IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 July 2022, accepted 9 August 2022, date of publication 18 August 2022, date of current version 19 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3199736

TOPICAL REVIEW

Thermography for Emotion Recognition Using Deep Learning in Academic Settings: A Review

FARDIAN FARDIAN[1,3,4], (Member, IEEE), MARTY MAWARPURY[2], KHAIRUL MUNADI[3,4], (Member, IEEE), AND FITRI ARNIA[3,4], (Member, IEEE)

[1]Doctoral School of Engineering Science, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[2]Psychology Study Program, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[3]Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[4]Telematics Research Center (TRC), Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

Corresponding author: Fitri Arnia (f.arnia@unsyiah.ac.id)

This work was supported in part by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under the 2021 Doctoral Research Grant (Hibah Penelitian Disertasi Doktor).

ABSTRACT Understanding students' emotional states during the learning process is one of the important aspects to improve learning quality. Measurements of emotion in an academic setting can be performed manually or automatically using a computer. However, developing an emotion recognition method using an imaging modality that is contactless, harmless, and illumination-independent is challenging. Thermography, as a non-invasive emotion recognition method, can recognize emotion variance during learning by observing the temperature distributions in a facial region. Deep learning models, such as convolutional neural networks (CNNs), can be used to interpret thermograms. CNNs can automatically classify emotion thermograms into several emotional states, such as happiness, anger, sadness, and fear. Despite their promising ability, CNNs have not been widely used in emotion recognition. In this study, we aimed to summarize the previous works and progress in emotion recognition in academic settings based on thermography and CNN. We first discussed the previous works on emotion recognition to provide an overview of the availability of modalities with their advantages and disadvantages. We also discussed emotion thermography potential for the academic context to find if there is any information in the available emotion thermal datasets related to the subjects' educational backgrounds. Emotion classification using the proposed CNN model was described step by step, including the feature learning illustration. Lastly, we proposed future research directions for developing a representative dataset in the academic settings, fed the segmented image, assigned a good kernel, and built a CNN model to improve the recognition performance.

INDEX TERMS Academic emotions, convolutional neural network, deep learning, emotion recognition, thermograms.

I. INTRODUCTION

Classroom is a place where students experience many types of emotion while doing activities, such as completing projects, taking exams, and building social relationships. Emotions, such as enjoyment, curiosity, interest, hope, pride, anger, anxiety, shame, confusion, frustration, and boredom frequently emerge during the learning process. Emotions experienced in educational settings have a strong correlation with students' academic achievement and personal growth. Experiencing positive emotions, such as enjoyment, while working on class projects can help students envision goals, improve creativity and problem solving, and support self-regulation [1], [2], [3]. On the other hand, experiencing negative emotions, such as anxiety, can hinder academic performance and negatively influence physical and psychological health [4]. The importance of emotions in education also equally applies to teachers, authorities, and administrators [5].

Emotions comprise a set of psychological processes, including affective, cognitive, physiological, motivational,

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

and expressive components [6]. Since emotions are mentally represented in the conscious mind and humans are able to communicate their feelings using verbal language, self-report has been widely used as a method to measure academic emotions [7]. Test anxiety, the first emotion method using self-report measurement, has been used since the 1930s [8]. It also has dominated emotion studies until the 1990s [3]. Later, researchers began to develop a method to measure other types of emotion.

However, self-report as a measurement instrument has several disadvantages. First, the assessment of emotional responses is limited to what is represented in the conscious mind [9]. Second, it has limited language preferences [10]. Last, it is difficult to maintain the respondents' emotions during the assessment. Self-report emotion has a possibility to produce a biased report [11]. Regarding the above issues, there is an opportunity to complement or substitute self-report with other methods to fill the gap. With the advancement of Affective Computing (AC) researchers are able to objectively measure academic emotions in a real-time manner, both in the conscious and subconscious mind [12].

AC is a multidisciplinary area that attempts to explore human affective experiences using computer technology combined with other disciplines, such as psychology, education, cognitive science, neuroscience, sociology, and psychophysiology. With AC, it is possible to detect, express, and create a system that is able to feel emotions [13], [14]. AC has great potential considering recent studies showing that emotional skill is one of the key factors that supports various activities, especially in critical related fields, such as health, security, and engineering [15]. AC studies are challenging since in humans, emotional states usually are less varied during activities, especially in learning [16].

The number of AC studies in education has steadily increased since 2010 [17]. There are various modalities that have been used, such as textual, visual, vocal, physiological, and multimodal, which indicate that various sensing technologies have been widely utilized. The advancement of computationally efficient devices and cheap sensing instruments have made it possible for an emotion recognition system to be massively implemented in the education sector.

Among other modalities, research focus on assessing human physiological signals to measure emotion in AC has significantly increased since 2011 [12]. Most measurement methods used were contact-based, such as to record skin conductance response, electroencephalography signals, facial expression recognition, and electrocardiogram measurement. However, contact-based methods can prevent elicitation in the subjects while wearing sensors [18].

Nevertheless, based on a review conducted by [17], body temperature measurement has not yet been explored. As warm-blooded beings, humans self-regulate their own body core and skin conditions to adapt to environmental changes and internal needs [19]. The self-regulating process involves physiological activities, and it has an impact on temperatures changes. These changes can be interpreted as signals to understand the human body and mind.

The human face has been widely chosen as a local area of emotion recognition because, as a part of the body, it is highly responsive to emotions [20]. It can express more than 30 emotional states [21], be easily recorded, and is naturally exposed to social stimuli [22]. This condition is suitable for a classroom setting where the face becomes the most exposed part during the learning process. Thermal changes on facial regions have also been dominantly explored in their relation with human affective states considering a human face consists of a number of micro-muscle units [23]. It causes temperature changes whenever they are activated [24].

Recently, several computer-based methods have been developed to recognize facial expression through thermograms [22], [25], [26], [27], [28], [29], [30], [31], [32]. Previous research shows that feature learning is still performed manually and not specifically designed for the education sector. However, so far, there is no study focusing on developing non-invasive emotional expression thermography using a Deep Neural Network (DNN), especially for the education sector.

Considering the current limited resources, it can be said that the work on emotion recognition using facial thermography based on DNN for the education sector is still at its early stage. Hence, significant effort is required to initiate the development of a reliable non-invasive technology to enable the recognition of emotional expressions for academic purposes. The study can be directed and focused on substantial issues identified during research to provide a better understanding of the most suitable approach to be implemented.

In this study, we aimed to review the current progress in emotion expressions recognition using Deep Learning (DL) and the use of thermography as a non-invasive approach. We also highlighted necessary future research directions to improve the accuracy of emotion recognition using thermal-imaging and DL for the academic context. The novelty and contributions of this study are arranged as follows:

- Section II presents review strategy on selecting references used on this paper
- Section III describes an overview of emotions in academic settings.
- Section IV presents the current measurements of emotions in academic settings.
- Section V presents the state-of-the-art of CNN as an image classifier in the DNN model for emotion recognition.
- Section VI discusses previous research on emotion classification using the available algorithms and CNN models.
- Section VII proposes recommendations for future works
- Section VIII summarizes future direction and its challenges to improve the accuracy and processing speed.

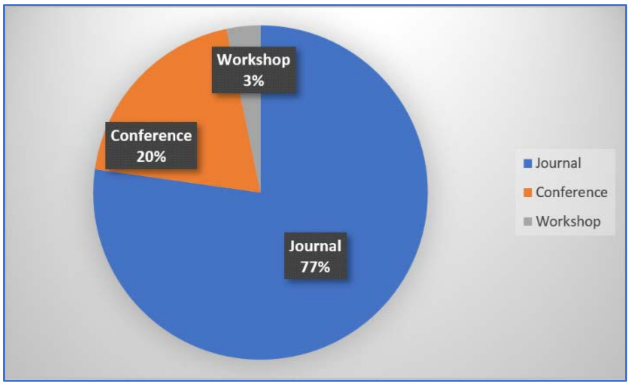

FIGURE 1. Distribution of studies based on publication source.

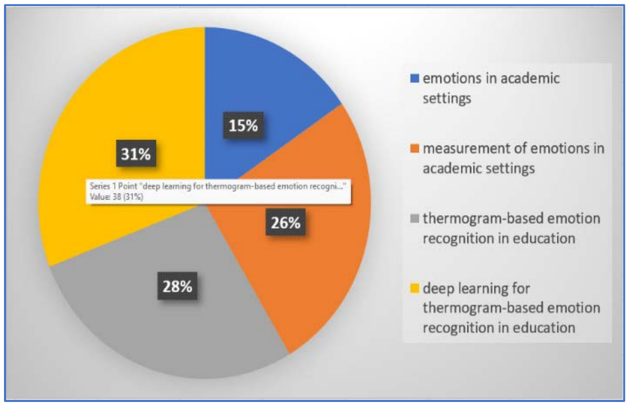

FIGURE 2. Distribution of studies categories.

II. REVIEW STRATEGY

In this study, we considered the articles from journals, conferences, and workshops published in the English language from 2010 to 2022. This period of time is chosen considering the term Affective Computing has been increasingly used in education sector since 2010 [12]. However, there is no effort that specifically focuses on the implementation of two recent potential technologies namely thermography and deep learning published from 2010 onwards.

A. STUDY SELECTION PROCESS

This review consisted of both manual and automatic search for selecting the references. We reviewed several digital databases including IEEE Explore, Springer Link, Science Direct, ISI Web of Knowledge. The main search keywords/phrase used in this study includes: "affective computing", "affective computing in education", "academic emotion", "emotion recognition", "thermal imaging", "artificial intelligence for thermal emotion", "emotion recognition database". Manual search was done for selecting the references to ensure that all relevant articles were retrieved for review. During the review process, if new articles were found, the search process was started again. The step repeated until no new article was found.

B. DATA EXTRACTION

The automatic search conducted on the selected digital libraries retrieved 232 studies. After manually checking the title, abstract, keywords, and conclusions of these studies, 157 studies excluded because there where not clearly relevant to our goal, leaving 123 studies. Figure 1 shows the distribution of studies based on publication source.

Figure 1 shows that the majority of sources were from journals 95 (77%), followed by conference 24 (20%), and workshop 4 (3%).

In this research, each paper was classified into one of four relevant categories: emotions in academic settings, measurement of emotions in academic settings, thermogram-based emotion recognition in education, and deep learning for thermogram-based emotion recognition in education.

Figure 2 demonstrates that the most common deep learning studies for thermogram-based emotion recognition in education 31% (38 articles), followed by thermogram-based emotion recognition in education 28% (34 articles), measurement of emotions in academic settings 26% (32 articles) and emotions in academic settings 15% (19 articles).

