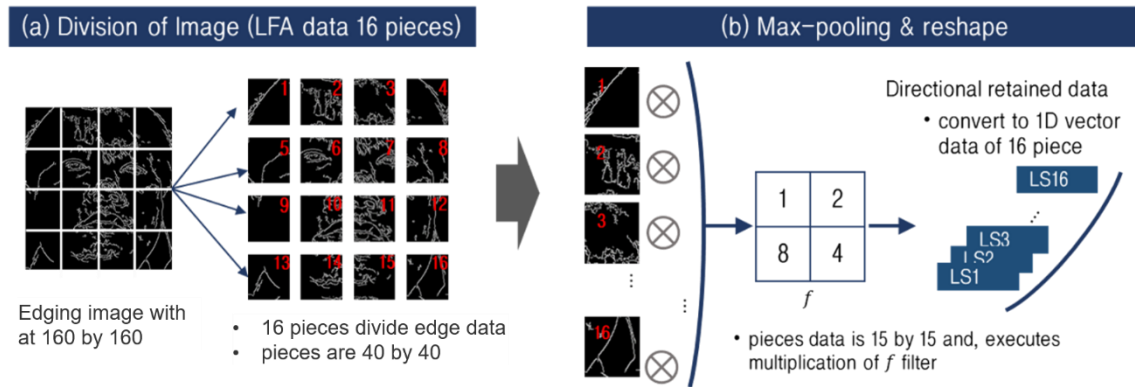


Figure 4. Driver facial contour-line data segmentation and line-segment type examination using f : (a) Division of image (LFA data 16 pieces); (b) Max-pooling and reshape.

Figure 4 shows the first LFA process. The contour line image calculated through pre-processing (detection) had a size of 160×160 , and this image was segmented into 16 parts, as shown in Figure 4a. This process is calculated as shown in Algorithm 1. The segmented parts have a size of 40×40 , and the segments are arranged in a way that does not modify the structure of the original image. These segments are max-pooled via the calculation shown in Figure 4b, and the arrangement of the segments is adjusted. This process is defined in Equation (1):

$$D_w, D_h = 4, 4, P_w = \frac{W}{D_w}, P_h = \frac{H}{D_h}, P[n, m] = x[n * P_w : (n + 1) * P_w, m * P_h : (m + 1) * P_h], \quad (1)$$

$$(0 \leq n \leq 4, 0 \leq m \leq 4) MP[n, m] = \max - \text{pooling}(P[n, m]),$$

Equation (1) is a calculation where the contour line image obtained during pre-processing is divided into 16 equal segments, and the divided segments are max-pooled. D_w and D_h denote the number of segments in the width and height, respectively, and P_w and P_h denote the size of the segmented data from dividing the contour line image by D_w and by D_h , respectively. P indicates the space for memorizing the segmented data, and the segmentation position is maintained through $P[n, m]$, in which n and m refer to a two-dimensional array index, having a value ranging between 0 and 4. MP memorizes the segmented data's max-pooling results. In every process, the sequence of the segmented images must not be lost. The sequence of the re-arranged segments must not be lost as well. Figure 4b shows the calculation where a convolution between the segment images and the filter is calculated: the parameters of the segmented images are converted, the sum of the parameters is calculated, and one-dimensional vector data are generated. The number of segmentations and the size of images in this process were the values selected through experiential selection, and after experimenting on various conditions, the optimal variables were calculated.

3.2.2. Line-Segment Aggregation-Based Reduced Data Generation for Pain Feature-Extracted Data Processing Load Reduction

The information from the line segment (LS) extracted (based on real-time streaming images) is matched with a unique number. The unique numbers are 1, 2, 4, and 8; they have a value that does not overlap another value, and the aggregate value deduces mutually different values. The LFA algorithm uses a 2×2 filter having a unique number for matching normal line-segment data. The LS has a value of 0 or 1, and where a filter consisting of a unique number is matched with the LS, only the areas having 1 as the unique number are displayed. A serial number is given to express information on segments, which is visual data in a series of information on numbers. That is converted to a series of information on numbers for easy counting of various segments (curve, horizontal line, and vertical line, etc.). Namely, visual data are converted into a series of patterns (numbers). Figure 5 shows the process where a segmented image is converted into 1D vector data.

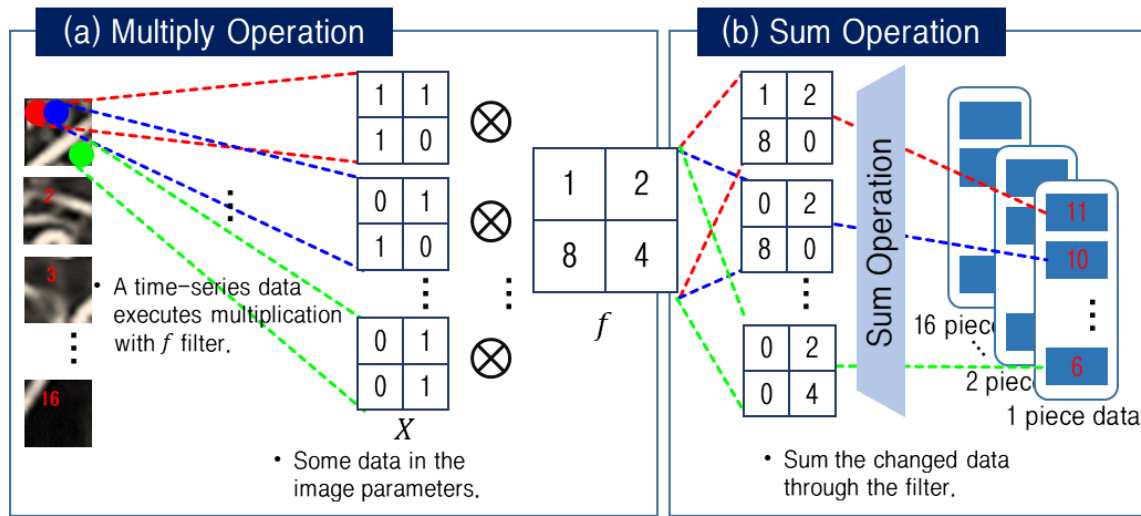


Figure 5. Conversion from a segment image to 1D vector data: (a) Multiply operation; (b) Sum operation.

