IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 October 2022, accepted 19 October 2022, date of publication 28 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217904

TOPICAL REVIEW

A Review of Facial Thermography Assessment for Vital Signs Estimation

SYAIDATUS SYAHIRA AHMAD TARMIZI[1], NOR SURAYAHANI SURIANI[1], AND FADILLA 'ATYKA NOR RASHID[2]

[1]Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor 86400, Malaysia
[2]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak 94300, Malaysia

Corresponding author: Nor Surayahani Suriani (nsuraya@uthm.edu.my)

This research was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/TK0/UTHM/02/12) and Universiti Tun Hussein Onn Malaysia (UTHM) through Tier1 (vot H756).

ABSTRACT Estimated vital signs might include a variety of measurements that can be used in detecting any abnormal conditions by analyzing facial images from continuous monitoring with a thermal video camera. To overcome the limitless human visual perceptions, thermal infrared has proven to be the most effective technique for visualizing facial colour changes that could have been reflected by changes in oxygenation levels and blood volume in facial arteries. This study investigated the possibility of vital signs estimation using physiological function images converted from thermal infrared images in the same ways that visible images are used, with a need for an efficient extractor method as correction procedures that have used datasets that include images with and without wearing glasses or protective face masks. This paper, summarize thermal images using advanced machine learning and deep learning methods with satisfactory performance. Also, we presented the evaluation matrices that were included in the assessment based on statistical analysis, accuracy measures and error measures. Finally, to discuss future gaps and directions for further evaluations.

INDEX TERMS Thermal images, features extractions, vital signs estimation, evaluation matrices.

I. INTRODUCTION

Many clinical applications have included vital signs estimation for monitoring an individual based on physical or mental health. These vital signs estimation, such as measurements based on pulse rate, heart rate (HR), blood pressure (BP), blood glucose, respiration (breathing) rate (RR), oxygen saturation (SpO2) and temperature variations are among the most important indicators to predict abnormal conditions, including the detection of chronic illness progression for identifying patients at risks [1], [2]. For this reason, delaying seeking appropriate treatment may have serious consequences not only for themselves but also for their families and communities. Previous studies have shown an increase in research trends particularly for vital signs estimation as an effective marker, specifically using facial imagery such as facial thermography [3], [4], [5]. The facial region is highly suggested because direct confrontation in social interaction and communication that makes it possible for measuring vital signs using the facial region has been identified to be the best way of conveying valuable information such as the state of our organs without stressing on the facial skin area [6].

For this purpose, varied vital signs estimation is essentially needed because the assessment of vital signs differs according to gender, age, body compositions and pathological developments in organ systems depending on the manner of their use [7], [8]. For instance, prior research has suggested developing monitoring systems that require the use of temperature variations in addition to the selection of vital signs measurement to achieve the highest accuracy readings in addressing human states [2]. In this case, instead of relying on the attachment of sensors to skin patients in any part of the body where blood vessels are close to the skin's surface which may result in pain, discomfort, stress, infection risk and invasiveness, most

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

TABLE 1. Infrared spectrum classifications.

Classification	Infrared Spectrum	Wavelength Range (μm)
RIRMV	Near-infrared	0.7-1.0
	Far-infrared	15-1000
	Short-wave infrared	1-3
TIRMV	Mid-wave infrared	3-5
	Long-wave infrared	8-15
	Thermal infrared	8-14

continuously vital signs monitoring nowadays are contactless and noninvasive sensors that are painless and more securely even for the infant's skin. Due to the fact that these contactless and noninvasive vital signs estimations are beneficially helpful for clinicians, medical practitioners or researchers from misdiagnosis of their patients, several factors including the cost and specifications of the equipment, effective approaches for enhancing the facial analysis, contribute in rarely being able to produce satisfactory results with more than 80% correctness and a lack of knowledge about temperature and humidity evaluation as well as expectations about uniform indoor environment conditions without realizing that there may be any spatial variances, have been the reason for limiting the use of infrared images compared to natural colour images [4], [9]. As a result, further research using infrared images which applied advanced approaches from the emergence of computer vision, Machine Learning (ML) and Deep Learning (DL) has given inspiration over time.

Facial thermography is known as facial thermal infrared imaging that was formed from the procedure of mapping the skin temperature of the body regions based on Planck's law of radiation, Wien's displacement law, Kirchhoff's radiation law and Stefan-Boltzmann's law [10], [11], [12]. Many existing infrared spectrums can be accessed from a distance and used for capturing facial thermography, including Near-Infrared (NIR), Far-Infrared (FIR), Short-Wave Infrared (SWIR), Mid-Wave Infrared (MWIR), Long-Wave Infrared (LWIR) and Thermal Infrared (TIR) [13], [14]. As shown in Table 1, these infrared spectrums can be classified as Reflected Infrared Machine Vision (RIRMV) and Thermal Infrared Machine Vision (TIRMV). However, most of the image representations using RIRMV are reflected on natural objects, making TIRMV easily observable using a specialized Forward-Looking Infrared (FLIR) video camera, which has significantly suggested better results. In general, a passive TIRMV does not transmit harmful ionize radiations, making it very sensitive and reconfigurable in complete darkness without the need for active illumination in both controlled, semi-controlled, or uncontrolled environments [15].

Fig. 1 shows an overview study of facial thermography using passive TIRMV. It can be seen that the uniqueness of thermal pattern has the highest correlations with temperature variations present on the face that is produced by an internal body temperature that is controlled by the automatic nervous system, whereas the heat increases the extensibility of soft tissue and causes hemoglobin to release oxygen into tissue [16].

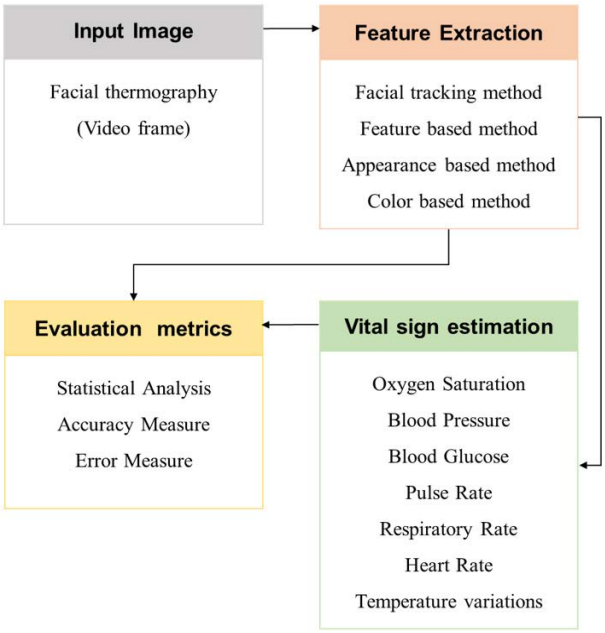


FIGURE 1. An overview study of facial thermography.

Besides, it can also be influenced by physical states, the external and internal environment (hot or cool), physiological factors, emotional distress, skin or fat thickness, blood flow and climate changes [11]. Thus, it is important to follow standard lab settings hence minimize the probability error as well as to maintain excessive movement to ensure consistency, which has a remarkable effect on vital signs estimations [7]. It's interesting to observe how various applications have demonstrated vital signs estimations that subjects must complete based on three scenarios: (1) the measurement is only performed once during the experiments; (2) the measurements are carried out before and after the experiments; and (3) the effects of performing in the before, during and after the experiments are studied. Most importantly, all assessments are monitored from continuous consecutive video frames to achieve the best estimation.

For a better understanding of how to make uses thermal infrared in the physiological function images to analyze pathophysiological abnormalities, various papers from different publications have been considered, including Google Scholars, IEEE Xplore, MDPI, Science Direct, Springer and others. As a result, after defined inclusion and exclusion criteria, the collection of papers for the study is based on ''facial thermography'', ''facial thermal datasets'', ''contactless physiological signs monitoring'', ''feature representations'' and ''deep learning applications in medical fields''.

The main purpose of this paper is to review existing facial thermography applications, as well as to extract facial feature extractor information and the accuracy of vital signs estimation. The remainder of this paper is structured as follows: section 2 discussed the feature extraction of thermal imaging to allocate shape and its facial Region of Interests (ROIs).

Section 3 explored vital signs estimations and their existing application. Section 4 suggests evaluation matrices to validate the studies. Section 5 analyzed the current research gaps. Section 6 discussed the future direction of works. Finally, section 7 draws a conclusion about the study.

II. FEATURE EXTRACTION OF THERMAL IMAGING

In this paper, it can conclude feature extraction procedures can be divided into three main categories: data preprocessing, feature extraction methods and computational models. Also, we highlight the importance of feature extraction methods in an attempt to solve face recognition problems specifically in thermal domain.

A. DATA PREPROCESSING

It is inspiring to observe feature extraction from existing and own datasets containing images of individual subjects with or without wearing glasses or protective face masks. Most existing experiments prefer to request their subjects to remove their makeup, glasses and protective face masks to facilitate the process of locating facial ROIs positions and prevent important information loss around specific facial areas [17]. Recent studies have shown insufficient evidence to support the implementation of extractor methods to solve the problems with subjects wearing glasses or protective face masks during the experiments. This could explain why several researchers have purposefully included their subjects wearing glasses or protective face masks to evaluate the preciseness of their purposed methods [18], [19].

Because the human face varies in shape and size, as well as head poses and orientations that make it difficult for extracting facial features from infrared images, data preprocessing is necessary for thermal imaging as the first step in the experiments. It makes sense when most of the texture, contexture, colour and edge information is completely lost, especially in low quality infrared images which could be the reason why specialized extractor methods are required in the thermal domain [20], [21]. A lot of effort is pulled in to coordinate the features in visible images by mapping them onto infrared images. The data fusion technique has been introduced in an attempt to combine both visible and thermal images to produce a more informative feature than the original thermal images [22], [23]. This technique has proven helpful although the percentage of visible images is lower than that of thermal images. But, it may become complicated if different wavelengths from different sensors are implemented. Another study presented in [24] used Gradient Weighted Class Activation Mapping (Grad-CAM) technique to maximize the visualization of thermal images for use in the facial feature extraction process. Different backgrounds were also employed in the experiments, although there was no significant difference in the results whether a neutral or non-neutral background was used, except for vital sign estimations to examine how the facial pattern changed [9], [25].

There are a total of twelve standard effective colour features that can be utilized to portray temperature distribution from thermal imaging [26]. Indeed, in an attempt to highlight which colours are most suited to exploit, it has been discovered that seven colour palettes are commonly selected for the representation of facial thermography, including contrast, iron, gray, lava, arctic, wheel and rainbow [27]. Because of these varying coloured infrared images, a normalization or binarization process was used to transform thermal images into grayscale images to reduce the effect of blurring, illumination, noise and to minimize computational time. As previously stated, Contrast Limited Adaptive Histogram Equalization (CLAHE) is a popular technique for the normalization process [28], [29]. The normalization or binarization process is performed because most texture extractor methods are preferred to use grayscale images to solve the issues of imbalanced displacement, rotation and scale in the image.

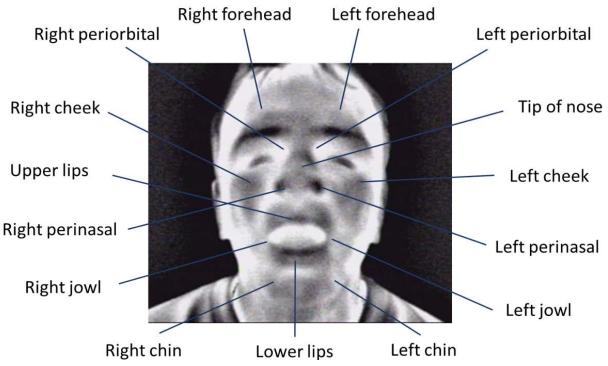


FIGURE 2. Facial region of interest.

B. FEATURE EXTRACTION METHODS

It has been observed that after correction procedures in data preprocessing phases have been applied, four different types of state-of-the-art extractor methods such as facial tracking method, feature based method, appearance based method and colour based method may be suggested to improve the extraction of facial features in the same ways as natural colour images have been used as presented in Table 2. Based on the analysis, it may be useful if facial imaging includes measurements of either globally features from the entire face or locally features focused on the segmented ROIs such as eyes, nose (maxillary), cheeks, mouth (open and close) and forehead (supraorbital) for human health conformation [28], [30]. These features can be extracted automatically with the use of a computational algorithm or manually using bounding box coordinates or sliding window methods [8], [29]. A review suggested that there are around a total of fifteen local facial features that can be identified from the four primary facial parts, as shown in Table 3. In general, it can be noted that these four main regions are frequently discussed by researchers as their primary facial parameters before further analysis is carried out. Fig. 2 illustrated local features in infrared images that were mapped or fused from visible images to compensate for the loss of important facial information.

