A Hierarchical Separation and Classification Network for Dynamic Microexpression Classification

Jordan Vice, Masood Mehmood Khan, *Member, IEEE*, Tele Tan, Iain Murray, *Senior Member, IEEE* and Svetlana Yanushkevich, *Senior Member*, *IEEE*

***Abstract*—Macrolevel facial muscle variations, as used for building models of seven discrete facial expressions, suffice when distinguishing between macrolevel human affective states but won't discretise continuous and dynamic microlevel variations in facial expressions. We present a hierarchical separation and classification network (HSCN) for discovering dynamic, continuous, and macro- and microlevel variations in facial expressions of affective states. In the HSCN, we first invoke an unsupervised cosine similarity-based separation method on continuous facial expression data to extract twenty-one dynamic facial expression classes from the seven common discrete affective states. The between-clusters separation is then optimized for discovering the macrolevel changes resulting from facial muscle activations. A following step in the HSCN separates the upper and lower facial regions for realizing changes pertaining to upper and lower facial muscle activations. Data from the two separated facial regions are then clustered in a linear discriminant space using similarities in muscular activation patterns. Next, the actual dynamic expression data are mapped onto discriminant features for developing a rule-based expert system that facilitates classifying twenty-one upper and twenty-one lower microexpressions. Invoking the random forest algorithm would classify twenty-one macrolevel facial expressions with 76.11% accuracy. A support vector machine (SVM), used separately on upper and lower facial regions in tandem, could classify them with respective accuracies of 73.63% and 87.68%. This work demonstrates a novel and effective method of dynamic assessment of affective states. The HSCN further demonstrates that facial muscle variations gathered from either upper, lower, or full-face would suffice classifying affective states. We also provide new insight into discovery of microlevel facial muscle variations and their utilization in dynamic assessment of facial expressions of affective states.**

***Index Terms*—Affective state assessment, cosine similarity-based separation, facial expression classification, hierarchical classification, microexpression detection, rule-based systems.**

Manuscript received 24 January 2023; revised 3 July 2023 and 21 August 2023; accepted 11 October 2023. Date of publication 16 January 2024; date of current version 31 May 2024. This work was supported by the Faculty of Science and Engineering, Curtin University Australia. The work of Masood Mehmood Khan and Jordan Vice was supported by the Curtin University under Grant APA18336230. The work of Svetlana Yanushkevich was supported in part by the Natural Sciences and Engineering Research Council of Canada under the Discovery Grant "Biometric intelligent interfaces." *(Corresponding author: Masood Mehmood Khan.)*

Jordan Vice is with the University of Western Australia, Perth, WA 6009 Australia.

Masood Mehmood Khan, Tele Tan, and Iain Murray are with Curtin University, Perth, WA 6102 Australia (e-mail: Masood.Khan@curtin.edu.au).

Svetlana Yanushkevich is with the University of Calgary, Calgary, AL T2N 1N4 Canada.

Digital Object Identifier 10.1109/TCSS.2023.3334823

I. Introduction

FACIAL expressions convey internal thoughts, feelings, and emotions and provide interpretable external signals that convey variations in affective states [1]. Variations in facial expressions result from conscious and subconscious processing of several internal and external stimuli, including any contextual biases and the interactions between past and present experiences [2]. Patterns of facial muscle movements provide reliable models for automated recognition of human affective states [3], [4], [5]. Facial muscle movement models have been used for realizing complex and difficult to identify affective states and have been tested using large datasets [5], [6], [7], [8].

Humans realize the dynamic and continuous affective states through assessment of facial expressions using intuitive knowledge and collective experiences [3]. Given the complexities of emotion elicitation and the neuro- and pathopsychological factors behind them, emotions cannot be regarded as static occurrences in time [5], [9]. Changes in facial expressions are regarded as continuous and time-dependent functions. Therefore, we posit that microlevel facial expressions should also be classifiable within a continuous prevailing space. This appears logical as microlevel facial expressions represent transient macrolevel expressions such as perceived expressions of anger, joy, and fear [10].

Affective computing literature cites a diverse range of related works on facial expression recognition and affective state assessment [11], [12], [13]. A large majority of the cited facial expression classifiers use Ekman's discrete models of affective states [14]. Ekman's seven distinct and discrete models of affective are based on significant macrolevel differences in facial muscle movement patterns [15], [16], [17]. However, these seven discrete models cannot represent microlevel facial expressions.

Microlevel variations in facial muscles can be seen as representing delicate, involuntary, and spontaneous changes in facial expressions and they often show true emotive experiences. A review of important microfacial expression recognition approaches [18] suggests that due to their transient nature and low intensity, microlevel expressions are not captured easily in real life situations [7], [18], [20], [21]. Because of difficulties in capturing micro expressions, several interesting approaches

have been proposed for their realization. For example, an accretive layer was added to a hybrid network in [19] for refining the features of facial expressions. Microfacial expressions were recognized in [22] using the accordion spatiotemporal data classified by the random forest (RF) algorithm. Similarly, in [7], an algorithm ensemble that exploited the handcrafted features and deep features was used for identifying faces and classifying microfacial expressions.

In order to assess human facial expressions, analysts would usually rely on discriminative spatiotemporal features. In [22], a hybrid deep learning approach was proposed. The method first uses a spatial convolutional neural network (CNN) for processing static facial images and then uses a temporal CNN for processing the optical flow in images to separate yet simultaneous learning of significant spatial and temporal features. These significant spatial and temporal features are taken to a deep belief network (DBN) model where the average pooling is performed to obtain the fixed-length global feature representation. Finally, a support vector machine (SVM) performs the classification [22]. The complexity and computational costs required for continuous dynamic assessment of affective states in [22] were also observed in several other works [23], [24], [25]. Despite the recent advances, the idea of extracting microexpressions from the discrete models of affective states has not been explored much.

For avoiding data acquisition problems and algorithmic complexities involved in building dynamic and occluded facial expression classifiers, this work proposes a novel method of classifying micro, macro, continuous, and dynamic-facial expressions. We use a rule-based expert system to exploit facial muscle movements for classifying micro- and macrolevel facial expressions. The hierarchical separation and classification network (HSCN) used in this work consists of three subsystems, viz., 1) a module for macrolevel affective state assessment using whole face data; 2) a module for upper facial region microexpression classification; and 3) a module for lower facial region microexpression classification. We used an unsupervised, cosine similarity-based separation method for exploiting mutual information in continuous facial expression data, facilitating the discovery of boundaries and regions within a multidimensional hyperplane. A following linear discriminant analysis (LDA) subsystem further separates and clusters facial expressions by identifying discriminant features and realizing multiple hyperplanes within the discriminant space [2], [26]. Compared with other statistical and neural multilabel data classification techniques, LDA is considered computationally efficient as it would reduce the dimensionality and requires lesser amount of training data [13], [26], [27].

Following the LDA transform, the HSCN uses a novel rule-based expert system for upper and lower facial region microexpression classification. The expert system uses data pertaining to the continuous muscle movements and the logic manifested in facial action coding system (FACS) [28]. Though in a different context, expert systems have previously been used for facial expression analyses. For example, in [29] a self-adaptive expert system used the facial feature contours localized in a static dual-view facial image to label the interpreted facial expressions. Authors [30] deployed a belief rule-based expert system that exploited the outputs of a CNN classifier for inferring the mental state of a person using the observed facial expressions. In both cases, rule-based expert systems were used to augment and improve the classifier performance. Building upon previous works, we provide another example showing how rule-based systems can be deployed for continuous and dynamic assessment of affective states.

A. Contributions

Building upon previous works, this article contributes the following.

1) We present a novel framework and an ensemble of classifiers enabling hierarchical separation and classification of macro- and microlevel expressions of affective states. We exploit well-tested generic algorithms in the context of a dynamic affective state assessment environment.
2) This work demonstrates use of a novel affective state assessment schema and introduces methods of capturing and monitoring continuous facial expressions.
3) We introduce a unified and systematic approach of separating the upper and lower facial regions in a dynamic environment and separately classifying them in tandem for continuous and dynamic classification of facial expressions.
4) We propose a novel methodology of developing and applying a rule-base for classifying continuous macrolevel expressions.
5) We detail implementation of a rule-based expert system that uses the dynamic linear discriminant features for realizing continuous microexpressions on upper and lower facial regions.
6) This work presents a systematic way of extracting microlevel facial expressions of affective states.

In relation to the first contribution stated above—please note that ensemble approaches are becoming popular in complex feature classification tasks and detection of implicit patterns [31], [32]. Such ensemble approaches allow for hierarchical processing of information signals and enable stepwise refinement of separable features. This work extends application of ensemble approaches by using an ensemble for microexpression classification.

This article is organized such that Section I introduces the work and Section II presents the relevant background information. Section III discusses related works and current trends in the literature. Section IV details methods used for unsupervised clustering and labeling, cosine similarity estimation, data separation, macro- and microlevel LDA, and construction of the rule-based expert system. Section V reports results pertaining to the aforementioned analyses and system validation. Section V also describes the rule-based expert system. Section VI concludes this work and provides directions for future research. Section VII informs on the funding source and ethics compliance.

II. Background Information

"Core affect" is a psychological construct that signifies the continuous nature of emotions and affective states [9]. The concept of "core affect" helps in interpreting complexities of human affective states and emotions by providing theoretical foundations for building a dynamic affective state assessment solution. The fluidity and multidimensional nature of expressions of affective states described in the core affect model highlights the need for a dynamic classifier capable of accounting for complex, multidimensional expressions.

Several models that represent "unique emotions" have been introduced in the literature. For example, the authors [4], [33] attempted to model disgust and anger. Such single-expression models provide well-defined ways of comprehending and classifying human expressions of affective states. One such model, the Hourglass model [34] followed by its enhanced version [35] highlights the continuous and nonstatic nature of human sentiment and their assessment. The hourglass of emotions, an emotion categorization model, was optimized for polarity detection. It was built using empirical data pertaining to sentiment analysis. Nonetheless, it could be used in the context of affective state classification. The hourglass model categorizes similar and dissimilar emotions and presents a dynamic model that appears more representative of human emotions compared with the discrete emotion models. Other continuous emotion models include the Plutchik spectrum and the three-factor model [36], [37]. Such continuous emotion models can delineate emotions and their macrolevel expressions [10]. In a somewhat similar manner, macrolevel expressions have also been assessed along multidimensional axes of valence, arousal, and intensity [38].