A segment of an image utilizing contour line data has a parameter of 0 or 1, as shown in Figure 5a. The involved segment is a line segment when this parameter is 1, and is background when the parameter is 0. Such segment data are calculated with the filter, f , in sequence. The segment data have a size of 20×20 , and filter f is 2×2 . The 2×2 window is used to calculate a convolution between the segment data and filter f . At this point, the window moves one pixel at a time (stride = 1) to scan the entire area of the segmented image. Each scanned area is calculated with filter f ; the parameter is changed, and the image's 1 parameter is replaced with the f parameter. The process in Figure 5b is calculated as shown in Algorithm 2.

Algorithm 2 1D Vector Conversion Algorithm

Input: $[x_1 = [p_1, p_2, \dots, p_{16}], x_2 = [p_1, p_2, \dots, p_{16}], \dots, x_n = [p_1, p_2, \dots, p_{16}]]$

def Convert image to a 1D vector

Label = [1, 2, 8, 4]

Y = List()

for x_i in $[x_1, x_2, \dots, x_n]$ **do**

// Sub1 is a list to save the result of a piece of the image.

sub1 = List ()

for p_i in x_i **do**

// Sub2 is a list to save the result of the image of the matched piece

// with label data.

sub2 = List ()

for w from 0 to $W-f_w+1$ **do**

for h from 0 to $H-f_h+1$ **do**

$p = p_i[w:w + f_w, h:h + f_h]$

$p = p.reshape(-1) * \text{Label}$

sub2.append(sum(p))

sub1.append(sub2)

Y.append(sub1)

Output: $Y[Y_1, Y_2, \dots, Y_n]$

Equation (2) shows the calculation between the segment image and filter f , in which f_w and f_h represent the size. At this point, f has a fixed parameter and a fixed size; x_i is a partial area of the

segment image, and segments it into pieces the same size as f . Once a convolution between such segmented data and f is calculated, the calculated results are added up and recorded in P_i .

$$f = [[1, 2], [8, 4]] f_w, f_h = 2, 2P_i = \sum_{w=0}^W \sum_{h=0}^H x_i[w : w + f_w, h : h + f_h] \otimes f, \quad (2)$$

For example, when the segment image has parameters set to $[[1, 0], [0, 1]]$, as a convolution between the segment image and f is calculated, the segment image's parameters are changed to $[[1, 0], [0, 4]]$. These changed parameters obtain different values according to the position of each parameter due to f . Adding them up shows a different result, according to the data expression of the scan area. Table 1 shows the type and sum of lines according to the scanned areas.

Table 1. Type and sum of lines according to the scanned areas.

Scanned Area	Summing Data	Line Type	Scanned Area	Summing Data	Line Type
0000	0	Non-Active	1000	8	Point
0001	1	Point	1001	9	Vertical
0010	2	Point	1010	10	Diagonal
0011	3	Horizontal	1011	11	Curve
0100	4	Point	1100	12	Horizontal
0101	5	Diagonal	1101	13	Curve
0110	6	Vertical	1110	14	Curve
0111	7	Curve	1111	15	Active (Side)

When scanned areas are expressed as 0, they are considered the background (as shown in Table 1), and the summed value is also expressed as 0. On the other hand, all the areas expressed as 1 are considered active, and the side acquires a value expressed as 15. Other areas, according to the position and number of 1s, are expressed as point, vertical, horizontal, or diagonal, and are given a unique number. Despite being identical line types, all data are assigned a different number according to the expressed position, and the summed value is the unique value. For example, vertical is one of the line types detected in areas expressed as 0110 or 1001. However, each summed value is either 6 or 9 and has a different unique value. This means that the same line types are considered different lines based on their line-expressed positions. In addition, each line type's total cannot exceed 15. The data calculated through such a process will tie and save the line types (total) calculated per segment as a 1D vector, and will create a total of 16 1D vectors. Each vector has a size of $(20 - 2 + 1) \times (20 - 2 + 1) = 841$, and each vector's parameter has a value ranging from 0 to 15.

3.2.3. Unique Number-Based Data Compression and Feature Map Generation for Image Dimensionality Reduction

The 16 one-dimensional vector data calculated through the process shown in Figure 5 consist of unique values according to the line type determined through the information calculated by segmenting the facial image into 16 parts and matching each part with a particular filter. Such vector data consist of parameters ranging from 0 to 15. Each parameter has a unique feature (line-segment information). This section describes how cumulative aggregate data are generated based on the parameter value owned by each segment. The term “cumulative aggregate data” refers to data generated through a process where a parameter value is utilized as an index to generate a 1D array having a size of 16. The involved array's factor increases by 1 every time each index is called. Figure 6 shows the process where cumulative aggregate data are generated.

Algorithm 3 Cumulative Aggregation Algorithm**Input:** $[x_1 = [p_1 = [v_1, v_2, \dots, v_m], p_2, \dots, p_{16}], x_2, \dots, x_n]$ **def** Cumulative aggregation used to make LFA data

Y = List()

for x_i in $[x_1, x_2, \dots, x_n]$ **do**

sub1 = List()

for p in x_i **do**

sub2 = array(16){0, ... }

for i from p **do**

sub2[i]++

sub1.append(sub2)

Y.append(sub1)

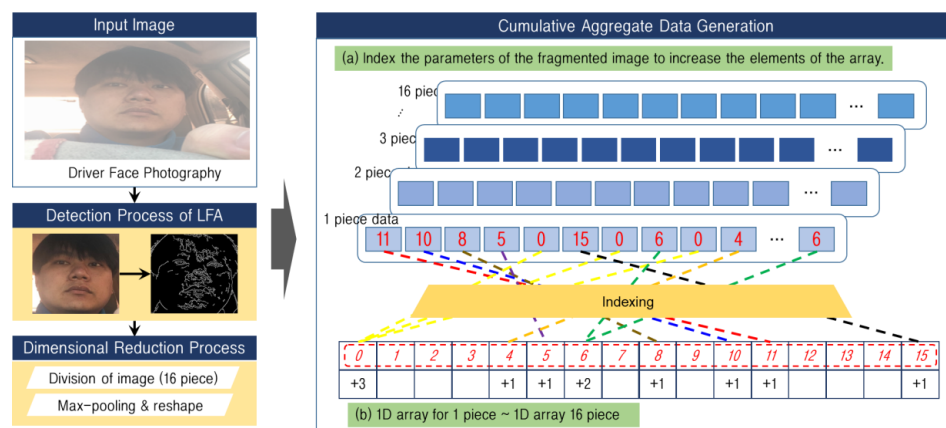
Output: Y

Figure 6. Process of cumulative aggregate data generation: (a) Index the parameters of the fragmented image to increase the elements of the array; (b) 1D array for 1 piece~1D array 16 piece.

