A Hierarchical Separation and Classification Network for Dynamic Micro-Expression Classification

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

***Abstract*—Models of seven discrete expressions developed using macro-level facial muscle variations would suffice identifying macro-level expressions of affective states. These models won't discretise continuous and dynamic micro-level variations in facial expressions. We present a Hierarchical Separation and Classification Network (HSCN) for discovering dynamic, continuous and macro- and micro-level 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 optimised for discovering the macro-level 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 micro-expressions. Invoking the random forest algorithm would classify twenty-one macro-level 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 micro-level facial muscle variations and their utilization in dynamic assessment of facial expressions of affective states.**

***Index Terms*—Cosine similarity-based separation, Hierarchical classification, Micro-expression detection, Affective state assessment, Facial expression classification, Rule-based systems.**

I. Introduction

FACIAL expressions convey internal thoughts, feelings, 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]–[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]–[8].

This work was funded by the Faculty of Science and Engineering, Curtin University Australia.

Manuscript received January 2023.

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 patho-psychological 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 micro-level facial expressions should also be classifiable within a continuous prevailing space. This appears logical as micro-level facial expressions represent transient macro-level expressions such as perceived expressions of anger, joy or fear [10].

Affective computing literature cites a diverse range of related works on facial expression recognition and affective state assessment [11]–[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 macro-level differences in facial muscle movement patterns [15]–[17]. However, these seven discrete models cannot represent micro-level facial expressions.

Micro-level variations in facial muscles can be seen as representing delicate, involuntary and spontaneous changes in facial expressions and they often show true emotive experiences. Relevant works on recognition of micro-level facial expressions and review of the employed approaches [18] are suggestive of the fact that transient nature and low intensity of micro-level expressions won't let easily capture them 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. Micro facial expressions were recognized in [22] using the accordion spatiotemporal data classified by the Random Forest algorithm. Similarly, in [7], an algorithm ensemble that exploited the handcrafted features and deep features was used for identifying faces and classifying micro facial expressions. 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. Micro facial expressions were recognized in [22] using the accordion spatiotemporal

data classified by the Random Forest algorithm. Similarly, in [7], an algorithm ensemble that exploited the handcrafted features and deep features was used for identifying faces and classifying micro facial expressions.

Assessment of human facial expressions is typically carried out using easy to discriminate spatiotemporal facial features. For example, in [22], the hybrid deep learning approach was employed such that first a spatial convolutional neural network (CNN) was used in order to process the facial image frames. This was followed by use of a temporal CNN to analyse the optical flow in images for simultaneous separating and learning of significant spatial and temporal features. These significant spatial and temporal features are then used in a Deep Belief Network (DBN) model that would perform average pooling in order to obtain the fixed-length global feature representation. In a final step, a Support Vector Machine (SVM) would perform classification [22]. The complexity and computational costs required for continuous dynamic assessment of affective states in [22] was also observed in several other works [23]–[25]. Despite the recent advances, the idea of extracting micro expressions 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 macro-level facial expressions. The hierarchical separation and classification network (HSCN) used in this work consists of three subsystems viz., (i) a module for macro-level affective state assessment using whole face data, (ii) a module for upper-facial region micro-expression classification and (iii) a module for lower-facial region micro-expression 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. The following Linear Discriminant Analysis (LDA) would further separate and cluster facial expressions by identifying discriminant features and discovering multiple hyperplanes within the linear discriminant space [2], [26]. Compared with other statistical and neural classification techniques, LDA is considered computationally efficient as it reduces the data dimensionality and needs smaller amounts of training data [13], [26], [27].

After invoking the LDA, our HSCN framework engages a domain-specific and innovative Rule-Based Expert System (RBES) that would assist in classifying micro facial expressions evident on upper and lower regions of the face. The RBES exploits 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 in [30], deployed a Belief Rule-Based Expert System that exploited the outputs of a convolutional neural network (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

Moving forward, this work contributes the following to the existing literature:

1) We present a novel framework and an ensemble of classifiers enabling hierarchical separation and classification of macro- and micro-level 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 macro-level expressions;
5) We detail implementation of a RBMS that uses the dynamic linear discriminant features for realizing continuous micro-expressions on upper- and lower-facial regions and;
6) Results of applying a new and systematic methodology for extracting micro-level facial expressions of affective states.

In relation to the contribution stated first in the above paragraphs, 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 step-wise refinement of separable features. This work extends application of ensemble approaches by using an ensemble for micro-expression classification.

This manuscript in Section I introduces our work. Section II provides important information and sets ground for this work. Section III discusses previous works and current trends in this domain. Section IV details how unsupervised clustering and labelling was performed. Discussions on cosine similarity estimation, data separation, macro- and micro-level linear discriminant analysis and construction of the RBMS are also included in Section IV. The following Section V presents results pertaining to the aforementioned analyses and details the system validation. Section V also explains our novel RBMS. Section VI concludes this work and suggest future research directions. Section VII inform readers on this project's funding source and ethical 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, [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] now has an enhanced version [35] that exploits the continuous and dynamic 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 categorises 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 macro-level expressions [10]. In a somewhat similar manner, macro-level 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, haemodynamic 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 EMFACS were discussed in [28] and their application details were presented in [16], [40]–[42]. These works categorise facial muscles movements through coding and action units and work as tools for the facial expression assessment and recognition. The EMFACS action units allow modelling 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 micro-expression classification is based on the muscle movements defined in the EMFACS. 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 micro-expression 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. Visualising 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 modelling 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 modelling the transient expressions and expression-intensity variations at the macro-level. It also demonstrates the transience of expressions by modelling continuous micro-level muscle movements in upper and lower facial regions and allows for classification of various micro-expressions. Furthermore, the HSCN builds upon the prevailing dynamic facial expression recognition systems [47] and proposes an alternate approach for continuous macro- and micro-expression 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 micro-expressions 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 micro-expression classifiers.

TABLE I
AFFECTIVE STATE MODELS AVAILABLE IN THE CK+ DATASET. THE FACIAL MUSCLE ACTIONS FOR EACH MODEL WERE REALIZED THROUGH FACS AND EMFACS INVESTIGATIONS.

