Smart sentiment analysis system for pain detection using cutting edge techniques in a smart healthcare framework

Anay Ghosh¹ · Saiyed Umer² · Muhammad Khurram Khan³ · Ranjeet Kumar Rout⁴ · Bibhas Chandra Dhara⁵

Received: 21 October 2021 / Revised: 12 January 2022 / Accepted: 13 January 2022 / Published online: 29 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

A sentiment analysis system has been proposed in this paper for pain detection using cutting edge techniques in a smart healthcare framework. This proposed system may be eligible for detecting pain sentiments by analyzing facial expressions on the human face. The implementation of the proposed system has been divided into four components. The first component is about detecting the face region from the input image using a tree-structured part model. Statistical and deep learning-based feature analysis has been performed in the second component to extract more valuable and distinctive patterns from the extracted facial region. In the third component, the prediction models based on statistical and deep feature analysis derive scores for the pain intensities (no-pain, low-pain, and high-pain) on the facial region. The scores due to the statistical and deep feature analysis are fused to enhance the performance of the proposed method in the fourth component. We have employed two benchmark facial pain expression databases during experimentation, such as UNBC-McMaster shoulder pain and 2D Face-set database with Pain-expression. The performance concerning these databases has been compared with some existing state-of-the-art methods. These comparisons show the superiority of the proposed system.

Keywords Smart · Sentiment analysis · Pain expression · Recognition · Healthcare

1 Introduction

In present days Internet of Things (IoT) has gained immense popularity. Thus, IoT has been applied in smart homes, automatic driving, and traffic congestion monitoring. Like the field mentioned above, the IoT can also be used in artificial intelligence (AI) with a healthcare framework. The healthcare Industries is an essential sector in employment, health-related facilities, revenue, and contributor for healthy civilization in smart cities. These

Industries include clinical trials, hospitals, medical equipment and devices, telemedicine, outside patient clinic, health insurance, doctors, nurses, and other medical professionals [1]. A new era of technologies has enriched healthcare industries by introducing mobile-Health, and electronic-Health facilities.

The daily healthcare supports about the medicines to the patients are provided by the mobile-Health system [1]. The electronic-Health services use the information and communication technology for delivering facilities through computers and drugs to both doctors and patients with

Saiyed Umer and Bibhas Chandra Dhara contributed equally to this work.

✉ Saiyed Umer
saiyedumer@gmail.com
Anay Ghosh
anay.ghosh1@gmail.com
Muhammad Khurram Khan
mkhurram@ksu.edu.sa
Ranjeet Kumar Rout
ranjeetkumarout@nitsri.net
Bibhas Chandra Dhara
bcdhara@gmail.com

¹ Department of Computer Science & Engineering, University of Engineering & Management, Kolkata 700156, India

² Department of Computer Science & Engineering, Aliah University, Kolkata 700156, India

³ Center of Excellence in Information Assurance, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

⁴ Department of Computer Science & Engineering, National Institute of Technology, Srinagar 190006, India

⁵ Department of Information Technology, Jadavpur University, Kolkata 700098, India

digitally processing. Hence, in smart cities, e-Healthcare systems are supported by both mobile-Health and electronic-Health with technological advancement. These e-Healthcare systems benefit all medical professionals, doctors, patients, and businesses to build a healthy civilization in smart cities. Here the e-Healthcare system works for the patients by localizing their facial expressions, speech, voice, and gesture movement to monitor their health status. This work aims to build a sentiment analysis system (SAS) for detecting pain levels in patients by analyzing their expressions on the facial region. This SAS will enrich the facilities of the e-Healthcare system to make it easier to decide by doctors and nurses [2]. This proposed system provides less costly and more efficient health services to the patients. During integrating the proposed SAS to the e-Healthcare system, the other healthcare system requirements such as privacy, authenticity, security, access, and sophisticated tools have also been considered.

The proposed SAS will work together with the help of IoT, AI, and facial expression [3] of a person that the intensity of pain can be analyzed using their sentiments. In the real-time scenario, the SAS detects and analyzes the facial images of patients for deriving feelings remotely [1]. The pain analysis [4] is crucial for providing legitimate patient care and evaluating its effectiveness within the clinical ambiance. In general, a patient's self-report is used as the 'gold standard' reporting pain with the help of non-automatic pain evaluation appliances, which includes a visual analog scale and verbal numerical scale. After all, human realization and verdict on pain are distinctive, and the scale report may differ surprisingly among individuals. A patient's behavioral consciousness, especially the utilization of facial expression, is regarded as the essential critical behavioral pointer of pain, has been considered as a valuable process for the assessment of pain [5], most notably when the ability of the patient to communicate to pain is harmed. Patient's who are passing away, rationally paralyzed [6], badly ill and composed [7], or having mental decay [8], cancer in head and neck or brain realignment [9, 10] are specifically weak and requiring technology for providing faithful and authentic alerts about their pain to busy psychoanalyst. There may be other SASs related to facial emotion analysis, depression analysis, Parkinson's disease, contextual semantic search, intent analysis, etc. These SASs are based on text, images, and videos. Several statistical and deep learning techniques are employed to analyze the patterns for classifying the sentiments. Hence, the employed methodology for building the proposed system can also be employed for the other SAS discussed above.

The sentiment on pain assessment for nonverbal patients [4] has been described by the American Society for Pain Management Nursing (ASPMN). A hierarchical structure

of pain estimation is noted in the description, which includes facial expression-based observation as a valid approach for pain estimation. McGuire et al. [6] concluded that for the non-intellectual population, pain may be under-treated and under-realized, mainly for those who are not having the proper capacity for communicating their discomfort. In a study of patients on procedural pain [10], it is observed that procedural pain has a strong relationship with behavioral responses like shutting of eyes, wincing, and grimacing. It is comparatively infrequent that no facial development is present in the face during some painful procedure. Payen et al. [7] highly concluded that facial development can be used as a pain appraiser for critically ill unruffled patients. A pain assessment [8] study on older patients having intense mental illness gives proof that patients having pain display more conspicuous response in comparison with patients having no pain. It is also notable that medical inspections of facial development are exact in assessing the appearance of pain for those patients who cannot verbally communicate due to progressive mental illness. From the research, it is observed that facial development gives accurate measures for pain across cultural varieties [5] and the entire human lifespan, and there found a tremendous compatible facial development for painful impulses.

Measurement of facial development for pain not only introduces increased value in the presence of verbal reports but also is a crucial appraiser for pain in non-interacting patients. Despite several pain detection algorithms, there is a scope for introducing better approaches for pain recognition. This paper presents a deep learning-based approach and statistical approaches to building a SAS in a healthcare framework for smart cities. This system can classify the intensity of pain of patients into three different categories such as high pain (HP), low pain (LP), and no pain (NP). To fulfill the above objectives, the contributions of this paper are as follows:

- A novel method of statistical-based feature extraction has been adopted, which has been applied in different ways of image partitioning the facial region. The features extracted from various segments are well represented from its local to global representation that performs well for predicting pain expression in the proposed SAS.
- An efficient convolutional neural network (CNN)-based end-to-end deep learning framework has been proposed for deriving patterns as texture information from the facial region to improve the generalization and robustness of the prediction model for pain analysis.
- The obtained predictive models detect and learn powerful high-level features from the input image and extract more distinctive and discriminant features that

give effective result for the proposed SAS under various illumination changes, pose and age variations artifacts in the facial region.

