Enhancing Facial Expression and Micro-Expression Recognition Through Hybrid Dimensionality Reduction and Machine Learning Classification

Viola Bakiasi (Shtino)

Phd, Faculty of Information Technology, Computer Science Department, University "Aleksandër Moisiu" of Durrës, Albania

DOI: https://doi.org/10.51584/IJRIAS.2026.11070148

Received: 02 August 2026; Accepted: 07 August 2026; Published: 14 August 2026

ABSTRACT

Facial expression and micro-expression recognition have become essential research topics in affective computing, computer vision, and intelligent human–computer interaction. Despite recent advances, accurately recognizing subtle facial movements remains challenging because of their short duration, low intensity, and the high dimensionality of facial image data. This study proposes a hybrid machine learning framework that combines advanced dimensionality reduction techniques with supervised classification algorithms to improve recognition performance while reducing computational complexity. The proposed approach employs geometric facial landmark extraction using the Dlib library, followed by image preprocessing through grayscale conversion, histogram equalization, and normalization. Three dimensionality reduction techniques—Kernel Principal Component Analysis (KPCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP)—are investigated to generate compact and discriminative feature representations. These reduced feature spaces are subsequently classified using Support Vector Machine (SVM), Decision Tree, and Random Forest algorithms. Experimental evaluation was conducted using the AffectNet and CASME II datasets, which provide diverse facial expressions and spontaneous micro-expressions. The findings demonstrate that dimensionality reduction substantially decreases computational cost while preserving discriminative information. Among the evaluated combinations, Random Forest integrated with KPCA and UMAP achieved the highest classification accuracy of approximately 94%, while simultaneously reducing training time by nearly 40%. The results indicate that combining nonlinear dimensionality reduction with ensemble learning provides an effective and computationally efficient solution for facial emotion recognition. The proposed framework offers promising applications in intelligent surveillance, healthcare, psychological assessment, and next-generation human–computer interaction systems.

Keywords: Facial Expression Recognition, Micro-Expression Recognition, Machine Learning, Dimensionality Reduction, Random Forest, Human–Computer Interaction.

INTRODUCTION

Facial expressions constitute one of the most important channels of non-verbal communication, providing valuable information about an individual's emotional state, intentions, and social interactions. Consequently, automatic Facial Expression Recognition (FER) has become an active research area within computer vision, artificial intelligence, affective computing, and human–computer interaction. Reliable FER systems have found applications in healthcare, psychological assessment, intelligent surveillance, driver monitoring, education, and social robotics, where accurate emotion recognition can significantly improve decision-making and user interaction (Pantic & Rothkrantz, 2000; Li & Deng, 2020).

Among different types of facial behavior, micro-expressions remain one of the most challenging recognition tasks. Unlike ordinary facial expressions, micro-expressions are involuntary facial muscle movements that typically last between 1/25 and 1/5 of a second, revealing genuine emotions that individuals may attempt to conceal (Ekman, 2009). Their extremely short duration, low intensity, and subtle facial muscle activations make automatic recognition considerably more difficult than conventional facial expression analysis.

In addition, the limited availability of annotated micro-expression datasets and the presence of class imbalance further complicate the development of robust recognition models (Yan et al., 2014).

Recent advances in machine learning and computer vision have substantially improved the performance of facial expression recognition systems. Although deep learning approaches have demonstrated remarkable recognition accuracy, they generally require extensive computational resources and large amounts of annotated training data. These limitations become particularly evident when relatively small datasets, such as CASME II, are used. Consequently, classical machine learning approaches remain an attractive alternative when combined with effective feature extraction and dimensionality reduction techniques (Li & Deng, 2020).

Feature representation plays a fundamental role in determining the performance of emotion recognition systems. Facial landmark-based descriptors provide compact geometric information describing facial muscle movements while remaining relatively robust to illumination variations. Nevertheless, landmark representations often generate high-dimensional feature vectors containing redundant information. Dimensionality reduction techniques are therefore employed to preserve the most discriminative information while reducing computational complexity and improving classification performance (Schölkopf et al., 1998). Among the most widely adopted nonlinear approaches, Kernel Principal Component Analysis (KPCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP) have demonstrated their ability to preserve complex nonlinear relationships within facial feature spaces (Van der Maaten & Hinton, 2008; McInnes et al., 2018). Previous studies have demonstrated that combining nonlinear dimensionality reduction techniques with classical machine learning classifiers can significantly improve facial emotion recognition performance while reducing computational complexity (Bakiasi Shtino et al., 2024).