Facial expression analysis and affective state classification are complex problems. Thus, many of the available solutions underperform in real life situations. Changes in expressions of affective states are causal, representing some response to a particular temporal event or a combination of multiple external and internal stimulating factors [5], [9]. Responding to certain stimuli and experiencing particular affective states would cause internal pathological and physiological changes in humans. Hence, fluctuations in cues like heart rate, skin conductance, and hormone balances have been used for affective state classification. Variations in affective states are also reflected through external cues like speech rate and/or volume and, hemodynamic changes on the face and facial expressions [39]. Recent affective computing and psychophysiology literature highlights limitations of discrete facial expression models and affective state assessment solutions.

The FACS and emotion facial action coding system (EMFACS) were discussed in [28] and their application details were presented in [16], [40], [41], [42]. These works categorize facial muscles movements through coding and action units and work as tools for the facial expression assessment and recognition. The EMFACS action units allow modeling feature fluctuations in time, as one's expressions change from one state to another. The rule-based expert system deployed in this work for upper and lower facial microexpression classification is based on the muscle movements defined in the EMFACS.

TABLE I
List of Affective States Present in the CK+ Dataset and Their Corresponding Facial Muscle Actions Identified in FACS and EMFACS Studies

State	Action Units	Physical Actions	Muscles
Happy	12 6, 7 26, 27	Lip corners raised Raised lower eyelids Open mouth	Zygomaticus major Orbicularis Oculi Orbicularis Oris
Surprise	1 5 26, 27	Raised eyebrows Raised upper eyelids Open mouth	Medial Frontalis Levator palpebrae superioris Orbicularis Oris
Anger	4 5 6, 7 23	Eyebrow frown Raised upper eyelids Raised lower eyelids Lip tightener	Corrugator Supercilii Levator palpebrae superioris Orbicularis Oculi Orbicularis Oris
Disgust	4 6, 7 9, 10	Eyebrow frown Raised lower eyelids Raised upper lip	Corrugator Supercilii Orbicularis Oculi Levator labii superioris
Fear	4 1 5 26, 27	Eyebrow frown Raised eyebrows Raised upper eyelids Open mouth	Corrugator Supercilii Medial Frontalis Levator palpebrae superioris Orbicularis Oris
Sadness	4 1 15	Eyebrow frown Raised eyebrows Lowered lip corners	Corrugator Supercilii Medial Frontalis Depressor Anguli Oris
Contempt	L12 or R12 L14 or R14	Slight lip corner raised (*asymmetrical*) Dimpler (*asymmetrical*)	Zygomaticus major Buccinator

It would be prudent to note at this point that the "upper facial region" classifies muscle movements emanating from eyes, eyelid, brow, and upper cheek, whereas the "lower facial region" classifies muscle movements emanating from nostril, mouth, lip, buccinator, and lower cheek [43], [44].

The HSCN we propose was trained on the extended Cohn-Kanade (CK+) dataset [45] for the dynamic assessment of affective states and microexpression classification. The CK+ dataset contains continuous facial expression information as actors transitioned from an inactive/neutral state to an activated state as outlined in Table I, along with their corresponding action units and muscle movements. Using the FACS, the continuous nature of the CK+ dataset allows for a continuous model of facial muscle movements in real-time. Visualizing changes in expressions as time-dependent functions and interpreting them using the EMFACS can help in the initial validation of the rule-based expert system being proposed.

Through dynamic modeling of expressions of affective states and using multiple features, our proposed HSCN aims to improve on prevailing facial expression classification systems [16], [46]. Considering variations in affective states as functions of time, the HSCN exploits continuous emotion models and attempts to further their static, discrete classification counterparts. As HSCN is based on continuous expression delineation, it goes beyond modeling the transient expressions and expression-intensity variations at the macrolevel. It also demonstrates the transience of expressions by modeling continuous microlevel

muscle movements in upper and lower facial regions and allows for classification of various microexpressions. Furthermore, the HSCN builds upon the prevailing dynamic facial expression recognition systems [47] and proposes an alternate approach for continuous macro- and microexpression analysis.

Table I makes it obvious that the upper facial muscles move in a different way compared to the lower facial muscles during discrete expressions of affective states. This fact raises a question—whether one facial region is more important than the other. Research in [48] answers this question by examining the human participants' response to video recordings in order to determine the relative importance of upper and lower facial regions for classifying facial expression. The study [48] reported that the importance of different facial regions is dependent on the affective state being expressed and that a full facial expression is always easiest to classify [48]. These findings were also supported by research conducted in [42], which looked at the impact of different facial regions while attempting to classify facial expressions. These studies suggest that for developing a comprehensive affective state assessment system, detection and classification of both upper and lower facial region microexpressions are important [44], [47]. In [48], authors reported that humans more accurately classify affective states using the lower facial region expressions compared with the upper facial region. This pattern was also observed while validating the performance of the HSCN's microexpression classifiers.

III. Related Works

In order to overcome limitations of a single cue-based, vision-supported affective state classifier, multiple-cue supported classifiers have been proposed. A recent survey [16] presents a corpus of affective state assessment solutions, focusing on the ones related to assessment of audio and visual cues. Research in [46] reported deployment of a prototype multimodal affective state assessment machine that used facial expressions and speech signals to improve the classification performances of a septenary classifier that could be compared to those discussed in [16]. For real-time classification of affective states, an active-camera system has been used to track changes in the face and integrating them in a classifier that exploits human face and lip features to describe muscle-based expressions of affective states [47].

Pfister et al. [49] suggested using temporal interpolation for feature mapping prior to implementing traditional machine learning classifiers like SVMs, multiple kernel learning, and RFs. Xu et al. [50] proposed a "facial dynamics map" which characterizes microexpression related movements using granular pixel features along with an algorithmic approach that was based on optical flow estimation. Xu et al. [50] employed a SVM classifier to identify and categorize different types of facial microexpressions. Polikovsky et al. [40] used the EMFACS for microexpression detection such that their method divided full facial images into smaller facial regions based on action unit locations. A histogram of oriented gradients (HOGs) approach was combined with the K-nearest neighbor (KNN) classifier for detecting microexpression and action unit activations.

Instead of exploiting the visual cues, the authors [41], [42] used facial thermal features for facial expression classification. In [41], facial thermal features were compared on the basis of muscle-activated temperatures along both, upper and lower facial regions. In [42], the authors reported differences in classifier performances when different subregions of the face were used for feature extraction. While thermal features are more innate and use biometric data, their use case in real-time systems is hampered by the cost and accessibility of thermal cameras.

In [51], authors deployed an extension of the popular bidirectional transformer (BERT) model called micron-BERT, which exploited attention maps to perceive the difference between two frames. Authors performed experiments using 3/4/7 classes [51] and proposed a solution. In this work, we used HSCN to model 21 microexpressions in both the upper and lower facial regions. Similarly, in [52], authors proposed an attention-based magnification-adaptive network (AMAN) to focus on the magnification levels of different microexpressions. Their proposed network deployed a magnification attention module which leverages a pretrained ResNet-18 model followed by a frame attention module which combines the five-class classification [52].

We collate the related works in Table II and summarize: 1) feature space; 2) deployed architecture; 3) microexpression classification; 4) macrolevel facial expression classification; 5) number of classes modeled; and 6) accuracy/performance metric. Through these works, we observed that feature maps and attention mechanisms were common in microexpression recognition literature (including thermal representations). In these works, the authors also stressed on the fleeting nature of microexpressions and the difficulty in modeling real-time changes in microexpressions. We therefore propose the HSCN framework as a vehicle to model activations in time and use upper and lower facial region microexpressions to explain the detected macrolevel expressions.

IV. Methods

Unsupervised learning approaches help in discovering effective states by separating and labeling patterns within a collection of unlabeled data. Therefore, unsupervised learning is widely used for pattern classification for assessing affective state. Previous works had treated continuous expression intensity estimation as an unsupervised learning problem. Generally, continuous expression sequences begin as a neutral expression and evolve to a fully activated and unique facial expression [12], [53]. The corpus of unsupervised learning algorithms is extensive, ranging from dimensionality reduction to manifold learning, to linear and nonlinear clustering techniques [54]. These methods rely on statistical foundations. They detect similarities within a set of unlabeled data and exploit either similarity or dissimilarity measures for the purpose of identifying trends and building clusters in a classification space [54].