- The proposed system can monitor different sentiments such as pain expression, anxiety, and depression to measure the treatment level in the smart healthcare framework.

This paper is organized as follows: Sect. 2 describes the related works for the proposed methodology with some basic terminology employed for the implementation of the proposed system. The implementation of the proposed system has been elaborated in Sect. 3. All the experiments with results and discussions have been reported in Sect. 4. Section 5 concludes this paper.

2 Related work

The following section describes the studies related to pain recognition from facial expressions, including the existing research, overall neural network, and general synopsis of deep learning methods. There exist several automatic pain detection methods, who are capable of recognizing pain from facial expressions. Some of them are non-deep learning methods, and the others are deep learning methods that have been introduced to detect pain from different facial expressions. The progress in this particular research field is also observed significantly over the past recent years. In terms of the general classification problem, the classification is done on the extracted facial features, where some traditional feature extraction techniques do feature extraction. Some supervised machine learning techniques like logistic regression (LR), support vector machine (SVM) and K-nearest neighbour have been used for classification purposes. In the case of traditional feature extraction methods, there exist some well-known methods like local binary pattern (LBP), active shape model (ASM), active appearance models (AAM), and Gabor wavelets employed for pain recognition systems. As an example [11, 12] used AAM based features while [13, 14] used Gabor wavelet and LBP features for detecting pain with SVM classifier.

For efficient identification of face and other objects from images as some deep learning-based CNN architectures have also been used in [15, 16]. In the case of CNN-based feature representation followed by classification for pain-type has been performed by [17]. In recent years it has been observed that the performances of CNN models are very high using GoogleNet [18], and AlexNet [19] CNN architectures. A new method of emotional analysis based on CNN-BiLSTM hybrid neural network has been proposed by Liu et al. [20]. A comparative study on bio-

inspired algorithms for sentiment analysis had been done by Yadav and Vishwakarma [21].

In computer vision, extracted features by pre-trained CNNs are used for various objectives such as object identification, emotion recognition, etc. It is also observed that the CNN pulled features' performances are pretty better than handcrafted features. In recent years the growth in the research field associated with deep learning provides solutions to several cutting edge problems. Deep learning-based algorithms are capable of revealing inherently obscured patterns in complicated datasets, thus for feature extraction and classification purposes [19]. A novel context-aware multimodal framework for Persian sentiment analysis [22] has been proposed by Dashtipour et al. Sagum [23] has proposed a method of measuring sentiments using emotion analysis based on the review comments on movies. Performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis has been built by Rustam et al. [24].

- *Local binary pattern, LBP* Reference [25] is a very well known and robust technique for texture identification from images. Previously the LBP technique was used to full fill the necessary measures of image contrast in the local scope. In the present scenario, the modification of image texture and local primitives LBP technique is used broadly. The working principle of the LBP operator is as follows: at first, consider an input image $\mathcal{I}$ and label every pixel with the help of a threshold value. Now select a particular pixel as a central pixel and consider a 3×3 neighborhood around that central pixel. Perform a comparison between each of the 3×3 neighborhood pixel values and central pixel values. Each cell value of the neighborhood is labeled by 0 or 1 depending on the cell value is less than or greater than that of the central value. Following this approach, binary bits will be generated based on that a bit string is considered. This binary string gives a local pattern of LBP codes computed for the entire image. Each LBP code for the entire input image $\mathcal{I}$ is considered as a micro-texton, and using these textons; a 256-bin histogram is prepared.
- *Histogram of oriented gradients, HoG* Reference [26] feature mainly retains the shape and texture oriented information. Further, it is successfully used for human detection; now, it is efficiently applied for improved face recognition. HoG [27] technique has similarity with scale-invariant feature transformation (SIFT) method. HoG feature techniques are determined using block-wise orientation with the help of Sobel edge detectors to measure horizontal and vertical gradient details. These details play an important role in determining whether feature representation's orientation and

Fig. 1 Block diagram of the proposed system

magnitude information correspond to an image. This feature extraction technique considers an image size 128×128 , then computes an 81-dimensional feature vector.

- Convolution neural network, CNN* from the above discussion, we have observed that the previous two techniques are mainly related to image feature extraction. To extract detailed and essential features from images, various researchers have introduced different feature extraction methodologies like binary robust independent elementary features (BRIEF), features from accelerated segment test (FAST), scale-invariant feature transform (SIFT), speeded up robust features (SURF), etc. [28]. Over time, it has been observed that these proposed feature extraction techniques mainly suffered from two types of problems: too over-engineered or too general as a result of that designing process of the filter becomes hard to generalize or too simple. To overcome these situations, some feature learning techniques have been proposed. Using these feature learning techniques, all the relevant and available features from the images are extracted automatically, which replaces the efforts of manual feature engineering. Neural networks are there to help in feature learning with multi-layer perceptron (MLP). Concerning image processing, MLP's have various drawbacks. For RGB images, the perceptrons are used by MLP's, and as a result, the weight amount within the network becomes unhealable within a short period. Another common problem of MLP is they are translation variants; thus, MLP's reaction to an input image and its shifted version is different. In the case of CNN mainly various kinds of filters are applied to capture relevant image features with the help of convolution operation. This convolution operation takes place throughout the image, and the filter is shifting pixel by pixel, starting from the top-left corner to the bottom-right corner. With the extracted feature for any

particular object within an image, the filters always guide how strongly a particular object can be found within an image. In that context, the location of the object and how many times the object appears within the image the performance does not get affected. In CNN, various layers are used, such as convolution, pooling, fully connected, and dense layers during the feature representation of an image.

3 Proposed method

In present days, cloud-based mobile applications associated with streaming data and real-time video processing are gaining popularity. In this type of application, we usually find two components related to it: front-end and back-end. Front-end components run within the mobile device itself, and the back-end components run on the cloud. In the cloud, the facilities of improved data processing and computing capabilities are available. Our proposed methodology helps for running complex applications smoothly on devices with available processing resources and power. In this section, we have presented a SAS for predicting pain intensity on a human facial region of a patient. This work may migrate cache and computer facilities near the end-user to provide quality modules for a smart healthcare framework. To carry out the implementation of the proposed system, this work has been divided into four components. The block diagram of the proposed SAS is shown in Fig. 1 as follows.

3.1 Image pre-processing

The actual research activity depends on the acquisition of pain-based facial expression as an image from video sources. The addition of pain-based facial expression always suffers from different noises during data acquisition. These noises affect the images heavily concerning the

Fig. 2 Image preprocessing task for the proposed sentiment analysis system

quality and information available within that image, resulting in inappropriate experimentation. In the image preprocessing task, a color image $\mathcal{I}$ is considered for extracting facial region for pain emotion detection. Here, for face detection, a tree-structured part model [29] has been employed, which works for all variants of face poses. This method computes 68 landmark points for the frontal face, while 39 landmark points have been extracted for the profile face. After face detection, the extracted facial region undergoes for normalization process. For this process, the bilinear image interpolation method has been employed to normalize the extracted face region $\mathcal{F}$ for a fixed image size of $N \times N$. The face detection process for the proposed SAS has been shown in Fig. 2.