As shown on the right side of Figure 6a, the parameters of the data segmented through the previous process are utilized as an array of the index, and a 1D array having a size of 16 is generated according to each segment. This array is shown in Figure 6b, and the factor value of the index position corresponding to each parameter of the one-dimensional array having a maximum size of 16 increases by 1. The process in Figure 6b is calculated as shown in Algorithm 3. Since this process is applied to each segment, an array having a size of 16 and corresponding to each segment is generated for each segment, and a total of 16 arrays are generated. These are known as LFA data and are shown in Figure 7a.

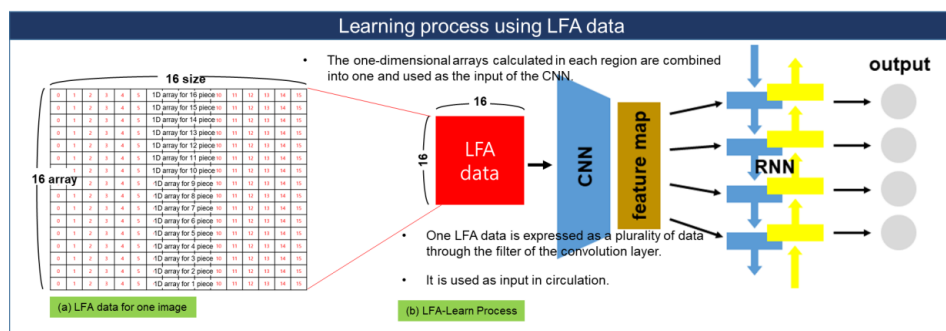


Figure 7. Learning process using LFA data: (a) LFA data for one image; (b) LFA-Learn process.

The LFA process in Figure 7a restructures each array generated through each segment image in the appropriate order (in the order prioritized based on the segmentation position). Through this,

for one image, the LFA data calculated through the LFA process are expressed as two-dimensional sequences with a size of 16×16 . This is used as input for the CRNN.

3.3. LFA-CRNN Model for Driver Pain Status Analysis

Once a feature map is generated, and the image is deduced through facial and contour line detection, the pre-processing of the given input images restructures them into two-dimensional arrays having a size of 16×16 through the LFA process. Specifically, the dimensionality is reduced through the LFA technique. Since LFA always has the same output size and consists of aggregate information on the line segment contained in the image, the reduced data themselves can be considered unique features. In addition, a learning model dedicated to LFA data is designed instead of a general CRNN learning model architecture for drivers' pain status, and the learning process is performed as well. Figure 8 shows the structure of the proposed LFA-CRNN model.

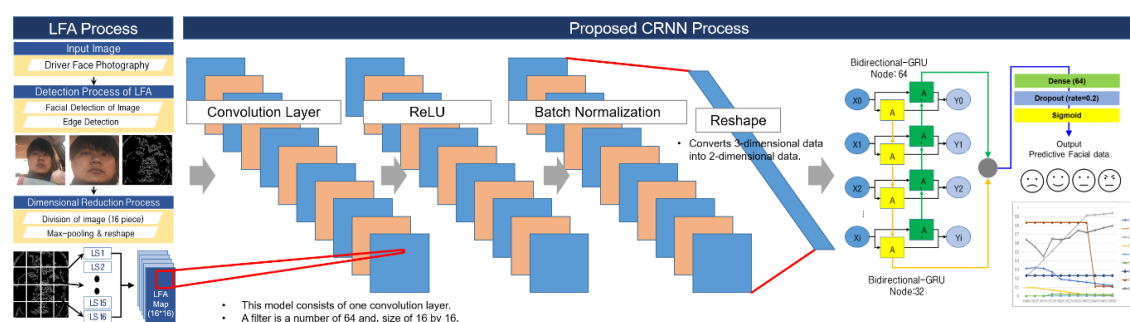


Figure 8. The proposed LFA-CRNN architecture.

The LFA-CRNN architecture is a CRNN learning model. It consists of one convolution layer, and expresses a feature map as sequence data through the reshaped layer. The features that changed into sequence data are transferred to the dense layer through two bidirectional gated recurrent units (BI-GRUs), and the sigmoid layer serves as the final layer before the results are output. Through the convolution layer's batch normalization (BN), the risk of depending on and overfitting the learning-speed improvement and initial weighted-value selection is reduced [36–38]. Since this learning model uses dimensionality-reduced LFA data, the compressed data themselves can be considered one feature. Accordingly, to express one major feature as a number of features in a convolution, the input-related expressions are diversely divided through a total of 64 filters having a size of 16×16 . The value deduced through such a process passes through the BN and generates a series of feature maps through the rectified linear unit (ReLU) layer. Such feature maps are restructured through the reshape layer into 64 sequence data having a size of 256 and are used as the RNN model's input. The RNN model consists of two BI-GRUs, one with 64 nodes and one with 32 nodes. The data deduced through this process are delivered to the sigmoid layer through the dense layer. At this point, the dropout layer is arranged between the dense layer and the sigmoid layer to prevent calculation volume reduction and overfitting [39–41]. Lastly, through the Sigmoid class, nine types of pain are classified. In this model, the pooling layer generally used in the pre-existing CNN and CRNN models is not used. Since the input LFA data themselves have a considerably small size of 16×16 , and consist of the cumulative number of line segments owned by the images when the involved data are compressed, the main features may be damaged or removed. In addition, in this model, BN and the dropout layer are arranged instead of the pooling layer, and the convolution's stride and padding are set to 1 and *same*, respectively. We used the convolution layer to get a variety of information about the expression of individual, highly-concentrated LFA data by designing the model like Figure 9. Thus, the filter of the convolution layer was set to 16×16 with stride = 1 and padding = "same." Through this, one LFA data size is maintained, and because of the weighted value of the filter, it can express a

lot of information. The data are used as input in each cycle of the RNN, and through the previous characteristics, strong characteristics are gradually detected from within.