III. EMOTIONS IN ACADEMIC SETTINGS

Academic emotions are defined as emotions experienced by students in a learning environment [33]. Academic emotions have a strong correlation with students' achievement in the learning process [34]. Achievement emotions are emotions related to the activities or outcomes based on competency set by certain standards [5]. In education, the activities are mostly related to academic activities, such as studying, doing exams and homework, having class discussions, doing student projects, succeeding or failing in these activities. The emotions can also be caused by cognitive loads of information and time taken to process the information related to knowledge-generating aspects of cognitive activities [35].

During a learning process, a student can experience various types of emotion depending on the focus of attention. In addition, emotion can be stimulated by the topic being discussed and influence students' and teachers' interest and motivation in an academic environment [36]. Lastly, social emotions have a strong influence on students' engagement during class interactions and emotions caused by the events outside school, such as problems in the family [37].

A. EMOTION COMPONENTS

Emotions are multicomponent structures that can be differentiated from one another. The structures help us know the emotions that play a role in learning and teaching, the emotions that should be encouraged and discouraged, and the ways to regulate emotions in educational settings [5].

Emotions consist of multiple components viz subjective feeling, action tendency, appraisal, motor activity, and physiological component [38], [39], [40]. Each component is associated with a different function. Subjective feeling is associated with a monitoring function, action tendency with

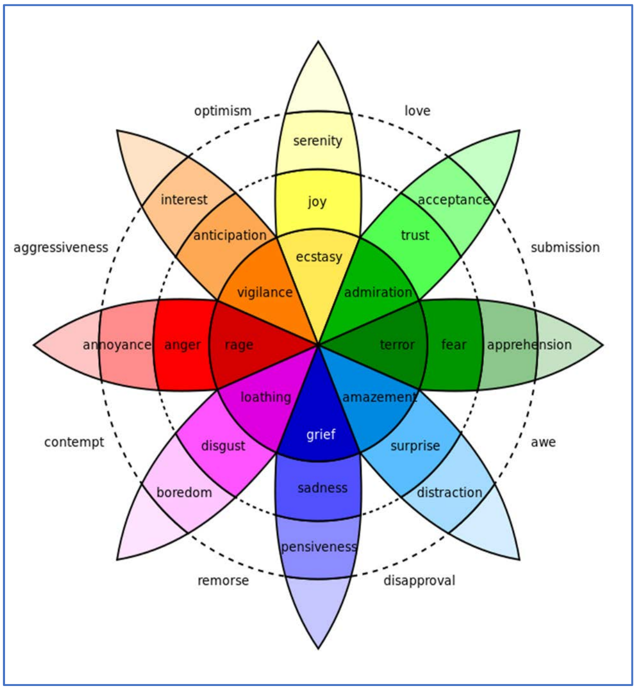

FIGURE 3. Plutchik model [45].

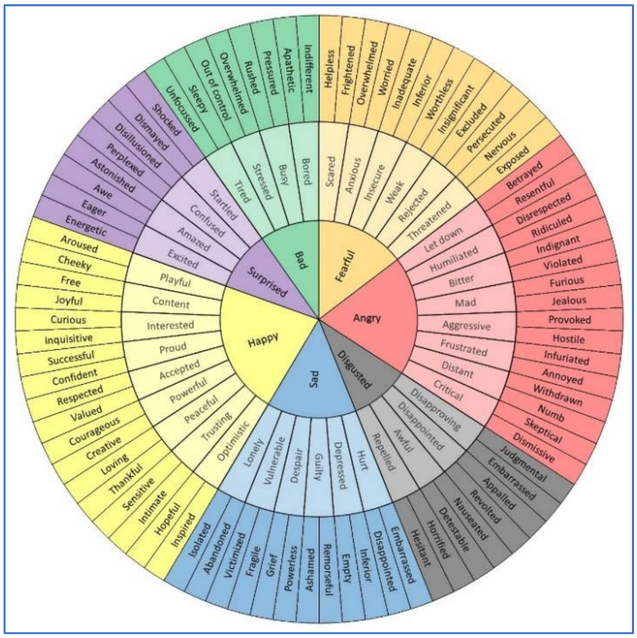

FIGURE 4. Wilcoxx model [46].

communicative function, appraisal with meaning-making function, and physiological with the support function of other components.

Several models have illustrated the structures of emotion, such as Plutchik's Circumplex Model [38], Scherer's Component Model [41], Geneva's Emotion Wheel [42], and Willcox's Feeling Wheel Model [43]. An attempt to connect emotion measurement with a computational system has been performed by Kelley [44], in which he used two emotional models, namely the Plutchik (Figure 3) and the Willcox model (Figure 4).

B. EMOTIONS IN EDUCATION: CONTENT DOMAIN, CONTEXT AND CULTURE

Emotions in education can be experienced differently in each content domain. Both teachers and students often have a complex interaction that requires a cognition process, stimulating positive or negative emotions. In addition, activities in a school subject often involve activities, such as problem solving, procedure handling, dealing with new concepts, adjusting to the learning standard defined in a curriculum, doing frequent evaluations, and adapting to various situations. These activities may stimulate different kinds of emotion [47].

School subjects, such as science education presenting in human pursuit, may also trigger certain kinds of emotion. During the teaching process, a student may experience more complex types of emotion than a teacher [48].

In educational settings, students frequently engage in reading and comprehending content materials through writing activities. These tasks involve organizing and communicating written thoughts [49]. Reading and writing activities involve positive and negative emotions which may cause anxiety [50].

Emotions may also appear in daily classroom life. Emotions during interrelationship between students and teachers have a central role in supporting learning achievement [51]. Cultural backgrounds may uniquely involve emotions depending on race, ethnicity, and identity during the learning process [52].

IV. MEASUREMENT OF EMOTIONS IN ACADEMIC SETTINGS

A. AVAILABLE MODALITIES

The number of AC studies in the education domain moderately has increased since 2010. They are grouped into five categories, namely textual, visual, vocal, physiological, and multimodal channels [17]. The methods used to assess emotional states vary from self-reporting and expert observation [53], [54], [55], [56]; facial expression, body poses, and gestures [57], [58], [59]; speech and intonation [60], [12], human organ system monitoring, such as electroencephalogram (EEG), electrocardiogram (ECG), heart rate variability (HRV), blood volume pulse (BVP), and eye-tracking [61], [62], [63], [64], to integration of different channels [65], [66], [67], [68], [69]. Most of the previous AC studies focused on negative emotions, in which the researchers attempted to find suitable techniques to manage negative emotions to improve learning quality [70], [71]. The available methods used in different modalities are presented in Table 1.

B. CURRENT MODALITIES: ADVANTAGES AND DISADVANTAGES

Textual modality has several advantages. First, it is easy to implement. Second, it does not depend on specific

TABLE 1. Methods used in the available modalities to measure emotions in academic settings.

Modality	Method
Textual	- Self-reporting - Expert Observation
Visual	- Facial expression - Head pose - Body gesture
Vocal	- Speech - Intonation
Physiological	- EEG - ECG - HRV - BVP - Eye Tracking
Multimodal	Integration of different modalities

TABLE 2. Advantages and disadvantages of the available modalities in measuring emotions in the academic settings.

Modality	Advantages	Disadvantages
Textual	- Easy to implement - Less dependent Specific instrument - Cheap - Meaningful feedback	- Not real-time - Lack of accuracy - Language barrier
Visual	- Naturally exposed - Visually observed - Practically used - Affordable equipment	- Noise - Image processing complexity - Privacy issues
Vocal	- Natural - Noticeable - Accurate - Practically deployed	- Dialog based system - Time-consuming - Culture obstacle
Physiological	- Closer access to body bio-signals - Real-time	- Visually less-observable - Uncomfortable - Privacy issues - Tight environment settings - Specialized and fragile instrument - Interpretation complexity
Multimodal	Combination of other modalities	- Multiple channel integration complexity - Data analysis complexity

instruments. Third, the instruments it requires are more cost-effective. Last, it can provide meaningful feedback. However, textual modality also has several disadvantages, such as not being real-time, having low accuracy and limited language preferences.

Visual channel also offers several benefits. First, it is naturally exposed. Second, it can be observed visually. Third, it is practical to use. Last, the equipment it requires is affordable. However, the noise, image processing complexity, and privacy issues have become the issues of this modality type.

Being natural, noticeable, accurate, practically deployable are the advantages of the vocal modality. However, it also has some limitations, such as using dialogue-based systems, being time- and resource-consuming, and having cultural and language differences.

There are two advantages of physiological signals. First, they have closer access to body bio-signals. Second, it can be implemented in a real-time manner. On the other hand, the use of the physiological instruments has several drawbacks, such as being less observable and uncomfortable, having privacy issues, requiring highly controlled environmental settings as well as specialized and fragile equipment, and being difficult to interpret.

Multimodal channel proposes better approaches to overcome the constraints of a single channel with great potential to generate a more accurate measurement. However, there are technical issues when integrating multiple channels and complexity in data analysis [17]. Table 2 sums up the advantages and disadvantages of current modalities.

C. THERMAL IMAGING AND VISUAL IMAGING: A COMPARISON

Capturing affect-related physiological signatures can be done in contactless manner such as body motion-based system [72] and voice-based system [73]. In addition, the signatures can be also performed via non-contact sensing devices such as visual cameras [74] and thermal cameras [75], [76].

In order to understand the advantages of thermal imaging over visible imaging, we need to understand how they work. Basically, visible cameras mimic how human eyes work that only sensitive to a narrow range of visible light of electromagnetic spectrum. They collect data from objects through the radiations in the visible spectrum objects' surface emits or reflects when hit by source of light [77]. This means that without emission from visible light sources such as the sun or incandescent bulbs, this vision system is generally unable to sense objects.

However, thermal cameras are designed to capture infrared radiations while visible cameras are not. According to Planck's law, every object above absolute zero temperature emits thermal radiation. Most of emitted radiations fall in the infrared spectrum range (0.9 – 14 μm) rather than visible spectrum range (380 – 780 nm) [78].

Since thermal and visual imaging work on different electromagnetic spectrum, thermal imaging could be more informative than visual imaging because:

1. Visible imaging suffers from illumination effects such as extremes of darkness and brightness due to sensor saturation or sensitivity [79], [80] while thermal imaging is less affected than those constraints [81].
2. Thermal imaging has less privacy issues rather than visual imaging [81].
3. Thermal imaging can penetrate smokes, aerosols, dust, and mist more effectively than visual imaging [82].

4. Thermal imaging is able to read more different types of physiological activities other than visual imaging ability [75], [83], [84].