Now, the normalized facial region $\mathcal{F}$ undergoes for feature computation task. Here for extracting valuable and distinctive features, (i) statistical-based feature representation followed by machine learning-based classification techniques and (ii) deep learning-based approaches have been employed. Both these approaches have been implemented individually. The responses generated from each approach are fused to make the final decision for the proposed system. The fusion of performance not only improves the performance but also increases the robustness of the proposed method.

3.2 Statistical based approach followed by machine learning techniques

The method that transforms an input image into a collection of features capable of representing and discriminating the input image is feature extraction. From the above face preprocessing steps shown in Fig. 2, the extracted face regions contain tones in regular or irregular patterns. We have analyzed these texture characteristics during feature computation to extract more discriminant and disjunctive

features that overcome intra-inter class similarity or dissimilarity and poor image quality problems due to different noise artifacts. For discriminatory feature representation techniques, the textures in the facial region are mainly analyzed using three approaches: (i) structural, (ii) statistical, and (iii) transformed-based approaches. Non-rigid facial muscular action exposes emotions. The feature extraction module works as an interpreter to extract admissible non-rigid information, which is very important for manifesting action units from the perceptions of computer vision. The non-rigid facial details can be customized broadly within the framework structure and a wealthy alternate of feature descriptors. There can be two categories of features such as appearance-based and geometric-based. Various geometric measurements related to coordinates and fiducial points are extracted in the geometric-based feature extraction technique. Similarly, the appearance-based approach extracts various features from pixel intensity values.

The statistical analysis of the texture pattern is more convenient and practical than the structural and transformed-based approaches. This analyzing approach evaluates more information from the pixel intensity values of the image. It helps to compile and present the appearance-based features from an image. Since the preprocessed face image $\mathcal{F}$ may have both regular and non-regular patterns [30]. So, the statistical-based approaches are more suitable to analyze regular as well as non-regular patterns. Here, during feature computation, both global and local to global feature representation schemes have been considered. During statistical-based feature computation, the facial image $\mathcal{F}$ of $N \times N \times 3$ is converted to a gray-scaled image $N \times N$. Then LBPHoG features are computed from whole $\mathcal{F}$ to obtain its global feature representation as f_g . To extract more local features, each $\mathcal{F}$ is divided into either two-equal halves (horizontally/vertically) ($f_{\mathcal{H}\mathcal{L}}/f_{\mathcal{V}\mathcal{L}}$), or four-equal halves ($f_{\mathcal{L}\mathcal{L}}$). Here, f_g indicates feature computed globally, $f_{\mathcal{H}\mathcal{L}}$ indicates feature computed horizontally local, $f_{\mathcal{V}\mathcal{L}}$ indicates feature computed vertically local, and $f_{\mathcal{L}\mathcal{L}}$ indicates feature computed locally to globally [31]. The $f_{\mathcal{H}\mathcal{L}}$ is a feature vector obtained by concatenation of features extracted from two horizontal halves of $\mathcal{F}$. Similarly, $f_{\mathcal{V}\mathcal{L}}$ is a feature vector obtained by concatenation of features extracted from two vertical halves of $\mathcal{F}$. $f_{\mathcal{L}\mathcal{L}}$ is a feature vector obtained by concatenation of features extracted from four vertical halves (horizontally and then vertically) of $\mathcal{F}$.

The feature extraction of these schemes have been demonstrated in Fig. 3 where f_g , $f_{\mathcal{H}\mathcal{L}}$, $f_{\mathcal{V}\mathcal{L}}$, and $f_{\mathcal{L}\mathcal{L}}$ feature vectors are described.

Now the features extracted from the different schemes (shown in Fig. 3) undergoes for machine learning classification task. The objective of the classification task is to

Fig. 3 Statistical based approach for feature representation from the facial region $\mathcal{F}$

recognize the pain-intensity types on the facial region of the patients. Due to the presence of various discriminant values in the extracted feature vectors, the classification task gives a predictive decision on the learned features by the classifier [32]. For classification purpose, the decision tree [33], SVM [32], K-nearest neighbour [34] and LR [35] classifiers have been employed. The functionality of these classifiers is based on (i) input, (ii) learning algorithm, and

(iii) target function (model). The mathematics behind these classifiers differ, and thus the different performances are obtained for the same training–testing datasets. In this work, the pain-detection on the facial region is 2-class [(NP), and Pain (P)] or 3-class [NP, LP, and HP] problem. So, both binary and multi-class classification techniques have been employed. The proposed system’s classification

performance due to these features is represented in terms of Accuracy [correct recognition rate (%)] and F1-Score.

3.3 Deep learning based approach

Recently in various fields, the deep learning-based approaches successfully developed discriminative features from images and videos using convolution neural network (CNN) techniques. The CNN approaches have gained much popularity and have shown remarkable efficiency for complex domain image analysis problems. In CNN, the neural network mainly uses various kernels for performing mixed convolution operations on images. After these convolution operations, the pooling technique is operated by a neural network with weight sharing for preparing different training parameters of the network [36]. Nowadays, it has been observed that different research-based problems such as scene understanding, object recognition, object detection, face recognition, texture classification, and many more from the computer vision field are looked over and worked out comfortably with the help of CNN models [18]. The modern state-of-the-art edge caching of generic CNN feature representation for facial expression recognition problems [37] could compete with computer vision's statistical and structural-based methods. These generic CNN features can cope with the articulation and occlusion face images captured in an unconstrained environment and obtain better performance. A SAS for analyzing facial expressions in a facial expression recognition with trade-offs between data augmentation and deep learning features has been proposed in [28].

In this work, we have employed deep learning-dependent convolution neural network (CNN) architectures [18] to perform feature learning followed by classification of images into two or three pain classes based on analysis of emotions on the facial region. Since the facial expression recognition system suffers from the over-fitting problem, the lack of sufficient training samples and bias cause expression variations such as age variation, head-pose, identity bias, and illumination variation. The proposed methods focus on these issues and overcome the computational complexity of the CNN model [38]. The CNN architectures' hybrid nature helps extract both the shape and texture information from the images. These shapes and textural information provide both geometrical and appearance features in feature descriptors in the proposed system. The usage of graphical processing units (GPUs) helps overcome substantial computational power, which is the main drawback of the CNN model. The GPUs are basically intended to accomplish huge amounts of computational tasks like numerous representation levels and abstraction to manage complex data patterns, transfer learning, pretraining, and fine-tuning methods that can be

applied to existing CNN models. To train the CNN architecture, a bulky database has been used and the weights of the network are adjusted as per the number of classes of the given problem [39].

Here, a CNNs architecture using a deep learning approach has been proposed to perform feature analysis to classify the discriminating patterns from the facial region. This proposed CNN architecture contains six to eight layers of deep image perturbation. Each layer of the architecture performs operations first convolution, then activation function (rectified linear unit) is applied on the convolved feature maps. To extract discriminant features from rectified feature maps, the max-pooling operations have been performed that bring most prominent features which undergo the batch normalization layer to introduce regularization within the standardized features. At the end of the CNN layers, two fully connected layers are added that flatten the extracted features from the top of the previous layers. To enhance the performance of the proposed CNN, some novel mechanisms such as fine-tuning, regularization, and transfer learning techniques are employed to reduce data imbalanced and over-fitting problems. The introduction of batch normalization, label smoothing, regularization, and the mixture of optimizers bring advantages to the used CNN architecture.