Affective State	Identified Action Units	Resulting Physical Actions	Engaged Musculature
Happy State	12	Raised Lip corners	Zygomaticus major
	6, 7	Lower eye-lids raised	Orbicularis Oculi muscle
	26, 27	Open mouth	Orbicularis Oris
Expression of Surprise	1	Eyebrows raised	Medial Frontalis
	5	Upper eye-lids raised	Levator palpebrae superioris muscle
	26, 27	Mouth opened	Orbicularis Oris
Expressed Anger	4	Eyebrow Frown	Corrugator Supercilii
	5	Upper eye-lids raised	Levator palpebrae superioris muscle
	6, 7	Raised lower eye-lids	Orbicularis Oculi muscle
	23	Lip Tightener	Orbicularis Oris muscle
Expression of Disgust	4	Eyebrow Frown	Corrugator Supercilii associated muscles
	6, 7	Lower eye-lids raised	Orbicularis Oculi muscle
	9, 10	Upper lip raised	Levator labii superioris muscle
Expression of Fear	4	Eyebrow Frown	Corrugator Supercilii muscle
	1	Eyebrows raised	Medial Frontalis muscle
	5	Upper eye-lids raised	Levator palpebrae superioris muscle
	26, 27	Mouth opened	Orbicularis Oris
Expression of Sadness	4	Eyebrow Frown	Corrugator Supercilii muscle
	1	Eyebrows raised	Medial Frontalis muscle
	15	Lowered lip corners	Depressor Anguli Oris muscle
Expression of Contempt	L12 or R12	Slightly raised lip corner (*asymmetrical*)	Zygomaticus major muscle
	L14 or R14	Dimpler (*asymmetrical*)	Buccinator muscle

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 support vector machines, multiple kernel learning and random forests. Xu et al., [50] proposed a "Facial Dynamics Map" which characterises micro-expression related movements using granular pixel features along with an algorithmic approach that was based on optical flow estimation. Xu et al., [50] employed a support vector machine classifier to identify and categorise different types of facial micro-expressions. Polikovsky et al., [40] used the EMFACS (Emotion Facial Action Coding System) for micro-expression detection such that their method divided full facial images into smaller facial regions based on action unit locations. A histogram of oriented gradients (HOG) approach was combined with the K-nearest neighbour classifier for detecting micro-expression and action unit activations.

Instead of exploiting the visual cues, [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 sub-regions 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 are hampered by the cost and accessibility of thermal cameras.

In [51], an extension of the popular Bidirectional Transformer (BERT) model called Micron-BERT was deployed in order to exploit attention maps for perceiving differences between two frames. Authors performed experiments using three, four and seven facial expression classes [51] for inferring a solution. Moving away from the largely used seven models of facial expressions, this work uses HSCN for modelling twenty-one micro-expressions of affective states realized around both, the upper and lower-facial regions. In a way, we built upon the work in [52], that employed an attention-based magnification and adaption network to focus on the magnification levels of different micro-expressions. The network in [52] leverages a pre-trained 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 summarise: (i) feature space, (ii) deployed architecture, (iii) micro-expression classification, (iv) macro-level facial expression classification, (v) number of classes modelled, (vi) accuracy/performance metric. Through these works, we observed that feature maps and attention mechanisms were common in micro-expression recognition literature (including thermal representations). In these works, the authors also stressed on the fleeting nature of micro-expressions and the difficulty in modelling real-time changes in micro-expressions. We therefore propose the HSCN framework as a vehicle to model activations in time and use upper and lower-facial region micro-expressions to explain the detected macro-level expressions.

IV. METHODS

Unsupervised learning approaches help in discovering effective states by separating and labelling patterns within a collection of unlabelled data. Therefore, unsupervised learning

TABLE II
A COMPARISON OF WORKS DISCUSSED IN SECTION III ON THE BASIS OF FEATURE SPACE CONSTRUCTION, DEPLOYED MODEL AND CLASSIFICATION PERFORMANCE USING UPPER, LOWER AND FULL-FACE DATA.

Author	**Features**	**Deployed Model**	**Upper-Face**	**Lower-Face**	**Full-Face**	**No. Classes**	**Classifcation Accuracy**
Vice [46]	Image space	InceptionV3, Xception	X	X	✓	7	75.88-92.78%
Oliver [47]	2D ‘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]	3D 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%

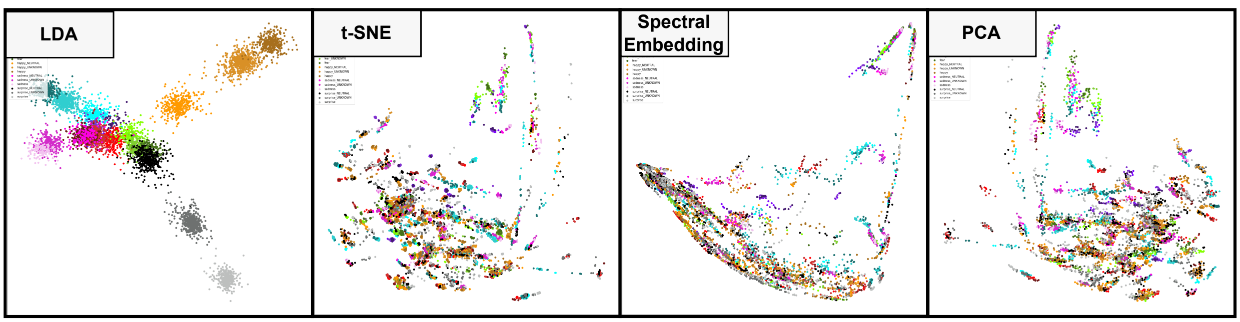

Fig. 1. A comparison of different projection /embedding techniques that were explored during HSCN’s early design stages. Included techniques were: (i) linear discriminant analysis, (ii) t-distribution-based stochastic neighbour embedding, (iii) spectral embedding and, (iv) principal component analysis. Through these projections, use of LDA was justified as the basis for the HSCN.