V. THERMOGRAM-BASED EMOTION RECOGNITION IN EDUCATION

This section discusses four main aspects of thermogram-based emotion recognition in education. First, it explains the thermography potential in terms of body heat generation from humans' physiological activity and its relationship with emotions. Second, it presents the advantages of using thermography for emotion recognition compared to other modalities. Third, it describes public dataset availability in the academic context. Last, it discusses the available techniques to extract thermal features related to emotions in facial regions.

A. POTENTIAL

A number of recent studies have shown a strong correlation between emotion response and automatic nervous system (ANS) activity. However, the level of specificity of ANS activation widely diverges, varying from undifferentiated arousal to clearly specific predictions of patterns for certain emotions [85]. Some studies show that physiological aspects are strongly related to ANS, such as cardiovascular, respiratory, perspiratory, and muscular activity. Signals generated from these physiological cues have been widely used to measure a person's affective states [86], [87], [88], [89]. Recent studies show that the use of choreography provides possibilities for thermal imaging to monitor physiological signatures from facial regions. The most widely implemented aspect to physiological thermal signals is temperature change triggered by activities related to cardiovascular activity [89], [90], [91], [92], [93].

Vasodilation and dilatation in cardiovascular activity induce thermal directional changes and have demonstrated temperature patterns mainly in the facial areas [89], [90], [91], [92], [94], [95]. Vasoconstriction causes a decrease in temperature, whereas vasodilation occurs in the opposite way. They work by narrowing or widening blood vessels, causing blood flow to decrease or increase. It also has a strong correlation with temperature changes. In addition to that, skin regions containing many sweat glands also cause either an increase or decrease in temperature [96], [84], [97].

Furthermore, air exchanges from the breathing cycle can be monitored using thermal imaging because it produces thermal patterns [98], [99], [100], [101], [102], [75]. Lastly, muscular activation can also be observed using thermal imaging and is closely be linked with behavioral changes related to human's affection [76].

B. EMOTION THERMOGRAPHY IN EDUCATION

Despite its great potential, thermography is still understudied. There is only a little amount of research devoted to thermography for emotion recognition in education. Thermography presents more advantages compared to the other listed methods.

First, as a non-invasive method, it provides a better opportunity to capture actual emotions. The use of a contact-based method may prevent elicitation of genuine emotions while wearing the device [18]. This is suitable for capturing emotions during the learning process. Second, it is a risk-free monitoring system. The use of other measurements, such as sound and magnetic force, can harm our health [103]. Third, it needs a low-cost thermal camera that has been available in the market, unlike other methods that require expensive equipment with electromagnetic spectra, such as gamma, x-rays, ultraviolet, and other higher ranges of frequency [104]. Last, thermography does not depend on the illumination effect because it only relies on thermal emission from an object where a visible camera is light-sensitive [105].

C. DATASET OF EMOTION THERMOGRAPHY IN EDUCATION

In deep learning, a dataset can be treated by a computer for analytic and prediction purposes. This paper attempts to explore the available datasets of emotion thermography to identify the correlation with education by investigating the educational backgrounds of human subjects used on the datasets. Table 3 presents the available emotion datasets of the human subjects with their educational backgrounds.

Table 3 shows that there are only two datasets that contain the information on the human subjects' educational background information, namely the USTC-NVIE and KTFE database. Although all datasets are made for general purposes, these two datasets are the readiest datasets to implement in the academic context. Having compared both datasets, we found that USTC-NVIE is superior to KTFE for several reasons. First, USTC-NVIE represents more general features because it has a greater number of participants. It also consists of 215 students while KTFE only has 26 students. Second, USTC-NVIE only has one age group (17-31 years old), whereas KTFE has more diverse age groups ranging from children to adults (12-32 years old). Children are not small adults. Unlike adults, children's neurological development is still actively growing [115].

D. FACIAL EMOTION THERMAL FEATURES

The main goal of feature extraction is to obtain the most relevant information from the original data and represent the information in a lower dimensionality shape [116]. For the computational process, when the data to be input to an algorithm are too large and have potential to be reduced, transforming them into a reduced representation set of features is necessary.

Recent studies reported that facial muscular thermal signature has a relation to human's affective states [24], [76], [107], [117], [118]. In addition, facial micro-muscle activations generate heat and contribute to the production of numerous emotional expressions.

Wang et.al [107] proposed the use of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to reduce the dimension and select informative features of the

TABLE 3. Emotion thermography dataset with educational background information.

Dataset	Pose	Emotions	Education Background	Age (year)
Equinox [106]	Posed	Smile, Frown, Surprise	n/a	n/a
USTC-NVIE [107]	Posed and Spontaneous	Happy, Angry, Neutral, Disgust, Fear, Sad, Surprise	215 students (157 males + 58 females)	17-31
UCHT Thermal Face [108]	Posed	Happy, Sad, Angry	n/a	n/a
KTFE Database [109]	Spontaneous	Neutral, Angry, Happy, Sad, Fear, Disgust, Surprise	3 post docs, 9 PhDs, 11 masters, 2 bachelors, 1 pupil	12-32
Iris [110]	Posed	Surprise, Laugh, Angry	n/a	n/a
RGB-D-T [111]	Posed	Neutral, Happy, Sad, Angry, Surprise	n/a	n/a
VIS-TH [112]	Posed	Neutral, Happy, Sad, Angry, Surprise	n/a	n/a
RWTH Aachen Univ [113]	Posed	Neutral, Happy, Sad, Angry, Surprise, Fear, Disgust, Contempt	n/a	n/a
Tufts Face Database [114]	Posed	Neutral, Smile, Shocked	n/a	n/a

activated facial action units and K-nearest neighbors is used as a classifier. Each emotion has particular thermographic patterns or characteristics in several parts of the human face, such as nose, mouth, eyes, forehead, and cheeks [119]. To retain temperature for data analysis [107], thermogram images are segmented manually into five regions to ensure consistent segmentation, as shown in Figure 5.

The three-step ANOVA analysis using five statistical parameters was used. The first step is to ensure which statistical parameter is the most useful to reflect temperature changes related to emotion changes. The second step aims to monitor which facial regions with different emotional states result in the greatest temperature change. The third step is

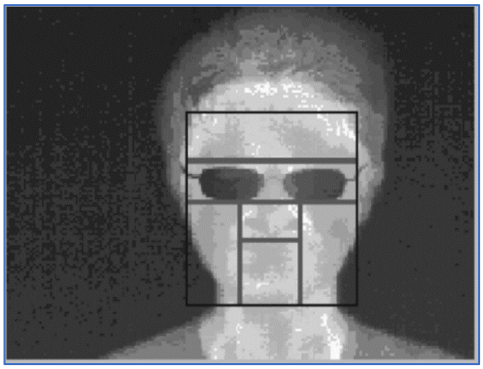

FIGURE 5. Face segmentation of five facial regions (forehead, eyes, nose, mouth, and cheeks) [107].

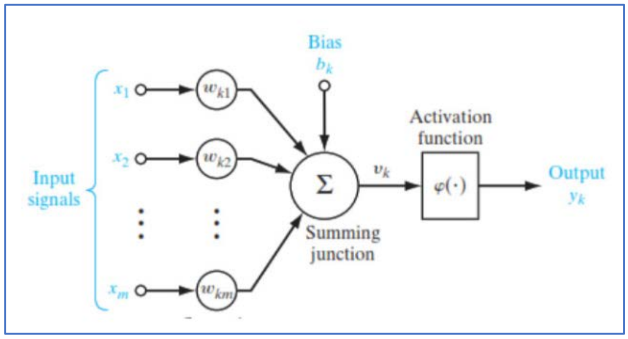

FIGURE 6. Neural network model mimics the human nervous system [121].

to analyze which emotional states differ most in each facial sub-region.

VI. DEEP LEARNING FOR THERMOGRAM-BASED EMOTION RECOGNITION IN EDUCATION

Artificial Neural Network (ANN) mimics the physiology and functioning of the human brain. Like the human brain, each neuron receives input and performs a dot operation with weights and biases. Weight describes the strength of the connection between two nodes, whereas bias is an external value that changes the network input of the activation function [120]. Nodes are described as individual processing units in each layer. Figure 6 illustrates the mathematical model of how NN operates.

An ANN comprises neurons as units with activation function $\varphi(\cdot)$ and parameter $\theta = \{W, B\}$, where W is the vector of weights (kernel) while B is the vector of biases. Equation (1) formulates the convolution operation [122].

$$y = \sum_i w_i x_i + b = \varphi(W^T x + B) \tag{1}$$

The activation function defines a linear combination of input x with respect to neurons and parameters, followed by element-wise non-linearity. The function also decides whether the neuron status is active or inactive based on the weighted sum of input signals.

The ANN learns the data to understand the process of data and data interpretation, and to predict future outcomes. Predictions do not require a probabilistic accuracy rate. However, high accuracy is necessary to ensure that decision making during learning is efficient.

ANN has some advantages in terms of learning ability, generalization, and robustness [123], [124]. Recently, studies in the neural networks have increased significantly, especially in Deep Neural Networks (DNNs) [125]. Deep Learning (DL) along with neural networks with multi hidden layers and massive training data aims to learn essential feature representation of the data by constructing high-level features from low-level pixels. Among other various DL techniques, Convolutional Neural Network (CNN) is the most widely used.

CNN is a DL algorithm that processes input images by assigning certain learnable weights and biases to map important features to differentiate one image from another. The output of CNN is the classification results. While performing data learning using CNN, three phases must be considered: dataset image pre-processing, feature learning, and classification steps. The classification may comprise several emotional states, such as happiness, anger, neutrality, disgust, fear, sadness, and surprise. In the next section, we will review the concepts and attempts in CNN implementation for emotion recognition classification of the dataset associated with the provided academic backgrounds of human subjects.

A. IMAGE PRE-PROCESSING AND FACIAL EXTRACTION

Image pre-processing is a step that aims to improve the quality of image data by eliminating the unwanted parts of the data and enhancing the important features to increase the performance of the NN model. In many cases, image pre-processing is crucial to support the learning process in terms of accuracy or timing process. Image pre-processing may be performed using mean subtractions, normalization, PCA whitening, and local contrast normalization [126].

Unlike visible images, thermal-based images comprise different characteristics of geometric, appearance, and texture [127]. Thermal-based images need different pre-processing methods for image enhancement and noise reduction, especially for facial extraction. Several studies have shown various methods to enhance thermal images and to extract facial regions, as shown in Table 4.

Table 2 shows various methods proposed for thermal image enhancement and facial extraction. In terms of the recognition performance, the best method for a general dataset still cannot be decided since each study was conducted using thermal cameras with different specifications, different environmental settings, and varied subjects' backgrounds. This statement is also strengthened by [138] that agrees there is no particular standard dataset for thermal facial emotion recognition imaging used consistently across the studies. However, considering the available datasets supported by the advancement of the current pre-processing techniques and various improved algorithms, there is still a great opportunity available to produce a system with more accurate measurement and lower computational cost in the future.