Our ultimate aim was to deploy a model that would represent upper, lower, and full facial expressions using linear representations and describe changes in micro- and macrolevel

TABLE II
COMPARISON OF WORKS DISCUSSED IN SECTION III, HIGHLIGHTING THE FEATURE SPACE, DEPLOYED MODEL, PERFORMANCES, AND CAPABILITIES REGARDING UPPER, LOWER, AND FULL-FACIAL CLASSIFICATION

Author	Features	Deployed Model	Upper Face	Lower Face	Full-Face	No. Classes	Classification Accuracy
Vice [46]	Image space	InceptionV3, Xception	X	X	✓	7	75.88%–92.78%
Oliver [47]	2-D "blob" features	Hidden Markov Model (HMM)	X	✓	X	5	95.95%
Pfister [49]	Active Shape Model (ASM) feature points	Multiple Kernel Learning (MKL), RF, SVM	X	X	✓	2	47.6%–83.0%
Xu [50]	Facial landmarks → Facial dynamics map	SVM	X	X	✓	2–3	41.96%–75.66%
Polikovsky [40]	3-D Orientation Gradients Histogram + AUs	K-means classifier + voting	✓	✓	X	47	68.34%–81.5%
Khan [41]	Thermal features → Principal components	PCA + LDA	X	X	✓	3–6	71.05% and 73.0%
Khan [42]	Thermal features → Principal components	PCA + LDA	✓	✓	✓	5	66.28% and 56.0%
Nguyen [51]	Image → Attention Map	Modified BERT	✓	✓	✓	3–7	32.54%–89.14%
Wei [52]	Image → Attention Map	Modified ResNet-18	X	X	✓	5	66.82%–79.87%
This work	Image space → Linear discriminants	LDA, SVM, RF	✓	✓	✓	3 x 21	73.63%–87.68%

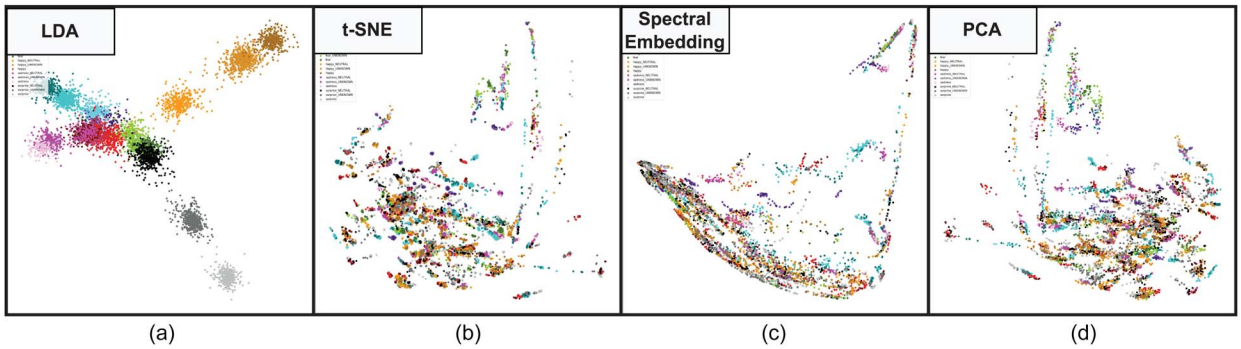

Fig. 1. Comparison of different projection/embedding techniques that were explored during HSCN's early design stages including: (a) LDA; (b) t-distributed stochastic neighbor embedding (t-SNE); (c) spectral embedding; and (d) principal component analysis (PCA). Through these projections, use of LDA was as the basis for the HSCN is justified.

expressions. To achieve this, we needed to separate discrete, septenary class data into a continuous representation that could describe more nuanced expressions of states using muscle movements.

In early design stages, we experimented with projection techniques like t-SNE, PCA, and spectral embedding, among others. We found that LDA resulted in the most optimal intra- and intercluster variance in two dimensions while using 100×100 pixel images. Results of these experiments are visualized in Fig. 1. These figures revealed that LDA produced well-defined clusters that would reveal cluster-to-cluster relationships and define a logical rule-based system. This influenced our decision to opt for LDA moving forward.

The HSCN combines two techniques as shown in Figs. 2 and 3. The initial unsupervised clustering and labeling approach is based on measurements of cosine similarity measures in continuous data that were projected onto an m-dimensional hyperplane. Similar approaches were used in [55], [56] for multiclass facial expression classification.

The LDA transform has been previously used for maximizing the separation between clusters [57]. Invoking LDA would project the high-dimensional data onto a lower dimensional linear discriminant (feature-based) space and would cluster data in a way that maximizes the *intercluster* variance and minimizes the *intracluster* variance. This method maximizes the separation between cluster centroids while minimizing the separation between samples that belong to the same class [58].

Analyzing the separated clusters enables modeling the state-to-state transitions and classification of macrolevel affective states in a continuous domain. Clusters formed at the macrolevel via LDA provide foundations for defining the microlevel clusters of lower and upper facial regions' data. As visualized in (1), these lower and upper facial regions' data relationships could have been more complex if other embedding techniques were deployed. Thus, the hierarchical clustering approach provides the structure to build the HSCN's rule-based expert system. Dimensionality reduction and clustering using

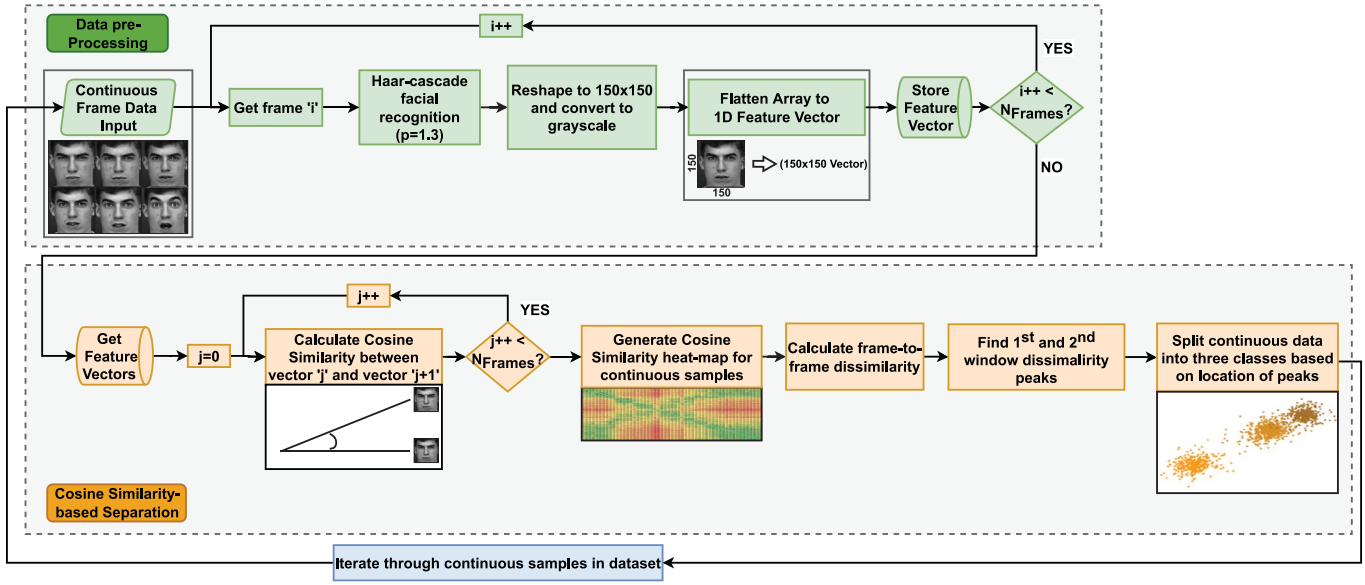

Fig. 2. Preprocessing of continuous expression data and their unsupervised cosine-similarity based separation. Note how continuous frame data from the CK+ dataset are input and preprocessed prior to undergoing the cosine similarity-based separation and clustering in the HSCN.

LDA were regarded as a set of logically apt optimization steps for this work.

Solution to the optimization problem was found in determining the linear discriminants, which corresponded to the largest eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$, noting that the number of linear discriminants required to solve an LDA problem depends on the number of labeled classes in a given set [58]. Facial expression data "$\mathbf{x_i}$" were then projected onto the discriminant function used for the classification tasks and for determining to which class "$\mathbf{k}$" an expression "$\mathbf{x_i}$" would belong to on the basis of similarity measures. For example

$$b'(\mathbf{x_i} - \bar{\mathbf{x_1}}) - b'(\mathbf{x_i} - \bar{\mathbf{x_k}}) - \cdots - b'(\mathbf{x_i} - \bar{\mathbf{x_{21}}}) < 0 \quad (1)$$

where "$\bar{\mathbf{x_k}}$" defines the $\mathbf{k}$th cluster centroid.

As shown in Figs. 2 and 3, projections onto the lower dimensional space were applied in two stages.

1) Projection of the cosine similarity-separated clusters onto a two-dimensional linear discriminant space, maximizing separation between cluster centroids to create the *macrolevel* facial expression classifier.
2) Projection of upper and lower facial region data onto two-dimensional space divided by hyperplanes. Using the rule-based expert system allowed for the systematic detection and classification of upper/lower facial region *microexpressions*.

The processes and subsystems contained within the HSCN framework are further discussed in the following sections.

A. Cosine Similarity-Based Separation

Some definitions of the used terms and concepts are given as follows for explaining the unsupervised separation and clustering methodology.

1) $\mathbf{x_i} = \{x_1, x_2, ..., x_m\}$ defines a pattern or feature vector, i.e., a flattened facial expression image containing "m" raw pixels/features.
2) $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$ defines a set of "$\mathbf{N}$" input patterns all containing "m" features. In this work, "$\mathbf{X}$" defines a continuous series of facial expression images ranging from neutral to activated, which have been projected onto an $m-$dimensional hyperplane.
3) $\mathbf{C} = \{c_1, c_2, ..., c_k\}$ defines the "k" class labels for the patterns contained in the pattern set "$\mathbf{X}$." As mentioned earlier, there are $k = 21$ classes for all micro- and macrolevel classifiers in the network.

Similarity measures have been used in both supervised and unsupervised learning problems [53]. Discovering similarity and dissimilarity measures across "$\mathbf{N}$" patterns in a continuous sample set "$\mathbf{X}$" allows for categorizing subsets of patterns based on similar features and mutual information. The HSCN is split into three major subsystems, the first is tasked with the autonomous extraction of dynamic, macrolevel affective state clusters from a set "$\mathbf{X}$" of data. Separation and initial clustering of patterns was based on mutual information extracted via cosine similarity measures. Separation of continuous data was done by comparing the cosine similarity between all images/patterns within an m-dimensional hyperplane.