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 non-linear clustering techniques [54]. These methods rely on statistical foundations. They detect similarities within a set of unlabelled 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 macro-level 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 several projection techniques including: t-distribution-based stochastic neighbour embedding (t-SNE), Principal Component Analysis (PCA) and Spectral Embedding. We found that LDA resulted in the most optimal intra- and inter-cluster variance in two dimensions while using 100×100 pixel images. Results of these experiments are shown in Fig. 1. The shown figures revealed that LDA would produce well-defined clusters for revealing cluster-to-cluster relationships and building a logical rule-base for an expert system. This influenced our decision to opt for LDA.

The HSCN combines two techniques as shown in Fig. 2 and Fig. 3. The initial unsupervised clustering and labelling 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 multi-class facial expression classification.

The LDA transform has been previously used for maximising 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 maximises the *inter*-cluster variance and minimises the *intra*-cluster variance. This method maximises the separation between cluster centroids while minimising the

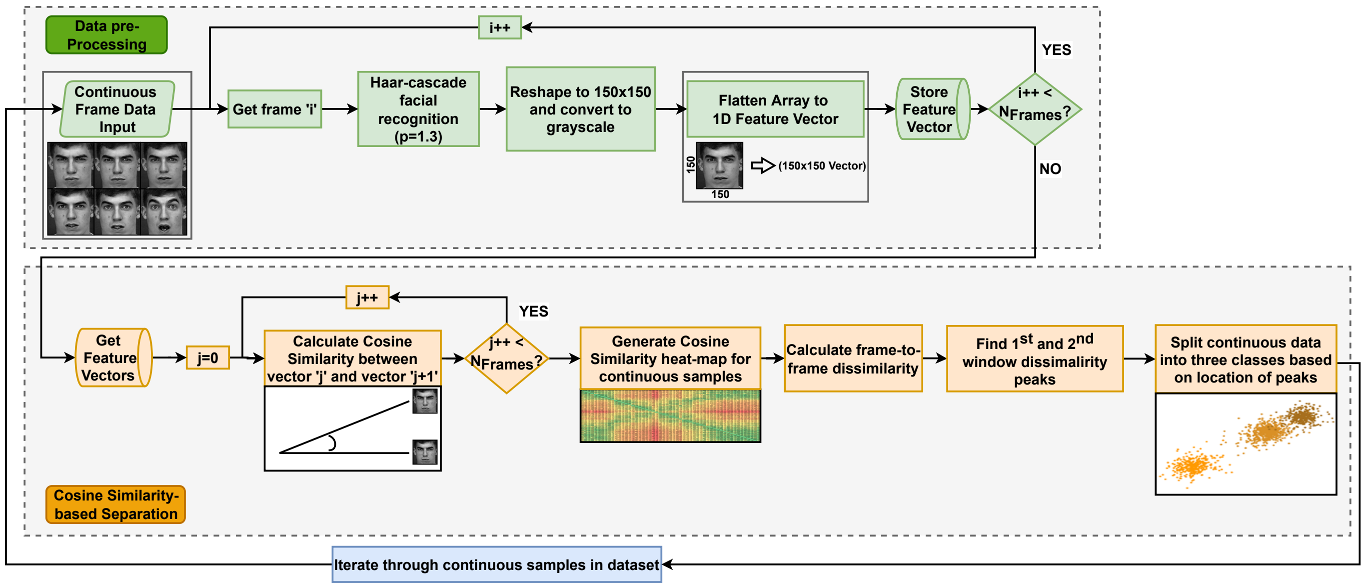

Fig. 2. Pre-processing of continuous expression data and their unsupervised cosine-similarity based separation. Note how continuous frame data from the CK+ dataset are input and pre-processed prior to undergoing the cosine similarity-based separation and clustering in the HSCN.

separation between samples that belong to the same class [58].

Analysing the separated clusters enables modelling the state-to-state transitions and classification of macro-level affective states in a continuous domain. Clusters formed at the macro-level via LDA provide foundations for defining the micro-level clusters of lower and upper facial regions' data. As visualised 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 LDA were regarded as a set of logically apt optimisation steps for this work.

Solution to the optimisation problem was found in determining the linear discriminants, which corresponded to the largest eigenvalues of $\mathbf{W^{-1}B}$, noting that the number of linear discriminants required to solve an LDA problem depends on the number of labelled 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:

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

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

As shown in Fig. 2 and Fig. 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, maximising separation between cluster centroids to create the *macro*-level 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 *micro*-expressions.

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 below for explaining the unsupervised separation and clustering methodology:

- $\mathbf{x_i} = \{x_1, x_2, ..., x_m\}$ defines a pattern or feature vector i.e., a flattened facial expression image containing 'm' raw pixels/features.
- $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$ 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, that have been projected onto an $m-$dimensional hyperplane.
- $\mathbf{C} = \{c_1, c_2, ..., c_k\}$ 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 macro-level 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 categorising 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, macro-level 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