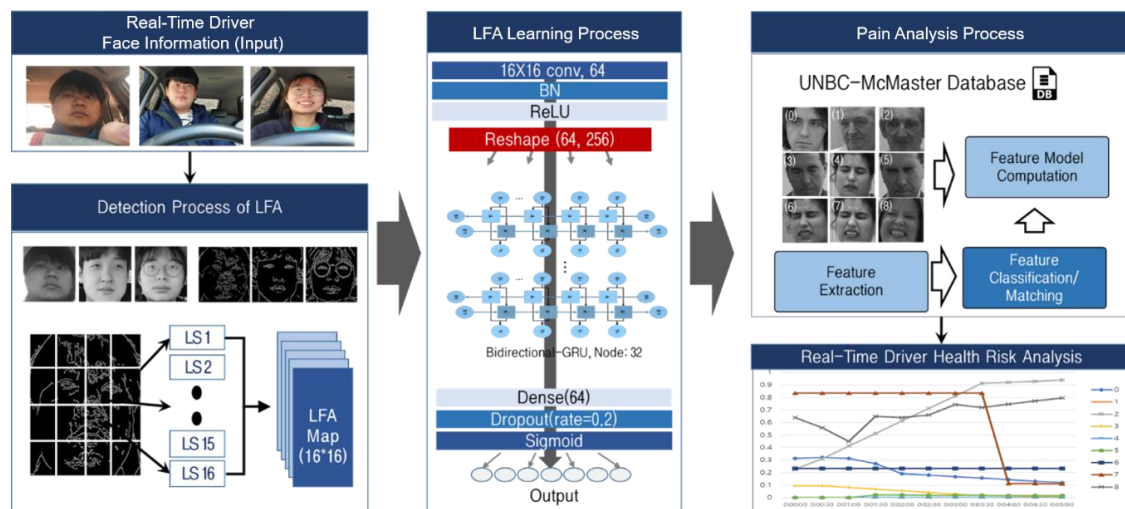


Figure 9. Driver pain status analysis process and its performance evaluation.

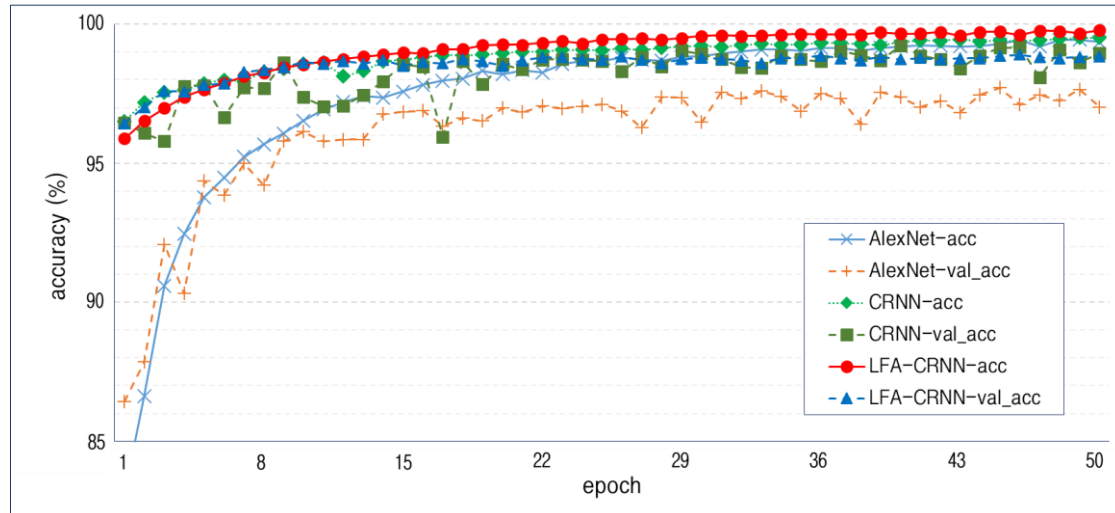
4. Simulation and Performance Evaluation

A simulation was conducted in the following environment: a Microsoft Windows 10 pro 64-bit O/S on an Intel Core(TM) i7-6700 CPU (3.40 GHz) with 16GB RAM, and an emTek XENON NVIDIA GeForce GTX 1060 graphics card with 6GB of memory. To implement this algorithm, we utilized OpenCV 4.2, Keras 2.2.4, and the Numerical Python (NumPy) library (version 1.17.4) based on Python 3.6. OpenCV was used to perform the Canny technique during pre-processing by the LFA, and the calculation of the queue generated in the LFA process was performed using the NumPy library. The neural network model was implemented through Keras. Figure 9 shows the process by which the driver's pain status is analyzed and under which the system's performance was evaluated.