The convolution operation is the primary operator in this network, the main building block of a CNN architecture. This layer is used to extract features from the images by performing convolution operations. The convolution operation is executed over the input image with the help of a $t \times t$ sized kernel or filter and then generates feature maps. The convolution operation is performed by sliding the filter followed by non-linearity over the input. At every location, matrix multiplication is performed and sums the result onto the feature map. Finally, we use a method of fine-tuning parameters to enhance the performance of the proposed SAS. The max-pool calculates the maximum value for patches of a feature map and uses it to create a down-sampled feature map. It is usually used after a convolutional layer. The benefits of the max-pooling layer are: (i) it is translation invariance; (ii) reduced computational costs; (iii) faster matching; and (iv) improved accuracy. A fully connected layer of a CNN architecture reduces the full feature maps to a single vector of class scores and produces a resulting vector of size of class labels. In a dense layer, all input layers are connected to the output layers by a weight value. Batch normalization acts as a regularizer and allows the model to use higher learning rates [40]. It is used in various image classification problems and achieves higher accuracy with fewer training steps. Batch normalization also has a beneficial effect on the gradient flow through the network by reducing the dependence of gradients on the scale of the parameters or

Fig. 4 The employed CNN architecture for the proposed system

their initial values. Therefore in this research work, we have proposed a CNN architecture using the above-discussed layers. The design of the proposed CNN architecture has been provided in Fig. 4. The detailed description of this CNN architecture with employed input–output hidden layers, output shapes of the convoluted image, input image size, and parameters generated at each layer have been shown in Table 1 respectively for better understanding and clarity about the models.

Here, the input layer always accepts the image as input data. On the other hand, different kernels (filter banks) perform the convolution operation to get the convoluted image (feature map) for individual filters. The convolution exertion enhances the difficulties with the improved image size, and with the incremented number of filters, the network will require massive computations [41]. Here for the filter sets, the parameters are regarded as fine-tuned weights. In max-pooling layers [42], the computational barrier gets reduced with the decrements in the number of parameters inside the network. For maximum occasion in the max-pooling layer, the size of the engaged filter for every individual feature map (fetched from the previous layer) is 2×2 . After that, a 2-down-sample step with the gained average or minimum or maximum values are calculated between 4 digits alongside horizontal and later on in a vertical direction. The next layer is the fully connected layer. In fully connected layer [39] all the features are transformed from the previous layer to the 1-dimensional vector. Another kind of fully connected layer is the dense layer [39] which performs a linear operation in it. In linear operations, input and output are connected, and as an outcome, it generates the probability scores with the help of an activation function like Softmax.

Here some factors such as fine-tuning and transfer learning affect the performance of the proposed CNN architecture. The fine-tuning allows the based CNN model to derive the higher-order feature representations from building more discriminant features for the proposed SAS. The fine-tuning techniques freeze some layers and respective parameters and retrain the model to reduce computational overheads. Transfer learning helps a trained

model of a specific problem be effectively used as a generic model for other related issues. The model that has been trained earlier is known as the pre-trained model. Our proposed deep learning convolution neural network model uses a transfer learning technique. The weights of the pre-trained model and or a set of layers from the pre-trained model are further used in the new model to solve such similar problems. The benefits of using transfer learning are that it reduces the training time and results in lower generalization error.

3.4 Fusion techniques

Hence to improve the performance of the proposed SAS for pain classification, the scores of the SVM classifier using features extracted from the Statistical based approach and the scores from the deep learning-based approach are fused with the help of different score level fusion techniques: *Sum*, *Product* and *Weighted-sum*. Let the post-classification scores using the statistical-based approach with LBP feature, the statistical-based approach with HoG feature, and the deep learning-based approach be δ_1 , δ_2 and δ_3 respectively. So, the rule-based techniques with (i) *Sum* is defined as $\delta = \delta_1 + \delta_2 + \delta_3$, (ii) *Product* is defined as $\delta = \delta_1 \times \delta_2 \times \delta_3$ and (iii) *Weighted-sum* is defined as $\delta = \omega_1 \times \delta_1 + \omega_2 \times \delta_2 + \omega_3 \times \delta_3$, where δ is considered as fused score on the other hand ω_1 , ω_2 and ω_3 are the corresponding weights so that $\omega = \omega_1 + \omega_2 + \omega_3$.

4 Experiments

In this research work, for the proposed SAS, Python language is used, and as an operating system, Windows-10 along with 16 GB RAM and Intel Core i5 processor 3.20 GHz has been used. At the time of implementation, we have used various packages of Python with Theano [43], and Keras [44] essential packages. For deep learning-oriented approaches like the proposed CNN architecture, the Theano and Keras packages have been employed, and for other work, various Python libraries have been employed.

Table 1 The proposed CNN architecture for the input size 96×96 with layers, output shape, image size and parameters

Layers		Output shape		Image size		Parameters			
Block-1									
Conv2D(3 × 3)@32		(n, n, 32)	(96, 96, 32)			896			
Batch norm		(n, n, 32)	(96, 96, 32)			128			
Activation ReLU		(n, n, 32)	(96, 96, 32)			0			
Maxpool2D(2 × 2)		(n1, n1, 32), n1 = n/2	(48, 48, 32)			0			
Dropout		(n1, n1, 32), n1 = n/2	(48, 48, 32)			0			
Layers		Parameters		Output shape		Image size		Parameters	
Block-2									
Conv2D(3 × 3)@32		(\$n_{\{1\}}, n_{\{1\}}, 32)	(48, 48, 32)	(3 × 3 × 3) + 1 × 32 = 9248	Conv2D(3 × 3)@96	(n3, n3, 96)	(12, 12, 96)	(3 × 3 × 64) + 1 × 96 = 55,392	
Batch norm		(\$n_{\{1\}}, n_{\{1\}}, 32)	(48, 48, 32)	4 × 32 = 128	Batch norm	(n3, n3, 96)	(12, 12, 96)	4 × 96 = 384	
Activation ReLU		(\$n_{\{1\}}, n_{\{1\}}, 32)	(48, 48, 32)	0	Activation ReLU	(n3, n3, 96)	(12, 12, 96)	0	
Maxpool2D(2 × 2)		(n2, n2, 32)n2 = n1/2	(24, 24, 32)	0	Maxpool2D(2 × 2)	(n4, n4, 96)n4 = n3/2	(6, 6, 96)	0	
Dropout		(n2, n2, 32)	(24, 24, 32)	0	Dropout	(n4, n4, 96)	(6, 6, 96)	0	
Block-3									
Conv2D(3 × 3)@64		(n2, n2, 64)	(24, 24, 64)	((3 × 3 × 32) + 1) × 64 = 18,496	Conv2D(3 × 3)@96	(n4, n4, 96)	(6, 6, 96)	((3 × 3 × 96) + 1) × 96 = 83,040	
Batch norm		(n2, n2, 64)	(24, 24, 64)	4 × 64 = 256	Batch norm	(n4, n4, 96)	(6, 6, 96)	4 × 96 = 384	
Activation ReLU		(n2, n2, 32)	(24, 24, 64)	0	Activation ReLU	(n4, n4, 96)	(6, 6, 96)	0	
Maxpool2D(2 × 2)		(n3, n3, 64)n3 = n2/2	(12, 12, 64)	0	Maxpool2D(2 × 2)	(n5, n5, 96)n5 = n4/2	(3, 3, 96)	0	
Dropout		(n3, n3, 64)	(12, 12, 64)	0	Dropout	(n5, n5, 64)	(3, 3, 96)	0	
Block-6									
Conv2D(3 × 3)@64		(n5, n5, 64)	(3, 3, 64)					55,360	
Batch norm		(n5, n5, 64)	(3, 3, 64)					256	
Activation ReLU~		(n5, n5, 64)	(3, 3, 64)					0	
Maxpool2D(2 × 2)		(n6, n6, 64), n6 = n5/2	(1, 1, 64)					0	
Dropout		(\$n_{\{6\}}, n_{\{6\}}, 64)	(1, 1, 64)					0	
Layers		Output shape		Image size		Parameters			
Flatten									
		(1, n6 × n6 × 64)	1, 64				0		
Dense		(1, 256)	(1, 256)				(1 + 64) × 256 = 16,640		
Batch normalization		(1, 256)	(1, 256)				1024		
Activation ReLU		(1, 256)	(1, 256)				0		
Dropout		(1, 256)	(1, 256)				0		
Dense		(1, 256)	(1, 256)				(256 + 1) × 256 = 65,792		