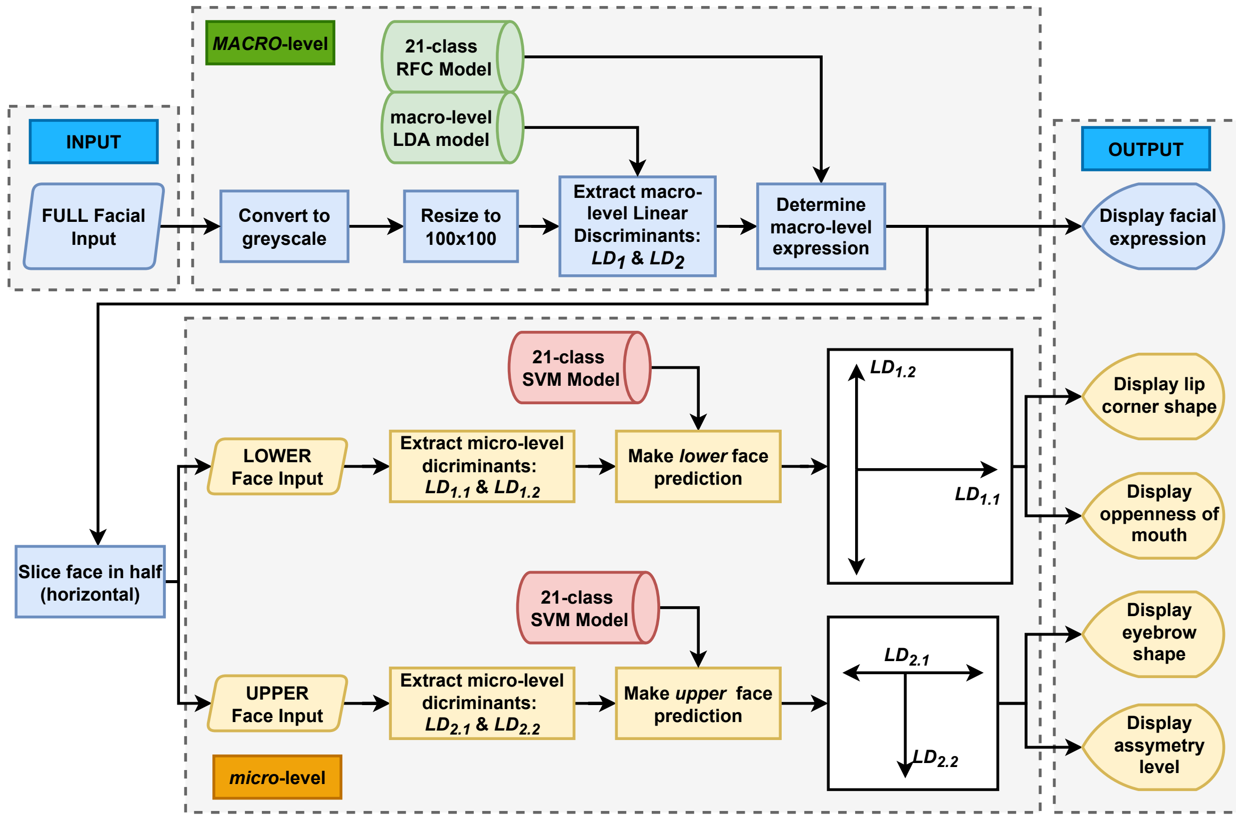

Fig. 3. Visual summary of the proposed HSCN containing a rule-based expert system. The flowchart shows how facial image data are processed from input to macro- and micro-level assessment stages, and how classification results are reported and displayed to the users. Combined with Fig. 2, one can see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

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:

$$S_{\cos\theta} = \frac{\mathbf{x_i} \cdot \mathbf{x_j}}{|\mathbf{x_i}|\,|\mathbf{x_j}|} \tag{2}$$

In the CK+ dataset, a large *inter*-cluster 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, macro-level facial expression clusters based on similar features, labelled as follows:

- Cluster 1: **Neutral**-dominated state
- Cluster 2: **Partially** activated state
- 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 modelling the continuous nature of affective states, this allowed for classification of twenty-one, transient macro-level 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: $N_{states} = N_{peaks} + 1$. Furthermore, this method is not limited to the image domain and could be deployed for the separation of affective speech and video data.

B. Macro-level Linear Discriminant Analysis

The initial clustering via the cosine similarity-based separation method were input into the second tier of the HSCN, where macro-level LDA clustering was performed. Please note that linear discriminant analysis has been extensively used to effectively separate and cluster labelled facial expressions

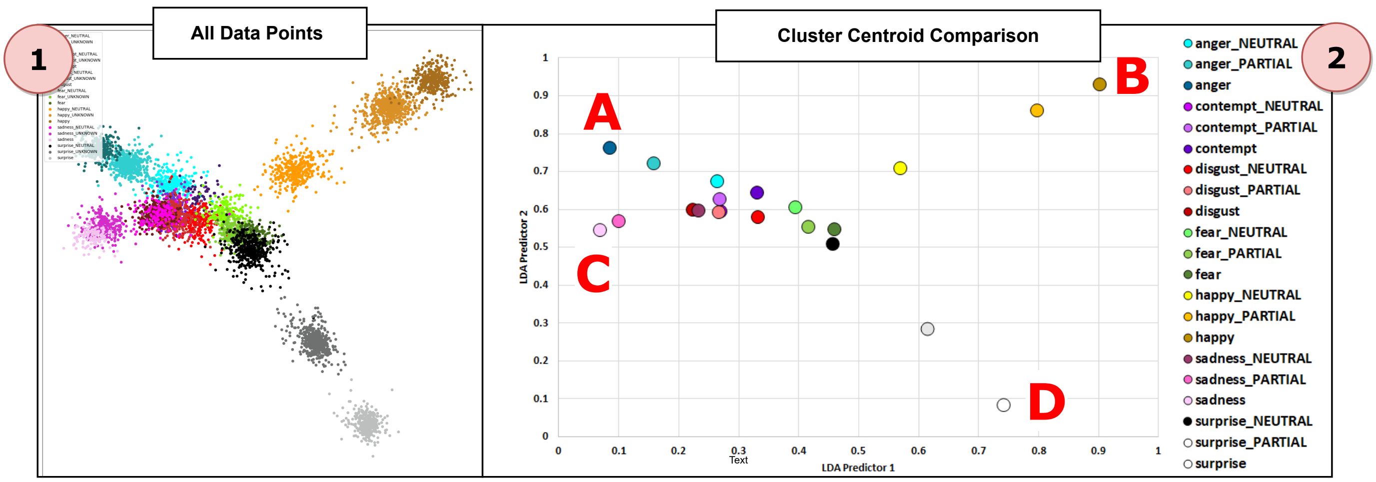

Fig. 4. Macro-level LDA clustering results. **Subplot 1** on the top left is the input from the initial, cosine similarity-based separation algorithm displaying all data samples extracted from the CK+ dataset. **Subplot 2** on the top right are the cluster centroids for each of the macro-level expressions 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.