TABLE 2. Local facial features.

Extractor Methods	Author	Database	Descriptors	Classifiers	Significances
Facial tracking method	[2]	10 subjects (7 male, 3 female)	RetinaFace	-	RetinaFace achieved 99.95% correctness in tracked face images.
	[18]	7920 facial images (7285 subjects with masks, 635 subjects without masks)	UltraLight, Retina Face, Yolov3, LFFD	-	Yolov3 outperformed with a mAP value of 99.3% and precision of 66.10%
	[33]	50 subjects	Ensemble of regression tree	-	The nostril tracking is employed to minimize mislabeled nose regions
	[3]	28 subjects (4 male, 24 female)	Particle filter-based object tracker	-	Speed up tracking region of measurement between frames
	[32]	12000 facial images (7 male, 12 female)	Harris corner detector, SIFT, SURF, ORB	-	The SIFT detector achieved the lowest displacement of the nostril area.
	[34]	30 subjects (18 male, 12 female)	Minimum Eigenvalue detection (Shi-Tomasi), KLT	-	The proposed methods achieved better results for nose localizations to monitor breathing patterns.
	[35]	15 subjects (14 male, 1 female)	TLD predator algorithm	-	The enhanced TLD tracker can trace the mouth and nose in the facial area.
Feature based method	[31]	IRDatabase[1] (2935 facial images) [37]	AAM, DAN, ShapeNet	RF	The DAN methods are the most accurate for detecting face landmarks with a precision of 65.7%.
	[38]	17 children (9 boys, 8 girls)	Viola Jones algorithms	-	The Viola-Jones with probability error is more accurate in locating the facial ROIs.
	[9]	IRDatabase[1] (2935 facial images) [37]	LBP, HOG, DSIFT	SVM, KNN, LDA, NB, RF, BDT	The proposed methods only yielded better results for four different emotions.
	[4]	12 subjects (7 male, 5 female)	Haar cascade algorithm	-	The Haar cascade algorithm can detect human faces and extract ROIs for further analysis.
	[39]	RGB-D-T based face recognition[2] (51 subjects, 15300 images) [40]	-	CNN	CNN classifiers outperform in all face recognition rates, including head pose (98%), expression (99.40%) and illumination (100%).
	[28]	KTFE database (26 subjects) [41]	Hu's moment invariants, histogram statistics	SVM	The SVM polynomial kernel has the highest accuracy of 87.50%
	[42]	EmotionTable datasets (10000x4 features)	HOG	SVM, KNN, Tree	The HOG and Fine Gaussian SVM outperform with 63.50% correctly
	[43]	50 subjects (500 facial images)	Adaptive Quantization of Local Directional Responses Pattern (AQLDRP)	-	The AQLDRP works better in preserving facial feature information at partitioning mode of 4x2
	[36]	Equinox HID face database (3244 facial images) [36]	MACE, OTSDF, PCA, normalized correlation, LFA (FaceIt)	-	The proposed methods outperform for no eyeglass thermal face images
Appearance based method	[44]	160 subjects (79 male, 81 female)	GLCM	SVM, LDA, KNN	SVM achieved the highest accuracy value of 89.37%
	[29]	EmotionTable datasets (460800x4 features)	GLCM	SVM, KNN, Tree	The Weight KNN performed better with best scores of 99.10%
	[45]	IRIS thermal/ visible face database[3] (20 subjects without glasses) [46]	Gabor filters, FastICA	SVM	Results reported that SVM (linear kernel) classifiers obtained an average recognition rate of more than 96%.
	[14]	17 subjects (578 thermal images), Terravic facial IR database[4] (20 subjects (hat, glasses)) [47]	Haar wavelet transform, LBP	ANN, minimum distance	Results reported that only own datasets have the highest accuracy of 95.09%
	[48]	IRIS thermal/ visible face database[3] (1320 images without glasses) [46], Own database (1320 images without glasses, 880 images with glasses)	Edge orientation histograms, co-occurrence matrices	BGMM	BGMM outperforms for benchmark dataset with 96.02% and its own collected datasets with 95.33%.
	[19]	Lab database (39 subjects with glasses), UND database (82 subjects without glasses)	Gabor wavelet transform (GWT), face pattern words (FPWs)	-	The GWT was utilized to generate the FPWs (FPW_1 FPW_2,). The hamming distance is then calculated to identify the two FPWs.
Colour based method	[27]	10 subjects (1000 facial images)	Statistical	-	The colour feature is adjusted to enhance image representation.

[1]https://github.com/marcinkopaczka/thermalfaceproject
[2]https://vap.aau.dk/rgb-d-t-based-face-recognition
[3]http://vcipl-okstate.org/pbvs/bench/Data/02/download.html
[4]http://vcipl-okstate.org/pbvs/bench/Data/04/download.html

1) FACIAL TRACKING METHOD

Similarity positions can be utilized in experiments to monitor a range of heads posed using frontal infrared images rather than non-frontal images such as left and right images, particularly for facial tracking to reduce positional biases [20], [31], [32]. The facial tracking method has been proposed to reduce the head motion caused by the subject's face movement. It is because certain experiments allow their subjects to freely move their head and body in any direction as long as the subject's head positions can still be captured by a thermal camera. The global facial tracking method is noticeable in the studies proposed in [2] and [18], with the use the DL-based object tracking method to track the entire face image with or without wearing a protective face mask with nearly 99% correctly as compared to local facial tracking that may use nose area in studies described in [3], [32], [33], and [34]. It differs from previous studies in [35], which deliberately included the nose and mouth regions because the airflow mainly gets in and out of those regions. In any case, the nose area is the most desirable in several applications, such as RR monitoring is likely to observe temperature variations in the nostril area.

A closer look reveals that the use of the local facial tracking itself is unsatisfactory. For example, as Hochhausen et al. [3] describe, the proposed tracker was coupled with a state transition model and an observation model to optimize the manual process of selecting the ROIs with a labeled region of measurement. Reference [32] uses an active contour algorithm to the expected location produced by four categories of detectors, namely Harris corner detector, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Oriented Fast and Rotated Brief (ORB). The use of four distinct detectors helps in the localization of ROIs in low quality thermal images with satisfactory results. According to Basu et al. [34], adopting Minimum Eigenvalue detection with Shi-Tomasi is insufficient unless combined with the Kanade Lucas-Tomasi (KLT) to track changes by minimizing the dissimilarity between consecutive face frames. Similarly, Chauvin et al. [35] discovered that the ROIs loss problem can only be overcome by combining the proposed detector module into the Tracking, Learning and Detection (TLD) predator algorithm, although it is time-consuming. Thus, this situation explains why tracking local features is considerably more difficult than tracking global features, although a lot of work on thermal imaging has been done.

2) FEATURE BASED METHOD

Because of the complex structural facial features in thermal images, advanced approaches are necessary, notably when the sample image size varies. It could explain why frontal images require advanced approaches to allow for the localization and classification of facial thermography specifically by using a feature based method. Instead of using shape template matching, which has been found in many works in thermal images, an effective solution based on facial thermal feature points, which is a set of points used to extract face image segments using the Dlib landmark detector can be recommended [38]. This procedure is more accurate than the Haar cascade algorithm, sensitive to head movements and perhaps limited to no head motions by locating all the connections up to 68-81 position points in the face from the forehead to the jaw, including detecting the eye, nose and mouth regions [20], [39]. In addition, it is interesting to see the experiments that have been conducted in either a restricted or unrestricted environment, but the reasons why most of these extractors will eliminate pixels that do not correspond to facial features, such as the background area or the hairy part of the face due to the acquisition of meaningful relationships.

TABLE 3. Feature extraction for facial thermography.

Main Facial Regions	Location of Facial Features
Head	Right forehead, left forehead, right chin, left chin
Eyes	Right periorbital, left periorbital, right cheek, left cheek
Nose	Tip of the nose, right perinasal , left perinasal
Mouth	Upper lips, lower lips, right jowl, left jowl

The implementation of extractor methods is not restricted to only one because several studies have successfully tested their datasets for global features with the use of two or more extractor methods. To discover which extractor method is most suited for appropriately shaping facial images, Kopaczka et al. [31] have chosen an Active Appearance Model (AAM), Deep Alignment Network (DAN) and Shape-Constrained Networks (ShapeNet) as their face detection algorithms. DAN method is outperformed with less runtime per frame because it is trained with the updated detecting face bounding box (bound-DAN) and the landmark feature points for the shape update (shape-DAN). For automatically classified emotions, features like Local Binary Pattern (LBP), Histogram Oriented Gradient (HOG) and Dense SIFT (DSIFT) have been applied [9]. As a result, DSIFT and linear SVM outperformed four basic emotions than HOG-SVM, which performed better for eight distinct emotions than the human classification rate. Reference [28] has fused the first five Hu's moment invariants and histogram statistics to classify the four different emotions. The findings demonstrate that the experiment can only perform better when combined with a multi-class SVM classifier. Other research in [36] considered Minimum Average Correlation Energy (MACE), Optimal Trade-Off Synthetic Discriminant Function (OTSDF), PCA, normalized correlation, Local Feature Analysis (LFA (FaceIt)) to address the problem with and without wearing eyeglass in low and high resolution thermal images. However, OTSDF outperformed both before and after eyeglass removal. For no eyeglass detection, OTSDF is more effective for low resolution face sizes than LFA (FaceIt) is best for higher resolution face sizes.

The suggested solution to the problem may lead to different results while aiming for the same goals of preserving facial feature information. Still, some extractors only obtained

approximately 60% accuracy in [31] and [42] and less than 50% in [9] when trained with a variety of parameters. These proposed methods are barely achieving satisfactory results because consideration of varying expression and the effect of head movement mainly in [9] and [31]. The reason is that the facial expressions are spontaneous and humans often misclassified the actual expressions into negative feelings [49]. Hence, a new modification approach is needed to accurately distinguish real emotions that reflect actual thoughts at that moment.

3) APPEARANCE BASED METHOD

Human faces have a broad range of appearance and expressiveness, which could explain why there is textural variability amongst them. Feature extractions under the appearance based methods are diverse and mostly contain multiple extractors with an effort to reduce the number of dimensions of feature space. The appearance based methods have been proposed to extract any relevant information from an image in addition to more correctly analyzing skin textures. These methods can be classified as statistical, structural, signal processing based and model based [50]. It also has advantages and drawbacks depending on the application domain. As a result, several criteria must be considered in selecting the optimum extractor methods, this should include a quality metric, which is an image quality features such as sharpness, tone and colour that may be used to quantify the quality score [51]. Most prominent appearance extractors such as Gray Level Co-Occurrence Matrix (GLCM), LBP, Gabor filters, Fast Independent Component Analysis (FastICA), Binarized Statistical Image Features (BSIF) and many others, prefer grayscale images rather than coloured infrared images for minimizing computational time. Recent studies have inspired the implementation of appearance based methods directly on colour visible imaging, which has resulted in better results with some modifications to obtain all possible colour channel relations [52]. Therefore, it would be a good practice to consider this method for colour thermal imaging in dealing with texture problems.

4) COLOUR BASED METHOD

Colour based method have been used in thermal imaging to monitor the smallest changes in heat energy to distinguish the presence of hotspots for several parts of facial ROIs. Notably, researchers are working with raw infrared images from thermal cameras and some examples have transformed the images using selective colormaps or heatmaps to extract colour information from the image. Meanwhile, [24] performed colour correction procedures by using statistical analysis to adjust the RGB colour channels to find the mean and standard deviation values to ensure that no overlapping colour exists to lessen the effects of environmental and illumination variations.

C. COMPUTATIONAL MODELS

The differentiating extractor methods are ineffective until combined with additional classifiers such as the advanced ML and DL methods in evaluation. Based on research trends, one possible explanation is that most existing computational models, such as ML algorithms have been widely used in feature extraction. According to studies, Support Vector Machine (SVM) classifiers either in linear, radial basis function, quadratic or polynomial kernel have been frequently implemented in object detection due to their ability to manage data with high correlation, faster classification and produce highly accurate results in the range of 80-100% accuracy [28], [44], [45]. As confirmations of the studies, classifiers such as Random Forest (RF), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Bayesian Generalized Gaussian Models (BGMM) and others have been employed to verify the effectiveness of extractor methods. Nonetheless, KNN and BGMM classifiers can also provide the highest accuracy rate of more than 90% correctly in [29] and [48]. It differs from the studies presented in [31], where their experimental analysis achieved less than 70% due to the reasons for the implementation of RF classifiers and additional extractor methods that they have proposed, even with using the same size datasets as proposed in [20].