Table 1 (continued)

Layers	Output shape	Image size	Parameters
Batch normalization	(1, 256)	(1, 256)	1024
Activation ReLU	(1, 256)	(1, 256)	0
Dropout	(1, 256)	(1, 256)	0
Dense	(1, 7)	(1, 7)	$(256 + 1) \times 7 = 1799$
Total parameters for the input image size			310,247
Total number of trainable parameters			308,455
Non-trainable parameters			1792

Here the pain intensity of a person has been analyzed using image data (image data contains pain emotion on faces).

4.1 Database used

In this work, for the proposed SAS, we have employed two image datasets. Our first employed dataset is UNBC-McMaster [45] shoulder pain expression archive database. This database comprises 129 participants (63 male and 66 female), where the participants have shoulder pain, and three physiotherapy clinics have identified their problems. The videos are captured on the campus of McMaster University. During data acquisition, the participants are assumed to have suffered from tendinitis, bursitis, rotator cuff injuries, arthritis, bone spur, dislocation, subluxation, impingement syndromes, capsulitis, and these causes shoulder pain. Here from each video, the images are extracted where each image is assigned a label from ‘No pain’ to ‘High-intensity pain’ classes. In this work, we have categorized these images into three categories, i.e., non-aggressive (NAG) (NP) ($\mathcal{U}_1$), covertly aggressive (CAG) (LP) ($\mathcal{U}_2$) and overtly aggressive (OAG) (HP) ($\mathcal{U}_3$). The description of the samples of these three classes has been demonstrated in Table 2. Some images of this database have been shown in Fig. 5. Our second employed database is 2D Face set database with Pain-expression Set [46] $\mathcal{D}_2$ database. This database is composed of 599 images from 10 men and 13 women subjects. Usually, this database has a 2-class problem. There are 298 images have ‘No-pain’ ($\mathcal{W}_1$) while 298 has ‘Pain’ ($\mathcal{W}_2$) expression. The image samples of this database are shown in Fig. 6 (Table 3).

4.2 Results and discussion

In this work, during image preprocessing, the tree-structured part model is employed for extracting the face region and then normalized the image into $\mathcal{F}$ of size $200 \times 200 \times 3$ from the input image. This work aims to identify the pain level among three classes ‘No-Pain’, ‘Low-Pain,’ and ‘High-Pain’. Here both statistical-based followed by machine learning approach and deep learning-based approach classification techniques have been employed individually for the proposed system. Then, the scores due to these techniques are fused to enhance the performance of the proposed system. In Sect. 3.2, the variants of statistical-based feature representation schemes have been demonstrated using HoG and LBP feature extraction techniques. Hence using these features, from the facial region $\mathcal{F} f_{\mathcal{F}}^{HoG} \in \mathbb{R}^{1 \times 81}$ and $\mathcal{F} f_{\mathcal{F}}^{LBP} \in \mathbb{R}^{1 \times 256}$ feature vectors are obtained using *Scheme*₁ (*P1*). The feature vectors extracted from the facial regions undergo different classifiers such as LR [35], K-nearest neighbour [34], decision

Table 2 Description of UNBC image database for the proposed system

3-Class problem	Sample
No-pain (NP, $\mathcal{U}_1$)	40,029
Low-pain (LP, $\mathcal{U}_2$)	2909
High-pain (HP, $\mathcal{U}_3$)	5460
2-Class problem	Sample
No-pain ($\mathcal{V}_1$)	40,029
Pain ($\mathcal{V}_2$)	8369

tree [33], Random Forest, and SVM [32]. Each employed dataset is divided into 50% of data in the training set and 50% of data in the testing set. Then, a 10-fold cross-

validation technique was used, and the average performance for testing data was reported for the proposed system. The performance of the proposed system for the UNBC 2-class problem due to these classifiers has been shown in Table 4.

From Table 4, it has been observed that among LR, K-nearest neighbour, decision tree, Random Forest, and SVM classifier, the proposed system has attained better performance for the SVM classifier. So, for further experiment the SVM classifier has been employed. Now using $Scheme_1$ (P_1), $Scheme_2$ (P_2), $Scheme_3$ (P_3), and $Scheme_4$ (P_4), from each facial region $\mathcal{F}$, using HoG features, $f_g^{HoG} \in \mathbb{R}^{1 \times 81}$, $f_H \mathcal{L}^{HoG} \in \mathbb{R}^{1 \times 162}$, $f_V \mathcal{L}^{HoG} \in \mathbb{R}^{1 \times 162}$, and $f_{\mathcal{L}} \mathcal{G}^{HoG} \in$

Fig. 5 Some samples of the employed UNBC database **Fig. 6** Some samples of the employed $\mathcal{D}_2$ database

Table 3 Description of image database $\mathcal{D}_2$ for the proposed system

2-Class problem	Sample
No-pain ($\mathcal{W}_1$)	298
Pain ($\mathcal{W}_2$)	298

$\mathbb{R}^{1 \times 324}$ feature vectors are obtained. Similarly, using LBP features, $f_{\mathcal{G}}^{LBP} \in \mathbb{R}^{1 \times 256}$, $f_{\mathcal{H}} \mathcal{L}^{LBP} \in \mathbb{R}^{1 \times 512}$, $f_{\mathcal{V}} \mathcal{L}^{LBP} \in \mathbb{R}^{1 \times 512}$, and $f_{\mathcal{L}} \mathcal{G}^{LBP} \in \mathbb{R}^{1 \times 1024}$ feature vectors are obtained from $\mathcal{F}$ facial region. The purpose of these feature extraction is to extract more and more local features such that the performance of the proposed system is increased. The performance due to these feature vectors for UNBC 2-class problem has been reported through the bar-graphs in Fig. 7 where the x-axis shows the schemes types while the y-axis shows the performance of the SVM classifier with respect to each feature vector scheme.