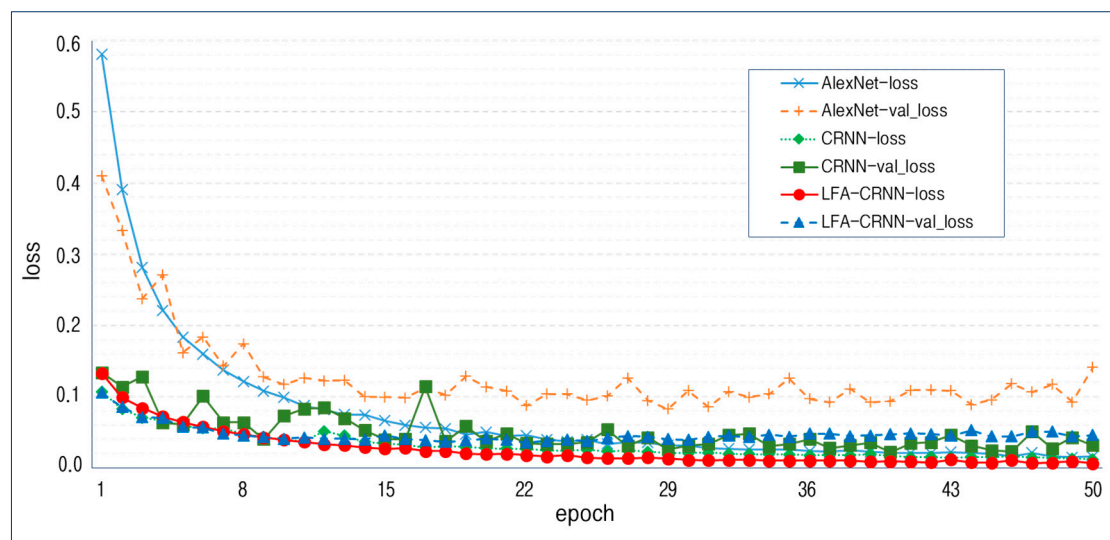
To evaluate the performance of the LFA-CRNN model-based face recognition (suffering and non-suffering expressions), the UNBC-McMaster database was used. In addition, a comparison was made with AlexNet and CRNN Models. The experiment of this paper chose the basic structure of the proposed model, the CRNN model and the AlexNet model generally well known for image classification to compare the performance. The UNBC-McMaster database classifies pain into nine stages (0~8) using the Prkachin and Solomon Pain Intensity (PSPI) scale, with data consisting of 129 participants (63 males and 66 females). The accuracy and loss measurement test were based on such data, calculated through pre-processing (face detection and contour line extraction). The LFA conversion process was used as the LFA-CRNN's input, and the CRNN [42] and AlexNet [43] for performance comparison used the data calculated through the face detection process. The test was conducted by taking 20% of the data from the UNBC-McMaster database [44] as the test data, and utilizing 10% of the remaining 80% as verification data. In the process of classifying data, to prevent data from leaning too much towards a particular class, the classification was undertaken by designating a specific percentage for each class. Specifically, 42,512 data units consisted of 29,758 learning data units, 3401 verification data units, and 8503 test data units.

Figure 10 shows the results of the accuracy and loss, using the UNBC-McMaster Database. As shown in Figure 10, the LFA-CRNN showed the highest accuracy, with AlexNet second and the CRNN third. AlexNet showed a large gap between the training data and verification data. The CRNN showed a continuous increase in the training data accuracy but showed a temporary decrease in the verification data accuracy due to overfitting. Although the LFA-CRNN proposed in this paper showed a bit of a gap between the learning and validation data, such a gap is not considered significant. Since no temporary decrease was shown in the validation data, it was confirmed that no learning overfitting

occurred; loss data showed the same patterns. AlexNet showed the highest gap between learning and validation data, in terms of loss. The CRNN showed a continuous decrease of loss in both learning and validation data, but showed a temporary increase in validation data. Therefore, the LFA-CRNN can be considered more reliable than both AlexNet and traditional CRNN.



(a) accuracy



(b) loss

Figure 10. Accuracy and loss measurement results using the University of Northern British Columbia (UNBC)-McMaster shoulder pain expression database.

Figure 11 shows the accuracy and loss achieved with the test data. As shown in the figure, the LFA-CRNN had the highest accuracy at approximately 98.92% and the lowest loss at approximately 0.036. The CRNN showed temporary overfitting during learning, and this was determined to be the reason why its accuracy was lower than the LFA-CRNN. Likewise, it was determined that AlexNet showed a performance decrease in its accuracy due to the verification data's wide gap. The test results shown in Figures 10 and 11 can be summarized as follows. As far as UNBC-McMaster-based learning is concerned, the LFA-CRNN model showed no rapid change in accuracy and loss, and it was confirmed that a stable graph was maintained as the epochs progressed (i.e., no overfitting or

large gap). In addition, compared to the basic models, the proposed method showed the highest performance with an accuracy of approximately 98.92%.

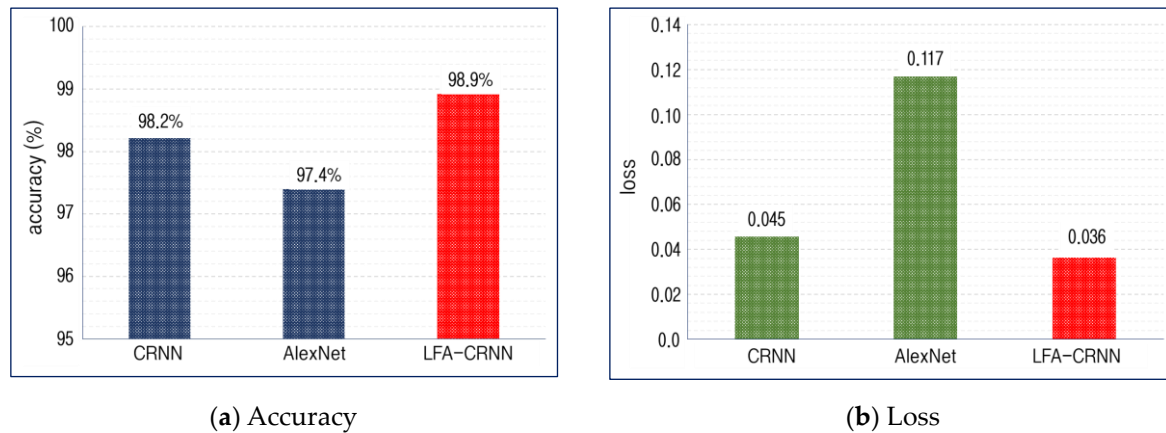


Figure 11. Accuracy and loss with the test data.

To measure the accuracy and reliability of the proposed algorithm, precision, recall, and the receiver operating characteristic (ROC) curve [45] were measured. Figure 12 shows the results achieved.