TABLE 4. **Emotion thermography datasets pre-preprocessing and facial extraction techniques.**

Study	Dataset	Method
Kopaczka et al [128]	RWTH Aachen Univ (2016)	- Unsharp mask for image enhancement Gaussian filter for image smoothing
Kopaczka et al [129]	RWTH Aachen Univ (2018)	HOG with SVM for face detection
Liu and Yin [130]	Their own dataset and USTC-NVIE	Unified model (Face detection, pose estimation, and landmark localization [132] + first frame of video to calculate head motion)
Wang et al [132]	USTC-NVIE	- Otsu Thresholding for image binarization - Facial edge detection from a binary image
Latif et al [133]	Their own dataset	Contrast Limited Histogram Equalization (CLHE) to improve image contrast
Mohd et al [134]	Their own dataset	- Viola-Jones boosting algorithm + Haar features to detect facial region - Bilateral filter for noise reduction and facial edge detection
Nguyen et al [135]	KTFE	Temperature space method to distinguish the facial region from the background
Trujillo et al [136]	IRIS	Bi-modal thresholding method to locate facial boundaries
Kolli et al [137]	Their own dataset	Region growing with a morphological operation for facial detection
Goulart et al [138]	Espirito Santo University	- Median and Gaussian filters to extract the facial region - Image binarization
Khan et al [139]	Their own dataset	- Median smoothing filter for blurring and noise reduction - Sobel filter for edge detection
Albarran et al [140]	Their own dataset	Image thresholding to separate the facial region from the background
Mostafa et al [141]	Their own dataset	A proposed tracking ROI method to track facial ROI

B. CONVOLUTIONAL NEURAL NETWORKS (CNNs) IN THERMAL FACIAL EMOTION RECOGNITION (FER)

The ability to shift from hand-crafted feature extraction to automatic learning through Neural Networks (NN) has brought some advantages for thermal image translation to visible image translation [142], [143], [144] and automated vector extraction of facial emotion recognition [145]. Early works on the implementation of thermal FER in Deep Learning (DL) began in 2014. Table 5 summarizes the studies of thermal FER in DL.

TABLE 5. Studies of thermal FER in deep learning.

Author / Year	Affect States	ROIs	DL Model	Dataset	Accuracy
Wang / 2014 [132]	Spontaneous	Whole face	DBM	USTC-NVIE	62.9%
Wu / 2016 [145]	Posed	Whole face	CNN	RGB-D-T	99.4%
Simon / 2016 [111]	Posed	Whole face	CNN	RGB-D-T	UNK
Cho / 2017 [146]	Stress	Nose	CNN	Custom Dataset	85.59%
Elbarawy / 2019 [147]	Posed	Whole face	CNN	IRIS	96.7%
Ilikci / 2019 [148]	Posed	Whole face	CNN	IRIS	92.72%
Kamath / 2019 [149]	Posed	Whole face	CNN	Tufts Face	96.2 %

Table 5 demonstrates that the majority of the DL models used were Convolutional Neural Networks (CNN). This finding shows that CNNs are still considered the most suitable DL technique for image recognition, especially for thermal FER because CNN is a deep network that imitates how the brain processes and recognizes images [150]. CNN enables feature extraction to learn patterns from high dimensional inputs performed automatically. As shown in Figure V, a CNN architecture consists of two main layers: a feature extraction layer and a fully connected layer.

1) FEATURE EXTRACTION LAYER

A feature extraction layer is a phase where input images are extracted to generate image features. This layer consists of two sub-layers: a convolutional layer and a pooling layer. The convolutional layer performs image conversion using convolution operation by applying digital filters (kernels). Raw FER images taken from a thermal camera are usually converted into visual images consisting of three-color channels (RGB), where these three channels correspond with three kernels. A kernel slides along the width and height of the input feature map, where each slide denotes the dot product operation of each part from the feature map with a suitable kernel value. For instance, an image transformed into a 4×4 2D feature image contains numbers. Then, a 2×2 convolution filter is applied to it.

The convolutional layer performs the multiplication of the feature image with the filter size of 2×2. This procedure is repeated until the whole input area is multiplied by the filter. The resulting values are then summed to generate one output called activation map. The number of feature maps depends on the sizes of the kernels.

In the convolution operation, the size of stride and padding must be taken into account. Stride is the parameter that determines the steps taken along the horizontal positions followed by vertical positions. For instance, if the stride size is 2, the kernel steps will consist of 2 pixels in a horizontal position and 2 pixels in a vertical position [126]. The smaller stride produces more detailed information retrieval. However, the smaller stride size is not always related to good performance.

Output dimension will always be smaller than the size of the input dimension, except the kernel size being 1×1 width and the stride size being 1×1. Since the output will be fed as input for the next layer, more information will be rendered unnecessary. To overcome this obstacle, a padding parameter is applied to the input. Padding is the parameter determining the number of pixels to be added at each side of the input to manipulate the output dimension of the feature map. By applying the padding to all input sides, the output dimension can be made equal. This allows a deeper convolutional layer to be applied, which results in more features being extracted. The padding step may improve the DNN performance by allowing the convolution filter to identify true information among zero values.

The feature map from the feature layer process is then fed into the pooling layer. The pooling layer comprises one filter with a certain size of stride. In the convolutional layer, the feature map is up-sampled. To avoid overfitting, in the pooling layer, the dimension of the feature map is reduced. There are two commonly used activation functions in this layer: max pooling and average pooling. The maximum value of the feature maps is selected in the max-pooling, whereas the average value of feature maps is selected in the average pooling.

CNN layers are commonly followed by a non-linear activation function. The activation function takes an input with a real value and transforms it into small ranges, such as [0,1] and [1,1]. The implementation of the activation function allows NNs to learn from non-linear mapping. It works like a switch that decides whether a neuron can be activated or not when provided with certain inputs. Sigmoid, Tanh, and ReLU activation functions are widely used in DNN [126].

In the learning features, CNNs iterate convolution and max-pooling processes several times to recognize the features of the input. Figure 7 illustrates the convolutional process using facial expression thermograms as the input images. Since each input has three channels (RGB), each kernel also comprises three kernels. The size of each kernel is determined by the number of feature maps.

Figure 8 illustrates the visual results of the convolutional phases of the NN in learning the features of the facial thermograms' affective states. The feature maps are stored in the pooling layer, and the position of one pixel in the activation function of one channel corresponds to the same position in the original image. Each tile in the grid of the feature map represents the convolution results of the input image with a particular kernel. Some feature maps provide important information about the input images. The interpretation of feature

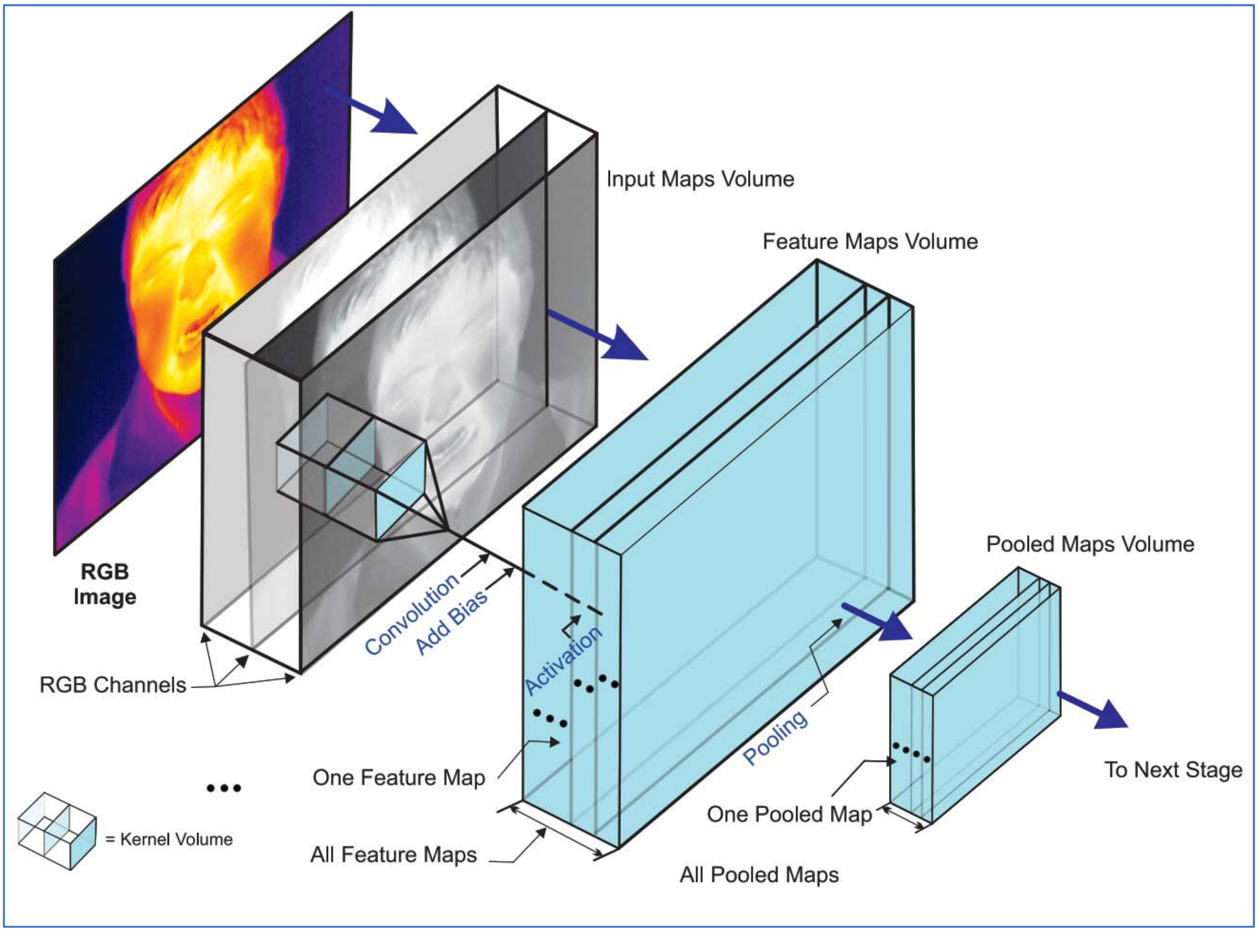

FIGURE 7. **Visualization of the convolutional process of a facial emotion thermogram; modified from a previous study [151].**

mapping results indicates that a suitable kernel confidently extracts the input features. Assigning a good kernel should reduce the training time to make the learning process perform rapidly.