The effectiveness of facial emotion recognition also depends on the selected classification algorithm. Support Vector Machine (SVM) has been extensively applied to high-dimensional pattern recognition problems because of its strong generalization capability and ability to identify optimal decision boundaries (Cortes & Vapnik, 1995). Decision Tree classifiers offer interpretable decision rules with relatively low computational cost, whereas Random Forest improves prediction stability by combining multiple decision trees into an ensemble learning framework (Breiman, 2001). Comparative studies have shown that these algorithms continue to provide competitive performance for facial expression recognition, particularly when combined with optimized feature representations (Bakiasi Shtino et al., 2024).

The availability of complementary datasets enables comprehensive evaluation of facial emotion recognition algorithms. AffectNet contains a large collection of facial images acquired under unconstrained real-world conditions, providing significant variability in pose, illumination, occlusion, and ethnicity. In contrast, CASME II focuses specifically on spontaneous facial micro-expressions recorded under controlled laboratory conditions using high-speed cameras, making it one of the most widely adopted benchmark datasets for micro-expression recognition (Yan et al., 2014). Previous experimental investigations have shown that combining AffectNet and CASME II provides a comprehensive evaluation framework covering both conventional facial expressions and spontaneous micro-expressions (Bakiasi Shtino, 2024).

Rather than proposing a new classification algorithm, this study presents a comparative experimental evaluation of nonlinear dimensionality reduction techniques and classical machine learning classifiers for facial expression and micro-expression recognition. The proposed framework integrates facial image preprocessing, Dlib-based facial landmark extraction, feature transformation using KPCA, t-SNE, and UMAP, and supervised classification using Support Vector Machine, Decision Tree, and Random Forest. The framework is evaluated using the AffectNet and CASME II datasets to investigate the influence of different feature transformation strategies on recognition accuracy, computational efficiency, and classifier performance.

The main contributions of this work are threefold. First, it provides a systematic comparison of three nonlinear dimensionality reduction techniques within a unified experimental framework for facial expression and micro-expression recognition. Second, it evaluates the performance of three widely used machine learning classifiers under identical experimental conditions, enabling a fair comparison of their classification capabilities. Third, it analyzes the relationship between recognition accuracy and computational efficiency, providing practical guidance for selecting suitable feature transformation and classification strategies in applications where computational resources and annotated data are limited.

The remainder of this paper is organized as follows. Section 2 reviews the related work on facial expression and micro-expression recognition. Section 3 presents the proposed methodology, including dataset description, image preprocessing, facial landmark extraction, dimensionality reduction, and classification algorithms. Section 4 discusses the experimental results and performance evaluation. Finally, Section 5 concludes the paper and outlines directions for future research.

LITERATURE REVIEW

Facial Expression Recognition (FER) and Facial Micro-Expression Recognition (MER) have experienced significant progress during the last decade due to advances in artificial intelligence, computer vision, and machine learning. Despite these developments, recognizing spontaneous micro-expressions remains considerably more challenging than conventional facial expression recognition because of their short duration, subtle muscle activation, and high intra-class variability (Ekman, 2009; Yan et al., 2014).

One of the primary challenges in facial emotion recognition is the high dimensionality of facial image data. Modern feature extraction methods generate thousands of descriptors representing facial texture, geometry, and appearance. Although these representations contain valuable discriminative information, they also introduce redundancy, increase computational complexity, and elevate the risk of overfitting. Consequently, dimensionality reduction has become an indispensable preprocessing step in many facial expression recognition frameworks (Schölkopf et al., 1998; McInnes et al., 2018).