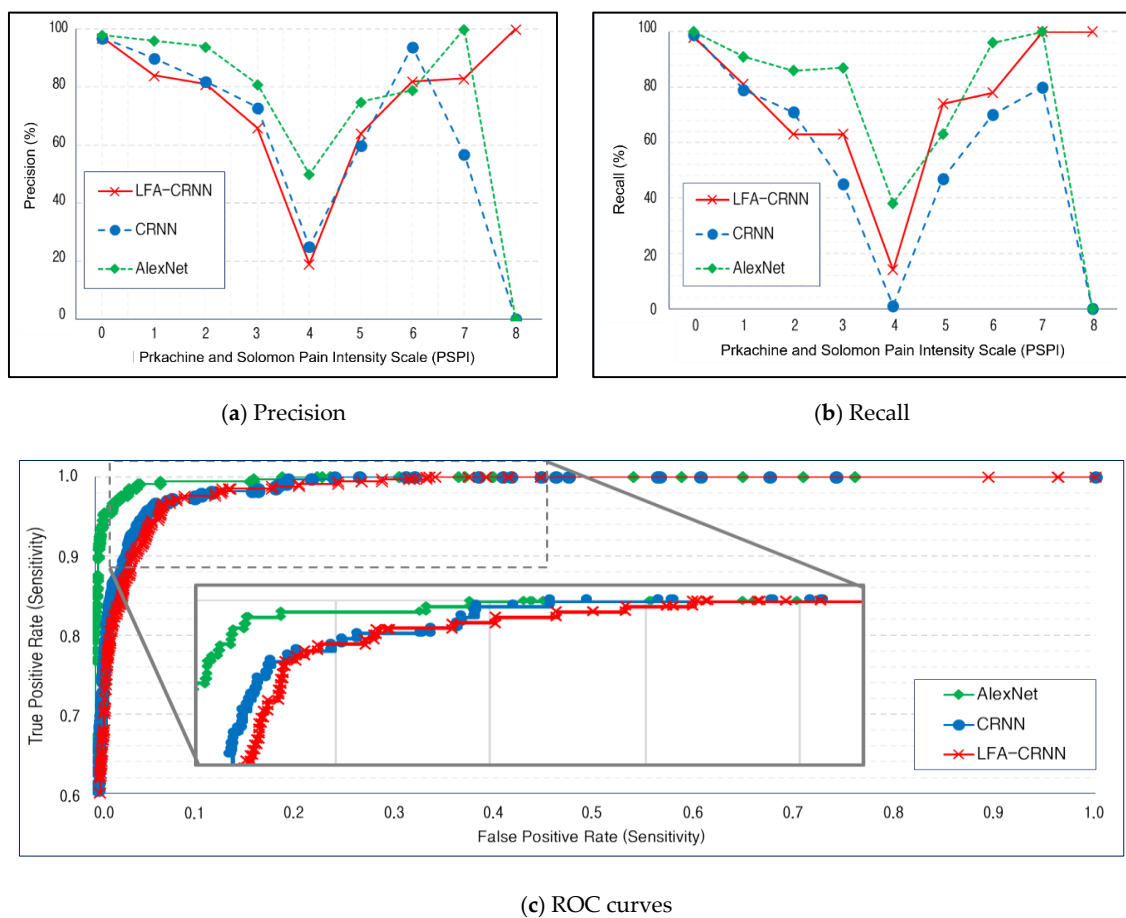


Figure 12. Results of precision and recall, plus the receiver operating characteristic (ROC) curve evaluation for each algorithm.

In Figure 12, the precision results show the percentage of the number of samples actually determined to be true out of the samples predicted to be true for each pain severity class. The

LFA-CRNN showed the following results: 0 = 98%, 1 = 81%, 2 = 63%, 3 = 63%, 4 = 19%, 5 = 74%, 6 = 78%, 7 = 100%, and 8 = 100%. Such results are quite poor, compared to the results achieved by AlexNet and the CRNN. It was determined that such results are attributable to the dimensionality reduction LFA technique. Since the dimensionality reduction technique itself either compresses the original image to generate new data or reduces the data size by using particular features consisting of strong features, it removes specific features and only uses strong features. However, only the LFA-CRNN was able to detect data having a PSPI of 8. In addition, as a result of confirming the average precision, both LFA-CRNN and AlexNet showed an average precision of 75%, while the CRNN showed an average precision of 56%. In addition, the recall measurements were similar to the precision results. The LFA-CRNN showed an average recall of 75%, AlexNet showed an average recall of 73%, and the CRNN showed an average recall of 56%. Based on this test, it was confirmed that it was difficult for all the models to detect data having a PSPI of 4, and that only the LFA-CRNN was able to detect data having a PSPI of 8. To sum up all experiments, the proposed LFA-CRNN model showed a stable graph in the learning process, and in the performance evaluation, using the test data, it showed the highest performance of 98.92%. In addition, its loss measurement showed the lowest result at approximately 0.036. Although the LFA-CRNN's precision and recall were quite poor, its average precision was 75% (as high as the precision by AlexNet), and it showed the highest average recall at 75%.

The LFA-CRNN proposed in this study showed higher accuracy, using fewer input dimensions than comparable models. We judge that this is because of the effect of the maximum removal of unnecessary regions. We examined the metadata necessary for analyzing test data of facial expressions and judged that the color and area (size) constituting the images were unnecessary elements. Thus, the remaining element was the information about segments, and we set up a hypothesis for a sentiment analysis algorithm through this. When people analyze facial expressions, they do not usually consider colors, and the color element was removed, using the understanding of emotions through the shapes of the mouth, eyes, and eyebrows. Also, the images with colors removed were similar to the images expressed with the outline. In learning with the neural network model, a big loss of data took place when images were reduced via max-pooling and stride in the processing, and the overfitting and wind-up phenomena occurred. Thus, we devised a method for reducing the size of the images, and that method is LFA. LFA maintained information about segments as much as possible to prevent data loss that might occur during processing, utilizing data with both color and unnecessary areas removed. In other words, when we extracted emotions, necessary elements were maintained as much as possible, and all other information was minimized. We judge that LFA-CRNN shows high accuracy for these reasons.