2) FULLY CONNECTED LAYER

A fully connected layer, also known as a dense layer, operates based on features of an image from the feature extraction layer and generates an output. Feature maps resulting from the convolutional layer are in the form of a multidimensional array. A fully connected layer reshapes the multidimensional array into one dimensional array (vector). Each input from the feature extraction layer is fully mapped to final outputs with the probability score of each class in a classification task. The final fully-connected layer usually has the same number of output nodes as that of classes [152]. Figure 9 demonstrates fully connected layers with the classification results of the recognition process described in the probability value of each output.

3) IMAGE CLASSIFICATION

Image classification is a process of categorizing and labelling images according to their visual content and specific rules. The training process where a thermogram with a given emotional state label is known as supervised learning [153].

CNN often produces the categories with different probabilistic values that will decide the types of emotions being displayed in the thermograms. The output categories will be an array of numbers between 0 and 1. One common type of output model is the soft-max function. The soft-max function works by calculating the probability of an output image over possible target classes [152].

4) BACKPROPAGATION

Backpropagation is performed in the final layer of CNN and is only used during the training process. With backpropagation, NNs learn from errors during training. This process iteratively updates weights and changes the biases' values to zero based on the differences in the target output and predicted output.

An optimization algorithm is needed to reduce loss. Recently, several algorithms applied as optimizers, such as stochastic gradient descent (SGD) [154], limited-BGFS [155], parallelized SGD [156], stochastic variance reduced gradient [157], and Adam optimizer [158].

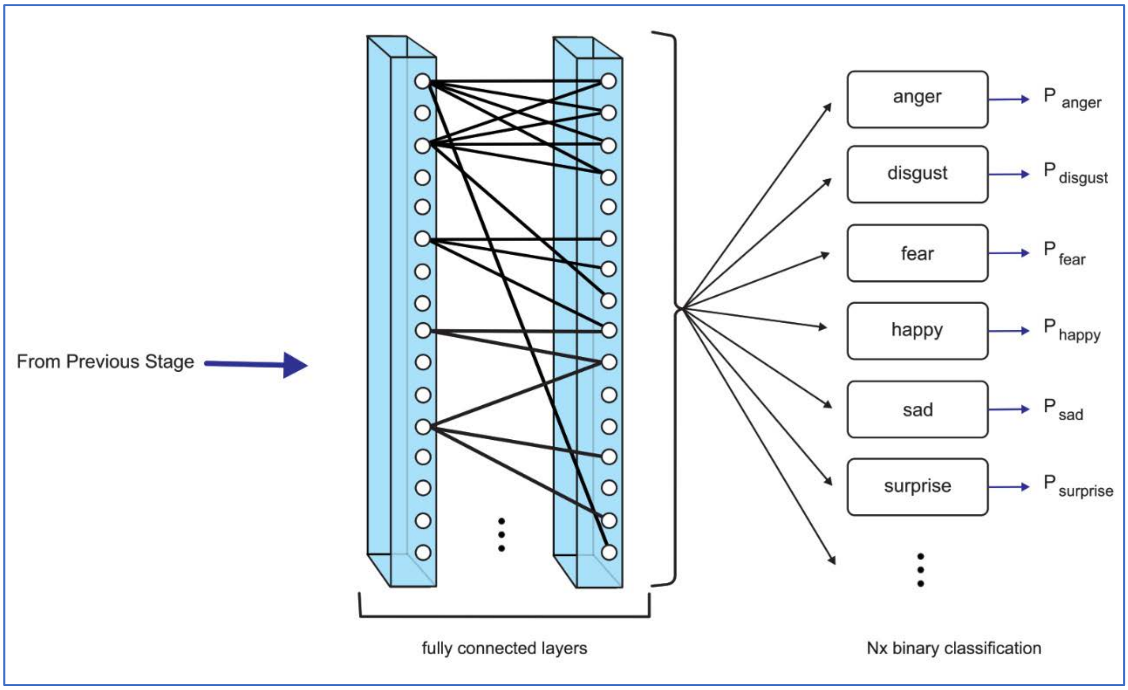

FIGURE 8. Visualization of segmentation process of a facial emotion thermogram; modified from a previous study [151].

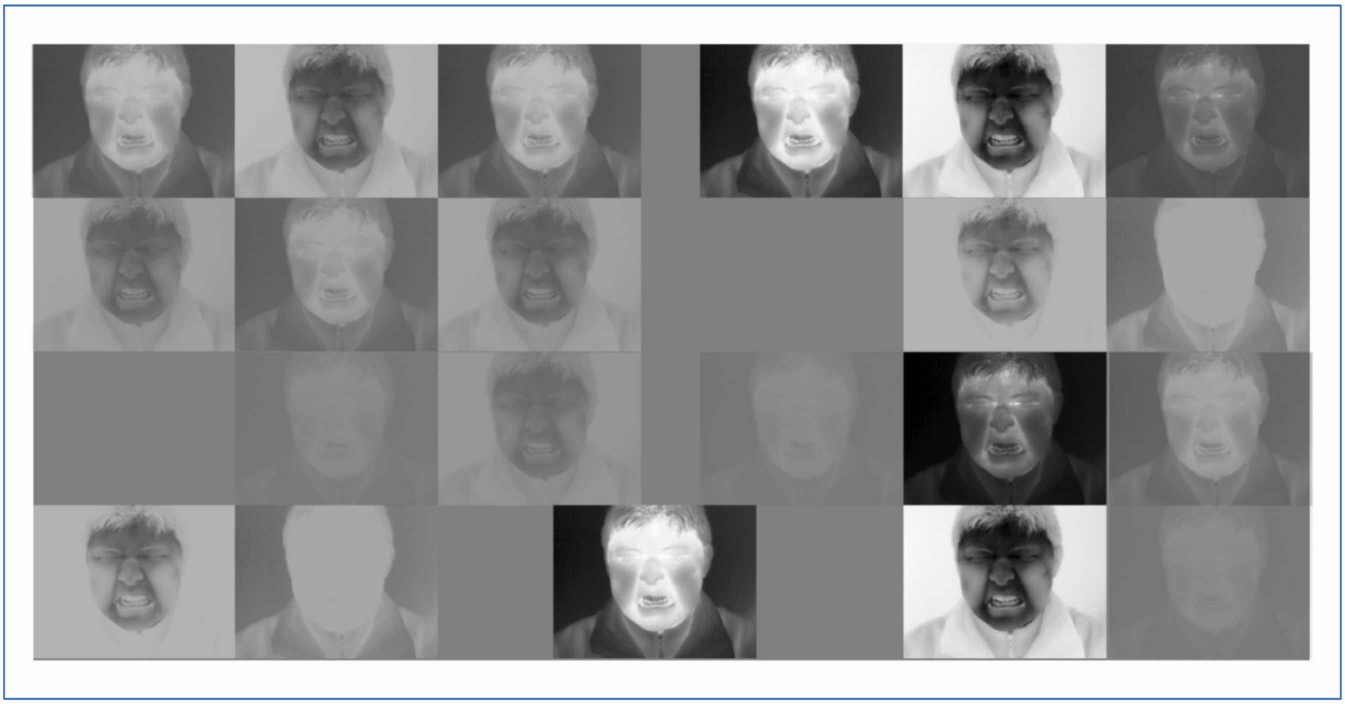

FIGURE 9. Feature mapping of facial emotion thermogram with a size of 244 × 244 into 32 feature maps.

VII. FUTURE DIRECTIONS

A. REPRESENTATIVE DATASET

The availability of a representative dataset is important for the training process. A good dataset will increase the robustness of training performance. Several factors must be considered when working with a certain dataset. The first factor is the quantity of the dataset. A large number of samples will provide more accurate mean values and reduce the margin error. The second is the quality of the dataset, which has been described in data reliability and feature representation [159].

The third is dataset domain specific. A good dataset is specifically built for a suitable case.

Based on the review of the emotion datasets shown in Table 3, the available datasets for thermogram emotion are made for the general context even though two of them (USTC-NVIE and KTFE) have the educational background information of their human subjects. Therefore, it is highly necessary to have datasets specifically designed for emotion recognition in the academic context. The datasets must represent content domain, context, and culture in education, as described in Section III.B.

Suitable dataset pre-processing to reduce unwanted data and enhancing important image features is also needed. For better performance, dataset pre-processing must suit the chosen CNN model architecture. The right pre-processing step will increase emotion recognition accuracy and decrease time process.

B. AUTOMATIC SEGMENTATION AND AUGMENTATION

Image segmentation is an important factor in image processing and computer vision. The segmentation process influences the training data, the choice of the network architectures, loss functions, training strategies, and performance results [159]. Image segmentation can include image denoising dan taking ROIs of a facial thermogram.

Another potential strategy is data augmentation. Data augmentation is an approach to deal with limited datasets. Thus, it is a useful technique to improve data learning, increase interpretation accuracy, and minimize the time needed for the emotion recognition process.

C. GOOD KERNEL

CNN enables automatic feature extraction. Automatic extraction simplifies the complexity provided by manual extraction. A good kernel is achieved by knowing important features in facial emotion thermograms. This information is important because we can shorten the feature learning process when designing a DL model. In addition, a convolutional calculation may be minimized, and the classification process can take place more efficiently.

D. LIGHTWEIGHT MODEL

Designing a simple CNN model with adequate layers and good kernels can speed up the convolution computation. A lightweight model will enable an emotion recognition system to be implemented widely in the academic context. The system can be used for self-evaluation using a mobile device or a low-cost computer with a minimum specification to allow the installation in every classroom to monitor students' and teachers' emotional states.

VIII. CONCLUSION

This study has presented a review of thermography for emotion recognition using deep learning in academic settings. We conclude that understanding emotional states during the learning process is one of the key aspects to developing a better learning system. Thermography has been proposed considering its advantages compared to other computer-based emotion recognition methods. Thermography enables emotion recognition to be interpreted from signals generated from internal physiological activities represented in thermal distribution.

Thermal distribution on facial regions can be evaluated using computer-assisted technology to measure emotional states. This technology can automatically perform feature extraction to minimize errors. Our review has shown that the current NN models have achieved higher accuracy rates in emotion recognition classification. Nevertheless, the performance of the NNs model still has to be improved.

Further research needs to work toward an improved classification of facial emotion thermograms in the academic context. This will require providing representative datasets, preparing suitable ROIs, assigning good kernels, and implementing lightweight models. These objectives will improve performance in terms of computation time efficiency and increase classification accuracy rates. A suitable method using thermography can be proposed for self-evaluation and the learning process in a classroom during learning.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for funding this research under the 2021 Doctoral Research Grant (Hibah Penelitian Disertasi Doktor).