From Fig. 7 it has been observed that for *Scheme*₄ (P_4), the performance is better for both HoG, and LBP features. Hence, the performance of the *Scheme*₄ (P_4) statistical-based feature representation scheme has been adopted for further experiment for UNBC 2-class, UNBC 3-class, and $\mathcal{D}_2$ 2-class dataset, and these performances are reported in Table 5. From this table, it has been observed that the proposed system has achieved 78.34% accuracy with 0.7723 F1-Score using HoG, 80.29% accuracy with 0.7977 F1-Score using LBP feature for UNBC 2-class problem. Similarly, the maximum accuracy of the UNBC 3-class problem is 79.14% using HoG and 80.08% using LBP features. For $\mathcal{D}_2$ database 2-class problem, the maximum accuracy is 61.40% using HoG, and 63.12% using the LBP feature.

4.3 Feature learning and classification

In feature learning with classification mechanism, there exist two particular activities such as (i) image preprocessing and (ii) feature learning with classification. This image preprocessing technique is the same as mentioned above (using the TSPM method). Then the normalized face

Fig. 7 Performance due to different image partitioning methods discussed in Sect. 3.2**Table 5** The proposed system's performance using HoG and LBP statistical based approaches

2-Class problem using UNBC database				
Scheme	HoG		LBP	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
P1	75.85	0.7405	76.87	0.7551
P4	78.34	0.7723	80.29	0.7977
3-Class problem using UNBC database				
Scheme	HoG		LBP	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
P1	77.85	0.7533	78.71	0.7702
P4	79.14	0.7751	80.08	0.7891
2-Class problem using $\mathcal{D}_2$ database				
Scheme	HoG		LBP	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
P1	58.19	0.5678	55.03	0.5218
P4	61.40	0.5954	63.12	0.6191

region $\mathcal{F}$ is resized to 96×96 image size. Then the resized face images from the training data set to go through the proposed CNN architecture (Fig. 4) where the input to the

Table 4 Performance due to *Scheme*₁ (P_1) feature extraction followed by different machine learning classifiers

2-Class problem using UNBC database				
Classifier	HoG		LBP	
	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
Logistic regression	74.28	0.7267	74.39	0.7266
K-nearest neighbour	71.33	0.7045	73.56	0.7285
Decision tree	75.44	0.7413	73.62	0.7182
Random Forest	74.45	0.7434	76.07	0.7597
Support vector machine	75.85	0.7405	76.87	0.7551

Table 6 The proposed system's performance using deep learning-based approach

2-Class problem using UNBC database			
Accuracy (%)	F1-Score	Training-time (s)	Testing-time (s)
81.54	0.7962	39.63	0.03
3-Class problem using UNBC database			
Accuracy (%)	F1-Score	Training-time (s)	Testing-time (s)
81.33	0.8193	41.29	0.03
2-Class problem using $\mathcal{D}_2$ database			
Accuracy (%)	F1-Score	Training-time (s)	Testing-time (s)
74.59	0.7391	23.31	0.03

architecture is $96 \times 96 \times 3$ image. The CNN architecture (Fig. 4) is trained in such a way that it would perform both feature computation and pain classification tasks. For a better understanding of the functionality of these architectures, at first, we have started the experiment with training the CNN architecture with training samples. In contrast, the performance of the trained CNN model is evaluated using the remaining testing samples. Learning the parameters in any CNN architecture is a crucial task, and it depends on two factors, i.e., epochs and batches. Both these factors affect the learning capabilities of the architecture during training the samples in the network. So, for this work, a trade-off between 700 epochs and 16 batches has been established that improves the performance of the proposed SAS. To improve the performance of the proposed system, both fine-tuning and transfer learning

techniques have been employed. Hence, the performance of the proposed system using UNBC 2-class, UNBC 3-class, and $\mathcal{D}_2$ 2-class have been demonstrated in Table 6. These performances have been reported in terms of Accuracy, F1-Score, Training-time, and Testing-time, respectively.

From Table 6, it has been observed that the proposed system attains 81.54% accuracy for UNBC 2-class problem, 81.33% accuracy for UNBC 3-class problem, and 74.59% accuracy for $\mathcal{D}_2$ 2-class problems. Compared to these performances with the performances achieved due to the statistical-based feature approach, the proposed system gives better results for the deep learning-based approach. Hence, the deep learning approach better represents facial features for pain analysis than the statistical-based features. In this work, the facilities of both these approaches are fused to enhance the performance of the proposed system. The fused performance is discussed below section.

4.4 Fusion

Here, the classification scores are obtained using statistical (HoG and LBP features) and deep-learning-based features. During fusion, post-classification techniques have been employed. These classification scores are generated by the corresponding machine learning classifiers and CNN models. In post-classification fusion, the *Sum-rule*, *Product-rule*, and *Weighted Sum-rule* methods are applied on the classification scores. The fused performance of the proposed system due to different combinations of the fusion methods has been demonstrated in Table 7. From this table it has been observed that among *Sum-rule*, *Product-rule*, and *Weighted Sum-rule* methods, the proposed

Table 7 Performance due to fusion techniques in the proposed sentiment analysis system for the pain detection

Method	2-Class UNBC Accuracy (%)	3-Class UNBC Accuracy (%)	2-Class $\mathcal{D}_2$ Accuracy (%)
HoG	78.34	79.14	61.40
LBP	80.29	80.08	63.12
CNN	81.54	81.33	74.59
Sum-rule (HoG–LBP)	80.39	80.13	63.40
Sum-rule (HoG–CNN)	81.54	81.27	69.52
Sum-rule (LBP–CNN)	81.45	81.11	70.81
Sum-rule (HoG–LBP–CNN)	81.33	81.21	72.65
Product-rule (HoG–LBP)	80.71	80.34	65.11
Product-rule (HoG–CNN)	81.02	81.56	71.54
Product-rule (LBP–CNN)	82.29	82.32	71.39
Product-rule (HoG–LBP–CNN)	82.54	83.71	75.67
Weighted Sum-rule (HoG–LBP)	79.92	78.93	71.03
Weighted Sum-rule (HoG–CNN)	80.71	80.21	73.85
Weighted Sum-rule (LBP–CNN)	81.29	81.23	74.45
Weighted Sum-rule (HoG–LBP–CNN)	81.67	81.23	74.19

Table 8 Performance comparison of the proposed image-based sentiment analysis system with the other competing methods using UNBC-McMaster shoulder pain database

Method	Accuracy (%)	Remarks
Vgg16 [47]	76.84	Class (3), Train/Test Split
ResNet50 [48]	79.32	Class (3), Train/Test Split
Inception-v3 [49]	79.64	Class (3), Train/Test Split
Werner et al. [50]	75.50	Class (3), Train/Test Split
Lucey et al. [45]	81.80	Class (3), Train/Test Split
Proposed	83.71	Class (3), Train/Test Split

Table 9 Performance comparison of the proposed image-based sentiment analysis system with the other competing methods using $\mathcal{D}_2$ database

Method	Accuracy (%)	Remarks
Vgg16 [47]	61.76	Class (2), Train/Test Split
ResNet50 [48]	64.05	Class (2), Train/Test Split
Inception-v3 [49]	63.19	Class (2), Train/Test Split
Werner et al. [50]	65.19	Class (2), Train/Test Split
Lucey et al. [45]	74.33	Class (2), Train/Test Split
Proposed	75.67	Class (2), Train/Test Split

system has better performance for *Product-rule* fusion method with fusion of classification scores due to LBP, HoG, and CNN feature representation schemes. Hence, the proposed system attains accuracy of 82.54% for 2-class UNBC, 83.71% for 3-class UNBC, and 75.67% for 2-class $\mathcal{D}_2$ database. These performances are considered for comparison purposes in the next section.