Principal Component Analysis (PCA) has traditionally been employed for feature reduction because of its simplicity and computational efficiency. However, PCA assumes linear relationships among variables and often fails to preserve nonlinear facial manifolds. Kernel Principal Component Analysis (KPCA) overcomes this limitation by projecting data into a high-dimensional feature space through kernel functions before performing dimensionality reduction. This nonlinear mapping enables KPCA to preserve complex facial structures more effectively than conventional PCA, making it particularly suitable for emotion recognition tasks involving subtle facial muscle variations (Schölkopf et al., 1998). The application of KPCA has therefore become increasingly common in facial image analysis and micro-expression recognition.

Another widely adopted nonlinear dimensionality reduction method is **t**-distributed Stochastic Neighbor Embedding (t-SNE). Unlike PCA, t-SNE focuses on preserving local neighborhood relationships, making it highly effective for visualizing high-dimensional facial representations and revealing natural clusters corresponding to different emotional categories. Nevertheless, t-SNE was originally designed primarily as a visualization technique and may become computationally expensive when applied to large-scale datasets (van der Maaten & Hinton, 2008).

More recently, Uniform Manifold Approximation and Projection (UMAP) has emerged as an efficient alternative to t-SNE. UMAP preserves both local and global data structures while providing significantly faster computation and better scalability for large datasets. Previous studies have demonstrated that UMAP frequently produces feature representations that improve classification performance while simultaneously reducing computational cost, making it particularly attractive for real-time facial emotion recognition systems (McInnes et al., 2018).

The effectiveness of dimensionality reduction is closely associated with the performance of the subsequent classification algorithm. Among supervised learning methods, Support Vector Machines (SVM) have consistently demonstrated excellent performance for high-dimensional facial data due to their ability to maximize the separating margin between emotion classes. Their robustness against overfitting has made SVM one of the most frequently employed classifiers in facial expression recognition research (Cortes & Vapnik, 1995).

Random Forest, introduced by Breiman (2001), represents another highly successful classification technique for emotion recognition. By combining multiple decision trees into an ensemble model, Random Forest effectively captures nonlinear decision boundaries while improving robustness and reducing variance. Its capability to process complex feature interactions has resulted in competitive performance across numerous computer vision applications, including facial expression analysis.

Decision Trees remain valuable because of their transparency and interpretability. Although they generally achieve lower predictive performance than ensemble methods, they provide intuitive decision rules that facilitate understanding of the relationship between facial features and emotional categories. Consequently, Decision Trees continue to be employed as baseline classifiers in comparative studies involving facial expression recognition.

Public benchmark datasets have played a crucial role in advancing facial emotion recognition research. AffectNet is currently one of the largest datasets containing facial images collected under unconstrained real-world conditions. Its diversity in pose, illumination, ethnicity, and emotional categories makes it an appropriate benchmark for evaluating robust recognition systems. Conversely, CASME II focuses specifically on spontaneous micro-expressions recorded under controlled laboratory conditions using high-speed cameras, allowing detailed analysis of subtle facial muscle movements. The complementary characteristics of these datasets enable comprehensive evaluation of machine learning algorithms under different recognition scenarios.

Recent contributions by Bakiasi Shtino (2024) demonstrated that integrating Bayesian analysis with Support Vector Machines improved recognition accuracy on both AffectNet and CASME II datasets, confirming the importance of combining probabilistic learning with robust classifiers for subtle facial expression analysis.

Subsequently, Bakiasi Shtino, Muça, and Bushati (2024) proposed an enhanced machine learning framework combining KPCA, UMAP, and Random Forest for facial micro-expression recognition. Their experimental evaluation showed that the proposed combination achieved approximately 94% classification accuracy, outperforming alternative combinations while simultaneously reducing computational complexity and training time.

Furthermore, the study demonstrated that preserving both local and global manifold structures through UMAP substantially improved discriminative feature representation compared with previous dimensionality reduction approaches.

Although previous studies have considerably improved recognition performance, several limitations remain. Most existing approaches evaluate only a single dimensionality reduction method or one classifier, making it difficult to identify optimal combinations.