REFERENCES

[1] G. L. Clore and J. R. Huntsinger, "How the object of affect guides its impact," *Emotion Rev.*, vol. 1, no. 1, pp. 39–54, Jan. 2009.
[2] B. L. Fredrickson, "The role of positive emotions in positive psychology: The broaden- and-build theory of positive emotions," *Amer. Psychol.*, vol. 56, no. 3, p. 218, 2001.
[3] R. Pekrun, T. Goetz, W. Titz, and R. P. Perry, "Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research," *Educ. Psychol.*, vol. 37, no. 2, pp. 91–105, Jan. 2002.
[4] M. Zeidner, "Test anxiety," *Corsini Encycl. Psychol.*, vol. 4, no. 1, pp. 1–3, Jan. 2010.
[5] R. Pekrun and L. Linnenbrink-Garcia, *International Handbook of Emotions in Education*. Evanston, IL, USA: Routledge, 2014.
[6] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivat. Emotion*, vol. 5, no. 4, pp. 345–379, Dec. 1981.
[7] R. Pekrun and M. Bühner, "Self-report measures of academic emotions," Routledge, New York, NY, USA, Tech. Rep., 2014, doi: 10.4324/9780203148211.
[8] C. H. Brown, "Emotional reactions before examinations: II. Results of a questionnaire," *J. Psychol.*, vol. 5, no. 1, pp. 11–26, Jan. 1938.
[9] S. D. Kreibig and G. H. E. Gendolla, "Autonomic nervous system measuring of emotion in education and achievement settings," Routledge, New York, NY, USA, Tech. Rep., 2014, doi: 10.4324/9780203148211.
[10] A. C. Frenzel, R. Pekrun, A.-L. Dicke, and T. Goetz, "Beyond quantitative decline: Conceptual shifts in adolescents' development of interest in mathematics," *Dev. Psychol.*, vol. 48, no. 4, p. 1069, 2012.
[11] A. C. Frenzel, "Teacher emotions," Routledge, New York, NY, USA, Tech. Rep., 2014, doi: 10.4324/9780203148211.
[12] C.-H. Wu, Y.-M. Huang, and J.-P. Hwang, "Review of affective computing in education/learning: Trends and challenges," *Brit. J. Educ. Technol.*, vol. 47, no. 6, pp. 1304–1323, Nov. 2016.
[13] R. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.
[14] R. A. Calvo, S. D'Mello, J. M. Gratch, and A. Kappas, *The Oxford Handbook of Affective Computing*. Oxford, U.K.: Oxford Univ. Press, 2015.

[15] T. A. Lashari, M. Alias, M. J. Kesot, and Z. A. Akasah, "An affective-cognitive teaching and learning approach for enhanced behavioural engagements among engineering students," *Eng. Educ.*, vol. 8, no. 2, pp. 65–78, Dec. 2013.
[16] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using 'emotional' data to improve learning in pervasive learning environment," *J. Educ. Technol. Soc.*, vol. 12, no. 2, pp. 176–189, 2009.
[17] E. Yadegaridehkordi, N. F. B. M. Noor, M. N. B. Ayub, H. B. Affal, and N. B. Hussin, "Affective computing in education: A systematic review and future research," *Comput. Educ.*, vol. 142, Dec. 2019, Art. no. 103649.
[18] S. Nayak, S. K. Panda, and S. Uttarkabat, "A non-contact framework based on thermal and visual imaging for classification of affective states during HCI," in *Proc. 4th Int. Conf. Trends Electron. Informat. (ICOEI)*, Jun. 2020, pp. 653–660.
[19] J. E. Hall and M. E. Hall, *Guyton and Hall textbook of Medical Physiology e-Book*. Amsterdam, The Netherlands: Elsevier, 2020.
[20] C. Goulart, C. Valadão, D. Delisle-Rodriguez, D. Funayama, A. Favarato, G. Baldo, V. Binotte, E. Caldeira, and T. Bastos-Filho, "Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction," *Sensors*, vol. 19, no. 13, p. 2844, Jun. 2019.
[21] C. Darwin and P. Prodger, *The Expression of the Emotions in Man and Animals*. Oxford, U.K.: Oxford Univ. Press, 1998.
[22] C. Filippini, D. Perpetuini, D. Cardone, A. M. Chiarelli, and A. Merla, "Thermal infrared imaging-based affective computing and its application to facilitate human robot interaction: A review," *Appl. Sci.*, vol. 10, no. 8, p. 2924, Apr. 2020.
[23] P. Ekman, "Facial expression and emotion," *Amer. Psychol.*, vol. 48, no. 4, p. 384, 1993.
[24] S. Jarlier, D. Grandjean, S. Delplanque, K. N'diaye, I. Cayeux, M. I. Velazco, D. Sander, P. Vuilleumier, and K. R. Scherer, "Thermal analysis of facial muscles contractions," *IEEE Trans. Affective Comput.*, vol. 2, no. 1, pp. 2–9, Jan./Jun. 2011.
[25] P. Marqués-Sánchez, C. Liébana-Presa, J. A. Benítez-Andrades, R. Gundín-Gallego, L. Álvarez-Barrio, and P. Rodríguez-Gonzálvez, "Thermal infrared imaging to evaluate emotional competences in nursing students: A first approach through a case study," *Sensors*, vol. 20, no. 9, p. 2502, Apr. 2020, doi: 10.3390/s20092502.
[26] C. Goulart, C. Valadão, D. Delisle-Rodriguez, D. Funayama, A. Favarato, G. Baldo, V. Binotte, E. Caldeira, and T. Bastos-Filho, "Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction," *Sensors*, vol. 19, no. 13, p. 2844, Jun. 2019, doi: 10.3390/s19132844.
[27] C. Filippini, E. Spadolini, D. Cardone, and A. Merla, "Thermal imaging based affective computing for educational robot," *Multidisciplinary Digit. Publishing Inst. Proc.*, vol. 27, no. 1, p. 27, Sep. 2019, doi: 10.3390/proceedings2019027027.
[28] J. Panasiuk, P. Prusaczyk, A. Grudzień, and M. Kowalski, "Study on facial thermal reactions for psycho-physical stimuli," *Metrol. Meas. Syst.*, vol. 27, no. 3, pp. 399–415, 2020, doi: 10.24425/mms.2020.134591.
[29] M. Kopaczka, L. Breuer, J. Schock, and D. Merhof, "A modular system for detection, tracking and analysis of human faces in thermal infrared recordings," *Sensors*, vol. 19, no. 19, p. 4135, Sep. 2019, doi: 10.3390/s19194135.
[30] "Human emotions recognition from thermal images using Yolo algorithm," in *Proc. IEEE Int. Conf. Commun. Signal Process. (ICCSP)*, Melmaruvathur, India, Jul. 2020, doi: 10.1109/ICCSP48568.2020.9182148.
[31] A. Basu, A. Dasgupta, A. Thyagharajan, A. Routray, R. Guha, and P. Mitra, "A portable personality recognizer based on affective state classification using spectral fusion of features," *IEEE Trans. Affect. Comput.*, vol. 9, no. 3, pp. 330–342, Jul. 2018, doi: 10.1109/TAFFC.2018.2828845.
[32] M. A. Hossain and B. Assiri, "Facial emotion verification by infrared image," in *Proc. Int. Conf. Emerg. Smart Comput. Inform. (ESCI)*. India: IEEE Pune Section, 2020, doi: 10.1109/ESCI48226.2020.9167616.
[33] R. Pekrun, "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," *Educ. Psychol. Rev.*, vol. 18, no. 4, pp. 315–341, Nov. 2006.
[34] K. Muramatsu, E. Tanaka, K. Watanuki, and T. Matsui, "Framework to describe constructs of academic emotions using ontological descriptions of statistical models," *Res. Pract. Technol. Enhanced Learn.*, vol. 11, no. 1, pp. 1–18, Dec. 2016.
[35] R. Pekrun and E. J. Stephens, "Academic emotions," in *APA Educational Psychology Handbook* (Individual Differences and Cultural and Contextual Factors), vol. 2. Washington, DC, USA: American Psychological Association, 2012, pp. 3–31.
[36] M. Ainley, "Being and feeling interested: Transient state, mood, and disposition," in *Emotion in Education*. Amsterdam, The Netherlands: Elsevier, 2007, pp. 147–163.
[37] L. Linnenbrink-Garcia, T. K. Rogat, and K. L. K. Koskey, "Affect and engagement during small group instruction," *Contemp. Educ. Psychol.*, vol. 36, no. 1, pp. 13–24, Jan. 2011.
[38] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.
[39] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychol. Rev.*, vol. 110, no. 1, p. 145, 2003.
[40] K. R. Scherer, "On the nature and function of emotion: A component process approach," *Approaches to Emot.*, vol. 2293, no. 317, p. 31, 1984.
[41] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005.
[42] K. R. Scherer, "Measuring the meaning of emotion words: A domain-specific componential approach," Oxford Univ. Press, New York, NY, USA, Tech. Rep., 2013, doi: 10.1093/acprof:oso/9780199592746.003.0002.
[43] G. Willcox, "The feeling wheel," *Transactional Anal. J.*, vol. 12, no. 4, pp. 274–276, Oct. 1982, doi: 10.1177/036215378201200411.
[44] D. J. Kelley, "Modeling emotions in a computational system," in *Google It*. New York, NY, USA: Springer, 2016, pp. 447–461.
[45] Machine Elf 1735. (2011). *Robert Plutchik's Wheel of Emotions*. Accessed: Jun. 9, 2022. [Online]. Available: https://en.wikipedia.org/wiki/File:Plutchik-wheel.svg
[46] G. Roberts, "The Wilcoxx feelings wheel," Taylor & Francis, London, U.K., Tech. Rep., 2015, doi: 10.1177/036215378201200411.
[47] G. A. Goldin, "Perspectives on emotion in mathematical engagement, learning, and problem solving," in *International Handbook of Emotions in Education*. New York, NY, USA: Routledge, 2014, pp. 391–414.
[48] G. M. Sinatra, S. H. Broughton, and D. Lombardi, "Emotions in science education," in *International Handbook of Emotions in Education*. Evanston, IL, USA: Routledge, 2014, pp. 425–446.
[49] J. L. Hagaman and R. Reid, "The effects of the paraphrasing strategy on the reading comprehension of middle school students at risk for failure in reading," *Remedial Special Educ.*, vol. 29, no. 4, pp. 222–234, Jul. 2008.
[50] C. M. Bohn-Gettler and D. N. Rapp, "Emotion during reading and writing," Routledge, New York, NY, USA, Tech. Rep., 2014, doi: 10.4324/9780203148211.
[51] D. K. Meyer, "Situating emotions in classroom practices," in *International Handbook of Emotions in Education*. Evanston, IL, USA: Routledge, 2014, pp. 468–482.
[52] R. Pekrun and L. Linnenbrink-Garcia, "The influence of culture on emotions: Implications for education," in *International Handbook of Emotions in Education*. Evanston, IL, USA: Routledge, 2014, pp. 549–568.
[53] A. Bellocchi, K. A. Mills, and S. M. Ritchie, "Emotional experiences of preservice science teachers in online learning: The formation, disruption and maintenance of social bonds," *Cultural Stud. Sci. Educ.*, vol. 11, no. 3, pp. 629–652, Sep. 2016.
[54] E. N. Castellar, J. Van Looy, A. Szmalec, and L. de Marez, "Improving arithmetic skills through gameplay: Assessment of the effectiveness of an educational game in terms of cognitive and affective learning outcomes," *Inf. Sci.*, vol. 264, pp. 19–31, Apr. 2014.
[55] K. R. Muis, C. Psaradellis, S. P. Lajoie, I. Di Leo, and M. Chevrier, "The role of epistemic emotions in mathematics problem solving," *Contemp. Educ. Psychol.*, vol. 42, pp. 172–185, Jul. 2015.
[56] I. Tjostheim, W. Leister, T. Schulz, and A. Larssen, "The role of emotion and enjoyment for QoE—A case study of a science centre installation," in *Proc. 7th Int. Workshop Quality Multimedia Exper. (QoMEX)*, May 2015, pp. 1–6.
[57] D. Urhahne, "Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion," *Teaching Teacher Educ.*, vol. 45, pp. 73–82, Jan. 2015.
[58] I. Behoora and C. S. Tucker, "Machine learning classification of design team members' body language patterns for real time emotional state detection," *Design Stud.*, vol. 39, pp. 100–127, Jul. 2015.
[59] V. J. Shute, S. D'Mello, R. Baker, K. Cho, N. Bosch, J. Ocumpaugh, M. Ventura, and V. Almeda, "Modeling how incoming knowledge, persistence, affective states, and in-game progress influence Student learning from an educational game," *Comput. Educ.*, vol. 86, pp. 224–235, Aug. 2015.