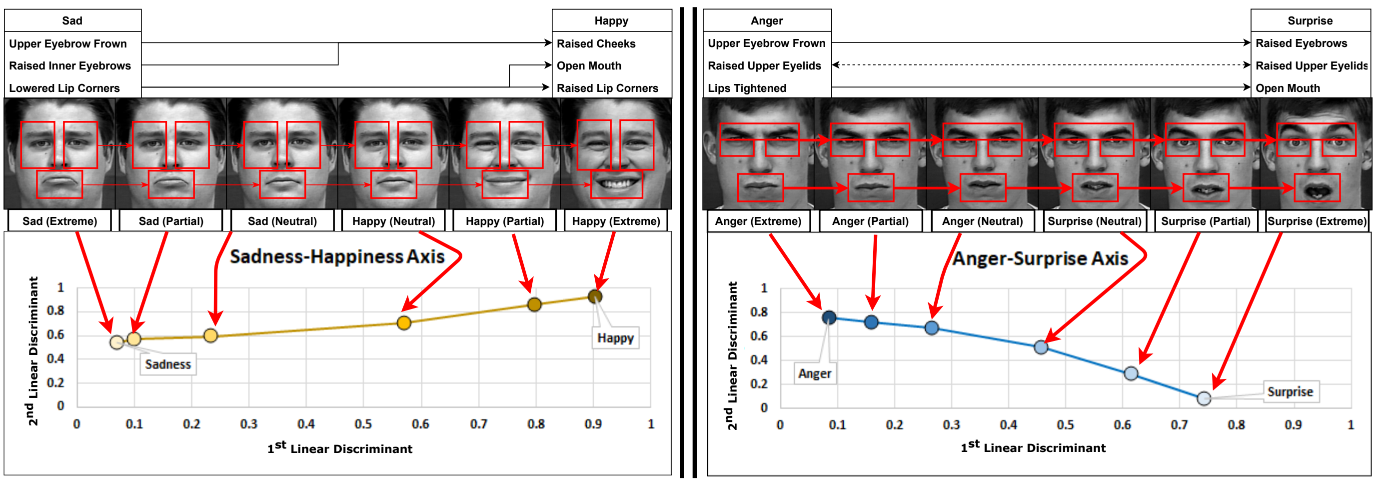

Fig. 5. **LEFT**: State-to-state transition along the **sadness-happiness** trend as visualised in Fig. 4. Examples of neutral and partial states are also shown. **RIGHT**: State-to-state transition along the **anger-surprise** trend as visualised in Fig. 4. Examples of neutral and partial states are also shown.

by discovering hyperplanes within a linear discriminant space [60]. The clustering was achieved by maximising inter-cluster variance in order to optimise cluster centroid separation. Fig. 4 highlights results of the macro-level LDA clustering algorithm when applied to a large volume of continuous facial expression data. Analysing the cluster centroids in subplot 2 of Fig. 4, we see emergence 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:

- Sadness (**C**) to happiness (**B**),
- Anger (**A**) to surprise (**D**).

We needed to understand "the theoretical underpinning of these axes?". The linear discriminant space visualised 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 modelled.

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 micro-expressions. By comparing sadness and happiness muscle activations in Table I, we were able to model the state-to-state transition and visualise 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:

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 Macro-level Facial Expression Clusters

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: (i) parallel relaxation of brows and raising of cheeks, (ii) 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 visualised 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:

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' activation. Therefore, the state-to-state transition could model the following transformations: (i) Eyebrows raise from an initial frowned/depressed position, (ii) Mouth opens from an initial tightened expression.

Expanding on the two rules formed thus far, we could postulate an initial hypothesis for explaining what axes X and Y (linear discriminants 1 and 2) would represent. Let the discriminant 'n' be LiD_n, the $|\Delta LiD_n|$ values represent states whose points $A \rightarrow D$ in subplot 2 of Fig. 4 are reported in Table III. Together with Fig. 6, these results provide basis for validating the rule-based micro-expression classifier as previously done in [61]. Given the evidence provided, we could define the following hypotheses and macro-expression rules:

1) The discriminant LiD_1 relates to the openness of the mouth pertaining to the lower region of the face given the following evidence:
   - Sadness and Anger shared a low ΔLiD_1. The two common actions between the states were: "upper eyebrow frown", "lips tightened/lowered corners".
   - Presence of two common actions would be troublesome if not for the presence of the surprise and happiness states, which also shared a low ΔLiD_1. The common action between surprise and happiness revolved around raised lip corners and ultimately, the open mouth.
2) Discriminant LiD_2 relates to the region around the eyes i.e., the eyelids and eyebrows – the upper facial region, evidenced by:
   - The Anger-Sadness transition evidences both a low ΔLiD_1 and ΔLiD_2. If the initial hypothesis is that LiD_1 is related to the mouth, then the second common action – "upper eyebrow frown" may be related to LiD_2, which supports the upper facial region relationship.
   - Analysing 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.
   - Large variance between happiness and surprise. Given that the open mouth was deduced as being referred to by LiD_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 LiD_2 relating to the upper facial region.

By inferring the above rules for the macro-level 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 macro-level facial expression classification.

C. Micro-level Linear Discriminant Analysis

We had previously defined the following rules:

1) LiD_1 relates to the shape of the mouth and the lower facial region.
2) LiD_2 relates to the region around the eyes, eyelids, and brows – the upper facial region.

These claims were substantiated through the necessary experiments. The micro-expression 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

TABLE III
AFFECTIVE STATE CHANGE-CAUSED VARIATIONS IN MAGNITUDES OF n^{th} LINEAR DISCRIMINANTS. THE DATA HIGHLIGHTED IN BOLD SHOW HIGH AND LOW $|\Delta LiD_n|$ VALUES.