Cosine similarity leans on measuring the angle between two image vectors $\{\mathbf{x_i}, \mathbf{x_j}\}$ projected onto a hyperplane of dimension "m" [59]. As the mutual information between the two vectors increases, the angle between them decreases, such that $\cos\theta = 1$ when $i = j$. The cosine similarity between two images was therefore calculated as such

$$S_{\cos\theta} = \frac{\mathbf{x_i} \cdot \mathbf{x_j}}{|\mathbf{x_i}|\,|\mathbf{x_j}|}. \quad (2)$$

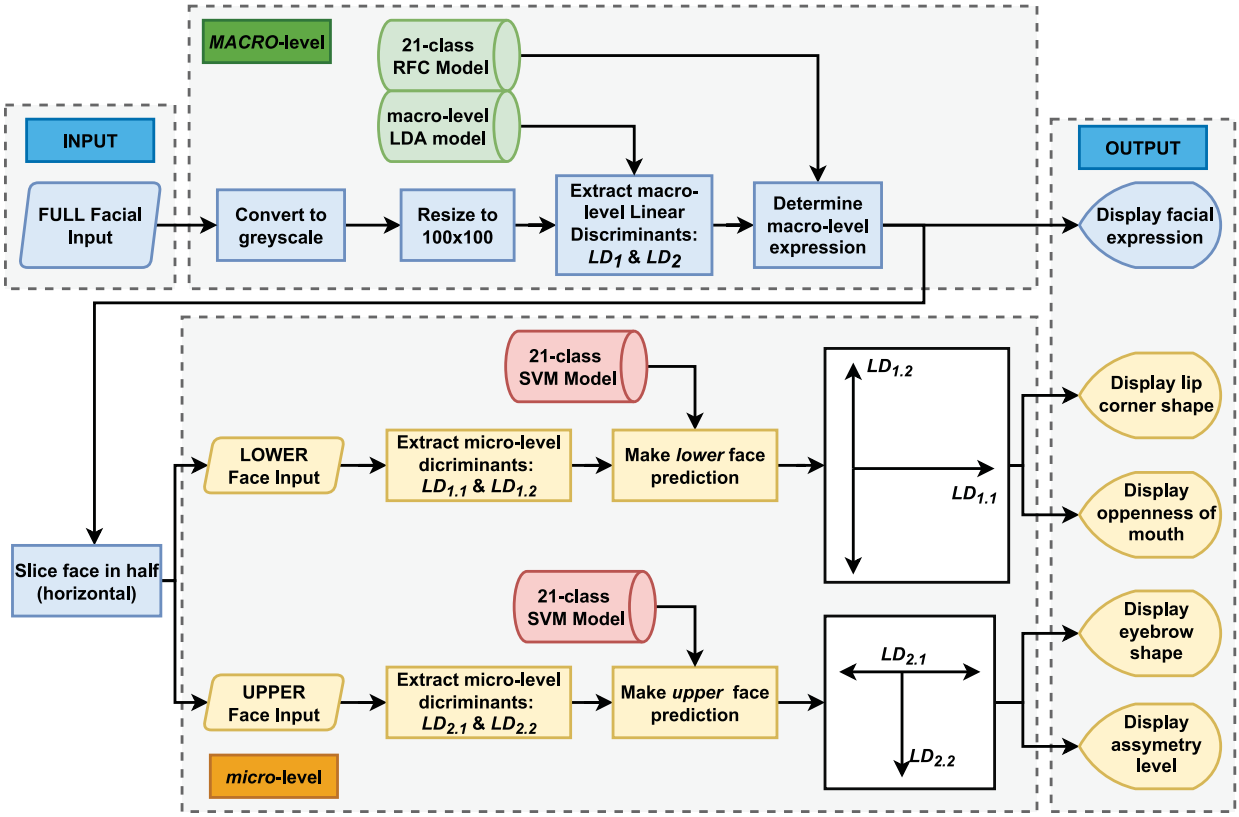

Fig. 3. Visual summary of the proposed HSCN containing a rule-based expert system. The flowchart shows how facial image data are processed from input to macro- and microlevel assessment stages, and how classification results are reported and displayed to the users. Combined with Fig. 2, one can see how the ensemble of RFC, LDA, and SVM models were deployed across various stages.

In the CK+ dataset, a large *intercluster* variance was observed using the cosine similarity approach. In this work, cosine similarity measures were used to detect a set of any two serial expressions showing higher levels of *dissimilarity*. Regarding the CK+ dataset; high levels of dissimilarity indicate a noticeable change in affective state expression intensity. For each continuous set of facial expression samples, the dissimilarity detection algorithm allowed for the separation of large clusters of images into three, macrolevel facial expression clusters based on similar features, labeled as follows:

1) Cluster 1: *Neutral*-dominated state;
2) Cluster 2: *Partially* activated state;
3) Cluster 3: *Fully* activated state.

This initial unsupervised separation process was performed through a "frame-to-frame gradient analysis" which iterates through continuous data and calculates the dissimilarity magnitude "$\Delta S_{\cos\theta}$" between facial expressions in the series. The gradient magnitude was calculated as

$$\Delta S_{\cos\theta} = S_{\cos\theta}(\mathbf{x_i}, \mathbf{x_i}) - S_{\cos\theta}(\mathbf{x_i}, \mathbf{x_{i+1}}) \tag{3}$$

with "$S_{\cos\theta}(\mathbf{x}, \mathbf{y})$" defining the similarity measurement between the two serial expressions. This equation is applied "$\mathbf{N} - 1$" times to define all frame-to-frame transitions in $\mathbf{X}$.

Dissimilarity magnitudes were used to detect locations of peak dissimilarity, which defined the cluster boundaries within the hyperplane. The deployed algorithm splits $\mathbf{X}$ into two equal length subsets, with the global maxima (peak dissimilarity) being defined in each half. By modeling the continuous nature of affective states, this allowed for classification of twenty-one, transient macrolevel facial expressions. The separation of CK+ image samples via the cosine similarity method is shown in Fig. 2 and is further explained through Algorithm 1.

Theoretically, this algorithm could be extended to increase the resolution of the transient facial expression classes. Increasing the number of dissimilarity peaks would correspond to an increase in the number of clusters extracted from a continuous sample such that: $N_{\text{states}} = N_{\text{peaks}} + 1$. Furthermore, this method is not limited to the image domain and could be deployed for the separation of affective speech and video data.

B. Macrolevel Linear Discriminant Analysis

The initial clustering via the cosine similarity-based separation method were input into the second tier of the HSCN, where macrolevel LDA clustering was performed. Please note that LDA has been extensively used to effectively separate and cluster labeled facial expressions by discovering hyperplanes within a linear discriminant space [60]. The clustering was achieved by maximizing intercluster variance in order to optimize cluster centroid separation. Fig. 4 highlights results of the macrolevel LDA clustering algorithm when applied to a large volume of continuous facial expression data. Analyzing the cluster centroids in subplot 2 of Fig. 4, we see emergence

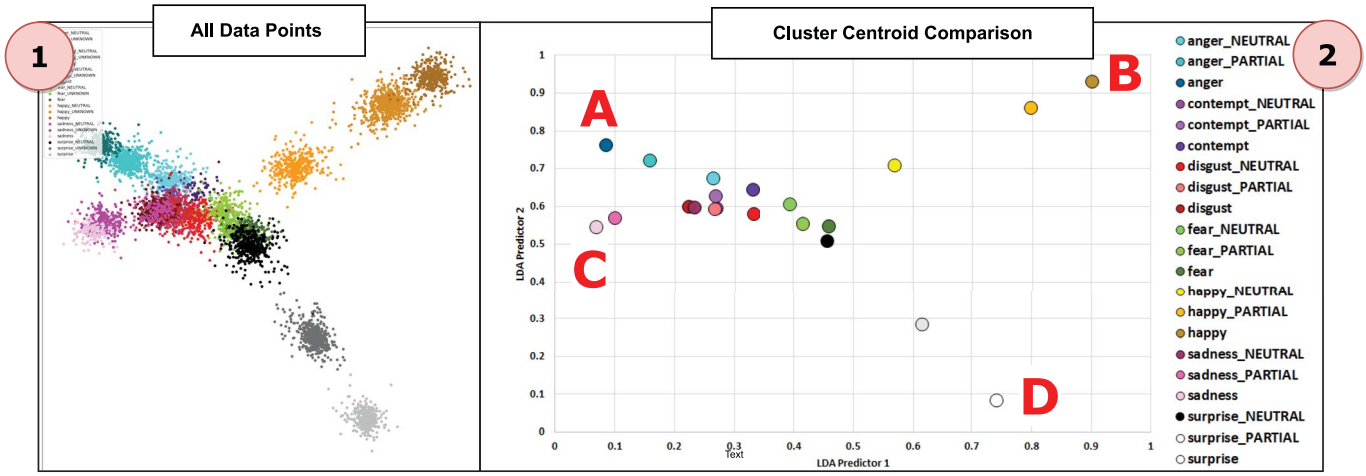

Fig. 4. Macrolevel LDA clustering results. *Subplot 1* on the top left is the input from the initial, cosine similarity-based separation algorithm displaying all data samples extracted from the CK+ dataset. *Subplot 2* on the top right are the cluster centroids for each of the macrolevel expressions with the centroid closest to (A) fully activated anger; (B) fully activated happiness; (C) fully activated sadness; and (D) fully activated surprise. We expand on these state transitions in another figure.

Algorithm 1: Cosine similarity-based separation

input: Continuous CK+ Dataset samples
Define $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$
for $\mathbf{x_i}$ *in* $\mathbf{X}$ **do**
  Extract Facial Image
  Reshape $\mathbf{x_i}$ to 150×150 pixels
  Convert $\mathbf{x_i}$ to greyscale
  Flatten $\mathbf{x_i}$, i.e. $\mathbf{x_i} = \{x_1, x_2, ..., x_{22500}\}$
end
Let "$\mathbf{x_i}$" $= i^{th}$ test facial expression vector
Let "$\mathbf{x_j}$" = comparison vector
for $i = 1;\ i \leq N;\ i = i + 1$ **do**
  for $j = 1;\ j \leq N;\ j = j + 1$ **do**
    $S_{\cos\theta} = \frac{\mathbf{x_i} \cdot \mathbf{x_j}}{|\mathbf{x_i}||\mathbf{x_j}|}$
  end
end
for $i = 1;\ i \leq N - 1;\ i = i + 1$ **do**
  $\Delta S_{\cos\theta} = S_{\cos\theta}(\mathbf{x_i}, \mathbf{x_i}) - S_{\cos\theta}(\mathbf{x_i}, \mathbf{x_{i+1}})$
end
Detect $\max[\Delta S_{\cos\theta}]$ in each half of $\mathbf{X}$
Define $3\times$ dynamic clusters per state
output: 21 Macrolevel Facial Expression Clusters

of linear trends from inactive (_NEUTRAL) expressions to partial expressions (_PARTIAL) to fully activated expressions of all affective states. Furthermore, using points **{A,B,C,D}**, we can begin to construct two continuous axes that separate these states:

1) sadness (**C**) to happiness (**B**);
2) anger (**A**) to surprise (**D**).