[60] X. Gu, Z. Wang, S. Zheng, and W. Wang, "Design and implementation of the emotional pedagogical agent," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, 2009, pp. 459–462.
[61] C.-M. Chen and Y.-C. Sun, "Assessing the effects of different multimedia materials on emotions and learning performance for visual and verbal style learners," *Comput. Educ.*, vol. 59, no. 4, pp. 1273–1285, Dec. 2012.
[62] C.-M. Chen and C.-H. Wu, "Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance," *Comput. Educ.*, vol. 80, pp. 108–121, Jan. 2015.
[63] R. Gil, J. Virgili-Gomá, R. García, and C. Mason, "Emotions ontology for collaborative modelling and learning of emotional responses," *Comput. Hum. Behav.*, vol. 51, pp. 610–617, Oct. 2015.
[64] C.-H. Wu, Y.-L. Tzeng, and Y.-M. Huang, "Measuring performance in leaning process of digital game-based learning and static E-learning," *Educ. Technol. Res. Develop.*, vol. 68, no. 5, pp. 2215–2237, Oct. 2020.
[65] J. Grafsgaard, J. Wiggins, K. E. Boyer, E. Wiebe, and J. Lester, "Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue," Educ. Data Mining, London, U.K., Tech. Rep., 2014.
[66] K. Muldner and W. Burleson, "Utilizing sensor data to model students' creativity in a digital environment," *Comput. Hum. Behav.*, vol. 42, pp. 127–137, Jan. 2015.
[67] S. Caballe, "Towards a multi-modal emotion-awareness e-Learning system," in *Proc. Int. Conf. Intell. Netw. Collaborative Syst.*, Sep. 2015, pp. 280–287.
[68] L. Shen, B. Xie, and R. Shen, "Enhancing user experience in mobile learning by affective interaction," in *Proc. Int. Conf. Intell. Environ.*, Jun. 2014, pp. 297–301.
[69] A. Kaklauskasa, A. Kuzminske, E. K. Zavadskas, A. Daniunas, G. Kaklauskas, M. Seniut, J. Raistenskis, A. Safonov, R. Kliukas, A. Juozapaitis, and A. Radzeviciene, "Affective tutoring system for built environment management," *Comput. Educ.*, vol. 82, pp. 202–216, Mar. 2015.
[70] S. D'Mello and R. A. Calvo, "Beyond the basic emotions: What should affective computing compute?" in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, 2013, pp. 2287–2294.
[71] F. Tian, P. Gao, L. Li, W. Zhang, H. Liang, Y. Qian, and R. Zhao, "Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems," *Knowl.-Based Syst.*, vol. 55, pp. 148–164, Jan. 2014.
[72] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 15–33, Jan. 2012.
[73] N. D. Lane, P. Georgiev, and L. Qendro, "DeepEar: Robust smartphone audio sensing in unconstrained acoustic environments using deep learning," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2015, pp. 283–294.
[74] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.
[75] C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, V. Blazek, and S. Leonhardt, "Remote monitoring of breathing dynamics using infrared thermography," *Biomed. Opt. Exp.*, vol. 6, no. 11, pp. 4378–4394, 2015.
[76] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 258–269, May 2007.
[77] C. Roychoudhuri, A. F. Kracklauer, and K. Creath, *The Nature of Light: What is a Photon*. Boca Raton, FL, USA: CRC Press, 2017.
[78] I. Pavlidis, P. Symosek, B. Fritz, M. Bazakos, and N. Papanikolopoulos, "Automatic detection of vehicle occupants: The imaging problemand its solution," *Mach. Vis. Appl.*, vol. 11, no. 6, pp. 313–320, May 2000.
[79] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 239–246.
[80] M. Aittala, T. Weyrich, and J. Lehtinen, "Two-shot SVBRDF capture for stationary materials," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 110–111, 2015.
[81] J. M. Lloyd, *Thermal Imaging Systems*. New York, NY, USA: Springer, 2013.
[82] K. J. Havens and E. Sharp, *Thermal Imaging Techniques to Survey and Monitor Animals in the Wild: A Methodology*. New York, NY, USA: Academic, s2015.
[83] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1418–1426, Aug. 2007.
[84] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin, and B. Bass, "Fast by nature–how stress patterns define human experience and performance in dexterous tasks," *Sci. Rep.*, vol. 2, no. 1, pp. 1–9, Dec. 2012.
[85] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.
[86] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
[87] S. A. Hosseini and M. A. Khalilzadeh, "Emotional stress recognition system using EEG and psychophysiological signals: Using new labelling process of EEG signals in emotional stress state," in *Proc. Int. Conf. Biomed. Eng. Comput. Sci.*, Apr. 2010, pp. 1–6.
[88] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *Proc. Int. Conf. Affective Comput. Intell. Interact.*, 2011, pp. 125–134.
[89] V. Engert, A. Merla, J. A. Grant, D. Cardone, A. Tusche, and T. Singer, "Exploring the use of thermal infrared imaging in human stress research," *PLoS ONE*, vol. 9, no. 3, Mar. 2014, Art. no. e90782.
[90] H. Genno, K. Ishikawa, O. Kanbara, M. Kikumoto, Y. Fujiwara, R. Suzuki, and M. Osumi, "Using facial skin temperature to objectively evaluate sensations," *Int. J. Ind. Ergonom.*, vol. 19, no. 2, pp. 161–171, Feb. 1997.
[91] C. K. L. Or and V. G. Duffy, "Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement," *Occupational Ergonom.*, vol. 7, no. 2, pp. 83–94, 2007.
[92] A. Di Giacinto, M. Brunetti, G. Sepede, A. Ferretti, and A. Merla, "Thermal signature of fear conditioning in mild post traumatic stress disorder," *Neuroscience*, vol. 266, pp. 216–223, Apr. 2014.
[93] E. Salazar-López, E. Domínguez, V. Juárez Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Gálvez, M. A. Rodríguez-Artacho, and E. Gómez-Milán, "The mental and subjective skin: Emotion, empathy, feelings and thermography," *Consciousness Cognition*, vol. 34, pp. 149–162, Jul. 2015.
[94] H. Veltman and W. Vos, "Facial temperature as a measure of operator state," *Found. Augment. Cogn.*, vol. 293, pp. 293–301, Nov. 2005.
[95] A. Yomna, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere, "Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–20, 2017.
[96] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis, "Imaging facial signs of neurophysiological responses," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 477–484, Feb. 2008.
[97] A. T. Krzywicki, G. G. Berntson, and B. L. O'Kane, "A non-contact technique for measuring eccrine sweat gland activity using passive thermal imaging," *Int. J. Psychophysiol.*, vol. 94, no. 1, pp. 25–34, 2014.
[98] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," in *Proc. IEEE Eng. Med. Biol. Soc.*, vol. 1, Sep. 2004, pp. 1196–1199.
[99] R. Murthy and I. Pavlidis, "Noncontact measurement of breathing function," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 3, pp. 57–67, May 2006.
[100] J. Fei and I. Pavlidis, "Thermistor at a distance: Unobtrusive measurement of breathing," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 988–998, Apr. 2009.
[101] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *Biomed. Eng. Online*, vol. 10, no. 1, pp. 1–17, 2011.
[102] G. F. Lewis, R. G. Gatto, and S. W. Porges, "A novel method for extracting respiration rate and relative tidal volume from infrared thermography," *Psychophysiology*, vol. 48, no. 7, pp. 877–887, 2011.
[103] M. F. Modest, *Radiative Heat Transfer*. New York, NY, USA: Academic, 2013.
[104] E. Villa, N. Arteaga-Marrero, and J. Ruiz-Alzola, "Performance assessment of low-cost thermal cameras for medical applications," *Sensors*, vol. 20, no. 5, p. 1321, Feb. 2020.
[105] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. G. Bulnes, "Infrared thermography for temperature measurement and non-destructive testing," *Sensors*, vol. 14, no. 7, pp. 12305–12348, 2014.
[106] D. A. Socolinsky, "Human identification at a distance database," Equinox Corp., USA, 2012. Accessed: Jun. 6, 2022. [Online]. Available: http://aserg.labsoft.dcc.ufmg.br/comets/dataset/equinox/
[107] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010, doi: 10.1109/TMM.2010.2060716.