Researchers have searched into other prominent computational models, such as DL methods which are powerful image processing methods that employ multilayer neural networks. It has achieved outstanding performance in the cases where the evaluation was performed during training and testing with varying datasets size. These DL methods have gained popularity because they can accelerate the process of automatically analyzing faces with high accuracy and low error rates while requiring less average training time. There are limits to distinguishing facial appearance with dissimilar poses, expressions and subjects wearing glasses or protective masks, which has encouraged researchers to select DL methods as the best possible solutions. Previous work has shown that the implementation of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) has proven the most effective in recent years [14], [39]. So, it can be concluded that even though similar datasets were used in the experiments, the proposed methods and the number of parameters used in the experiments could demonstrate varying performance depending on the type of classifiers employed to continue improving the accuracy of the analysis.

D. THE IMPORTANCE OF FACIAL FEATURE EXTRACTION FOR THERMAL IMAGES

As discussed above, facial feature extraction is necessary for the designed experiment. Automatically detecting global features from the whole face or local features will be very difficult with an inability to find faces from a wide variety of input thermal images that are typically in low quality images that are prone to be blurry. At the same time, it is challenging when the complexity remains within the face images due to the unique features of each human that vary from person to person with different size and shape even if it has the standard anatomical positions ROIs. There are many available features method that still suffers more on non-frontal images the

unrestricted experiments since they can only achieve highly robust performance when they are used only on frontal view image. There is also an interest in the pixel size of the actual standard input image that needs to be taken into consideration. Because many studies have reshaped the input images into a smaller uniform size due to the consequences of the processing time required for the training process [24]. Here, it is worth mentioning that the characteristics of an image and the advantages of extractor methods that employ ML and DL are among the most important element in the process whereby many strategies for image enhancement can be adopted.

Many feature extraction methods have been devised to accurately extract facial thermography. In all experiments, the algorithm has been improved to resolve the face recognition issue. As we can see, each of the categorized extraction methods such as the facial tracking method, feature based method, appearance based method and colour based method has its effective ways of addressing the issues dealing with faces by proposing different approaches for different tasks. It may be understandable that based on the comparison of these methods, many papers suggest the reasons why adopting a different method could be a good practice to allow more contribution given for a better more accurate analysis.

As opposed to single feature based methods, a great deal of progress has been made by implementing multi-feature based methods to find out if it may be the best solution in obtaining textural information and vital signs information. Unfortunately, the issues concerning the current success of these multi-feature based methods on how they can offer a promising alternative to the nature of the thermal imaging problem that required more work to be done to improve image representation. This excellent performance may not be accurate as visible imaging with 100% accuracy, however, it can be used as a possible solution in addressing facial features problems. The combination of single feature methods like appearance, colour and shape deformations through the use of facial tracking and feature based methods as multi-feature based can provide more precise findings and there are remain under research. Previously, the fusion of appearances and shape deformations was formerly the popular multi-feature based method, as opposed to the fusion of appearances and colour features. Even though some works claimed there was no significant difference for the inclusion of colour features in appearance, other works demonstrate the association of both colour features and appearance can help in achieving satisfactory result. It may be concluded that the multi-feature based is the best effort in figuring out the optimal feature combinations that can be utilized to increase the face recognition score.

III. VITAL SIGNS ESTIMATION

Literature searches for vital signs estimation using thermal infrared imaging with selected vital signs parameters and additional biosensors signal processing methods seem to be highly limited from previous works. However, with advanced technology and computational algorithms such as DL methods as well as continuous research, it is technically possible to address these challenges with the purpose to extract meaningful patterns of estimated vital signs in classifying an individual's physical or mental health, as presented in Fig. 3. This section will summarize existing research work that can provide suggestions for the future direction of the research.

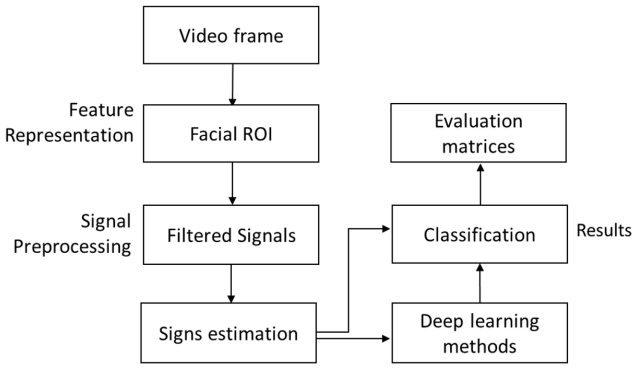


FIGURE 3. Process flow of vital signs estimations.

A. PHYSIOLOGICAL SIGNS MEASUREMENTS

The advantages of thermal infrared imaging in estimating contactless and noninvasive vital signs are not limited to measuring temperature variation, it can also be used in combination with other measurements such as pulse rate, HR, BP, blood glucose, RR and SpO2 depending on the purpose of the studies. However, temperature variation, HR and RR are the most commonly used as an indicator of severe illness in thermal infrared applications [53]. Other measurements of vital signs are also significant in detecting abnormal changes, but the reasons why it has continuously been the least frequently measured are still unclear. All of these estimations may incorporate physiological feature changes in human emotions, stress, environmental effects, before and after performed workload activities as well as security measures that can be utilized to detect fake faces for liveness detection [4], [22], [54]. Humans are homeotherms with the capability to maintain a constant temperature [8]. A constant body temperature can be achieved by balancing heat generation and heat loss. The normal body temperature is lower in the morning (35.5°C) than in the evening (37.7°C) and it rises when there is a fever. Human skin temperature contrasts with core body temperature because the highest temperatures are recorded on the head, neck and torso, followed by the lowest temperature from the limbs to the acral regions.

Visualizing the temperature colour distribution in correlation to vital signs such as HR, which is responsible for pumping blood throughout the body. The process of oxyhemoglobin (HbO2) and de-oxyhemoglobin (Hb) generated by heartbeats during the measurements might cause Heart Rate Variability (HRV) to be higher or lower. BP can be observed when there are substantial differences in blood volume. Normal BP was reported to be 90/60 mm Hg to

120/80 mm Hg. It may reflect blood flow when the heart is contracting (Systolic Blood Pressure (SBP)) and relaxing (Diastolic Blood Pressure (DBP)) which were extensively required in clinical assessment [1].

Pulse rate refers to the number of heartbeats per minute, which is considered to be between 60 to 100. It is related to BP in conditions when there are subtle changes in skin colour resulting from facial arteries that carry pulse signals from the heart. Surprisingly, pulse signals can travel slowly inside the face regions [54]. It may be explained that when some facial images will have a higher temperature in the forehead and another set of facial images will have a higher temperature in the chin areas.

The RR is another form of vital sign estimation. The number of breaths per minute is used to analyze the breathing pattern [55]. Normal breathing rates for healthy adults averaged around 12 to 18 breaths per minute. However, it is often employed in respiratory analysis, for instance to verify whether they are normal, apnea, bradypnea or tachypnea [33], [53]. This RR should be measured consistently and accurately when other vital signs are investigated. SpO2 has a significant important role in common individual survival and is extensively utilized in evaluating respiratory dysfunction. Normal blood oxygenation levels range from 95%-100%, but when oxygen saturation drops below 90%, it is considered low and may result in shortness of breath or hypoxemia [1]. Blood glucose is a biochemical analysis that measures the sugar concentration in a blood sample [44]. It may be affected by the pancreas, which can either produce insulin or not in the body which can be leading to hypoglycemia and hyperglycemia of blood glucose levels [56].

B. BIOSENSORS SIGNAL PROCESSING METHODS

Biosensors signal processing methods, both invasive and noninvasive, have been applied for vital signs extraction to ensure the accuracy of measurements, as illustrated in Table 4. In these cases, invasive biosensors are placed directly on specific body parts, whereas noninvasive biosensors are used without any contact and are typically placed in locations that allow for continuous vital signs signal extractions. HR signals can be extracted using noninvasive biosensors methods with the use of the thermal camera, laser Doppler blood-flow meter [57], Photoplethysmography (PPG) sensor [2], [5] as well as invasive methods like the finger-tip photo reflector [58], and Electrocardiography (ECG) sensor [59]. In contrast, RR estimated from multiple biosensors has been used to prove the validity of vital signs predictions. As invasive methods for RR, the Encephalan respiratory belt sensor [59], force-based chest respiration belt [2], thoracoabdominal respiration sensor [60], respiratory monitor belt (Vernier RMB) [53], [61] and Philips IntelliVue MP30 monitor [3] can be adopted. Then, there are noninvasive RR biosensors such as the thermal camera, Helios spirometer [34], 10-GHz respiration radar [57], 10-GHz microwave radar [58] and MS Kinect depth sensors [62]. In the experiments involving SpO2 measurements, invasive methods such as pulse oximeter sensor [58] or Philips IntelliVue MP30 monitor [3] were utilized.

TABLE 4. Biosensors methods.

Vital Signs	Biosensors signal processing methods
HR	Thermal camera, finger-tip photo-reflector, laser Doppler blood flow meter, ECG sensor, PPG sensor
RR	Thermal camera, Encephalan respiratory belt sensor, force-based chest respiration belt, thoracoabdominal respiration sensor, respiratory monitor belt (Vernier RMB), Helios spirometer, 10-GHz respiration radar, 10-GHz microwave radar, MS Kinect depth sensors, Philips IntelliVue MP30 monitor
SpO2	Pulse oximeter sensor, Philips IntelliVue MP30 monitor
Pulse rate	PPG sensor
BP	Sphygmomanometer, PPG sensor
Blood glucose	Glucometer (blood sample), Olympus auto analyzer, hyperinsulinaemia glucose clamping procedure, continuous subcutaneous glucose monitor
Temperature	Thermal camera, COZIR temperature/ humidity sensors, microbolometer sensor, Atmo-Tube sensors

For pulse rate estimation, it may use an invasive PPG sensor placed on the index finger [59]. The same apparent for estimated BP, which may favor the need for an invasive PPG sensor [63] and a sphygmomanometer [64] to analyze the level of BP. In general, standard biochemical invasive methods such as glucometer can be used to collect blood glucose [44]. It differs from the research proposed in [64], which might include an Olympus auto analyzer and the research proposed in [56], which uses a hyperinsulinemia glucose clamping procedure and a Continuous Subcutaneous Glucose Monitor (CGM) in their biochemical analysis. Measuring temperature variations using noninvasive biosensors was not limited to the use of the thermal camera. It can be found in studies that have included additional sensors such as COZIR temperature/ humidity sensors that were used to monitor the ambient conditions that were close to the subjects [4], microbolometer sensors to record thermal patterns [24] and Atmo-Tube sensors to measure environmental humidity [65].

C. VITAL SIGNS EXTRACTION

As shown in Table 5, various studies have presented the use of only one or multi-parameter vital signs measurements to measure physiological changes in adults or children using thermal infrared technology. Importantly, their research findings highlight how they attempt to solve problems and how their proposed solutions may improve results with a lower error rate. Existing clinical applications based only on statistical analysis or DL approaches will be discussed in this section.

1) HEART RATE EXTRACTION

a: STATISTICAL ANALYSIS

Contactless heart rate measurements using facial thermal imaging can be found in [66] whereas the Laplacian pyramid has been used as a spatial filtering method to boot the signal-to-noise ratio for each video frame. Next, the Fast Fourier

TABLE 5. Vital signs estimation and its applications.