We needed to understand "the theoretical underpinning of these axes?" The linear discriminant space visualized in Fig. 4 was a low-dimensional linear discriminant representation of facial expressions, a mapping that corresponded with certain feature changes and variations in facial expressions at a higher level. We also noticed that the other three activated states (contempt, disgust, and fear) centroids resided on the two defined axes, with contempt existing at the intercept of the two axes. This was predictable as it was the most "neutral" expression relative to the other affective states being modeled.

Henceforth, defining rules on the basis of the logical foundations of the EMFACS and Table I became very important. Comparing changes in muscle activations from one state to the other state would help in determining what these linear relationships actually represented in real-life. This would also provide foundations for building a rule-based expert system capable of detecting and classifying microexpressions. By comparing sadness and happiness muscle activations in Table I, we were able to model the state-to-state transition and visualize how expressions changes were based on muscle movements as shown in Fig. 5. Given the common facial muscles involved in changing the expression from sadness to happiness, we could define an axis rule.

Sadness–happiness axis rule: Sadness and happiness share common facial muscle groups surrounding the mouth, cheek, and eyelid regions. The formed axis would model the following transformations: 1) parallel relaxation of brows and raising of cheeks; and 2) raising of lip corners and mouth from an initial down-turned expression.

Similarly, comparing anger and surprise muscle activations in Table I, we could model the transition from anger to surprise as visualized in Fig. 5. Note that in this case, both states evidenced "raised upper eyelids" which was useful when attempting to derive a clearer relationship. Given this second example and the common muscle groups and facial regions that were activated (eyebrow and mouth region), we could define a second axis rule as the following.

Anger–surprise axis rule: Anger and surprise share common facial muscle movements surrounding the mouth and eyebrow regions and share a consistent "raised upper eyelid"

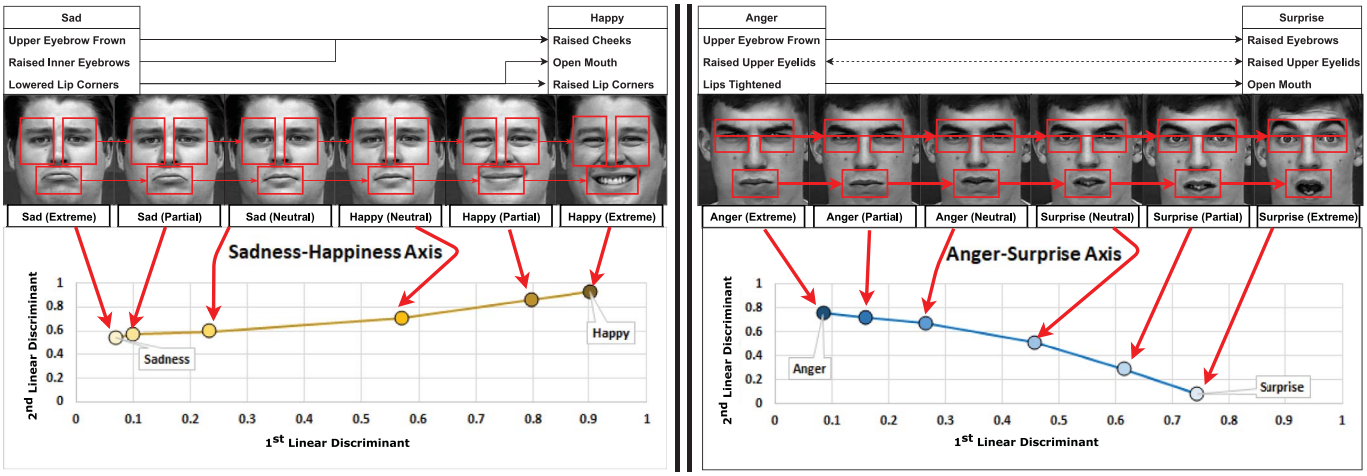

Fig. 5. Left: State-to-state transition along the *sadness—happiness* trend as visualized in Fig. 4. Examples of neutral and partial states are also shown. Right: State-to-state transition along the *anger–surprise* trend as visualized in Fig. 4. Examples of neutral and partial states are also shown.

TABLE III
STATE-TO-STATE COMPARISON SHOWING DIFFERENCES BETWEEN THE nTH LINEAR DISCRIMINANTS. THE HIGHLIGHTED NUMERICAL VALUES DISPLAY HIGH AND LOW $|\Delta LD_n|$ VALUES

$P_X \rightarrow P_Y$	State $\rightarrow$ State	$\lvert\Delta LD_1\rvert$	$\lvert\Delta LD_2\rvert$
A $\rightarrow$ B	Anger $\rightarrow$ Happy	**0.8170**	**0.1683**
A $\rightarrow$ C	Anger $\rightarrow$ Sadness	**0.0159**	0.2163
A $\rightarrow$ D	Anger $\rightarrow$ Surprise	0.6563	0.6789
B $\rightarrow$ C	Happy $\rightarrow$ Sadness	**0.8329**	0.3846
B $\rightarrow$ D	Happy $\rightarrow$ Surprise	**0.1606**	**0.8471**
C $\rightarrow$ D	Sadness $\rightarrow$ Surprise	0.6722	0.4626

activation. Therefore, the state-to-state transition could model the following transformations: 1) eyebrows raise from an initial frowned/depressed position; and 2) mouth opens from an initial tightened expression.

Expanding on the two rules that have been formed thus far, we might postulate an initial hypothesis in regard to what the X and Y axes (i.e., linear discriminant 1 and 2, respectively) represented in this case. Let the linear discriminant "n" be denoted by LD_n, the $|\Delta LD_n|$ values when comparing states, i.e., points $A \rightarrow D$ in subplot 2 of Fig. 4 are reported in Table III. Together with Fig. 6, these results serve as the basis for proving the validity of the rule-based microexpression classifier [61]. Given the evidence provided, we could define the following hypotheses and macroexpression rules.

1) LD_1 relates to the openness of the mouth and the lower region of the face given the following evidence.
   a) Sadness and anger shared a low ΔLD_1. The two common actions between the states were: "upper eyebrow frown" and "lips tightened/lowered corners."
   b) Presence of two common actions would be troublesome if not for the presence of the surprise and happiness states, which also shared a low ΔLD_1. The common action between surprise and happiness revolved around raised lip corners and ultimately, the open mouth.

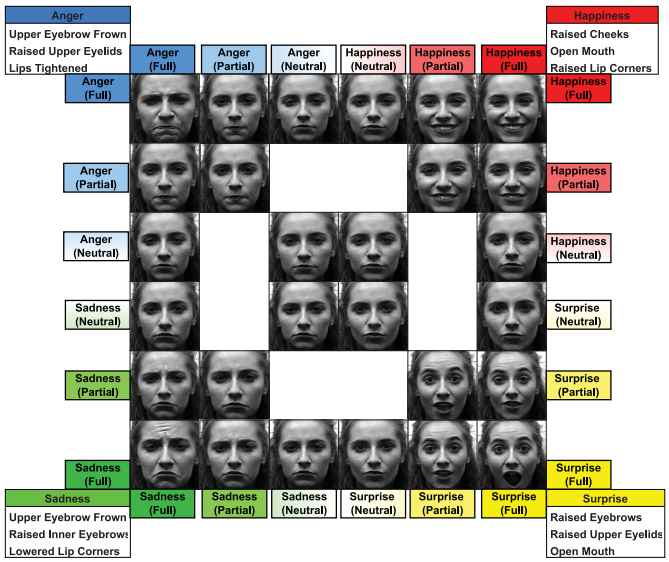

Fig. 6. Visualized state-to-state transition for anger, happiness, sadness, and surprise expressions. This figure assists in mapping LD_n features to real-world features.

2) LD_2 relates to the region around the eyes, i.e., the eyelids and eyebrows—the upper facial region, evidenced by the following.
   a) The anger–sadness transition evidences both a low ΔLD_1 and ΔLD_2. If the initial hypothesis is that LD_1 is related to the mouth, then the second common action—"upper eyebrow frown" may be related to LD_2, which supports the upper facial region relationship.
   b) Analyzing Fig. 6 and the transition from anger to happiness, we saw that the eyes remained the same shape, with the largest variance evident between *full* to *partial* anger states, when the frown was relaxed slightly. Removing the lower half of the face, we could observe that there were similarities between the brow/eye region of the two states.

c) The large variance between happiness and surprise. Given that the open mouth was deduced as being referred to by LD_1, we could identify the difference between happiness and surprise frames in Fig. 6 through the upper region of the face, specifically the brow and eye regions, thus providing further evidence toward LD_2 relating to the upper facial region.

By inferring the above rules for the macrolevel LDA clustering approach, we were able to define a relationship that allowed mapping statistical features with real-world features vis–à–vis providing a vehicle for transient macrolevel facial expression classification.

C. Microlevel Linear Discriminant Analysis

We had previously defined the following rules.

1) LD_1 relates to the shape of the mouth and the lower facial region.
2) LD_2 relates to the region around the eyes, eyelids, and brows—the upper facial region.

These claims were substantiated through the necessary experiments. The microexpression LDA clustering subsystem aimed to prove the validity of the two claims, while providing a deeper analysis of dynamic facial expressions and focusing on the upper and lower facial regions.