5. Conclusions

With this paper's proposed method, health risks due to an abnormal health status that may occur while someone is driving are determined through facial expressions, a representative medium capable of confirming a person's emotional state based on external clues. The purpose of this study was to construct a system capable of preventing traffic accidents and secondary accidents resulting from chronic diseases, which are increasing as our society ages. Although automated driving systems are being mounted on vehicles and are commercialized based on vehicle technology advancements, such systems do not take into consideration driver status. If abnormal health status in a driver is detected while the vehicle is in motion, it may operate normally, but the drivers might not be able to meet the required "golden time" to address any health problem that arises. Our system checks the driver's health status based on facial expressions in order to resolve to a certain extent problems related to chronic diseases. To do so, in this paper, an LFA dimensionality reduction algorithm was used to reduce the size of input images, and the LFA-CRNN model receiving the reduced LFA data as input was designed and used to classify the status of drivers as being in pain or not. The LFA is a method where a series of filters is used to assign a unique number to the line-segment information that makes up a facial image, and then, the input image is converted into a two-dimensional array having a size of

16 × 16 by adding up the unique numbers. As the converted data are learned through the LFA-CRNN model, facial expressions indicating pain are classified. To evaluate performance, a comparison was made with pre-existing CRNN and AlexNet models. The UNBC-McMaster database was used to learn pain-related expressions. As far as the accuracy and loss calculated through learning are concerned, the LFA-CRNN showed the highest accuracy at 98.92%, a CRNN alone showed accuracy of 98.21%, and AlexNet showed accuracy of 97.4%. In addition, the LFA-CRNN showed the lowest loss at approximately 0.036, the CRNN showed a loss of 0.045, and AlexNet showed a loss of 0.117. Although the LFA-CRNN's precision and recall measurement results were quite poor, average precision was 75%, which is as high as the 75% precision achieved by AlexNet.

We optimized the facial expressions and the data sources for the LFA-CRNN, and intend to compare the processing times of several models and improve the accuracy in the future. The proposed LFA-CRNN algorithm shows high dependency on the outline detection method. This is self-evident, because LFA is based on segment analysis. We are devising an outline detection technique that can optimally be applied to LFA based on this fact. In addition, the LFA performance process generates a one-dimensional sequence before the production of a two-dimensional LS-Map. It is expected that by converting this, a class can be produced that can be used in the neural network model. Through this improvement process, we will combine the LFA-CRNN model with a system for recognition of facial expressions and motions that can be used in services like smart homes and smart health care, and we plan to apply that to mobile edge computing systems and video security.

Author Contributions: K.C. and R.C.P. conceived and designed the framework. E.J.H. and C.-M.K. implemented LFA-CRNN model. R.C.P. and C.-M.K. performed experiments and analyzed the results. All authors have contributed in writing and proofreading the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 20CTAP-C157011-01).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yeem, M.J.; Song, H.J. The Effect of facial emotion Recognition of Real-face Expression and Emoticons on Interpersonal Competence: Mobile Application Based research for Middle School Students. *J. Emot. Behav. Disord.* **2019**, *35*, 265–284.
2. Olderbak, S.G.; Wilhelm, O.; Hildebrandt, A.; Quoidbach, J. Sex differences in facial emotion perception ability across the lifespan. *Cogn. Emot.* **2018**, *33*, 579–588. [\[CrossRef\]](#)
3. Poria, S.; Majumder, N.; Mihalcea, R.; Hovy, E.; Majumder, N.; Mihalcea, R. Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. *IEEE Access* **2019**, *7*, 100943–100953. [\[CrossRef\]](#)
4. Kang, X.; Ren, F.; Wu, Y. Exploring Latent Semantic Information for Textual Emotion Recognition in Blog Articles IEEE/CAA. *J. Autom. Sin.* **2018**, *5*, 204–216.
5. Guo, J.; Lei, Z.; Wan, J.; Avots, E.; Hajarolasvadi, N.; Knyazev, B.; Kuharenko, A.; Junior, J.C.S.J.; Baró, X.; Demirel, H.; et al. Dominant and Complementary Emotion Recognition from Still Images of Faces. *IEEE Access* **2018**, *6*, 26391–26403. [\[CrossRef\]](#)
6. Perlovsky, L.; Schoeller, F. Unconscious emotions of human learning. *Phys. Life Rev.* **2019**, *31*, 257–262. [\[CrossRef\]](#)
7. Chung, K.; Park, R.C. P2P-based open health cloud for medicine management. *Peer-to-Peer Netw. Appl.* **2019**, *13*, 610–622. [\[CrossRef\]](#)
8. Kim, J.; Jang, H.; Kim, J.T.; Pan, H.-J.; Park, R.C. Big-Data Based Real-Time Interactive Growth Management System in Wireless Communications. *Wirel. Pers. Commun.* **2018**, *105*, 655–671. [\[CrossRef\]](#)
9. Kim, J.-C.; Chung, K. Prediction Model of User Physical Activity using Data Characteristics-based Long Short-term Memory Recurrent Neural Networks. *KSII Trans. Internet Inf. Syst.* **2019**, *13*, 2060–2077. [\[CrossRef\]](#)