Physiological signs	Applications	Author	Database	DL approaches	Matrices	Significances	Limitations
HR	Effective state detection	[66]	22 male subjects (aged 22-55 years)	-	Accuracy measures	Only whole faces have a 92.46% accuracy rate for HR readings	The estimated HR appears to be less accurate than the referenced HR for several subjects.
		[67]	512 images (19 male, 13 female)	-	Accuracy measures	HR signal is best filtered using the enhancement filtered methods	Excessive movements could result in the unidentified vessel segments
RR	Remote medical diagnostics	[61]	SC3000 dataset (40 subjects, 3600 images), Lepton dataset (31 subjects, 720 images)	CNN-based SR	Statistical analysis, error measures	CNN-based SR was proposed to improve the accuracy of RR estimations.	A larger sample size with automatic ROIs selection is necessary to validate the study.
	PACU monitoring	[3]	28 subjects (4 male, 24 female)	-	Statistical analysis	Three different breathing rate were used to measure oxygen insufflation and no insufflation.	Difficulties in distinguishing breathing patterns in particular situations.
	Stress recognition	[68]	8 subjects (3 female)	CNN	Statistical analysis	Deep CNN with 5-layer networks achieved an accuracy of 56.52% for multi-class and binary accuracy of 84.59% for stress assessment.	The inclusion of several stress parameters could contribute to a reduced accuracy rate
	Hyperventilation detection	[34]	30 subjects (18 male, 12 female)	-	Statistical analysis	The spirometer results were used to distinguish between normal and hyperventilating breathing patterns	The hyperventilating assessment's side effects.
	In-home telerehabilation	[35]	15 subjects (14 male, 1 female)	-	Statistical analysis	Breathing rates were monitored in four different conditions, each with different tasks.	The difficulty in obtaining consistent readings due to the untracked face features
	Apnea detection	[53]	16 subjects (16 for the first experiment, 12 for the second experiment)	-	Statistical analysis, error measures	The selected RR estimators were used to evaluate RR signals based on two different biosensor methods.	The thermal gradient between the human body and the environment was the essential thermal-based method.
Temperature	Anomaly detection	[69]	Male subjects (4976 normal, 195 anomaly)	VAE	Statistical analysis	The proposed VAE was successful in detecting abnormal skin temperature.	Only the male gender ages are being examined.
		[70]	Male subjects (1920 normal)	VAE	Statistical analysis, accuracy measures	The proposed methods were effective in detecting abnormalities for daily assessment at each hour	A large number of data is required for validation.
	Autism spectrum detection	[71]	100 subjects (400 images (50 autistic, 50 non-autistic))	Customized CNN, ResNet 50	Accuracy measures	The customized CNN is the best for ASD classifications with an accuracy of 96% and sensitivity of 100%.	Because the deeper the networks, a larger number of sample datasets is desired for evaluation
	Healthcare monitoring	[27]	10 subjects (1000 facial images)	GoogLeNet	Statistical analysis, accuracy measures	The facial thermal feature has the highest accuracy for both high and too low body temperatures	Few subjects did not accurately predicted due to the misclassification problem
	Anti-spoofing detection	[72]	100 subjects (60 male, 40 female)	-	Accuracy measures	Spoofing attacks have been successfully detected in liveness detection using facial images	The temperature readings in the mouth area is less accurate because of the beard
	Effective state detection	[24]	IRDatabase[1] (1782 facial images) [37], Tufts Face Database (565 images, 113 subjects) [73]	IRFacExNet	Accuracy measures	The IRFacExNet model with the Snapshot ensemble achieved 88.43% for IRDatabase and 97.06% accurately for Tufts Face Database	Additional techniques must be considered for improving the accuracy rates and image misclassifications
		[74]	USTC-NVIE dataset (105 subjects) [75]	ET-CycleGAN	Accuracy measures	The ET-CycleGAN can recognize emotions by preserving the expression of the mouth.	The findings are limited by the small sample sizes in the datasets
		[76]	USTC-NVIE database (1891 samples) [75], Equinox HID face database (1264 samples) [36], MAHNOB laughter database (10364 samples) [77]	DBM	Accuracy measures	The proposed DBM outperforms the mixDataset with 68.2% accuracy, while the only NVIE database has 62.9%. accuracy.	Huge sample size datasets are needed to enhance the accuracy of the emotional classifications.
		[78]	12 subjects (6 male, 6 female)	ANN	Statistical analysis, accuracy measures	The ANN classifiers successfully identified the workload tasks with 98.9% accuracy for individual subjects and 81% for all subjects.	The deeper networks are necessary if additional samples are added to produce the best classification results.
HR, RR, pulse rate	Emotion recognition	[59]	50 subjects (aged 18-35 years, women=65.0%)	FFNNs, RF	Statistical analysis, accuracy measures, error measures	The FFNNs produced the best average accuracy of (70% ± 0.8%) when compared to RF classifiers	The experiments included a smaller sample size of the participants
HR, BP	Healthcare monitoring	[63]	MIMIC-II datasets (900 data instances) [79], IRDatabase[1] (2935 facial images) [37]	FCN	Statistical analysis, error measures	The PPG signals successfully estimated HR and BP with the addition of FCDNN	Various signal preprocessing methods are needed to enhance the raw PPG signals.
RR, temperature	ICUs monitoring	[80]	Datasets 1 (26 subjects, 3900 images), Dataset 2 (6 subjects, 1920 images)	YOLOv4, YOLOv4-Tiny	Statistical analysis, accuracy measures, error measures	The YOLOv4-Tiny performed better when analyzing the impact of different levels of disturbance.	The challenge in extracting RR signals as a result of movement disturbances.
	Physical activity	[62]	25 experiments (per individual)	The two-layer neural network	Statistical analysis	The mean regression coefficients for a lower temperature is -0.162°C/min and for frequency load is -0.72 bpm.	Temperature changes may be longer due to an increase in air flow volume
HR, RR, temperature	Physiological motoring	[81]	L-CAS thermal physiological monitoring dataset[5] (5 subjects, 3000 images) [82]	-	Statistical analysis, error measures	The proposed systems were effective in detecting any physiological changes in unrestricted conditions.	The RR and HR showed less satisfactory results because of the head movements during the experiments.
HR, RR, SpO2, temperature	Infection disease and fever screening radar systems	[58]	83 subjects (35 influenza patients, 48 normal control)	NN, k-means clustering method	Statistical analysis, accuracy measures	Various vital sign estimations were included to assess the level of influenza infection	Only the male gender is being examined.
Blood glucose, BP, pulse rate, temperature	Diabetes Mellitus	[56]	24 subjects (10 aware, 14 unaware)	-	Statistical analysis	Only aware group has a higher hypoglycemia symptom score and higher adrenaline (epinephrine) levels	The aware group has a higher glucose level compared to the unaware group.
Blood glucose, BP, temperature		[64]	62 subjects (27 male, 35 female)	-	Statistical analysis, accuracy measures, error measures	The SBP of the DM group positively correlates with age and negatively with skin temperature of the forehead, eye and ear.	Only seven body regions have positive values for the 95% confidence interval (CI) measure under the lower bound group

[5]https://lcas.lincoln.ac.uk/wp/research/data-sets-software/lcas-thermal-physiological-monitoring-dataset/.

Transform (FFT) and zero crossing were used in the spatio-temporal filtering to filter HR signals. The findings were compared with the ground truth obtained from PPG signals and the Complement of Absolute Normalized Difference (CAND) index based on the measurements from the detected ROIs in the whole face, the upper half of the facial and the supraorbital area.

Reference [67] tracked the vessel segments in the forehead region using iterative closest point tracking methods for the HR estimations. Three filtering methods such as FFT, wavelet-based, and a combination of wavelet-based and blood perfusion enhancements were utilized to filter HR signals based on the physiological responses to normal, mild pain and mild exercise. The reference ECG signals with the estimated HR from facial thermal-based were analyzed to find the optimum filtering methods for the estimated HR.

2) RESPIRATION RATE EXTRACTION

a: DEEP LEARNING METHODS

Kwasniewska et al. [61] retrieved raw RR signals from low-resolution thermal images using the average and skewness aggregation operators before filtering them with the fourth-order Butterworth Filter (BF). To enhance the accuracy of the estimated RR for both datasets, the CNN-based Super Resolution (SR) models were proposed and evaluated with the Eulerian Video Magnification (EVM) algorithm and bicubic interpolation.

Cho et al. [68] exploited RR signals to automatically assess stress levels based on breathing patterns from collected thermal video using a thermal gradient flow tracking algorithm. A Power Spectral Density (PSD) function was included to recover RR signals from a short period video before it was then augmented using a unidirectional sliding copper with a square window and evaluated using CNN based models. The 5 layers of CNN and Shallow Learning (single layer network with varying level RR) were used to evaluate the performance of the DL approaches.

b: STATISTICAL ANALYSIS

Hochhausen et al. [3] have monitored patients in Post-Anesthesia Care Units (PACUs) to assess RR signals derived from video thermal and body surface ECGs. The second-order BF was applied to filter raw RR signals and was computed using the three estimators function. However, only the estimated RR from video thermal is better with lower correlation coefficient values.

Basu et al. [34], demonstrated estimated RR by manually selecting facial ROIs using a bounding box and facial tracking detector. The raw RR signals are then filtered using a Low Pass Finite Impulse Response (LPFIR) filter to minimize noise caused by the head, body movements or atmospheric disturbances in the thermal images. The estimated RR is computed based on the number of peaks (exhalations) and crests (inhalations) in one complete breath cycle. The normal RR was obtained from the Slow Vital Capacity (SVC) test and hyperventilation was recorded from the Maximum Ventilator Volume (MVV) test.

A similar is apparent for Chauvin et al. [35] used bounding box and facial tracking detectors to manually locate facial ROIs for RR monitoring. A low resolution of FFT with a quadratic interpolation is applied to compute RR from thermal video using the Hann window approach. The acquired results were then compared with the respiration belt data using analog voltage readings to analyze the breathing rates.

Rumiński [53] proposed detecting apnea by extracting raw RR signals with fourth-order BF filters. To analyze the RR pattern signals, four types of RR estimators were utilized to determine the frequency value and evaluate the respiration frequency spectrum based on thermal-based as estimated measurements and pressure belt-based as reference measurements.

3) TEMPERATURE VARIATION EXTRACTION

a: DEEP LEARNING METHODS

The findings, which primarily use basic parameters such as temperature variations have been observed in anomaly detection applications using the deep generative models, namely Variational Autoencoder (VAE) [69], [70]. This model is comprised of an encoder and a decoder, which help minimize overfitting and achieve a high recall. The gradient decent method was used for VAE parameter learning, while Adam was employed as an optimization algorithm. Normal samples were chosen as the primary indicators for analyzing the abnormality temperature patterns. In contrast in reference [70], the Hotelling theory is also utilized to assess the performance of anomaly detection.

In individuals diagnosed with or without Autism Spectrum Detection (ASD), both emotional expression information and temperature changes were studied [71]. The temperature obtained in response to emotional arousal was computed to determine the minimum, maximum and average values. To evaluate which model is most suited for automatic classification, the performances of the customized CNN with ReLU function were compared to the ResNet 50, that includes ReLU and soft-max function.

Wang et al. [27] implemented GoogLeNet with an inception network structure in an attempt to increase accuracy and reduce computational time in a non-contact healthcare system. Four categories of facial images were evaluated and classified under five distinct temperature variations to determine which facial images are the most suitable in predicting health status.

Other researchers have presented an effective state detection to address the possible effect of emotional state and discomfort in controlled or uncontrolled experiments. For example, Bhattacharyya et al. [24] used the InfraRed Facial Expression Network (IRFacExNet) with the help of residual and transformation units to classify eight different facial expressions based on the facial temperature changes. A ReLU function is applied to bring the non-linearity and batch

normalization to make convergence faster and have a moderate generalization effect. The performance of the proposed IRFacExNet is evaluated using different sample size datasets to compare which datasets are more accurately classified.

Pons et al. [74] developed an Emotion-Guided Thermal CycleGAN (ET-CycleGAN) for classifying sixth facial emotions. To prevent overfitting, batch normalization was used. The performance of the ET-CycleGAN and ET-CycleGAN without involving the loss function was examined to see which classifiers are preferable for predicting emotional classifications.

Wang et al. [76] exploited the Otsu threshold algorithm to remove the highest frequency infrared spectrums while minimizing the influence of temperature changes. For accurately classified emotional states, a Deep Boltzmann Machine (DBM) model with layerwise pretraining (using Gaussian-binary RBM and binary RBM layer network) and joint training have been proposed, and two types of datasets have been used: samples from only one datasets and samples from mix datasets. The results show that DBM performs better when a large number of sample data are used.

Another finding published in [78] used a noninvasiveness thermal camera to automatically analyze facial skin temperature changes based on the three levels of mental state workload. The ANN models with softmax activation function were employed to accurately classify the cognitive workload assessment, which included low, moderate and high levels. The results were compared between individual assessments and assessments of all participants.

b: STATISTICAL ANALYSIS

In security applications, various research has been introduced in an attempt to strengthen security measures by analyzing any possibility of spoofing attacks using facial recognition systems. For example, Singh and Arora [72] proposed detecting real and fake facial in liveness detection by utilizing facial thermography in their research. Temperature measurements were used as the main indicators in this study, for highlighting which facial ROIs had the average minimum and maximum temperature readings, which were used to detect spoofing attacks.

4) MULTI-PARAMETER VITAL SIGNS EXTRACTION

a: DEEP LEARNING METHODS

Filippini et al. [59] proposed end-to-end emotion recognition using HR, RR and pulse rate. The warping procedure based on the Local Weighted Mean (LWM) algorithm was utilized to track the selected ROIs. The infinite impulse response filters such as the third-order BF were used to filter ECG signals for HRV, PPG signals for pulse rate and the Encephalan respiratory belt sensor for RR signals. The Feedforward Neural Networks (FFNNs) results were compared to RF algorithms results, and the best classifiers were chosen as the final classification.

Another paper showed that estimated HR and BP from thermal images can be done by transforming the raw PPG signals to the frequency domain using FFT and applying Inverse FFT (IFFT) to transform frequency domain to time domain [63]. The preprocessed PPG signals are then filtered using the Short Time Fourier Transform (STFT) for HR and BP using the Infinite Impulse Response (IIR) Chebyshev Type 2 Filter (CT2F). The BP was then predicted using a Fully Connected DNN (FCN) with Nadam optimizer, which may be classified as SBP or DBP.