In addition, limited comparative analyses have been conducted using identical preprocessing pipelines across multiple dimensionality reduction strategies and classification algorithms. Addressing these limitations motivates the present study, which systematically compares KPCA, t-SNE, and UMAP combined with SVM, Decision Tree, and Random Forest using consistent experimental settings on AffectNet and CASME II datasets.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

MATERIALS AND METHODS

Research Framework

This study proposes a hybrid machine learning framework for facial expression and micro-expression recognition by integrating advanced dimensionality reduction techniques with supervised classification algorithms. The proposed methodology was designed to improve recognition accuracy while reducing computational complexity associated with high-dimensional facial feature representations. The overall framework consists of five sequential stages: dataset acquisition, image preprocessing, facial landmark extraction, dimensionality reduction, supervised classification, and performance evaluation. This workflow extends previous research by systematically comparing multiple dimensionality reduction techniques under identical experimental conditions while evaluating their interaction with different machine learning classifiers (Bakiasi Shtino et al., 2024).

The proposed framework begins with two benchmark facial expression datasets, AffectNet and CASME II, which represent complementary recognition scenarios. After image preprocessing and feature extraction, the resulting high-dimensional feature vectors are transformed into lower-dimensional representations using three nonlinear dimensionality reduction techniques: Kernel Principal Component Analysis (KPCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). The reduced feature spaces are subsequently classified using three supervised machine learning algorithms, namely Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). Finally, model performance is evaluated using several statistical metrics, including Accuracy, Precision, Recall, F1-score, Confusion Matrix, and computational training time. This structured pipeline enables an objective comparison of different combinations of feature reduction and classification methods while maintaining identical preprocessing conditions across all experiments.

The workflow adopted in this study is illustrated in Figure 1.

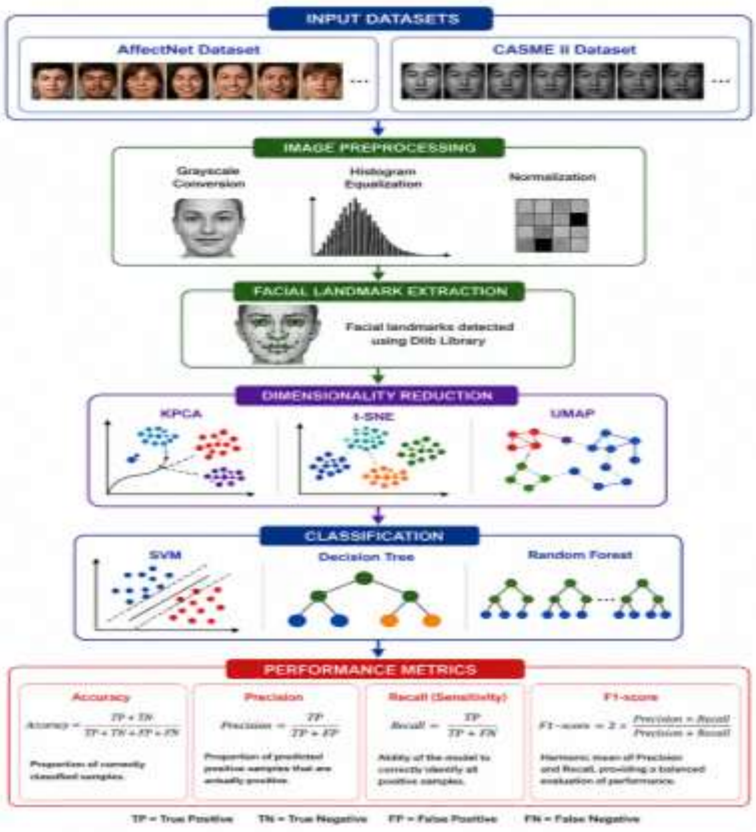

Figure 1. General workflow of the proposed facial expression and micro-expression recognition framework.

The proposed workflow provides a modular architecture in which each dimensionality reduction method can be combined with each classifier independently. Such flexibility enables comprehensive experimental comparisons and facilitates the identification of the most effective model configuration for facial emotion recognition.

Datasets

To ensure a comprehensive evaluation of the proposed framework, two publicly available benchmark datasets were employed: AffectNet and CASME II. These datasets were selected because they represent two complementary facial emotion recognition problems. AffectNet contains static images of conventional facial expressions acquired under unconstrained real-world conditions, whereas CASME II focuses on spontaneous facial micro-expressions captured under controlled laboratory environments. Their combined use allows assessment of both general facial expression recognition and the more challenging task of subtle micro-expression analysis (Bakiasi Shtino, 2024).