An automated function was implemented to slice the CK+ images in half (horizontally), allowing to focus on both the upper and lower facial regions independently. An additional LDA clustering approach was then applied to the new image vectors in an attempt to validate the above hypotheses, thus allowing for the classification of microexpressions in upper and lower facial regions. If the initial hypotheses were correct, then there would be a very discernible trend between states at the microlevel as this would indicate that the projected feature "LD_n" is related to a particular group of muscles.

Let us describe the *macrolevel* linear discriminant features as "LD_n"—i.e., LD_1 = the lower facial region and LD_2 = upper facial region. Moving to the *microlevel*, let "m" describe the microlevel features contained within the higher, "nth" level regions, i.e., "$LD_{n.m}$." For example, $LD_{1.1}$ and $LD_{1.2}$ describe microexpressions in the lower facial region.

The clusters shown in Fig. 7(a) share a similar $LD_{2.2}$ value, with the largest variance being in the direction of $LD_{2.1}$. We could also see that the clusters moved linearly from anger to surprise along the $LD_{2.1}$ axis. Note that fear and contempt states display large variances in the $LD_{2.2}$ axis, sharing common $LD_{2.1}$ values (cluster centroid coordinates). Furthermore, it could be noted that as per se contempt is an asymmetrical expression could explain why it was such an outlier. But besides these two states, we saw that the majority of expressions existed along the $LD_{2.1}$ axis.

The lower facial region (LD_1) is more sparsely clustered compared to the upper facial region. In Fig. 7(b), it is evident that most states reside on one side of the spectrum, sharing a similar $LD_{1.1}$ feature value, with happiness and its substates displaying the largest variance in $LD_{1.1}$. The notable trend observed in the lower region of the face could be attributed to the microlevel feature $LD_{1.2}$—the y-axis, showing disgust (top) and surprise (bottom) as the two extremes.

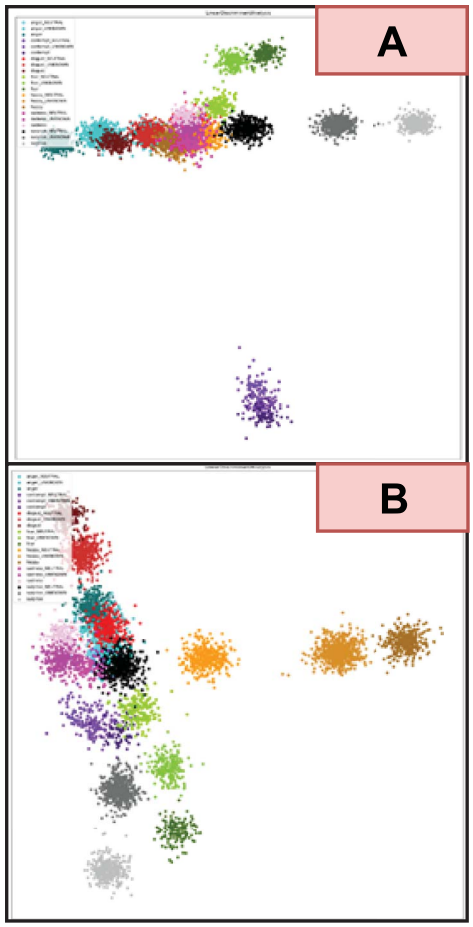

Fig. 7. (a) Two-dimensional linear discriminant space representation of the upper facial region, showing $\{LD_{2.1}, LD_{2.2}\}$ microexpression features. (b) Two-dimensional linear discriminant space representation of the lower facial region, showing $\{LD_{1.1}, LD_{1.2}\}$ microexpression features.

V. Results

A. Defining the Rule-Based Expert System

Throughout this work, we have emphasized on mentioning how similarity and dissimilarity measures can be exploited to improve classifier performances and the capabilities of decision-making systems. We can also use distance-based similarity (specifically Euclidean distances) to validate the rule-based system derived for the HSCN. Previously, we defined the macrolevel linear discriminant features as LD_n and the microlevel features as $LD_{n.m}$. Combined with Fig. 5, Table III proves the validity of our macrolevel facial expression rules. We observed that anger, happiness, sadness, and surprise exist as extreme points in a 2-D linear discriminant space and by referring back to the EMFACS muscle movements discussed previously, it allowed establishing that LD_1 relates to the shape of the mouth and the lower facial region and LD_2 relates to the region around the eyes, eyelids, and brows—upper facial region.

TABLE IV
MICROLEVEL LINEAR DISCRIMINANT CENTROID COORDINATES FOR UPPER AND LOWER FACE CLUSTERS. SUBSCRIPTS REFER TO NEUTRAL-DOMINATED, PARTIAL, AND FULLY ACTIVATED LABELS

	Lower Face		**Upper Face**	
Cluster	$LD_{1.1}$	$LD_{1.2}$	$LD_{2.1}$	$LD_{2.2}$
$Anger_N$	0.178	0.613	0.278	0.757
$Anger_P$	0.163	0.683	0.159	0.760
$Anger_A$	0.144	0.720	0.054	0.739
$Cont._N$	0.107	0.457	0.542	0.128
$Cont._P$	0.158	0.428	0.557	0.087
$Cont._A$	0.210	0.408	0.565	0.056
$Disg._N$	0.183	0.664	0.384	0.788
$Disg._P$	0.120	0.847	0.283	0.761
$Disg._A$	0.081	0.942	0.192	0.740
$Fear_N$	0.257	0.463	0.448	0.828
$Fear_P$	0.329	0.323	0.488	0.941
$Fear_A$	0.350	0.165	0.568	0.960
$Happy_N$	0.421	0.592	0.412	0.758
$Happy_P$	0.745	0.607	0.358	0.745
$Happy_A$	0.911	0.630	0.342	0.739
Sad_N	0.143	0.569	0.381	0.737
Sad_P	0.069	0.592	0.381	0.781
Sad_A	0.075	0.647	0.371	0.805
$Surp._N$	0.216	0.570	0.519	0.774
$Surp._P$	0.210	0.268	0.746	0.783
$Surp._A$	0.184	0.069	0.942	0.787

Deriving relationships between muscle movements of the full-facial region allowed to further explore the upper and lower facial regions and conduct similar experiments. Performing the same analysis in Table IV, we can establish relationships between features in the upper facial region and lower facial region using the EMFACS action units. The neutral, partial, and fully activated microexpression cluster centroids reported in Table IV are used to supplement the findings visualized in Fig. 7.

The *anger–surprise axis rule* modeled two transformations: 1) eyebrows raise from an initial frowned/depressed position; and 2) mouth opens from an initial tightened position. When combined with the hypothesis "LD_2 relates to the region around the eyes, eyelids, and brows—the upper facial region." Thus, we could state that the microlevel feature $LD_{2.1}$ referred to a translation of the eyebrows from an initial frowned/down-turned position to a raised position, making use of the *medial frontalis*, *levator palpebrae superioris,* and *corrugator supercilii* facial muscles.

Fig. 7(b) exhibits an axis that has been derived previously. Only in this instance, it has been mapped from one feature space to another, the sadness–happiness axis, which varies in regard to the $LD_{1.1}$ feature. The macrolevel, *sadness–happiness axis rule* modeled the following two transformations: 1) parallel relaxation of brows and raising of cheeks; and 2) raising of lip corners and mouth from a depressed initial condition. The second transformation relates to the lower facial region feature. This transformation, pertaining to the lip corner movements, may be the causal link between the microlevel feature $LD_{1.1}$ and the real world. Analyzing the variations in $LD_{1.2}$, we see disgust and surprise on opposite sides of the spectrum. The immediate hypothesis is that $LD_{1.2}$ models the openness of the mouth, and the manipulation of the central lip muscles—*orbicularis oris* and *levator labii superioris*.

The evidence provided by the microlevel LDA-derived clusters prove the validity of the macrolevel inferences. Moving into the microexpression (mth level), we can define each axis as follows.

1) $LD_{1.1}$: Lip corner muscles—modeling a translation from a down-turned shape to a lifted shape.
2) $LD_{1.2}$: Openness of the mouth—modeling the manipulation of the central lip muscles from closed to open.
3) $LD_{2.1}$: Models a translation of the eyebrows from an initial frowned position to a raised position.
4) $LD_{2.2}$: Used to identify expressions such as fear and contempt and asymmetrical movements outside of the spectrum from anger to surprise.

Establishing this rule-base provided us with the foundations on which the upper and lower facial region microexpression detection and classification systems were built.

Deploying the HSCN in a rule-based expert system would allow for the continuous monitoring and assessment of macrolevel expressions of affective states and microexpressions, which allow for the modeling of specific muscle movements in the upper and lower facial regions. In the previous sections, the macro- and microexpressions were modeled using linear discriminant features defined by nth and mth level subscript notation, i.e., $LD_{n.m}$. Fig. 3 visualizes the rule-based expert system and shows how facial image data are processed from input to output stages. Through classification, the proposed system is capable of assessing various levels of facial expressions and muscle movements using the rules derived earlier.

The rule-based expert system facilitates hierarchical detection of macrolevel facial expressions as well as detection of upper and lower facial region microexpressions as each of the three classifiers models some unique states. Using the microlevel rules defined in the previous section, the HSCN is able to detect changes in: 1) lip corner muscles; 2) openness of the moth; 3) translation of the eyebrows; and 4) level of asymmetry.