The RR and temperature measures are essential indicators used to monitor patients in Intensive Care Units (ICUs) [80]. To improve feature detection accuracy, two proposed DL networks, YOLOv4 and YOLOv4-Tiny were employed. The optical flow with a temporal filter algorithm was implemented for RR extractions, whereas the relative deviation was used for temperature extractions. YOLOv4-Tiny performed better due to the model pruning effect, which reduced the number of unnecessary parameters.

Reference [62] have also recommended analyzing RR and facial temperature in monitoring an individual's physical activity using home a exercise bike. A two-layer neural network comprised of the sigmoidal and softmax functions was deployed to measure the minimum and maximum temperature. The finite impulse response filter was chosen to eliminate noise and filtered RR signals from thermal video (using the fixed and moving ROIs) and MS Kinect depth sensors. Both estimated temperature and RR were studied to analyze the physiological changes during physical activity and resting situations.

In the studies presented by Sun et al. [58], physiological indications such as HR, RR, temperature and SpO2 are included to validate the studies. The estimated measurements such as HR, RR and temperature are analyzed using a neural network with Kohonen's Self-Organizing Map (SOM) and a k-means clustering method. The estimated SpO2 was then used as a reference measurement and was conducted using the chi-squared test. Both estimated and reference measurements are used to classify patients into three influenza groups: non-influenza, lower-risk influenza and higher-risk influenza.

b: STATISTICAL ANALYSIS

Healthcare screening for the elderly with mild cognitive impairment has been proposed in [81], which includes the temperature, RR and HR measurements. A morphological closing operation is performed to improve the representation of the facial contours and the average temperature variations are recorded. A Hamming windows is used to select facial ROIs for the extraction of the HR and RR signals before using FFT analysis to normalize the temperature signal from the thermal video. The results were compared to analyze the physiological differences between the static position and the effects of head movements.

Estimated blood glucose from biochemical analysis is insufficient in determining diabetic patients in Diabetes Mellitus (DM) applications. Sivanandam et al. [64] indicated that blood glucose, BP and temperature are significantly important in diagnosing their patients to verify the results,

but in research presented by Sejling et al. [56], included estimated pulse rate as an addition to get the best result in screening tests. The ECG signals were extracted to analyze the maximum and minimum pulse rates for five respiration cycles. The decrease in facial skin temperature caused by high insulin resistance can be detected with a noninvasive thermal camera. As a result, the uniqueness of temperature patterns was examined to analyze the minimum and maximum temperature presented on facial areas.

D. RELATION OF VISION BASED ON THE SUCCESSIVE VITAL SIGN MEASUREMENTS

Because most invasive methods nowadays can be replaced with noninvasive ones, many areas of fields have started to cover all possible vital signs measurements as an essential biomarker to measure physiological responses of the body. The curiosity has leads to if the proposed noninvasive thermal vision based approach with minimally use of biosensors method based on image sensors which can be acquired using facial thermal imaging is only applicable to common situations with the included selection of vital signs parameter. Concerning the input datasets, participants in experiments typically no longer include average citizens or students, but their usage is relatively limited and almost no studies have made use of more than two datasets either by using their own collected datasets or public datasets. The existing studies use a small number of subjects that might be biased in age and gender whereas less focus exclusively on elders or children subjects in their thermophysiological data analysis. For such reasons, sensitivity may have a critical effect on the measurement due to human anatomical responses to the environment are not always the same even for identical twins and may differ according to the subject's geographical location and skin types.

The challenges in this discussion are concerning the labeled region of measurements for each face frame. For this, monitoring a common region that predominates in the vital sign measurements might produce better results with faster processing time. In several existing studies, it is preferable to eliminate the subjects based on exclusionary criteria, such as excessive movements with less precise physiological readings that may interfere with the reasons for higher accuracy. Nevertheless, the physiological signal acquisition that simultaneously read through the use of thermal imaging has become crucial in analyzing vital signs measurements at the time of examination. Examples of physiological signal acquisition that may be acquired from the face directly, such as the exploration of temperature variations that accurately portray the important physiological changes as well as HR estimations that are derived from the maximum and minimum RR pattern, are useful for estimating blood pressures and pulse rates. However, most of the physiological signal acquisition relies on the proposed band-pass filtering method, which was used to filter the frequency of the physiological signal. For instance, RR patterns often have a narrow and low frequency, in contrast to HR patterns, which typically have a higher frequency since that region has the most blood flow [2], [68].

Reference [83] reveals that video duration and the accuracy of the vital sign readings are two conditions that influence the motivation for use. It is worthwhile to note, that longer videos than necessary are not always practicable in some applications and may result in less accurate reading when compared to preferred shorter videos that produce more accurate readings. It is important to remark that the screening system may require the verification of several physiological indications to address human states. As we can see, the experiment could be carried out to analyze the referenced measurements using ground truth collected from measurements of interest with at least one possibly invasive or noninvasive biosensor and estimated measurements that observed from noninvasive biosensors directly on the face at a minimum distance from the thermal camera. Besides, the incremental sample size is required to improve the training process in some circumstances whereas it helped monitor a participant under unrestricted conditions with less physiological signal, which mainly may provide the lowest accuracy outcomes.

For better obtaining valuable information regarding human activities and physiology, the proposed methods based on DL approach and statistical analysis have been highlighted in this paper. The use DL approach is still receiving less attention as opposed to the use ML approach in real-time although it is proven to have superior performance in various studies. In the last few years, the use of the DL approach has become demanding more with updated versions of the DL approach using different training techniques and optimizers. From this, we believe that it deserves further consideration in better understanding any potential underlying pathological process to obtain the most superior accuracy and it remains uncovered in research for a particular application.

IV. EVALUATION MATRICES

Research suggested three primary categories of evaluation matrices that may be utilized in validating feature extractor methods and vital signs estimation in terms of statistical analysis, accuracy measures and error measures as shown in Table 6. Based on the analysis, the implementation of the evaluation matrices is depending on how the researchers verify their research to produce the best results. Overall, various applications may have only performed statistical analysis or been fused with accuracy measures or error measures, while others may have only used accuracy measures or been fused with the error measures as illustrated in Fig. 4. Based on the analysis in Table 5, the use of evaluation matrices from 23 different research articles from years 2012 to 2022 can be classified into five categories, with the highest percentage contributions of 31% for statistical analysis and the lowest percentage contributions of 13% for the fusion of statistical analysis, accuracy measures and error measures. Most available applications have rarely mentioned the fusion of all three categories of evaluation matrices due to the reasons

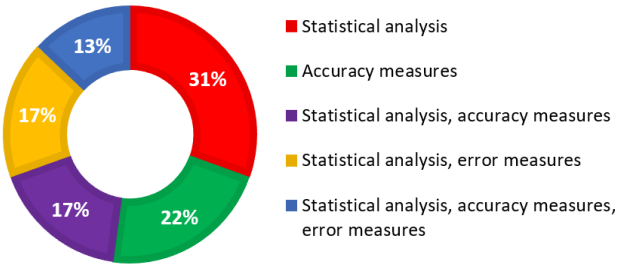


FIGURE 4. The evaluation matrices in the previous studies.

TABLE 6. Evaluation matrices.

Evaluation Matrices	Measurements
Statistical analysis	Mean, standard deviations, variance, median, minimum, maximum, total, contrast, correlation, energy, entropy, homogeneity, Mean Average Precision (mAP), coefficient of variation, covariance
Accuracy measures	Accuracy, sensitivity, specificity, precision, recall, F-measure, kappa values, ROC area, AUC, TPR, FPR
Error measures	MAE, MSE, RMSE, RAE, RRSE

for their proposed works, which may have offered promising satisfactory results.

For the statistical analysis, the mean and standard deviation are two possible statistical measurements that had been used frequently in most published works specifically in calculating the vital signs estimations for summarizing data variability. Other measurements have also been considered, for example in the experiment described in [44] where statistical analysis was performed by including twenty two measurements to compute their extractor methods, in contrast to the proposed works in [6], which only comprised four types of measurements. Table 7 shows the list of the statistical analysis that had been used in the prior research.

Previous studies highlight the importance of selecting alternative accuracy measures and error measures to verify the efficiency of the proposed methods more accurately with the deployment of computational algorithms such as machine learning and deep learning methods. To achieve optimal classifier performance, the sample datasets were tested and trained to accurately classify the thermal images using cross-validation, percentage splits method, holdout and no validation [29]. Table 8 presents the confusion matrix that was used to generate the accuracy measures as shown in Table 9. The confusion matrix is divided into two classes: the rows represent the actual values and the columns represent the predicted values, both of which are used to evaluate the model. To ensure that the results are valid and reliable, only a value which is close to 1 from the standard accuracy measures is selected because they can indicate perfect accuracy results. This review also presents the deployment of error measures that are significantly important as alternatives in the case whereas the successive classifiers have demonstrated the similarity of accuracy rate with minimal computational time

TABLE 7. Statistical analysis.

Statistical Analysis	Descriptions	Formula
Mean	The average of all the observations	$\frac{\sum x_i}{N}$
Standard deviations (std)	The average amount of variability in the dataset	$\sqrt{\frac{\sum(x_i-\mu)^2}{N-1}}$
Variance	Calculate the variation from the average and mean	$\frac{\sum(x_i-\mu)^2}{N}$
Median	The middlemost number or center value in the set of data	$\frac{n+1}{2}$
Minimum	The low outlier limit which is the smallest value in a set of values	$\mathrm{Min}(x)$
Maximum	The high outlier limit which is the highest value in a set of values	$\mathrm{Max}(x)$
Total	The total amount of number obtained through addition	$\sum x_i$
Contrast	Measure the dissimilarity between various image features	$\sum_{i,j=0}^{N-1} P_{ij}(i-j)^2$
Correlation	The statistical measure that expresses how closely two variables are linearly related.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$
Energy	Measures the uniformity of an image	$\sum_{i,j=0}^{N-1} p_{i,j}^2$
Entropy	Measures the randomness of the intensity distribution in the image	$\sum_{i,j=0}^{N-1} P_{ij}(-lnP_{ij})$
Homogeneity	Used to describe a set of data with the same variance	$\sum_{i,j=0}^{N-1} \frac{p_{ij}}{1+(i-j)^2}$
Mean Average Precision (mAP)	Measuring the accuracy of object detectors. Calculates the average precision value for recall values ranging from 0 to 1	$\frac{1}{N}\sum_{i=1}^{N} AP_i$
Coefficient of variation	Measurement of the relative variability or consistency of data.	$\frac{\sigma}{\mu}$
Covariance	The relationship between the two variables direction	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$

TABLE 8. The confusion matrix.

Predicted Values	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	True positive (TP)	False positive (FP)
Negative (0)	False negative (FN)	True negative (TN)

to further improve the exactness of the proposed methods as illustrated in Table 10. Therefore, in the result, it can be stated that which one of the selected approaches is the best to use in comparative analysis with minimizing output error as well as may result in better classification.

V. RESEARCH GAPS

In this section, some of the research gaps will be discussed in order to address the limitations of the facial thermography studies.

TABLE 9. Accuracy measures.

Accuracy Measures	Descriptions	Formula
Accuracy	Defined as a correctly classified measure	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	Referred to as sensitivity or true positive rate (TPR), which is the receiver operating characteristic (ROC) curve used to clarify whether the correctly classified is genuinely positive.	$\frac{TP}{TP+FN}$
Specificity	It was known as false positive rate (FPR) is used to indicate that the correctly classified is genuinely negative.	$\frac{FP}{FP+TN}$
Kappa values	Determined by adjusting the sensitivity and specificity values.	$\frac{Accuracy-\ p_e}{1-p_e}$
Area Under The Curve (AUC)	Used to summarize the performance of the classifiers.	$\frac{1}{2}\times\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$
Precision	Measure for correctly classified as positive.	$\frac{TP}{TP+FP}$
F-measure	Based on a combination of the recall and precision measures	$\frac{2TP}{2TP+FN+FP}$

TABLE 10. Error measures.