$P_X \rightarrow P_Y$	State $\rightarrow$ State	$\lvert\Delta LiD_1\rvert$	$\lvert\Delta LiD_2\rvert$
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

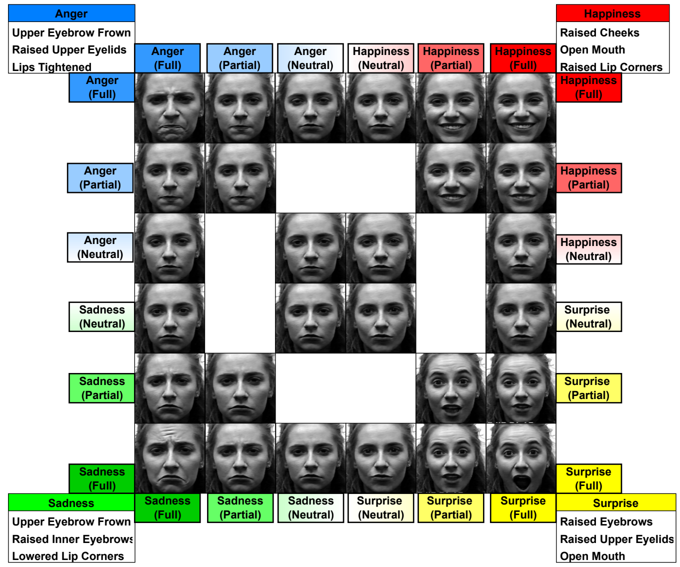

Fig. 6. Visualised state-to-state transition for anger, happiness, sadness, and surprise expressions. This figure assists in mapping LiD_n features to real-world features.

allowing for the classification of micro-expressions in upper and lower facial regions. If the initial hypotheses were correct, then there would be a very discernible trend between states at the micro-level as this would indicate that the projected feature 'LiD_n' is related to a particular group of muscles.

Let us describe the *macro*-level linear discriminant features as 'LiD_n' – i.e., LiD_1 = the lower facial region and LiD_2 = upper facial region. Moving to the *micro*-level, let 'm' describe the micro-level features contained within the higher, 'n^{th}' level regions i.e., '$LiD_{n.m}$'. For example, $LiD_{1.1}$ and $LiD_{1.2}$ describe micro-expressions in the lower facial region.

The set of multiple clusters shown in Fig. 7A shares a similar $LiD_{2.2}$ value (see Table IV), with the largest variance observed along the direction of the discriminant $LiD_{2.1}$ axis. We could also see that the clusters move linearly from expression of anger to the expression of surprise along the discriminant $LiD_{2.1}$ (Table IV). Expressions of fear and contempt display larger variances along the discriminant $LiD_{2.2}$ (Table IV) and have coordinates of cluster centroids in a close proximity. We also noted that the expression of contempt is an asymmetrical expression and so it was an outlier. Besides these two affective states, a majority of facial expressions resided along the $LiD_{2.1}$ axis (Table IV).

The lower facial region (LiD_1) is more sparsely clustered compared to the upper facial region. In Fig. 7B, it is evident that most states reside on one side of the spectrum, sharing a similar $LiD_{1.1}$ feature value (see Table IV), with happiness and its sub-states displaying the largest variance in $LiD_{1.1}$ (Table IV). The notable trend observed in the lower region of the face could be attributed to the micro-level feature $LiD_{1.2}$ (as obvious in Table IV); the y-axis shows disgust (top) and surprise (bottom) as two extreme values.

V. RESULTS

A. Defining the Rule-Based Expert System

This work attempts to highlight how similarity and dissimilarity measures could help in improving the classifier performance and decision-making capabilities of a system. We used Euclidean distance-based similarity to validate the inferences of our rule-based system. As previously defined, the macro-level linear discriminant features LiD_n and the micro-level features as $LiD_{1.1}$, $LiD_{n.m}$, combined with Fig. 5 and the given Table III validate our macro-level facial expression method. We observed that states of Anger, Happiness, Sadness and Surprise exist as extreme points in a 2D linear discriminant space and by referring back to the EMFACS muscle movements discussed previously, we could establish that the discriminant LiD_1 relates to the shape of the mouth (the lower facial region) and LiD_2 relates to the region around eyes, eyelids and brows (the upper facial region).

Deriving relationships between muscle movements on the entire face i.e., the two facial regions, allowed to further explore the upper and lower facial regions and conduct similar experiments. Performing the same analysis on Table IV, we can establish relationships between features in the upper and lower facial regions using the EMFACS-explained action units. The neutral, partial and fully-activated micro-expression cluster centroids reported in Table IV supplement our findings visualised in Fig. 7.

The **Anger-Surprise** axis Rule was modelled using two transformations: (i) Eyebrows raise from an initially frowned/depressed position, (ii) Mouth opens from an initially closed position. Combined with the hypothesis, "LiD_2 relates to the region around eyes, eyelids, and brows referring to the upper facial region". Thus, we could assume that the micro-level feature $LiD_{2.1}$ would refer to the translation of the eyebrows from an initially frowned/down-turned position to finally a raised position, using the musculatures involving *medial frontalis*, *levator palpebrae superioris* and *corrugator supercilii* muscles.

Fig. 7B 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 $LiD_{1.1}$ feature. The macro-level, **Sadness-Happiness Axis Rule** modelled the following two transformations: (i) parallel relaxation of brows and raising of cheeks, (ii) 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 micro-level feature $LiD_{1.1}$ and the real world. Analysing the variations in

$LiD_{1.2}$, we see disgust and surprise on opposite sides of the spectrum. The immediate hypothesis is that $LiD_{1.2}$ models the openness of the mouth, and the manipulation of the central lip muscles - *orbicularis oris* and *levator labii superioris*.

The evidence gathered via the micro-level expression (LDA-derived) clusters establishes validity of our macro-level expression related inferences. Moving into the micro-expression (at m^{th} level), we could define each axis as:

- $\mathbf{LiD_{1.1}}$ - lower face: Muscles around lip corners - provide a model for translation from a down-turned shape to a lifted shape.
- $\mathbf{LiD_{1.2}}$ - lower face: Muscles related to opening of the mouth - provide a model for the manipulation of the central lip muscles from closed to open.
- $\mathbf{LiD_{2.1}}$ - upper face: Models a translation of the eyebrows from an initial frowned position to a raised position.
- $\mathbf{LiD_{2.2}}$ - upper face: Models expressions of fear and contempt and explains asymmetrical movements outside the spectrum - from anger to surprise.

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

Deploying the HSCN in a rule-based expert system would allow for the continuous monitoring and assessment of macro-level expressions of affective states and micro-expressions, which allow for the modelling of specific muscle movements in the upper and lower facial regions. In the previous sections the macro- and micro-expressions were modelled using linear discriminant features defined by n^{th} and m^{th} level subscript notation i.e., $LiD_{n.m}$. Fig. 3 visualises the rule-based expert system and shows how facial image data is 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 macro-level facial expressions as well as detection of upper- and lower-facial region micro-expressions as each of the three classifiers models some unique states. Using the micro-level rules defined in the previous section, the HSCN is able to detect changes in: (i) lip corner muscles, (ii) openness of the moth, (iii) translation of the eyebrows and, (iv) level of asymmetry.