[108] G. Hermosilla, J. R.-del-Solar, R. Verschae, and M. Correa, "A comparative study of thermal face recognition methods in unconstrained environments," *Pattern Recognit.*, vol. 45, no. 7, pp. 2445–2459, Jul. 2012.
[109] H. Nguyen, K. Kotani, F. Chen, and B. Le, "A thermal facial emotion database and its analysis," in *Proc. Pacific-Rim Symp. Image Video Technol.*, 2013, pp. 397–408.
[110] (2021). OTCBVS. *OTCBVS Benchmark Dataset Collection*. Accessed: Dec. 1, 2021. [Online]. Available: https://Vcipl-okstate.org
[111] M. O. Simón, "Improved RGB-DT based face recognition," *IET Biometrics*, vol. 5, no. 4, pp. 297–303, 2016.
[112] K. Mallat and J.-L. Dugelay, "A benchmark database of visible and thermal paired face images across multiple variations," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.
[113] M. Kopaczka, J. Nestler, and D. Merhof, "Face detection in thermal infrared images: A comparison of algorithm-and machine-learning-based approaches," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2017, pp. 518–529.
[114] K. Panetta, Q. Wan, S. Agaian, S. Rajeev, S. Kamath, R. Rajendran, S. P. Rao, A. Kaszowska, H. A. Taylor, A. Samani, and X. Yuan, "A comprehensive database for benchmarking imaging systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 509–520, Mar. 2020. [Online]. Available: http://tdface.ece.tufts.edu/
[115] A. M. Passarotti, B. M. Paul, J. R. Bussiere, R. B. Buxton, E. C. Wong, and J. Stiles, "The development of face and location processing: An fMRI study," *Develop. Sci.*, vol. 6, no. 1, pp. 100–117, Feb. 2003.
[116] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *Proc. 4th Int. Conf. Adv. Comput. Commun. Technol.*, Feb. 2014, pp. 5–12, doi: 10.1109/ACCT.2014.74.
[117] Z. Liu and S. Wang, "Emotion recognition using hidden Markov models from facial temperature sequence," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2011, pp. 240–247.
[118] A. Wesley, P. Buddharaju, R. Pienta, and I. Pavlidis, "A comparative analysis of thermal and visual modalities for automated facial expression recognition," in *Proc. Int. Symp. Vis. Comput.*, 2012, pp. 51–60.
[119] J. Clay-Warner and D. T. Robinson, "Infrared thermography as a measure of emotion response," *Emotion Rev.*, vol. 7, no. 2, pp. 157–162, Apr. 2015.
[120] D. Graupe, *Principles of Artificial Neural Networks*, vol. 7. Singapore: World Scientific, 2013.
[121] D. Groupe, *Principle of Artificial Neural Networks*. Singapore: World Scientific, 2007.
[122] S. Haykin, *Neural Networks and Learning Machines, 3/E*. London, U.K.: Pearson, 2010.
[123] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.
[124] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 114–132, Jan. 1997.
[125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2017.
[126] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, "A guide to convolutional neural networks for computer vision," *Synth. Lectures Comput. Vis.*, vol. 8, no. 1, pp. 1–207, 2018.
[127] M. M. M. Al Qudah, A. S. A. Mohamed, and S. L. Lutfi, "Affective state recognition using thermal-based imaging: A survey," *Comput. Syst. Sci. Eng.*, vol. 37, no. 1, pp. 47–62, 2021.
[128] M. Kopaczka, K. Acar, and D. Merhof, "Robust facial landmark detection and face tracking in thermal infrared images using active appearance models," in *Proc. 11th Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2016, pp. 150–158.
[129] M. Kopaczka, R. Kolk, and D. Merhof, "A fully annotated thermal face database and its application for thermal facial expression recognition," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2018, pp. 1–6.
[130] P. Liu and L. Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–6.
[131] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2879–2886.
[132] S. Wang, M. He, Z. Gao, S. He, and Q. Ji, "Emotion recognition from thermal infrared images using deep Boltzmann machine," *Frontiers Comput. Sci.*, vol. 8, no. 4, pp. 609–618, Aug. 2014.
[133] M. H. Latif, M. H. Yusof, S. N. Sidek, and N. Rusli, "Texture descriptors based affective states recognition-frontal face thermal image," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2016, pp. 80–85.
[134] M. N. H. Mohd, M. Kashima, K. Sato, and M. Watanabe, "Mental stress recognition based on non-invasive and non-contact measurement from stereo thermal and visible sensors," *Int. J. Affect. Eng.*, vol. 14, no. 1, pp. 9–17, 2015.
[135] H. Nguyen, K. Kotani, F. Chen, and B. Le, "Estimation of human emotions using thermal facial information," in *Proc. 5th Int. Conf. Graphic Image Process. (ICGIP)*, Jan. 2014, p. 90690.
[136] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic feature localization in thermal images for facial expression recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2005, p. 14.
[137] A. Kolli, A. Fasih, F. A. Machot, and K. Kyamakya, "Non-intrusive car driver's emotion recognition using thermal camera," in *Proc. Joint INDS ISTET*, Jul. 2011, pp. 1–5.
[138] C. Goulart, C. Valadão, D. Delisle-Rodriguez, E. Caldeira, and T. Bastos, "Emotion analysis in children through facial emissivity of infrared thermal imaging," *PLoS ONE*, vol. 14, no. 3, Mar. 2019, Art. no. e0212928.
[139] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *ACM Trans. Appl. Perception*, vol. 6, no. 1, pp. 1–22, Feb. 2009.
[140] I. A. Cruz-Albarran, J. P. Benitez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, "Human emotions detection based on a smart-thermal system of thermographic images," *Infr. Phys. Technol.*, vol. 81, pp. 250–261, Mar. 2017.
[141] E. Mostafa, A. Farag, A. Shalaby, A. Ali, T. Gault, and A. Mahmoud, "Long term facial parts tracking in thermal imaging for uncooperative emotion recognition," in *Proc. IEEE 6th Int. Conf. Biometrics: Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–6.
[142] K. Mallat, N. Damer, F. Boutros, A. Kuijper, and J.-L. Dugelay, "Cross-spectrum thermal to visible face recognition based on cascaded image synthesis," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.
[143] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. Mizginov, "Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 1–19.
[144] C. Chen and A. Ross, "Matching thermal to visible face images using a semantic-guided generative adversarial network," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–8.
[145] Z. Wu, M. Peng, and T. Chen, "Thermal face recognition using convolutional neural network," in *Proc. Int. Conf. Optoelectronics Image Process. (ICOIP)*, Jun. 2016, pp. 6–9.
[146] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, "DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 456–463.
[147] Y. M. Elbarawy, N. I. Ghali, and R. S. El-Sayed, "Facial expressions recognition in thermal images based on deep learning techniques," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 10, pp. 1–7, Oct. 2019.
[148] B. Ilikci, L. Chen, H. Cho, and Q. Liu, "Heat-map based emotion and face recognition from thermal images," in *Proc. Comput., Commun. IoT Appl. (ComComAp)*, 2019, pp. 449–453.
[149] S. K. KM, R. Rajendran, Q. Wan, K. Panetta, and S. S. Agaian, "TERNet: A deep learning approach for thermal face emotion recognition," in *Proc. SPIE*, vol. 10993, May 2019, Art. no. 1099309.
[150] P. Kim, "Convolutional neural network," in *MATLAB Deep Learning*. New York, NY, USA: Springer, 2017, pp. 121–147.
[151] R. C. Gonzalez, "Deep convolutional neural networks [lecture notes]," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 79–87, Nov. 2018.
[152] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imag.*, vol. 9, pp. 611–629, Aug. 2018.
[153] O. Simeone, "A brief introduction to machine learning for engineers," 2017, *arXiv:1709.02840*.
[154] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[155] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proc. ICML*, Washington, DC, USA, 2011.
[156] M. Zinkevich, M. Weimer, A. J. Smola, and L. Li, "Parallelized stochastic gradient descent," in *Proc. NIPS*, vol. 4, no. 1, 2010, p. 4.

[157] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 315–323.

[158] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[159] (2019). Google. *The Size and Quality of a Data Set*. Accessed: Dec. 2, 2021. [Online]. Available: https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality

FARDIAN FARDIAN (Member, IEEE) was born in Banda Aceh, Indonesia, in 1979. He received the bachelor's degree in electrical engineering from Universitas Syiah Kuala, in 2003, and the master's degree in computer science from the University of Birmingham, U.K., in 2011. He is currently pursuing the Ph.D. degree with the Doctoral School of Engineering Science, Universitas Syiah Kuala. He has been working as a Lecturer at the Department of Electrical and Computer Engineering, since 2003. His research interest includes ICT development in the education sector. He is an Active Member of the Editorial Board of a nationally accredited journal, *Jurnal Komputer*, *Informasi Teknologi*, and *Dan Elektro.*

MARTY MAWARPURY received the M.Psi. degree in clinical psychology from Universitas Gadjah Mada (UGM) and the Doctoral degree from the Faculty of Psychology, UGM, in 2017. She has been a member of the Psychology Department of Medicine Faculty, Universitas Syiah Kuala (USK), since 2008. Beside her experience working with mental health and mental disorder, she also has many experiences working with human resource development. Her research interests include applied clinical psychology, mental health, management, and health care systems.

KHAIRUL MUNADI (Member, IEEE) received the B.Eng. degree in electrical engineering from the Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 1996, and the M.Eng. and Ph.D. degrees in electrical engineering from Tokyo Metropolitan University (TMU), Japan, in 2004 and 2007, respectively. From 1996 to 1999, he was a System Engineer at Alcatel Indonesia. Since 1999, he has been working as a Lecturer at the Electrical and Computer Engineering Department, Universitas Syiah Kuala (USK), Banda Aceh, Indonesia. He has become a Professor at USK, in 2019. From March 2007 to March 2008, he was a Visiting Researcher in information and communication systems of engineering at the Faculty of System Design, TMU. He was also a Visiting Scholar at the Department of Computer Engineering, Suleyman Demirel University (SDU), Isparta, Turkey, in 2016. His research interests include multimedia signal processing, knowledge-based management, and disaster management. He is a member of APSIPA.

FITRI ARNIA (Member, IEEE) received the B.Eng. degree from the Universitas Sumatera Utara (USU), Medan, in 1997, the master's degree from the University of New South Wales (UNSW), Sydney, Australia, in 2004, and the Ph.D. degree from Tokyo Metropolitan University (TMU), Tokyo, Japan, in 2008. She has been with the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala (USK), since 1999, where she is currently a Professor. She was a Visiting Scholar at TMU, in 2013 and at Suleyman Demirel University (SDU), Isparta, Turkey, in 2017. Her research interests include signal, image, and multimedia information processing. She is a member of ACM and APSIPA.