Error Measures	Descriptions	Formula
Mean Absolute Error (MAE)	The measure for the average of the absolute value between the predicted and actual value.	$\frac{\sum_{i=1}^{n}\lvert p_i-a_i\rvert}{n}$
Mean Square Error (MSE)	Defined as the predicted value of the estimator squared deviation from the actual value	$\frac{\sum_{i=1}^{n}(p_i-a_i)^2}{n}$
Root Mean Square Error (RMSE)	Measurement of the dissimilarity between the predicted and actual value. It employs a standard deviation for the prediction errors operation.	$\sqrt{\frac{\sum_{j=1}^{n}(P_{ij}-T_j)^2}{\sum_{j=1}^{n}(T_j-\bar{T})^2}}$
Relative Absolute Error (RAE)	The ratio used to divide the absolute error by the magnitude of the actual value.	$\frac{\sum_{i=1}^{n}\lvert p_i-a_i\rvert}{\sum_{i=n}^{n}\lvert\bar{a}-a_i\rvert}$
Root Relative Squared Error (RRSE)	measure that relative to the error for the average of the actual values	$\sqrt{\frac{\sum_{j=1}^{n}(b_i-a_i)^2}{\sum_{j=1}^{n}(a_i-\bar{a})^2}}$

A. INFLUENCE FACTORS

Several studies have addressed the influence factors of the thermal infrared in detecting human physiological dysfunction. These factors must be further analyzed to ensure proper measurements can be taken into account. For example, most thermal imaging always suffers from low resolutions as compared to visible images that are often acquired with a high resolution. The blurring problem with the thermal image always occurs because the thermal camera was too costly and has various sensor characteristics with multiple image colour palettes. The distance between skin and camera as well as recording time can also contribute to the unclear appearances that could hinder the thermal interpretation that may reduce the accuracy of vital sign measurements.

Additionally, previous research has shown gender bias when a large number of studies use male participants rather than females in the experiments. It will be doubted whether such research is reasonable if the exclusion of the female gender is consistently recorded unless a strong argument has been made. For example, [84] explained that males are more willing than females to participate in medical research studies based on study risk level due to a lack of trust and privacy concerns. In clinical applications, the inclusion of gender-balanced should be considered so that the experiment can be conducted to highlight gender-related differences and their physiological characteristics.

To study the thermal sensitivity and autonomic responses to environmental impacts in indoor and outdoor settings, a collection of thermal images may be obtained in a controlled, semi-controlled and uncontrolled environment. In this setting, the important parameters such as temperature distribution (hot and cold), weather, humidity levels (relative and absolute), airflow, clothing insulation and metabolic rate must take priority to prevent less accurate estimation [4], [85], [86]. Because of this, inclusion criteria such as demographic information including gender, age and ethnicity must be addressed when validating the study to overcome measurement bias. The efficacy of the proposed algorithms to enhance the performance of thermal images in detecting intra-class variations such as illumination variation (changes in the lighting), facial expression (watching emotion-induced videos), change of head poses, voice commands and obstacles (glasses, hat, face masks) have been discussed in all studies. Most algorithms may perform well in this scenario if a large number of intra-class variations sample images are observed.

B. FACIAL THERMOGRAPHY DATASETS

There are a few existing facial datasets that may include thermal imaging that is publicly available for open or restricted access, allowing for ongoing contributions to improve previous research works. The number of available datasets for natural colour images has grown rapidly even in recent years when compared to thermal images, encouraging researchers to make decisions by collecting their thermal datasets to test their approaches. However, most of their thermal datasets are not publicly available due to privacy issues. It can be seen that with the latest advanced technology, the quality of thermal image representation has improved better than the existing thermal images, which may consist of low quality images that have caused the loss of informative features. Table 11 summarizes some head poses, illumination and expression obtained from a publicly available facial thermography dataset in either an indoor or outdoor setting.

Dataset such as Equinox HID face database contains grayscale images of 91 subjects of 115 subjects after removing corrupted images from 24 subjects. The thermal images

are monitored by SWIR, MWIR and LWIR cameras. Subjects speaking, expressions (frown, surprise, smile) and illumination variations (frontal, lateral (right, left)) are among the images captured over two days in indoor settings. However, 3 of the 91 individuals are only wearing glasses. The database is no longer open access, which may result in a decrease in the number of available thermal datasets.

USTC-NVIE dataset consists of 215 subjects (157 males, 58 females) aged from 17 to 31 who were recorded while watching emotion-induced videos for 3-4 minutes using a SAT-HY6850 thermal camera. It has six different expressions (angry, disgusted, fearful, sad, happy and surprised) and illumination variations (frontal, lateral (right, left)). IRIS thermal/ visible face database, a publicly dataset was conducted in controlled settings involving 30 subjects (28 men, 2 women) using a Raytheon Palm-IR-Pro thermal camera and Panasonic WV-CP234 visible camera. Each subject was captured from eleven different angles with three facial expressions (surprise, happy, angry) under five illumination conditions (left light on, right light on, both light on, both light off, dark room). Some of the facial thermal images have a dark colour in the nostril because of breathing. Only 10 of the 30 subjects wear glasses.

Another accessible thermal imaging is the KTFE database. The datasets involved 26 subjects (16 males, 10 females) aged from 11 to 32 with different nationalities such as Vietnamese, Thai and Japanese, and seven expressions including neutral emotion in controlled settings using NEC R300 thermal camera. It has 14 subjects who wear glasses. To help subjects return to a state of neutral feeling, instrumental music was played before and after each session. The longest designed datasets for long periods such as UND-X1 datasets provide facial images from 82 subjects taken at various sessions over 10 weeks with three facial expressions (neutral, smile, laugh) and different lighting changes (FERET lighting, mugshot). The problem is that thermal images which have a lower resolution and noise when compared to high resolution visible images.

High resolution thermal images can be found in IRDatabase that were recorded using an Infratec HD820 infrared camera from 90 subjects with eight different facial expressions (fear, anger contempt, disgust, happy, neutral, sad and surprise) and nine head poses (upper left, upper frontal, upper right, frontal right, full frontal, frontal left, lower left, lower frontal, lower right). Terravic facial IR database contains grayscale thermal images collected from 20 subjects (19 men, 1 woman) in both indoor and outdoor settings with different variations (frontal, lateral (right, left)) using a Raytheon L-3 Thermal-Eye 2000AS. However, only 22784 thermal images from 18 subjects can be used as another image from two subjects was corrupted, with a total of 2 subjects without wearing glasses and a hat, 7 subjects wearing glasses and the remaining 9 subjects wearing glasses and a hat.

RGB-D-T based face recognition acquisitions in the indoor setting using AXIS Q1922 thermal camera and Kinect RGB

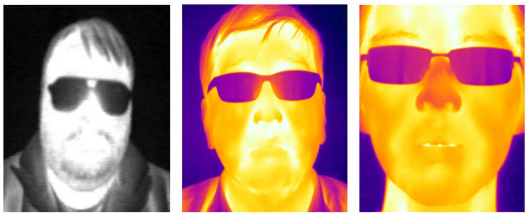


FIGURE 5. Examples of facial images with glasses [47], [87], [88].

camera. A total of 51 college students participated in the experiments. The obtained images contain head poses (frontal view, four side turns of the head, one for turning the head up and down), facial expressions (neutral, happy, sad, angry, surprised) and varying illumination whereas five lamps are used as light sources, each lamp being sequentially turned on and off by the operator. L-CAS thermal physiological monitoring dataset comprising rosbag files of 5 videos for 5 subjects, every 2 minutes with the ground truth of the respiration rate and heart rate using Optris PI-450 thermal camera. It also includes varying head poses (up, down, forward, backward, right, left) with 10 seconds for each position after one minute remains still.

The largest SpeakingFaces dataset, consisting of 142 subjects (74 males, 68 females) ranging in age from 20 to 65, was recorded from nine different angles positions that were acquired in semi-controlled laboratory settings with a fixed intensity of the light source using a FLIR T540 thermal camera and Logitech C920 Pro HD web-camera. Each subject was asked to read 100 English phrases or imperative commands displayed on a video screen. The TFW dataset provides thermal images of 147 subjects with 16509 labeled faces collected in indoor (controlled, semi-controlled) and uncontrolled outdoor (images taken during sunny and cloudy weather at different locations) settings, with some of them containing subjects from SpeakingFaces datasets using a FLIR T540 thermal camera. The images with and without face masks were collected for semi-controlled indoor and uncontrolled outdoor environments.

It can be observed the available datasets contained several subjects who were either only males or both males and females to analyze their physiological signs estimations. The images have a variety of head poses that are not only frontal and lateral (right, left), but also up, down, forward and backward, as seen in the L-CAS thermal physiological monitoring dataset. The illuminations variation such as frontal, lateral, dark room and lighting on and off sequentially at a time are included in an attempt to study the impact of lighting changes on facial recognition tasks. It also has existing datasets with vowel frames, such as the Equinox database, SpeakingFaces and TFW dataset which asked their participants to speak some phrases. Then, expression frames like IRDatabase, which has up to eight emotional moods including fear, anger contempt, disgust, happy, neutral, sad and surprise.

Several datasets have the collection of facial images of a subject wearing glasses such as Equinox database,

TABLE 11. Facial thermography datasets.

Datasets	Number of subjects	Total number of images	Thermal	Visible	Poses	Illumination	Expressions
Equinox HID face database [25], [36], [76], [89]	91 subjects	3244 images (320x240) pixels	✓	✓		✓	✓
USTC-NVIE [74]–[76]	215 subjects (157 males, 58 females)	N/A Thermal image (320x240) pixels Visible image (704x480) pixels	✓	✓	✓	✓	✓
IRIS thermal/ visible face database[3] [23], [46], [48], [90]	30 subjects (28 men, 2 women)	4228 images (320x240) pixels	✓	✓	✓	✓	✓
KTFE database [28], [41]	26 subjects (16 males, 10 females)	126 GB video (320x240) pixels	✓	✓			✓
UND-X1 [19], [91], [92]	82 subjects	4584 images Thermal image (312x239) pixels Visible image (1600x1200) pixels	✓	✓	✓	✓	✓
IRDatabase[1] [9], [20], [24], [31], [37]	90 subjects	2935 images (1024x768) pixels	✓		✓		✓
Terravic facial IR database[4] [14], [47], [93], [94]	20 subjects (19 men, 1 woman)	22784 image (320x240) pixels	✓		✓		
RGB-D-T based face recognition[2] [39], [40], [95]	51 subjects	45900 images Thermal images (384x288) pixels Visible images (640x480) pixels	✓	✓	✓	✓	✓
L-CAS thermal physiological monitoring dataset[5] [81], [82], [96]	5 subjects	3000 image (382x288) pixels	✓		✓		
SpeakingFaces[6] [87], [88]	142 subjects (74 males, 68 females)	4581595 image Thermal images (464x348) pixels Visible images (768x512) pixels	✓	✓	✓		
TFW dataset [87]	147 subjects	9982 images (464x348) pixels	✓		✓	✓	

[6]https://github.com/IS2AI/SpeakingFaces

USTC-NVIE, KTFE database, Terravic facial IR database, SpeakingFaces and TFW dataset as shown in Fig. 5. For the Equinox HID face database and USTC-NVIE, those wearing glasses are requested to repeat the experiment twice, with and without glasses to make it more practical. As a result, it was effective for facial tracking in liveness applications under unrestricted conditions or environmental settings. However, USTC-NVIE, UND-X1, RGB-D-T based face recognition and IRDatabase are among publicly available but with restricted access. It is difficult to find available datasets that may include various signs estimations for physiological monitoring. The L-CAS thermal physiological monitoring dataset is recommended because it includes the ground truth for RR and HR estimations derived from thermal images. The appropriate approaches are also essential because the varying size of datasets includes may affect the computational complexity.

C. VARYING VITAL SIGNS ESTIMATIONS

Most of the clinical applications that make use of thermal images have included the measurement of temperature variations. Motivated by past studies, it is possible that multi-parameter vital signs estimations can be measured from thermal images using additional biosensors signal processing methods to improve the preciseness of the results. Although most of these biosensor methods have been effectively deployed using RGB imaging, the current studies have demonstrated that these biosensors may also work very well in thermal video cameras since they are more visible to humans visual. It is beneficially helpful in vital signs estimation because of limitations in human visual perceptions to see facial colour changes reflected by changes in oxygenation levels and blood volume caused by the heartbeats. In these cases, varying thermal colour palettes can be used to emphasize facial colour changes, particularly rainbow colour

palettes for monitoring the smallest change of heat energy caused by blood passing through facial arteries from the jaw to the forehead in thermal images [27], [44]. Interestingly, because the forehead and cheeks have the highest blood flow, it makes sense that most prior studies include them in their comparative analysis, along with other face regions such as the nose and mouth areas where airflow continues in and out [35], [97]. A countless number of studies have also mentioned the eye and its inner canthus of the eye region as having the highest correlation with internal core body temperature in their experiments [98], [99].

D. DEEP LEARNING METHODS

Over the last few years, extensive studies have revealed the trending topics about the implementation of DL methods either to employ a hybrid or stacking methods in their research instead of ML algorithms in a variety of image classification tasks to cope with the information loss and accuracy begins to saturate and eventually degrades. These issues emerge because of the need to evaluate a wide range of image features from a large number of input images at the same time to extract relevant information for further analysis. Prior research has mainly used the most recent CNN networks with varying deeper of the network to produce more effective multiclass facial detection either globally or locally that were analyzed from a high or low resolution of thermal imaging. Other studies also have investigated the uses of transfer learning in DL methods to identify whether there are any new existing tasks but similar problems that need to be addressed. As discussed above, it is acceptable to conclude that many DL methods have been successfully utilized due to lowering the rate of image misclassification, computational complexity and processing time.