B. Hierarchical Classifier Performance

Both macro- and micro-level algorithms developed for this work were validated using the Random Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbour (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 100x100 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:

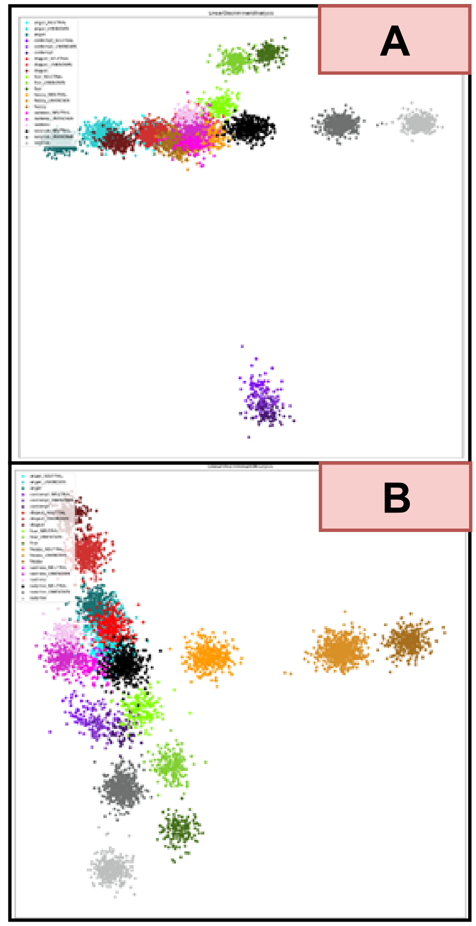

Fig. 7. **(A)** Two-dimensional Linear discriminant space representation of the upper facial region, showing $\{LiD_{2.1}, LiD_{2.2}\}$ micro-expression features. **(B)** Two-dimensional Linear discriminant space representation of the lower facial region, showing $\{LiD_{1.1}, LiD_{1.2}\}$ micro-expression features.

$$Acc_{val}(\%) = \frac{N_{(correctguesses)}}{N_{(totalguesses)}} \times 100\% \quad (4)$$

Classification performance of all models are reported in Table V. We used the Random Forest classifier for macro-expression classification, predicting twenty-one transient affective state expressions, across seven independent state axes with 76.11% accuracy. For the upper and lower facial micro-expression classifiers, we deployed SVMs, capable of predicting twenty-one variations of upper and lower facial muscle movements based on m^{th} level '$LiD_{n.m}$' 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 micro-expression 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

TABLE IV
COORDINATES OF UPPER- AND LOWER-FACE CLUSTER CENTROIDS IN OUR MICRO-EXPRESSION, LINEAR DISCRIMINANT SPACE. SUBSCRIPTS N, P AND A RESPECTIVELY REFER TO NEUTRAL-DOMINATED, PARTIAL AND FULLY ACTIVATED STATE LABELS.

	Lower facial region		**Upper facial region**	
Cluster	$LiD_{1.1}$	$LiD_{1.2}$	$LiD_{2.1}$	$LiD_{2.2}$
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

TABLE V
THE OVERALL CLASSIFICATION PERFORMANCE OF HSCN WHEN TRAINED AND VALIDATED USING THE CK+ DATASET. THE 80/20 (TRAIN/TEST) SPLIT WAS INVOKED ON TWENTY-ONE EXPRESSION CLASSES PER FACIAL REGION.

Region of the face	Model used	Validation Accuracy (%)
Entire-face (macro)	**RF**	**76.11**
Entire-face (macro)	SVM	72.95
Entire-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

states. When we compare our results with the works discussed in section III, we find that their range was between 32.94-95.95% 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 micro-expression 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. 7A and Fig. 7B, one would see why this might be the case. The lower facial region micro-expression clusters show a larger separation across $LiD_{1.1}$ and $LiD_{1.2}$ axes compared to the upper facial region micro-expressions. This shows variations along the $LiD_{2.1}$ 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 novel and new information on the respective roles of the facial musculature specific to the upper and lower facial regions in exhibiting and recognizing facial expressions of affective states. We were also able to explain how micro- and macro-level 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 categorising 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 micro-expression classes. The twenty-one affective state classification is performed using a rule-based expert system and a convincing level of accuracy was achieved during the training and testing stages. The reported HSCN shows competitive performance compared to recent state-of-the-art systems reported earlier in Table II.

The ability to detect and classify micro-expressions 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 macro-level facial muscle movement modelling approach would be useful in complex affective state assessment situations. The presented HSCN approach was able to predict twenty-one macro-level transient expressions vis-à-vis twenty-one upper and lower facial region micro-expressions. 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 (i) unsupervised cosine similarity-based separation, (ii) LDA-based clustering and, (iii) 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 modelling changes in affective *speech* expressions as continuous and time-dependent functions to be used for affective state assessment.

Acknowledgments

S. Yanushkevich acknowledges the partial support of the Natural Sciences and Engineering Research Council of Canada, through the Discovery Grant "Biometric intelligent interfaces".

VII. Compliance with Ethical Standards

Curtin University's ethics approval No. HRE2019-0722 dated 23.10.2019 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 Curtin University. Dr Masood M Khan and Jordan Vice received funding (APA18336230) from Curtin University to carry out this work. All authors hereby declare that they have no conflict of interest.

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," *Knowledge-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. of 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," *Scientific Reports*, vol. 12, no. 1, pp. 1-16, Jul. 2022.

[4] M. M. Khan, "Cluster analytic detection of disgust-arousal," in *2009 9th Int. Conf. on Intell. Syst. Design and 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 University 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. of 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. of Personality and 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. on Ind. Inform.*. vol 18, no. 12, pp. 9112-9121, Jul.2022.

[11] B. Fasel and J. Luettin, 2003, "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. on Pattern Anal. and Mach. Intell.*," vol. 22, no. 12, pp.1424-1445, Dec. 2000.