B. Hierarchical Classifier Performance

Both macro- and microlevel algorithms developed for this work were validated using the RF, SVM, and KNN classification approaches. Classifiers were trained using the clustered data that had been defined through the separation and clustering subsystems of the HSCN. The CK+ facial expression images used to train the classifiers were resized to 100 × 100 pixels and were then flattened, providing 5842 image samples. The 80/20, train/test split was used for validating performance of each classifier. The validation accuracies in this work show the percentages of correct guesses with respect to the total number of guesses made, defined as

$$\text{Acc}_{\text{val}}(\%) = \frac{N_{(\text{correctguesses})}}{N_{(\text{totalguesses})}} \times 100\%. \tag{4}$$

Classification performance of all models are reported in Table V. We used the RF classifier for macroexpression classification, predicting twenty-one transient affective state expressions, across seven independent state axes with 76.11%

TABLE V
HSCN PERFORMANCE WHEN TRAINED AND VALIDATED ON THE CK+ DATASET USING AN 80/20 TRAIN/TEST SPLIT INVOKED ON 21 UNIQUE CLASSES PER FACIAL REGION

Facial Region	Model	Validation Acc. (%)
Full-face (macro)	**RF**	**76.11**
Full-face (macro)	SVM	72.95
Full-face (macro)	KNN	54.11
Upper face (micro)	RF	70.89
Upper face (micro)	**SVM**	**73.63**
Upper face (micro)	KNN	72.67
Lower face (micro)	RF	86.37
Lower face (micro)	**SVM**	**87.68**
Lower face (micro)	KNN	87.26

accuracy. For the upper and lower facial microexpression classifiers, we deployed SVMs, capable of predicting twenty-one variations of upper and lower facial muscle movements based on mth level "$\text{LD}_{n.m}$" linear discriminant features and the rules defined in Section IV. We achieved 73.63% and 87.68% classification accuracies respectively for upper and lower facial region microexpression classification with the SVM classifier.

These observed results are comparable with the recent discrete affective state assessment solutions. Looking at the facial expression classifiers reported in [16], for example, we see that the accuracies of systems in [16] range between 41% and 88% while classifying a lesser number of affective states. When we compare our results with the works discussed in Section III, we find that their range was between 32.94% and 95.95% accuracies. This shows that our classifiers reside in the upper bounds of the performance metrics. The observed HSCN classifiers prove that the resolution and dimensionality of a recognition system can be improved without hindering classifier performances. Furthermore, our results show that continuous affective state assessment solutions could perform as well as discrete models supported systems would.

Performances of the lower and upper facial microexpression classifiers were consistent with the human observations made in [42], [48], stating that the classification of lower facial expressions is on average, more accurate than that of upper facial expressions. Looking at Fig. 7(a) and 7(b), one would see why this might be the case. The lower facial region microexpression clusters show a larger separation across $\text{LD}_{1.1}$ and $\text{LD}_{1.2}$ axes compared to the upper facial region microexpressions. This shows variations along the $\text{LD}_{2.1}$ axis for most states. In reality, the reason for this outcome could be the prominence and relative size of the mouth and lips in the lower facial region. Relatively speaking, the mouth is a larger facial feature compared to other features. Also, muscle activations around the mouth region would generally have a larger impact compared to muscular changes around the eyes or brow region. This explains why it may be easier to classify lower facial region expressions compared to upper facial expressions.

The rules derived in this work along with the previously discussed classifier results allow supporting the real-world phenomena with statistical findings. Through these findings, the importance of both upper and lower facial region muscle movements for facial expression classification has been highlighted and reinforced.

VI. CONCLUSION

This work provides new, novel, and useful information on the respective roles of upper and lower muscles in facial expression classification. We were also able to explain how micro- and macrolevel facial muscle movements could be used to build a robust set of features. Hence, this work should be regarded as a step forward from discrete affective state classification systems, capable of classifying one of "n" discrete affective states to the continuous affective state classification systems. The proposed HSCN is a new and powerful classifier ensemble that exploits separation and clustering for categorizing affective state expressions in a dynamic manner. The HSCN transforms seven independent facial expression clusters into twenty-one transient facial expression clusters and classifies twenty-one upper and twenty-one lower facial region microexpression classes. The twenty-one states classification is performed using a rule-based expert system. A high degree of accuracy was achieved in training and testing stages. The reported HSCN shows competitive performance when compared to other state-of-the-art approaches, reported earlier in Table II.

The ability to detect and classify microexpressions would help affective state assessment in demanding dynamic conditions. However, the complex nature of continuous expression signals would require developing comprehensive models of affective state-caused variations in facial features. Our proposed HSCN approach demonstrates how the micro- and macrolevel facial muscle movement modeling approach would be useful in complex affective state assessment situations. The presented HSCN approach was able to predict twenty-one macrolevel transient expressions vis–à–vis twenty-one upper and lower facial region microexpressions. As discussed previously, the reported validation performance of the HSCN makes it comparable with several previously reported affective state classification systems.

The rule-based, expert system-supported HSCN was built upon the theoretical foundations of the EMFACS and other continuous affective state expression models. Through a combination of: 1) unsupervised cosine similarity-based separation; 2) LDA-based clustering; and 3) traditional supervised learning classifiers, the HSCN's predictive capabilities suggest that it could be used as a quantitative assessment tool that is supported by a theory-driven back-end.

Our future research will focus on integration of the HSCN into a multimodal affective state assessment system. The goal will be to develop a dynamic assessment tool capable of detecting and classifying transient facial expressions in real-time. We intend to expand and modify the HSCN architecture for incorporating human speech as well. This will allow modeling changes in affective *speech* expressions as continuous and time-dependent functions to be used for affective state assessment.

ACKNOWLEDGMENT

All authors hereby declare that they have no conflict of interest.

Compliance With Ethical Standards

Curtin University's ethics under Approval HRE2019-0722, dated 23 October 2019, obtained for this work required adhering to the highest levels of ethical standards. No animals were involved in this research. All data, hardware, software, and related software code remain the intellectual property of the Curtin University.

References

[1] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach," *Knowl.-Based Syst.*, vol. 110, pp. 1–14, Oct. 2016.

[2] A. Cernea and A. Kerren, "A survey of technologies on the rise for emotion-enhanced interaction," *J. Vis. Lang. Comput.*, vol. 31, pp. 70–86, Dec. 2015.

[3] S. Shvimmer, R. Simhon, M. Gilead, and Y. Yitzhaky, "Classification of emotional states via transdermal cardiovascular spatiotemporal facial patterns using multispectral face videos," *Sci. Rep.*, vol. 12, no. 1, pp. 1–16, Jul. 2022.

[4] M. M. Khan, "Cluster analytic detection of disgust-arousal," in *Proc. 9th Int. Conf. Intell. Syst. Des. Appl.*, Pisa, Italy, Nov. 2009, pp. 641–647.

[5] J. A. Coan and J. J. Allen, Eds., *Handbook of Emotion Elicitation and Assessment*. New York, NY, USA: Oxford Univ. Press, 2007.

[6] W. J. Yan, Q. Wu, J. Liang, Y. H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *J. Nonverbal Behav.*, vol. 37, pp. 217–230, Jul. 2013.

[7] U. Saeed, "Facial micro-expressions as a soft biometric for person recognition," *Pattern Recognit. Lett.*, vol. 143, pp. 95–103, Mar. 2021.

[8] S. A. Khan, A. Hussain, and M. Usman, "Facial expression recognition on real world face images using intelligent techniques: A survey," *Optik*, vol. 127, no. 15, pp. 6195–6203, Aug. 2016.

[9] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant," *J. Personality Social Psychol.*, vol. 76, no. 5, pp. 805–819, May 1999.

[10] F. Becattini, F. Palai, and A. Del Bimbo, "Understanding human reactions looking at facial microexpressions with an event camera," *IEEE Trans. Ind. Inform.*, vol. 18, no. 12, pp. 9112–9121, Dec. 2022.

[11] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2003.

[12] X. Zhao and S. Zhang, "A review on facial expression recognition: Feature extraction and classification," *IETE Tech. Rev.*, vol. 33, no. 5, pp. 505–517, Jan. 2016

[13] M. Pantic and L. J. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.

[14] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, nos. 3–4, pp. 169–200, May 1992.

[15] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016.

[16] C. H. Wu, J. C. Lin, and W. L. Wei, "Survey on audiovisual emotion recognition: Databases, features, and data fusion strategies," *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. 12, pp. 1–18, Nov. 2014.

[17] G. R. Alexandre, J. M. Soares, and G. A. P. Thé, "Systematic review of 3D facial expression recognition methods," *Pattern Recognit.*, vol. 100, pp. 1–16, Apr. 2020.

[18] Y. H. Oh, J. See, A. C. Le Ngo, R. C. Phan, and V. M. Baskaran, "A survey of automatic facial micro-expression analysis: Databases, methods, and challenge," *Front. Psychol.*, vol. 9, no. 1128, pp. 1–21, Jul. 2018.

[19] L. Lei, J. Li, T. Chen, and S. Li, "A novel graph-TCN with a graph structured representation for micro-expression recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2237–2245.

[20] X. Ben et al., "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5826–5846, Sep. 2022.

[21] H. Guerdelli, C. Ferrari, W. Barhoumi, H. Ghazouani, and S. Berretti, "Macro-and micro-expressions facial datasets: A survey," *Sensors*, vol. 22, no. 4, pp. 1–34, Feb. 2022.

[22] S. T. Liong et al., "Evaluation of the spatio-temporal features and GAN for micro-expression recognition system," *J. Signal Process. Syst.*, vol. 92, no. 7, pp. 705–725, Jul. 2020.

[23] J. Tang et al., "Quantitative study of individual emotional states in social networks," *IEEE Trans. Affective* Comput., vol. 3, no. 2, pp. 132–144, Apr.–Jun. 2012.

[24] A. Metallinou, A. Katsamanis, and S. Narayanan, "Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information," *Image Vis. Comput.*, vol. 31, no. 2, pp. 137–152, Feb. 2013.

[25] Y. Tie and L. Guan, "A deformable 3-D facial expression model for dynamic human emotional state recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 142–157, Jan. 2013.

[26] A. Abdulhafedh, "Comparison between common statistical modeling techniques used in research, including: Discriminant analysis vs logistic regression, ridge regression vs LASSO, and decision tree vs random forest," *Open Access Library J.*, vol. 9, no. 2, pp. 1–19, Jan. 2022.

[27] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: An experimental investigation," *Knowl. Inf. Syst.*, vol. 10, no. 4, pp. 453–472, Mar. 2006.