VI. FUTURE WORKS

For future works, it can recommend contactless or noninvasive monitoring instruments which are widely used in a variety of applications for continuously monitoring an individual's health as well as their physiological affective states. Furthermore, the findings of this review suggest that it is worth considering various vital signs estimations and a range of symptoms in validating the results to allow many broader applications to be introduced from the use of facial thermography. These estimations which are based on machine learning or DL methods as well as an additional signal processing that demonstrates highly good performance, can be deployed in embedded devices or mobile devices that are running in real time and can assist users in monitoring their health status on their own with the assistance of an intelligent personal assistant. Because of the limitation with available infrared datasets that are focusing on emotional state classifications, it may be better in the future if there are existing datasets that can allow more contributions whereas covering the healthy applications that include disease diagnosing from rare features that are easily observable biometric images using facial areas, rather than other parts of the body that have already been seriously injured, amputation or blindness (eyes). Furthermore, this may be particularly important to highlight in efforts to improve the feature extraction with the purpose of utilizing computational algorithms that can enhance the representation of facial images that include subjects who are wearing eyeglasses or protective face masks in the real-time analysis. Importantly, these reviews can provide a possible reason for future research to address the main challenge in predicting human abnormal conditions using thermal imaging, which can be potentially applied in very similar ways to the uses of RGB images.

VII. CONCLUSION

Contactless or noninvasive vital signs estimation from FLIR video recordings of facial thermography has been introduced and presented in this paper, along with the discovery of feature extraction methods and the selection of vital sign estimations with the use of computational algorithms and biosensors signal processing methods, which was then validated using the selection evaluation matrices. It is not surprising that when compared to natural colour images that are less sensitive to temperature changes, facial redness in thermal imaging is more recognizable to the human visual, which is not only limited due to physiological features changes, excessive activities, environmental conditions and many other factors, but it can also be an early sign of abnormal health issues that require intentions. Measuring only one vital sign is ineffective for evaluating an individual's health in both adults and children unless with other important vital sign estimations that might potentially be extended and monitored in medical examination. As a result of these studies focusing on assessing vital signs that can undoubtedly be monitored using facial thermography, countless feature extractor methods have been established to mitigate the challenges caused by facial misalignment resulting from inaccurate facial localizations. However, whether global or local features are performed depends on the experiment that has been proposed, as numerous research studies may consider using local features instead of global features to perform a comparative analysis on selecting the specific ROIs. Most importantly, it has shown that ML and DL computational methods are increasingly being employed to improve the performance of the feature extractor methods, which can be applied to infrared images to enhance facial feature representations. Overall, the findings of this study can provide guidelines for scholars to encourage and recommend the use of infrared images in evaluating vital signs in a wide range of clinical applications.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/TK0/UTHM/02/12) and Universiti Tun Hussein Onn Malaysia (UTHM) through Tier1 (vot H756).

REFERENCES

[1] M. Elliott and A. Coventry, "Critical care: The eight vital signs of patient monitoring," *Brit. J. Nursing*, vol. 21, no. 10, pp. 621–625, May 2012.

[2] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless measurement of vital signs using thermal and RGB cameras: A study of COVID 19-related health monitoring," *Sensors*, vol. 22, no. 2, pp. 1–17, 2022.

[3] N. Hochhausen, C. B. Pereira, S. Leonhardt, R. Rossaint, and M. Czaplik, "Estimating respiratory rate in post-anesthesia care unit patients using infrared thermography: An observational study," *Sensors*, vol. 18, no. 5, pp. 1–12, 2018.

[4] D. Li, C. C. Menassa, and V. R. Kamat, "Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography," *Energy Buildings*, vol. 176, pp. 246–261, Oct. 2018.

[5] D. Perpetuini, A. Di Credico, C. Filippini, P. Izzicupo, D. Cardone, P. Chiacchiaretta, B. Ghinassi, A. Di Baldassarre, and A. Merla, "Is it possible to estimate average heart rate from facial thermal imaging?" *Eng. Proc.*, vol. 8, no. 10, pp. 1–4, 2021.

[6] M. H. Abd Latif, H. Md. Yusof, S. N. Sidek, and N. Rusli, "Implementation of GLCM features in thermal imaging for human affective state detection," *Proc. Comput. Sci.*, vol. 76, pp. 308–315, Jan. 2015.

[7] J. G. Chester and J. L. Rudolph, "Vital signs in older patients: Age-related changes," *J. Amer. Med. Directors Assoc.*, vol. 12, no. 5, pp. 337–343, Jun. 2011.

[8] Z. Wang, Z. Yang, and T. Dong, "A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time," *Sensors*, vol. 17, no. 2, pp. 1–36, 2017.

[9] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard, and D. Merhof, "A thermal infrared face database with facial landmarks and emotion labels," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1389–1401, May 2018.

[10] N. Jasti, S. Bista, H. Bhargav, S. Sinha, S. Gupta, S. K. Chaturvedi, and B. N. Gangadhar, "Medical applications of infrared thermography: A narrative review," *J. Stem Cells*, vol. 14, no. 1, pp. 35–53, 2019.

[11] A. A. Sarawade and N. N. Charniya, "Infrared thermography and its applications: A review," in *Proc. 3rd Int. Conf. Commun. Electron. Syst. (ICCES)*, Oct. 2018, pp. 280–285.

[12] G. J. Tattersall, "Infrared thermography: A non-invasive window into thermal physiology," *Comparative Biochemistry Physiol. A, Mol. Integrative Physiol.*, vol. 202, pp. 78–98, Dec. 2016.

[13] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, "Infrared machine vision and infrared thermography with deep learning: A review," *Infr. Phys. Technol.*, vol. 116, pp. 1–19, Aug. 2021.

[14] D. Bhattacharjee, A. Seal, S. Ganguly, M. Nasipuri, and D. K. Basu, "A comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern," *Comput. Intell. Neurosci.*, vol. 2012, pp. 1–12, Jan. 2012.

[15] J. Ruiz-Del-Solar, R. Verschae, G. Hermosilla, and M. Correa, "Thermal face recognition in unconstrained environments using histograms of LBP features," in *Local Binary Patterns: New Variants and Applications*, vol. 506. Berlin, Germany: Springer, 2014, pp. 219–243.

[16] G. Machin, D. Brettle, S. Fleming, R. Nutbrown, R. Simpson, R. Stevens, and M. Tooley, "Is current body temperature measurement practice fit-for-purpose?" *J. Med. Eng. Technol.*, vol. 45, no. 2, pp. 136–144, Feb. 2021.

[17] R. S. Ghiass, O. Arandjelovic, H. Bendada, and X. Maldague, "Illumination-invariant face recognition from a single image across extreme pose using a dual dimension AAM ensemble in the thermal infrared spectrum," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 2781–2790.

[18] N. Glowacka and J. Rumiński, "Face with mask detection in thermal images using deep neural networks," *Sensors*, vol. 21, no. 19, pp. 1–16, 2021.

[19] Y. Zheng, "A novel thermal face recognition approach using face pattern words," *Biometric Technol. Hum. Identif. VII*, vol. 7667, pp. 21–32, Apr. 2010.

[20] 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. (I2MTC)*, May 2018, pp. 1–6.

[21] F. Liu, P. Han, Y. Wang, X. Li, L. Bai, and X. Shao, "Super resolution reconstruction of infrared images based on classified dictionary learning," *Infr. Phys. Technol.*, vol. 90, pp. 146–155, May 2018.

[22] C. Kant and N. Sharma, "Fake face recognition using fusion of thermal imaging and skin elasticity," *Int. J. Comput. Sci. Commun.*, vol. 4, no. 1, pp. 65–72, Mar. 2013.

[23] A. Seal, D. Bhattacharjee, and M. Nasipuri, "Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images," *AEU Int. J. Electron. Commun.*, vol. 70, no. 8, pp. 1041–1049, Aug. 2016.

[24] A. Bhattacharyya, S. Chatterjee, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar, "A deep learning model for classifying human facial expressions from infrared thermal images," *Sci. Rep.*, vol. 11, no. 1, pp. 1–17, Dec. 2021.

[25] G. Hermosilla, J. Ruiz-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.

[26] R. L. Olalia, Jr., J. A. Olalia, and M. G. F. Carse, "Evaluating infrared thermal image's color palettes in hot tropical area," *J. Comput. Commun.*, vol. 9, no. 11, pp. 37–49, 2021.

[27] Z.-H. Wang, G.-J. Horng, T.-H. Hsu, C.-C. Chen, and G.-J. Jong, "A novel facial thermal feature extraction method for non-contact healthcare system," *IEEE Access*, vol. 8, pp. 86545–86553, 2020.

[28] A. Basu, A. Routray, S. Shit, and A. K. Deb, "Human emotion recognition from facial thermal image based on fused statistical feature and multi-class SVM," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–5.

[29] M. H. Latif, H. Md Yusof, S. N. Sidek, N. Rusli, and S. Fatai, "Emotion detection from thermal facial imprint based on GLCM features," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 1, pp. 345–350, 2016.

[30] F. Gioia, A. Greco, A. L. Callara, and E. P. Scilingo, "Towards a contactless stress classification using thermal imaging," *Sensors*, vol. 22, no. 3, pp. 1–16, 2022.

[31] 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, pp. 1–13, 2019.

[32] A. Kwasniewska and J. Ruminski, "Real-time facial feature tracking in poor quality thermal imagery," in *Proc. 9th Int. Conf. Human Syst. Interact. (HSI)*, Jul. 2016, pp. 504–510.

[33] P. Jagadev and L. I. Giri, "Non-contact monitoring of human respiration using infrared thermography and machine learning," *Infr. Phys. Technol.*, vol. 104, pp. 1–13, Jan. 2020.

[34] A. Basu, A. Routray, R. Mukherjee, and S. Shit, "Infrared imaging based hyperventilation monitoring through respiration rate estimation," *Infr. Phys. Technol.*, vol. 77, pp. 382–390, Jul. 2016.

[35] R. Chauvin, M. Hamel, S. Brière, F. Ferland, F. Grondin, and D. Létourneau, "Contact-free respiration rate monitoring using a pan–tilt thermal camera for stationary bike telerehabilitation sessions," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1046–1055, Jul. 2016.

[36] J. Heo, M. Savvides, and B. V. K. Vijayakumar, "Advanced correlation filters for face recognition using low-resolution visual and thermal imagery," in *Proc. Int. Conf. Image Anal. Recognit.*, 2005, pp. 1089–1097.

[37] *IRDatabase*. Accessed: Apr. 28, 2022. [Online]. Available: https://github.com/marcinkopaczka/thermalfaceproject

[38] 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, pp. 1–24, 2019.

[39] 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.

[40] RGB-D-T based Face Recognition. *Visual Analysis & Perception Lab*. Accessed: Jun. 5, 2022. [Online]. Available: https://vap.aau.dk/rgb-d-t-based-face-recognition/

[41] 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.*, 2014, pp. 397–408.

[42] M. H. Abd Latif, H. Md. Yusof, S. N. Sidek, and N. Rusli, "Thermal imaging based affective state recognition," in *Proc. IEEE Int. Symp. Robot. Intell. Sensors (IRIS)*, Oct. 2015, pp. 214–219.

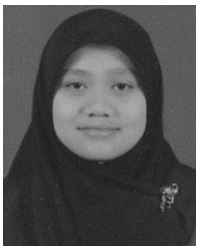
[43] Z. Xie, Z. Wang, and G. Liu, "Adaptive quantization of local directional responses for infrared face recognition," in *Proc. Int. Conf. Intell. Comput.*, 2015, pp. 199–207.

[44] U. Thirunavukkarasu, S. Umapathy, K. Janardhanan, and R. Thirunavukkarasu, "A computer aided diagnostic method for the evaluation of type II diabetes mellitus in facial thermograms," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 871–888, Sep. 2020.

[45] G. Majumder and M. K. Bhowmik, "Gabor-fast ICA feature extraction for thermal face recognition using linear kernel support vector machine," in *Proc. Int. Conf. Comput. Intell. Netw.*, Jan. 2015, pp. 21–25.