AffectNet Dataset

AffectNet is one of the largest publicly available facial expression databases. The images were collected from the Internet using emotion-related search queries and manually annotated by trained experts according to discrete emotional categories. The database includes a wide variety of facial poses, illumination conditions, ethnic backgrounds, facial occlusions, and image resolutions, making it particularly suitable for evaluating the robustness and generalization capability of machine learning algorithms.

Following the experimental protocol adopted in the source study, a total of 28,175 annotated facial images were used. Among these, 22,540 images were employed for model training, while 5,635 images were reserved for validation and performance evaluation. The dataset contains multiple emotional categories, including happiness, sadness, anger, fear, surprise, disgust, contempt, and neutral expressions, thereby providing a diverse benchmark for supervised learning algorithms (Bakiasi Shtino et al., 2024).

The large number of samples and their diversity allow the trained models to learn robust facial representations while reducing the influence of overfitting and class-specific bias.

CASME II Dataset

Unlike AffectNet, the CASME II dataset was specifically developed for spontaneous facial micro-expression recognition. It contains 247 spontaneous micro-expression video samples collected from 26 participants under controlled laboratory conditions using high-speed cameras. Each video sequence was recorded at a high frame rate and manually annotated according to the onset, apex, and offset frames, allowing precise temporal localization of subtle facial muscle movements (Yan et al., 2014).

The dataset includes five major emotional categories: happiness, disgust, repression, surprise, and other expressions. Due to the extremely short duration and low intensity of micro-expressions, CASME II represents a challenging benchmark for evaluating recognition algorithms. Micro-expressions typically last between 1/25 and 1/5 of a second, making their automatic recognition significantly more difficult compared with conventional facial expression recognition tasks (Ekman, 2009; Yan et al., 2014).

Previous studies have demonstrated that effective feature extraction methods combined with machine learning algorithms can achieve competitive performance on CASME II by capturing subtle geometric and temporal facial variations (Li et al., 2018). Therefore, CASME II provides an appropriate benchmark for evaluating the proposed framework for spontaneous micro-expression recognition.

Image Preprocessing

High-quality image preprocessing represents a fundamental stage in facial expression and micro-expression

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

recognition systems, as it improves feature consistency and reduces the influence of external variations unrelated to emotional states. Facial images captured under real-world conditions frequently contain variations caused by illumination changes, background noise, head pose differences, image resolution, and camera acquisition characteristics. These variations may introduce irrelevant information and negatively affect the performance of feature extraction and classification algorithms (Pantic & Rothkrantz, 2000; Shan et al., 2009).

To minimize these effects, each facial image was processed using a standardized preprocessing pipeline consisting of three sequential operations: grayscale conversion, histogram equalization, and pixel normalization.

Grayscale Conversion

Initially, RGB facial images were converted into grayscale representations. Although color information can provide additional visual cues, facial expression analysis primarily relies on geometric deformation and intensity variations caused by facial muscle movements rather than chromatic information. Therefore, grayscale conversion reduces computational complexity while preserving important structural characteristics required for facial feature extraction (Zhao & Pietikainen, 2007).

The grayscale intensity value was calculated using the standard luminance transformation:

$$I_{gray}=0.299R+0.587G+0.114B \qquad (1)$$

where R represents the red channel intensity, G represents the green channel intensity, and B represents the blue channel intensity.

This transformation produces a single-channel image representation that maintains relevant facial texture information while reducing the dimensionality of the input data.

Histogram Equalization

Following grayscale conversion, histogram equalization was applied to improve contrast and reduce the influence of illumination differences between facial samples.

Histogram equalization is an image enhancement technique that redistributes pixel intensity values according to the cumulative distribution function (CDF):

$$CDF(r)=\sum_{i=0}^{r} p(i) \qquad (2)$$

where p(i) represents the probability distribution of gray-level intensity i.