[28] P. Ekman and E. L. Rosenberg, Eds., *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 3rd ed. New York, NY, USA: Oxford Univ. Press, 2020.

[29] M. Pantic and L. J. Rothkrantz, "Self-adaptive expert system for facial expression analysis," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Nashville, TN, USA, Oct. 2000, pp. 73–79.

[30] T. U. Ahmed, M. N. Jamil, M. S. Hossain, K. Andersson, and M. S. Hossain, "An integrated real-time deep learning and belief rule base intelligent system to assess facial expression under uncertainty," in *Proc. Joint 9th Int. Conf. Inform., Electron. Vision (ICIEV), 4th Int. Conf. Imag., Vision Pattern Recognit. (ICIVPR)*, Kitakyushu, Japan, Aug. 2020, pp. 1–6.

[31] Z. Yu, L. Li, J. Liu, and G. Han, "Hybrid adaptive classifier ensemble," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 177–190, Feb. 2014.

[32] L. Ali, C. Chakraborty, Z. He, W. Cao, Y. Imrana, and J. J. Rodrigues, "A novel sample and feature dependent ensemble approach for Parkinson's disease detection," *Neural Comput. Appl.*, vol. 35, pp. 15997–16010, Mar. 2022.

[33] F. Burkhardt, T. Polzehl, J. Stegmann, F. Metze, and R. Huber, "Detecting real life anger," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 4761–4764.

[34] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Cognitive Behavioral Systems*, vol. 7403, A. Esposito, A. Vinciarelli, R. Hoffmann, and V. Muller, Eds., Heidelberg, Germany: Springer-Verlag, 2012, pp. 144–157.

[35] Y. Susanto, A. Livingstone, B. C. Ng, and E. Cambria, "The hourglass model revisited," *IEEE Intell. Syst.*, vol. 35, no. 5, pp. 96–102, Sep./Oct. 2020.

[36] J. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Personality*, vol. 11, no. 3, pp. 273–294, Sep. 1977.

[37] R. Plutchik, "Chapter 1—A general psychoevolutionary theory of emotion," in *Theories of Emotion*, vol. 1, R. Plutchik and H. Kellerman, Eds., New York, NY, USA: Academic, 1980, ch. 1, pp. 3–33.

[38] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affective Comput.*, vol. 10, no. 1, pp. 18–31, Jan.–Mar. 2019.

[39] N. Samadiani et al., "A review on automatic facial expression recognition systems assisted by multimodal sensor data," *Sensors*, vol. 19, no. 8, pp. 1–27, Apr. 2019.

[40] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expression detection in hi-speed video based on facial action coding system (FACS)," *IEICE Trans. Inf. Syst.*, vol. 96, no. 1, pp. 81–92, Jan. 2013.

[41] M. M. Khan, R. D. Ward, and M. Ingleby, "Toward use of facial thermal features in dynamic assessment of affect and arousal level," *IEEE Trans. Affective Comput.*, vol. 8, no. 3, pp. 412–425, Jul.–Sep. 2017.

[42] M. M. Khan, M. Ingleby, and R. D. Ward, "Automated facial expression classification and affect interpretation using infrared measurement of facial skin temperature variations," *ACM Trans. Auton. Adaptive Syst. (TAAS)*, vol. 1, no. 1, pp. 91–113, Sep. 2006.

[43] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 4057–4069, 2020.

[44] L. Zhang, B. Verma, D. Tjondronegoro, and V. Chandran, "Facial expression analysis under partial occlusion: A survey," *ACM Comput. Surv.*, vol. 51, no. 2, pp. 1–49, Mar. 2019.

[45] P. Lucey et al., "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR Workshops)*, San Francisco, CA, USA, 2010, pp. 94–101.
[46] J. Vice, M. Khan, and S. Yanushkevich, "Multimodal models for contextual affect assessment in real-time," in *Proc. IEEE 1st Int. Conf. Cognit. Mach. Intell. (CogMI)*, Los Angeles, CA, USA, Dec. 2019, pp. 87–92.
[47] N. Oliver, A. Pentland, and F. Bérard, "LAFTER: A real-time face and lips tracker with facial expression recognition," *Pattern Recognit.*, vol. 33, no. 8, pp. 1369–1382, Aug. 2000.
[48] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *J. Personality Social Psychol.*, vol. 37, no. 11, pp. 2049–2058, 1979.
[49] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expression," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 1449–1456.
[50] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affective Comput.*, vol. 8 no. 2, pp. 254–267, Apr.–Jun. 2017.
[51] X. B. Nguyen, C. N. Duong, X. Li, S. Gauch, H. S. Seo, and K. Luu, "Micron-BERT: BERT-based facial micro-expression recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Vancouver, BC, Canada, Jun. 2023, pp. 1482–1492.
[52] M. Wei, W. Zheng, Y. Zong, X. Jiang, C. Lu, and J. Liu, "A novel micro-expression recognition approach using attention-based magnification-adaptive networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Singapore, May 2022, pp. 2420–2424.
[53] R. Xu, J. Chen, J. Han, L. Tan, and L. Xu, "Towards emotion-sensitive learning cognitive state analysis of big data in education: Deep learning-based facial expression analysis using ordinal information," *Computing*, vol. 102, no. 3, pp. 765–780, Mar. 2020.
[54] R. O. Duda, P. E. Hart, and D. G. Stork, Eds., "Unsupervised learning and clustering," in *Pattern Classification*. New York, NY, USA: Wiley, 2001, ch. 10, pp. 517–601.
[55] H.-L. Nguyen, Y.-K. Woon, and W.-K. Ng, "A survey on data stream clustering and classification," *Knowl. Inf. Syst.*, vol. 45, no. 3, pp. 535–569, Dec. 2015.
[56] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Proc. Asian Conf. Comput. Vis.*, Queenstown, New Zealand, Nov. 2010, pp. 709–720.
[57] D. Wu, R. Yang, and C. Shen, "Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm," *J. Intell. Inf. Syst.*, vol. 46, no. 1, pp. 1–23, Feb. 2021.
[58] M. Kuhn and K. Johnson, Eds., "Discriminant analysis and other linear classification models," in *Applied Predictive Modelling*. New York, NY, USA: Springer-Verlag, 2013, pp. 275–326.
[59] S. Saha et al., "Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm," *Appl. Sci.*, vol. 10, no. 8, pp. 1–22, Apr. 2020.
[60] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino, and J. Torresen, "A facial expression recognition system using robust face features from depth videos and deep learning," *Comput. Elect. Eng.*, vol. 63, pp. 114–125, Oct. 2017.
[61] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affective Comput.*, vol. 13, no. 3, pp. 1195–1215, Jul.–Sep. 2022.

Jordan Vice received the B.Eng. (first class Hons) degree in mechatronic engineering from Curtin University, Perth, Australia, in 2019. He received the Ph.D. degree in mechatronic engineering from the same university, in 2022.

His research interests include applied artificial intelligence, explainable AI (XAI), machine learning, real-time, multimodal assessment of affective states, and affective computing.

Dr. Vice received the 2019 Proxima Consulting Prize for Most Outstanding Final Year Project in mechatronic engineering. His works attracted media attention in 2019, 2020, and 2022.

Masood Mehmood Khan (Member, IEEE) received the B.E. degree in mechanical, from NED University of Engineering & Technology, Karachi, Pakistan, the M.S. degree in systems engineering, from Colorado State University, Fort Collins, CO, USA, and the Ph.D. degree in computational engineering from The University of Huddersfield, Huddersfield, U.K.

He worked with the National University of Computer and Emerging Sciences Karachi, Karachi, Pakistan, Jefri Bolkiah College of Engineering, Kuala Belait, Brunei Darussalam, and American University of Sharjah, Sharjah, UAE, before joining the Faculty of Science and Engineering at Curtin University, Perth, Western Australia. His research interests include artificial intelligence and machine learning applied to affective computing, computer vision human–computer interaction, and robotics.

Dr. Khan is a fellow of the Higher Education Academy, U.K., and a fellow of the Academy of Engineering, Pakistan.

Tele Tan received the Ph.D. degree in electronics engineering from Surrey University, Surrey, U.K., in 1993.

He is currently the John Curtin Distinguished Professor in engineering with Curtin University, Perth, Australia. His research interests include computer vision and pattern recognition, which he applies to areas in mining and resources, medicine, and healthcare.

Iain Murray (Senior Member, IEEE) received the B.Eng. (Hons) degree in computer systems engineering from Curtin University, Perth, Australia, and the Ph.D. degree in computer systems engineering from Curtin University, Perth, Australia.

He has worked in the field of assistive technology for more than 25 years. Currently, he is the Curriculum Lead–Engineering with the School of Electrical Engineering, Computing & Mathematical Sciences, Curtin University, Perth, Australia. His research interests include learning environments for people with vision impairment, embedded sensors in health applications, IoT, and assistive technology. He founded the Cisco Academy for the Vision Impaired in 2002 to deliver ICT training to vision impaired people. He has supervised over 20 research students and has published in excess of 110 peer reviewed articles.

Dr. Murray is a Member of the Order of Australia, a fellow of the Australian Computer Society, and a Curtin Academy Fellow.

Svetlana Yanushkevich (Senior Member, IEEE) received the Dr.Sci. (Habilitation) degree in information sciences from the Technical University of Warsaw, Warsaw, Poland, in 1999.

She is a Professor with the Department of Electrical and Software Engineering (ESE), Schulich School of Engineering, University of Calgary, Calgary, Canada. She directs the Biometric Technologies Laboratory, University of Calgary, Calgary, Canada, the only research facility dedicated to biometric systems design at Canada. She was with the West-Pomeranian University of Technology, Szczecin, Poland, prior to joining the ESE Department, University of Calgary, in 2001. She has been contributing to the area of artificial intelligence for digital design and biometrics, since 1996. Most recently, she and her team have developed novel risk, trust, and bias assessment strategies based on machine reasoning, with applications in biometric-enabled border control, forensics, and healthcare.