10. Baek, J.-W.; Chung, K. Context Deep Neural Network Model for Predicting Depression Risk Using Multiple Regression. *IEEE Access* **2020**, *8*, 18171–18181. [[CrossRef](#)]
11. Baek, J.-W.; Chung, K. Multimedia recommendation using Word2Vec-based social relationship mining. *Multimed. Tools Appl.* **2020**, 1–17. [[CrossRef](#)]
12. Kang, J.-S.; Shin, D.H.; Baek, J.-W.; Chung, K. Activity Recommendation Model Using Rank Correlation for Chronic Stress Management. *Appl. Sci.* **2019**, *9*, 4284. [[CrossRef](#)]
13. Chung, K.; Kim, J. Activity-based nutrition management model for healthcare using similar group analysis. *Technol. Health Care* **2019**, *27*, 473–485. [[CrossRef](#)]
14. Haz, H.; Ahuja, S. Latest trends in emotion recognition methods: Case study on emotiw challenge. *Adv. Comput. Res.* **2020**, *10*, 34–50. [[CrossRef](#)]
15. Song, X.; Chen, Y.; Feng, Z.-H.; Hu, G.; Zhang, T.; Wu, X.-J. Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning. *Pattern Recognit.* **2019**, *88*, 127–138. [[CrossRef](#)]
16. Nassih, B.; Amine, A.; Ngadi, M.; Hmina, N. DCT and HOG Feature Sets Combined with BPNN for Efficient Face Classification. *Procedia Comput. Sci.* **2019**, *148*, 116–125. [[CrossRef](#)]
17. Lenc, L.; Kral, P. Automatic face recognition system based on the SIFT features. *Comput. Electr. Eng.* **2015**, *46*, 256–272. [[CrossRef](#)]
18. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2014; pp. 1701–1708.
19. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
20. Luttrell, J.; Zhou, Z.; Zhang, C.; Gong, P.; Zhang, Y.; Iv, J.B.L. Facial Recognition via Transfer Learning: Fine-Tuning Keras_vggface. In Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2017; pp. 576–579.
21. Sun, Y.; Wang, X.; Tang, X. Deep Learning Face Representation by Joint Identification-Verification. *arXiv* **2014**, arXiv:1406.4773.
22. Sun, Y.; Liang, D.; Wang, X.; Tang, X. DeepID3: Face Recognition with Very Deep Neural Networks. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
23. Khan, R.U.; Zhang, X.; Kumar, R. Analysis of ResNet and GoogleNet models for malware detection. *J. Comput. Virol. Hacking Tech.* **2018**, *15*, 29–37. [[CrossRef](#)]
24. Muhammad, G.; Alsulaiman, M.; Amin, S.U.; Ghoneim, A.; Alhamid, M.F. A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities. *IEEE Access* **2017**, *5*, 10871–10881. [[CrossRef](#)]
25. Lim, K.-T.; Won, C. Face Image Analysis using Adaboost Learning and Non-Square Differential LBP. *J. Korea Multimed. Soc.* **2016**, *19*, 1014–1023. [[CrossRef](#)]
26. Kang, H.; Lim, K.-T.; Won, C. Learning Directional LBP Features and Discriminative Feature Regions for Facial Expression Recognition. *J. Korea Multimed. Soc.* **2017**, *20*, 748–757. [[CrossRef](#)]
27. Jabon, M.E.; Bailenson, J.N.; Pontikakis, E.; Takayama, L.; Nass, C. Facial expression analysis for predicting unsafe driving behavior. *IEEE Pervasive Comput.* **2010**, *10*, 84–95. [[CrossRef](#)]
28. Agbolade, O.; Nazri, A.; Yaakob, R.; Ghani, A.A.; Cheah, Y.K. 3-Dimensional facial expression recognition in human using multi-points warping. *BMC Bioinform.* **2019**, *20*, 619. [[CrossRef](#)] [[PubMed](#)]
29. Park, B.-H.; Oh, S.-Y.; Kim, I.-J. Face alignment using a deep neural network with local feature learning and recurrent regression. *Expert Syst. Appl.* **2017**, *89*, 66–80. [[CrossRef](#)]
30. Wang, Y.; Li, Y.; Song, Y.; Rong, X. Facial Expression Recognition Based on Random Forest and Convolutional Neural Network. *Informatics* **2019**, *10*, 375. [[CrossRef](#)]
31. Jeong, M.; Ko, B.C. Driver's Facial Expression Recognition in Real-Time for Safe Driving. *Sensors* **2018**, *18*, 4270. [[CrossRef](#)]
32. Ra, S.T.; Kim, H.J.; Lee, S.H. A Study on Deep Learning Structure of Multi-Block Method for Improving Face Recognition. *Inst. Korean Electr. Electron. Eng.* **2018**, *22*, 933–940.
33. Facereader. Available online: <https://www.noldus.com/facereader/> (accessed on 16 December 2019).

34. Neighbor System of Korea. Available online: <http://www.neighbor21.co.kr/> (accessed on 3 January 2020).
35. Chung, K.; Shin, D.H.; Park, R.C. Detection of Emotion Using Multi-Block Deep Learning in a Self-Management Interview App. *Appl. Sci.* **2019**, *9*, 4830. [[CrossRef](#)]
36. Yuan, Q.; Xiao, N. Scaling-Based Weight Normalization for Deep Neural Networks. *IEEE Access* **2019**, *7*, 7286–7295. [[CrossRef](#)]
37. Pan, S.; Zhang, W.; Zhang, W.; Xu, L.; Fan, G.; Gong, J.; Zhang, B.; Gu, H. Diagnostic Model of Coronary Microvascular Disease Combined with Full Convolution Deep Network with Balanced Cross-Entropy Cost Function. *IEEE Access* **2019**, *7*, 177997–178006. [[CrossRef](#)]
38. Zhang, S.; Wang, Y.; Liu, M.; Bao, Z. Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM. *IEEE Access* **2017**, *6*, 7675–7686. [[CrossRef](#)]
39. Hu, Y.; Jin, Z.; Wang, Y. State Fusion Estimation for Networked Stochastic Hybrid Systems with Asynchronous Sensors and Multiple Packet Dropouts. *IEEE Access* **2018**, *6*, 10402–10409. [[CrossRef](#)]
40. Liu, L.; Luo, Y.; Shen, X.; Sun, M.; Li, B. β -Dropout: A Unified Dropout. *IEEE Access* **2019**, *7*, 36140–36153. [[CrossRef](#)]
41. Peng, D.; Liu, Z.; Wang, H.; Qin, Y.; Jia, L. A Novel Deeper One-Dimensional CNN with Residual Learning for Fault Diagnosis of Wheelset Bearings in High-Speed Trains. *IEEE Access* **2019**, *7*, 10278–10293. [[CrossRef](#)]
42. Shi, B.; Bai, X.; Yao, C. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2298–2304. [[CrossRef](#)]
43. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote. Sens.* **2017**, *9*, 848. [[CrossRef](#)]
44. Lucey, P.; Cohn, J.F.; Prkachin, K.M.; Solomon, P.E.; Matthews, I. Painful data: The UNBC-McMaster shoulder pain expression archive database. *Face Gesture* **2011**, 57–64. [[CrossRef](#)]
45. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).