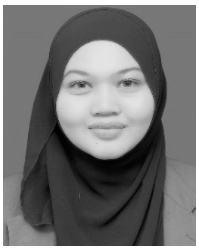
[46] (2012). *IRIS Thermal/Visible Face Database*. IEEE OTCBVS WS Series Bench. Accessed: Jun. 1, 2022. [Online]. Available: http://vciplokstate.org/pbvs/bench/Data/02/download.html
[47] R. Miezianko. (2006). *Terravic Facial Infrared Database*. Accessed: Jun. 1, 2022. [Online]. Available: http://vciplokstate.org/pbvs/bench/Data/04/download.html
[48] T. Elguebaly and N. Bouguila, "A Bayesian method for infrared face recognition," in *Machine Vision Beyond Visible Spectrum*, vol. 1. Berlin, Germany: Springer, 2011, pp. 123–138.
[49] M. F. Hashmi, B. K. K. Ashish, V. Sharma, A. G. Keskar, N. D. Bokde, J. H. Yoon, and Z. W. Geem, "LARNet: Real-time detection of facial micro expression using lossless attention residual network," *Sensors*, vol. 21, no. 4, pp. 1–23, 2021.
[50] T. Shu, B. Zhang, and Y. Y. Tang, "An extensive analysis of various texture feature extractors to detect diabetes mellitus using facial specific regions," *Comput. Biol. Med.*, vol. 83, pp. 69–83, Apr. 2017.
[51] D. Yang, V.-T. Peltoketo, and J.-K. Kamarainen, "CNN-based cross-dataset no-reference image quality assessment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3913–3921.
[52] M. Benco, R. Hudec, P. Kamencay, M. Zachariasova, and S. Matuska, "An advanced approach to extraction of colour texture features based on GLCM," *Int. J. Adv. Robot. Syst.*, vol. 11, no. 7, pp. 1–8, 2014.
[53] J. Rumiński, "Analysis of the parameters of respiration patterns extracted from thermal image sequences," *Biocybernetics Biomed. Eng.*, vol. 36, no. 4, pp. 731–741, 2016.
[54] S. Huynh, R. K. Balan, J. Ko, and Y. Lee, "VitaMon: Measuring heart rate variability using smartphone front camera," in *Proc. 17th Conf. Embedded Networked Sensor Syst.*, Nov. 2019, pp. 1–14.
[55] A. Al-Naji, K. Gibson, S. H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15776–15790, 2017.
[56] A.-S. Sejling, K. H. W. Lange, C. S. Frandsen, S. S. Diemar, L. Tarnow, J. Faber, J. J. Holst, B. Hartmann, L. Hilsted, T. W. Kjaer, C. B. Juhl, B. Thorsteinsson, and U. Pedersen-Bjergaard, "Infrared thermographic assessment of changes in skin temperature during hypoglycaemia in patients with type 1 diabetes," *Diabetologia*, vol. 58, no. 8, pp. 1898–1906, Aug. 2015.
[57] Y. Yao, G. Sun, T. Matsui, Y. Hakozaki, S. van Waasen, and M. Schiek, "Multiple vital-sign-based infection screening outperforms thermography independent of the classification algorithm," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 1025–1033, May 2016.
[58] G. Sun, T. Matsui, Y. Hakozaki, and S. Abe, "An infectious disease/fever screening radar system which stratifies higher-risk patients within ten seconds using a neural network and the fuzzy grouping method," *J. Infection*, vol. 70, no. 3, pp. 230–236, Mar. 2015.
[59] C. Filippini, A. Di Crosta, R. Palumbo, D. Perpetuini, D. Cardone, I. Ceccato, A. Di Domenico, and A. Merla, "Automated affective computing based on bio-signals analysis and deep learning approach," *Sensors*, vol. 22, no. 5, pp. 1–19, 2022.
[60] G. Sun, T. Negishi, T. Kirimoto, T. Matsui, and S. Abe, "Noncontact monitoring of vital signs with RGB and infrared camera and its application to screening of potential infection," in *Non-Invasive Diagnostic Methods*. IntechOpen, 2018, pp. 43–51.
[61] A. Kwasniewska, J. Ruminski, and M. Szankin, "Improving accuracy of contactless respiratory rate estimation by enhancing thermal sequences with deep neural networks," *Appl. Sci.*, vol. 9, no. 20, pp. 1–17, 2019.
[62] A. Procházka, H. Charvátová, O. Vyáta, J. Kopal, and J. Chambers, "Breathing analysis using thermal and depth imaging camera video records," *Sensors*, vol. 17, no. 6, pp. 1–10, 2017.
[63] K. S. Nair and S. Sarath, "Illumination invariant non-invasive heart rate and blood pressure estimation from facial thermal images using deep learning," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2021, pp. 1–7.
[64] S. Sivanandam, M. Anburajan, B. Venkatraman, M. Menaka, and D. Sharath, "Medical thermography: A diagnostic approach for type 2 diabetes based on non-contact infrared thermal imaging," *Endocrine*, vol. 42, no. 2, pp. 343–351, Oct. 2012.
[65] A. Scarano, F. Inchingolo, and F. Lorusso, "Facial skin temperature and discomfort when wearing protective face masks: Thermal infrared imaging evaluation and hands moving the mask," *Int. J. Environ. Res. Public Heal.*, vol. 17, no. 13, pp. 1–9, 2020.
[66] K. Hamedani, Z. Bahmani, and A. Mohammadian, "Spatio-temporal filtering of thermal video sequences for heart rate estimation," *Exp. Syst. Appl.*, vol. 54, pp. 88–94, Jul. 2016.
[67] T. R. Gault and A. A. Farag, "A fully automatic method to extract the heart rate from thermal video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 336–341.
[68] 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.
[69] A. Masaki, K. Nagumo, B. Lamsal, K. Oiwa, and A. Nozawa, "Anomaly detection in facial skin temperature using variational autoencoder," *Artif. Life Robot.*, vol. 26, no. 1, pp. 122–128, Feb. 2021.
[70] A. Masaki, K. Nagumo, Y. Iwashita, K. Oiwa, and A. Nozawa, "An attempt to construct the individual model of daily facial skin temperature using variational autoencoder," *Artif. Life Robot.*, vol. 26, no. 4, pp. 488–493, Nov. 2021.
[71] K. Ganesh, S. Umapathy, and P. T. Krishnan, "Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging," *Proc. Inst. Mech. Engineers, H, J. Eng. Med.*, vol. 235, no. 10, pp. 1113–1127, Oct. 2021.
[72] M. Singh and A. S. Arora, "Computer aided face liveness detection with facial thermography," *Wireless Pers. Commun.*, vol. 111, no. 4, pp. 2465–2476, Apr. 2020.
[73] 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.
[74] G. Pons, A. El Ali, and P. Cesar, "ET-CycleGAN: Generating thermal images from images in the visible spectrum for facial emotion recognition," in *Proc. Companion Publication Int. Conf. Multimodal Interact.*, Oct. 2020, pp. 87–91.
[75] 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, Jul. 2010.
[76] 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.
[77] S. Petridis, B. Martinez, and M. Pantic, "The MAHNOB laughter database," *Image Vis. Comput.*, vol. 31, no. 2, pp. 186–202, Feb. 2013.
[78] J. Stemberger, R. S. Allison, and T. Schnell, "Thermal imaging as a way to classify cognitive workload," in *Proc. Can. Conf. Comput. Robot Vis.*, 2010, pp. 231–238.
[79] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L. W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, 2011.
[80] S. Lyra, L. Mayer, L. Ou, D. Chen, P. Timms, A. Tay, P. Y. Chan, B. Ganse, S. Leonhardt, and C. H. Antink, "A deep learning-based camera approach for vital sign monitoring using thermography images for ICU patients," *Sensors*, vol. 21, no. 4, pp. 1–18, 2021.
[81] S. Cosar, Z. Yan, F. Zhao, T. Lambrou, S. Yue, and N. Bellotto, "Thermal camera based physiological monitoring with an assistive robot," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 5010–5013.
[82] (2018). *Lincoln Centre for Autonomous Systems L-CAS Thermal Physiological Monitoring Dataset*. The University of Lincoln's Cross-Disciplinary Centre in Robotics Research. Accessed: Apr. 25, 2022. [Online]. Available: https://lcas.lincoln.ac.uk/wp/research/data-sets-software/lcas-thermal-physiological-monitoring-dataset/
[83] Y. Yu, P. Raveendran, and C. Lim, "Dynamic heart rate measurements from video sequences," *Biomed. Optrics Exp.*, vol. 6, no. 7, pp. 2466–2480, 2015.
[84] A. Otufowora, Y. Liu, H. Young, K. L. Egan, D. S. Varma, C. W. Striley, and L. B. Cottler, "Sex differences in willingness to participate in research based on study risk level among a community sample of African Americans in north central Florida," *J. Immigrant Minority Health*, vol. 23, no. 1, pp. 19–25, Feb. 2021.
[85] M. Indraganti and M. A. Humphreys, "A comparative study of gender differences in thermal comfort and environmental satisfaction in air-conditioned offices in Qatar, India, and Japan," *Build. Environ.*, vol. 206, pp. 1–15, Dec. 2021.
[86] V. Cheng, E. Ng, C. Chan, and B. Givoni, "Outdoor thermal comfort study in a sub-tropical climate: A longitudinal study based in Hong Kong," *Int. J. Biometeorol.*, vol. 56, no. 1, pp. 43–56, Jan. 2012.

[87] A. Kuzdeuov, D. Aubakirova, D. Koishigarina, and H. A. Varol, "TFW: Annotated thermal faces in the wild dataset," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1–11, 2022.
[88] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis, and H. A. Varol, "SpeakingFaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams," *Sensors*, vol. 21, no. 10, pp. 1–20, 2021.
[89] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, 2006.
[90] 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) Workshops*, Sep. 2005, p. 14.
[91] M. S. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 426–438, 2017.
[92] X. Chen, P. J. Flynn, and K. W. Bowyer, "Visible-light and infrared face recognition," in *Proc. Workshop Multimodal User Authentication*, 2003, pp. 48–55.
[93] A. Ibrahim, T. Gaber, T. Horiuchi, V. Snasel, and A. E. Hassanien, "Human thermal face extraction based on superpixel technique," in *Proc. 1st Int. Conf. Adv. Intell. Syst. Inform.*, vol. 407, 2016, pp. 163–172.
[94] A. Ibrahim, A. Tharwat, T. Gaber, and A. E. Hassanien, "Optimized superpixel and AdaBoost classifier for human thermal face recognition," *Signal, Image Video Process.*, vol. 12, no. 4, pp. 711–719, May 2018.
[95] O. Nikisins, K. Nasrollahi, M. Greitans, and T. B. Moeslund, "RGB-D-T based face recognition," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1716–1721.
[96] F. Condrea, V.-A. Ivan, and M. Leordeanu, "In search of life: Learning from synthetic data to detect vital signs in videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1207–1216.
[97] K. Alghoul, S. Alharthi, A. El Saddik, and H. Al Osman, "Heart rate variability extraction from videos signals: ICA vs. EVM comparison," *IEEE Access*, vol. 5, pp. 4711–4719, 2017.
[98] Z. M. Lazri, Q. Zhu, M. Chen, M. Wu, and Q. Wang, "Detecting essential landmarks directly in thermal images for remote body temperature and respiratory rate measurement with a two-phase system," *IEEE Access*, vol. 10, pp. 39080–39094, 2022.
[99] D. Haputhanthri, G. Brihadiswaran, S. Gunathilaka, D. Meedeniya, S. Jayarathna, M. Jaime, and C. Harshaw, "Integration of facial thermography in EEG-based classification of ASD," *Int. J. Autom. Comput.*, vol. 17, no. 6, pp. 837–854, 2020.



SYAIDATUS SYAHIRA AHMAD TARMIZI received the B.S. degree in computer science and the M.S. degree in information systems (intelligent systems) from Universiti Teknologi Mara, in 2018 and 2019, respectively. She is currently pursuing the Ph.D. degree in electrical engineering with Universiti Tun Hussein Onn Malaysia. Her research interests include the expert systems, data mining applications, and image processing.

NOR SURAYAHANI SURIANI received the B.S. degree in computer and communication engineering from Universiti Putra Malaysia, in 2003, the M.S. degree in electrical and electronics engineering from Universiti Teknologi Malaysia, in 2006, and the Ph.D. degree in electrical, electronic, and system engineering from Universiti Kebangsaan Malaysia, in 2015. She is currently working as a Senior Lecturer with the Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, UTHM, Malaysia.



FADILLA 'ATYKA NOR RASHID was born in Batu Pahat, Johor, Malaysia, in May 1991. She received the bachelor's degree in computer science from Universiti Teknologi Malaysia, in 2014, and the master's degree in telecommunication and information engineering from Universiti Teknologi MARA, in 2016. She is currently pursuing the Ph.D. degree in electrical engineering with Universiti Tun Hussein Onn Malaysia. She is currently a Lecturer with Universiti Malaysia Sarawak (UNIMAS). Her research interests include computer vision and machine learning.