[14] P. Ekman, "An argument for basic emotions," in *Cognition and Emotion*, vol. 6, no. 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. on Pattern Anal. and Mach. Intell.*. vol. 38, no. 8, pp. 1548-1568, Jan. 2016..

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

[17] G. R. Alexandre and J. M. Soares, "Thé GAP. 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," *Frontiers in Psychol.*, vol. 9, no. 1128, pp. 1-21, Jul. 2018.

[19] Lei, L., Li, J., Chen, T. and Li, S., "A novel graph-tcn with a graph structured representation for micro-expression recognition" *Proc. of the 28th ACM International Conference on Multimedia*, pp. 2237-2245, 2020.

[20] X. Ben, Y. Ren, J. Zhang, S. J. Wang, K. Kpalma, W. Meng and Y. J. Liu, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 44, no. 9, pp. 5826-5846, Mar. 2021

[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. of 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. on Affective Comput.*, vol. 3, no. 2, pp. 132-144, Jul. 2011.

[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 and Vision Comput.*, vol. 31, no. 2, pp. 137-152, Feb. 2013.

[25] Y. Tie and L. Guan, L., "A deformable 3-D facial expression model for dynamic human emotional state recognition.," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 23, no. 1, pp. 142-157, Jun. 2012.

[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. and 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 University Press, 2020.

[29] M. Pantic and L. J. Rothkrantz, "Self-adaptive expert system for facial expression analysis," in *Proc. of 2000 IEEE Int. Conf. on Systems, Man and 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 *2020 Joint 9th Int. Conf. on Inform., Electronics & Vision (ICIEV) and 4th Int. Conf. on 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. on Cybernet.*, vol. 45, no. 2, pp. 177-190, May 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. and Appl.*, pp.1-14, Mar. 2022.

[33] F. Burkhardt, T. Polzehl, J. Stegmann, F. Metze and R. Huber, "Detecting real life anger," in *2009 IEEE Int. Conf. on Acoust., Speech and 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, 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, Oct. 2020.

[36] J. Russell and A. Mehrabian, "Evidence for a Three-Factor Theory of Emotions," in *J. of Res. in 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 Press, 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. on Affective Comput.*, vol. 10, no. 1, pp. 18-31, Aug. 2017.

[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. on Inf. and 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. on Affective Comput.*, vol. 8, no. 3, pp. 412-425, Jul. 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. on Autonomous and 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. on Image Process.*, vol. 29, pp. 4057-4069, Jan. 2020.

[44] L. Zhang, B. Verma, D. Tjondronegoro and V. Chandran, "Facial expression analysis under partial occlusion: A survey," *ACM Comput. Surveys (CSUR)*, 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 *2010 IEEE Comput. Soc. Conf. on Comput. Vis. and 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 *2019 IEEE 1st Int. Conf. on Cogn. Mach. Intell. (CogMI)*, Los Angeles, CA, USA, Dec. 2019, pp. 87-92.

[47] N. Oliver, A. Pentland and F. Bérard F, "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. of Personality and 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 *Int. Conf. on 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. on Affective Comput.*, vol. 8 no. 2, pp. 254-267, 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. of the IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, Vancouver, 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 *IEEE Int. Conf. on Acoust., Speech and 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: John Wiley & Sons, 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," *Knowledge and 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 *Asian Conf. on Comput. Vis.*, Queenstown, New Zealand, Nov. 2010 pp. 709-720.

[57] D. Wu, R. Yang R and C. Shen, "Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm," *J. of 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, 2013, pp. 275-326.

[59] S. Saha et al. "Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm," *Applied Sciences*, 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," *Computers & Elect. Eng.*, vol. 63, pp. 114-125, Oct. 2017.

[61] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. on Affective Comput.*, vol. 13, no. 3, pp. 1195-1215, Mar. 2020.

JORDAN VICE is a post-doctoral researcher at the University of Western Australia. He completed his PhD in Mechatronic Engineering in 2023 and his Bachelor of Engineering (First Class Honours) in 2019. He was placed on the Vice Chancellor's list in recognition of his postgraduate research.

MASOOD MEHMOOD KHAN (Member, IEEE) received Bachelor degrees is Industrial Technology and Mechanical Engineering, MS in Systems Engineering and Ph.D. in Engineering. Currently, Masood is working in the field of Artificial Intelligence. He applies his skills to design and implement affective computing systems, human-computer interaction mechanisms and robotic systems. He is a fellow of the Higher Education Academy (UK) and a fellow of the Academy of Engineering (Pakistan).

TELE TAN Tele Tan earned his Ph.D. in electronics engineering from the University of Surrey, U.K., in 1993. Presently, he holds the position of John Curtin Distinguished Professor in Engineering at Curtin University. His research focuses on signal processing, pattern recognition, machine learning and computer vision, and he applies his expertise to various domains, including mining and resources, as well as health and medicine.

SVETLANA YANUSHKEVICH (Senior Member, IEEE) directs the Biometric Technologies Laboratory in the Schulich School of Engineering (SSE) at the University of Calgary (UofC). She holds a Dr Habilitated in Tech Sci from the Technical University of Warsaw. She joined the UofC in 2001, and she is currently a Full professor in the Department of Electrical and Software Engineering and an Associate Dean of Research, Innovation and Strategic Partnership at the SSE. Dr. Yanushkevich's team is developing novel decision support and risk assessment strategies based on machine reasoning, with applications to biometric-enabled access control (risk assessment of misidentification) and healthcare monitoring systems (gait abnormalities detection, action/activity recognition, face attribute analysis). Dr. Yanushkevich chaired the Biometric Taskforce in the IEEE Computational Intelligence Society in 2017-2020, and currently is a vice-chair of this Task Force, and a member of the Intelligent Systems Applications Technical Committee.

IAIN MURRAY is with the School of Electrical Engineering, Computing & Mathematical Sciences at Curtin University. He earned B.Eng degree with Honours and PhD in Engineering. He has been developing assistive technology solutions for more than 25 years and has published more than 110 peer reviewed articles. He is a Member of the Order of Australia, a Fellow of the Australian Computer Society and a Curtin Academy Fellow.