4.5 Comparison

During the comparison, we have compared the performance of this proposed system with some existing state-of-the-art methods for the 3-class UNBC-McMaster shoulder pain database and 2-class $\mathcal{D}_2$ database. We have implemented Vgg16 [47], ResNet50 [48], Inception-v3 [49], Werner et al. [50], and Lucey et al. [45] methods and obtain the performance of the proposed system under the same training–testing protocol as used in the proposed system. The comparison of the proposed system for 3-class UNBC and for 2-class $\mathcal{D}_2$ database have been reported in Tables 8 and 9. From these performances, it has been observed that the proposed system has obtained better performance than the other competing methods concerning both the employed databases.

5 Conclusion

The IoT-based low-power sensors improve the functionality of smart healthcare systems. A smart healthcare system intelligently handles the patient monitoring system remotely. Smart devices compatible with the implanted sensors under the skin continuously monitor glucose levels in a diabetic patient. Similarly, a smart imaging device can monitor different sentiments such as anxiety, depression, and drug addiction to measure the dose level. This paper proposes a SAS by capturing the facial region to predict the pain intensity level as sentiments. The implementation of the proposed system has four components. In the first component, the face region is detected from the input image. The extracted face region undergoes both statistical and deep learning-based feature analysis in the second component. A prediction model based on statistical feature analysis derives scores; similarly, the deep learning-based prediction model computes scores for the pain intensities (NP, LP, and HP) on the facial region. These scores are used to obtain the performance of the proposed system in the third component. Finally, the scores due to the statistical feature analysis and deep feature analysis are fused to enhance the performance of the proposed method in the fourth component. For the experimental purpose, two databases are used: UNBC-McMaster shoulder pain and the 2D Face-set database with Pain-expression. The performance of the proposed system has been compared with some existing methods for these databases. Comparing the performance of the competing methods and the proposed method shows the superiority of the proposed approach.

Acknowledgements The work of Dr. Muhammad Khurram Khan is supported by Researchers Supporting Project Number (RSP-2021/12), King Saud University, Riyadh, Saudi Arabia.

References

1. 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* **5**, 10871–10881 (2017)
2. Abu-Saad, H.H.: Challenge of pain in the cognitively impaired. *Lancet (Lond. Engl.)* **356**(9245), 1867–1868 (2000)
3. Arif-Rahu, M., Grap, M.J.: Facial expression and pain in the critically ill non-communicative patient: state of science review. *Intensive Crit. Care Nurs.* **26**(6), 343–352 (2010)
4. Herr, K., Coyne, P.J., Key, T., Manworren, R., McCaffery, M., Merkel, S., Pelosi-Kelly, J., Wild, L.: Pain assessment in the nonverbal patient: position statement with clinical practice recommendations. *Pain Manag. Nurs.* **7**(2), 44–52 (2006)
5. de Williams, A.: Facial expression of pain: an evolutionary account. *Behav. Brain Sci.* **25**(4), 439–455 (2002)
6. McGuire, B., Daly, P., Smyth, F.: Chronic pain in people with an intellectual disability: under-recognised and under-treated? *J. Intellect. Disabil. Res.* **54**(3), 240–245 (2010)

7. Payen, J.-F., Bru, O., Bosson, J.-L., Lagrasta, A., Novel, E., Deschaux, I., Lavagne, P., Jacquot, C.: Assessing pain in critically ill sedated patients by using a behavioral pain scale. *Crit. Care Med.* **29**(12), 2258–2263 (2001)
8. Manfredi, P.L., Breuer, B., Meier, D.E., Libow, L.: Pain assessment in elderly patients with severe dementia. *J. Pain Symptom Manag.* **25**(1), 48–52 (2003)
9. Hadjistavropoulos, T., Herr, K., Turk, D.C., Fine, P.G., Dworkin, R.H., Helme, R., Jackson, K., Parmelee, P.A., Rudy, T.E., Beattie, B.L., et al.: An interdisciplinary expert consensus statement on assessment of pain in older persons. *Clin. J. Pain* **23**, 1–43 (2007)
10. Puntillo, K.A., Morris, A.B., Thompson, C.L., Stanik-Hutt, J., White, C.A., Wild, L.R.: Pain behaviors observed during six common procedures: results from Thunder Project II. *Crit. Care Med.* **32**(2), 421–427 (2004)
11. Ashraf, A.B., Lucey, S., Cohn, J.F., Chen, T., Ambadar, Z., Prkachin, K.M., Solomon, P.E.: The painful face-pain expression recognition using active appearance models. *Image Vis. Comput.* **27**(12), 1788–1796 (2009)
12. Lucey, P., Cohn, J., Howlett, J., Lucey, S., Sridharan, S.: Recognizing emotion with head pose variation: identifying pain segments in video. *IEEE Trans. Syst. Man Cybern. B* **41**(3), 664–674 (2011)
13. Littlewort-Ford, G., Bartlett, M.S., Movellan, J.R.: Are your eyes smiling? Detecting genuine smiles with support vector machines and Gabor wavelets. In: *Proceedings of the 8th Joint Symposium on Neural Computation*. Citeseer (2001)
14. Shan, C.: Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit. Lett.* **33**(4), 431–437 (2012)
15. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708 (2014)
16. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2892–2900 (2015)
17. Hu, F., Xia, G.-S., Hu, J., Zhang, L.: Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **7**(11), 14680–14707 (2015)
18. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012)
20. Liu, Z.-X., Zhang, D.-G., Luo, G.-Z., Lian, M., Liu, B.: A new method of emotional analysis based on CNN-BiLSTM hybrid neural network. *Clust. Comput.* **23**(4), 2901–2913 (2020)
21. Yadav, A., Vishwakarma, D.K.: A comparative study on bio-inspired algorithms for sentiment analysis. *Clust. Comput.* **23**(4), 2969–2989 (2020)
22. Dashtipour, K., Gogate, M., Cambria, E., Hussain, A.: A novel context-aware multimodal framework for Persian sentiment analysis. *arXiv preprint* (2021). [arXiv:2103.02636](https://arxiv.org/abs/2103.02636)
23. Sagum, R.A.: An application of emotion detection in sentiment analysis on movie reviews. *Turk. J. Comput. Math. Educ.* **12**(3), 5468–5474 (2021)
24. Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., Choi, G.S.: A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. *PLoS ONE* **16**(2), 0245909 (2021)
25. Medjahed, S.A.: A comparative study of feature extraction methods in images classification. *Int. J. Image Graph. Signal Process.* **7**(3), 16 (2015)
26. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893. IEEE (2021)
27. Kobayashi, T.: BFO meets HoG: feature extraction based on histograms of oriented PDF gradients for image classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 747–754 (2013)
28. Umer, S., Rout, R.K., Pero, C., Nappi, M.: Facial expression recognition with trade-offs between data augmentation and deep learning features. *J. Ambient Intell. Humaniz. Comput.* (2021). <https://doi.org/10.1007/s12652-020-02845-8>
29. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886. IEEE (2012)
30. Umer, S., Dhara, B.C., Chanda, B.: Face recognition using fusion of feature learning techniques. *Measurement* **146**, 43–54 (2019)
31. Umer, S., Dhara, B.C., Chanda, B.: An iris recognition system based on analysis of textural edgeness descriptors. *IETE Tech. Rev.* **35**(2), 145–156 (2018)
32. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, Berlin (2013)
33. Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **21**(3), 660–674 (1991)
34. Manocha, S., Girolami, M.A.: An empirical analysis of the probabilistic K-nearest neighbour classifier. *Pattern Recognit. Lett.* **28**(13), 1818–1824 (2007)
35. Hosmer, D.W., Jr., Lemeshow, S., Sturdivant, R.X.: *Applied Logistic Regression*, vol. 398. Wiley, New York (2013)
36. Saxena, A.: Convolutional neural networks: an illustration in TensorFlow. *XRDS Crossroads ACM Mag. Stud.* **22**(4), 56–58 (2016)
37. Tian, Y.-I., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2), 97–115 (2001)
38. Hossain, S., Umer, S., Asari, V., Rout, R.K.: A unified framework of deep learning-based facial expression recognition system for diversified applications. *Appl. Sci.* **11**(19), 9174 (2021)
39. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481 (2018)
40. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, PMLR, pp. 448–456 (2015)
41. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* **8**(1), 98–113 (1997)
42. Liu, M., Li, S., Shan, S., Chen, X.: AU-aware deep networks for facial expression recognition. In: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–6. IEEE (2013)
43. Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a CPU and GPU math expression compiler. In: *Proceedings of the Python for Scientific Computing Conference (SciPy)*, Austin, TX, vol. 4, pp. 1–7 (2010)
44. Gulli, A., Pal, S.: *Deep Learning with Keras*. Packt Publishing Ltd., Birmingham (2017)

45. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Matthews, I.: Painful data: the UNBC-McMaster shoulder pain expression archive database. In: 2011 IEEE International Conference on Automatic Face and Gesture Recognition (FG), pp. 57–64. IEEE (2011)
46. Hancock, P.: Psychological image collection at stirling (pics) (2008). <http://pics.psych.stir.ac.uk>
47. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint (2014). [arXiv: 1409.1556](https://arxiv.org/abs/1409.1556)
48. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
49. McNeely-White, D., Beveridge, J.R., Draper, B.A.: Inception and ResNet features are (almost) equivalent. *Cogn. Syst. Res.* **59**, 312–318 (2020)
50. Werner, P., Al-Hamadi, A., Limbrecht-Ecklundt, K., Walter, S., Gruss, S., Traue, H.C.: Automatic pain assessment with facial activity descriptors. *IEEE Trans. Affect. Comput.* **8**(3), 286–299 (2016)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Anay Ghosh received his B.Tech. Degree in Information Technology from the Institute of Science and Technology, Chandrakona Town, WB, India in 2012. After that, he did his M.Tech. in Computer Science and Engineering from the University of Engineering and Management, Newtown Kolkata, 2017. Currently, he is working as an Assistant Professor in University of Engineering and Management, Newtown Kolkata, India. His research

area and interest include image processing, pattern recognition, and video processing.

Saiyed Umer is currently an Assistant Professor with the Department of Computer Science and Engineering, Aliah University, Kolkata. He received the B.Sc. Degree (Hons.) in Mathematics from Vidyasagar University, India, in 2005, the M.C.A. Degree from the West Bengal University of Technology, India, in 2008, the M.Tech. Degree from the University of Kalyani, India, in 2012, and the Ph.D. Degree from the Department of Inform-

mation Technology, Jadavpur University, Kolkata, India. He was a Research Personnel with the Indian Statistical Institute (ISI), Kolkata, from November 2012 to April 2017. His research interests include computer vision, machine learning, and deep learning.

Muhammad Khurram Khan is a Global Thought Leader and Influencer in Cybersecurity. He is Founder and CEO of the 'Global Foundation for Cyber Studies and Research', an independent and non-partisan cybersecurity think-tank in Washington, DC, USA. He has contributed cyber policy works for the G20 in shaping a safer cyberspace for children, protection of masses and vulnerable populations in cyberspace, and empowering and enabling

women in the cybersecurity profession. He is also a Professor of Cybersecurity at the Center of Excellence in Information Assurance (CoEIA), King Saud University, Kingdom of Saudi Arabia. He is one of the Founding Members of CoEIA and has served as Manager R&D from March 2009 until March 2012. He, along with his team, developed and successfully managed the Cybersecurity Research Program at CoEIA, which turned it into one of the best centers of excellence in the region. In 2019, he has played an instrumental role as a cybersecurity subject expert for a USD 6 Million Series B Investment in a South Korean startup 'SecuLetter', which has received a corporate valuation of USD 40 Million (2020). His research areas of interest are Cybersecurity, digital authentication, IoT security, biometrics, multimedia security, cloud computing security, cyber policy, and technological innovation management. He is a Fellow of the IET (UK), Fellow of the BCS (UK), Fellow of the FTRA (Korea), Senior Member of the IEEE (USA), Senior Member of the IACSIT (Singapore), Member of the IEEE Consumer Electronics Society, Member of the IEEE Communications Society, Member of the IEEE Computers Society, Member of the IEEE Technical Committee on Security and Privacy, Member of the IEEE IoT Community, Member of the IEEE Smart Cities Community, and Member of the IEEE Cybersecurity Community. He has been Vice Chair of IEEE Communications Society Saudi Chapter. He is a Distinguished Lecturer of the IEEE (CESoC). His detailed profile is available at <http://www.professorkhurram.com>.

Ranjeet Kumar Rout is currently serving as Assistant Professor in the Department of Computer Science and Engineering, National Institute of Technology Srinagar, Hazratbal, India. He received his Ph.D. Degree from the Department of Information Technology of Indian Institute of Engineering Science and Technology Shibpur, West Bengal, India in the year of 2018. He earned Post Graduate and Bachelor Degree in Computer Science and Engineering

from Biju Patnaik University of Technology, Odisha, India, in 2010 and 2005 respectively. Prior to working at NIT Srinagar, Dr. Ranjeet had some useful research and teaching experience from Amity University Noida, National Institute Technology Jalandhar, and Indian Statistical Institute (ISI) Kolkata, India. His research interests include machine learning, and deep learning, visual cryptography and computational biology. He has published several papers in peer reviewed international and scientific journals in the field of non-linear Boolean functions and computational biology.

Bibhas Chandra Dhara received B.Sc. (Hons.) Degree in Mathematics and a B.Tech. Degree in Computer Science and Engineering from the University of Calcutta, India, in 1997 and 2000, respectively. He earned M.Tech. and Ph.D. in Computer Science from Indian Statistical Institute in 2002 and 2008, respectively. Currently, he is working as a Professor in the Department of Information Technology, Jadavpur University, India. His research area and

interest include image processing, pattern recognition, video processing, and audio processing.