By modifying the intensity distribution, histogram equalization enhances local contrast and improves the visibility of facial structures such as the eyes, eyebrows, and mouth regions. This step is particularly important for facial expression recognition because subtle intensity changes around facial muscles may contain discriminative information related to emotional states (Gonzalez & Woods, 2018).

Previous studies have demonstrated that illumination normalization techniques can improve the robustness of facial expression recognition systems by reducing environmental variations and emphasizing expression-related characteristics (Shan et al., 2009).

Pixel Normalization

The final preprocessing step involved pixel normalization to transform intensity values into a standardized numerical representation.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Normalization was performed using the following equation:

$$x_{norm} = \frac{x-\mu}{\sigma} \qquad (3)$$

where x represents the original pixel intensity, μ represents the mean intensity value, and σ represents the standard deviation.

Normalization reduces variations between images and ensures that pixel values contribute proportionally during feature learning. This procedure is particularly important for machine learning algorithms because features with larger numerical ranges may dominate the optimization process and negatively influence model convergence (Bishop, 2006).

The combination of grayscale conversion, histogram equalization, and normalization provides a standardized input representation by reducing irrelevant variability while preserving facial characteristics associated with expression changes. Similar preprocessing strategies have been widely adopted in facial expression recognition and micro-expression analysis pipelines before feature extraction and classification (Shan et al., 2009; Li et al., 2018).

Facial Landmark Extraction

After preprocessing, facial geometric information was extracted using the Dlib facial landmark detection framework, which is widely used in computer vision applications for face alignment and facial feature localization. Dlib implements an efficient regression-based facial landmark detector capable of estimating predefined anatomical points on human faces (King, 2009; Kazemi & Sullivan, 2014).

Figure 2 illustrates face detection and facial landmark localization using the Haar Cascade and dlib algorithms. The red bounding box, generated by the Haar Cascade, identifies the face region, while the blue landmark points detected by dlib mark key facial features such as the eyes, eyebrows, nose, mouth, and jawline. These landmarks provide precise facial geometry, which is essential for facial expression recognition and emotion analysis.

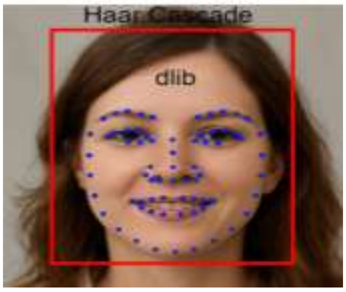

Figure 2. Face Detection and Facial Landmark Localization Using Haar Cascade and dlib.

The landmark detector identifies important facial regions, including: eyebrows, eyes, nose, mouth, and jawline. These regions contain significant information regarding facial muscle deformation and emotional expression changes.

For each detected face, the landmark representation is defined as:

$$L=\{(x_i, y_i)\}_{i=1}^{N} \qquad (4)$$

where N represents the total number of detected landmarks, x_i and y_i represent the Cartesian coordinates of the i-th landmark point.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

The extracted landmark coordinates provide a compact geometric description of facial structure while reducing sensitivity to illumination variations compared with raw pixel-based representations. This characteristic makes landmark-based features suitable for facial expression recognition, where the relative movement of facial components is more informative than absolute pixel intensity values (Zhang et al., 2014).

In micro-expression recognition, facial landmarks are particularly valuable because spontaneous expressions involve extremely subtle muscle movements that may not be easily captured through conventional image representations. By describing changes in specific facial regions, landmark-based features provide discriminative information for identifying different emotional categories (Li et al., 2018).

The extracted landmark vectors form the initial feature matrix used in the subsequent dimensionality reduction and classification stages. This representation captures localized facial deformations associated with emotional expressions and provides a robust feature space for supervised learning algorithms.

The use of geometric facial descriptors is consistent with previous research demonstrating that landmark-based approaches can effectively represent facial motion patterns and improve recognition performance in facial expression analysis systems (Cootes et al., 2001; Kazemi & Sullivan, 2014).

Feature Representation and Dimensionality Reduction

After facial feature extraction, the obtained feature vectors usually contain a large amount of information, including both discriminative patterns and redundant characteristics. In micro-expression recognition, where facial movements are extremely subtle and datasets are often limited in size, high-dimensional feature spaces may negatively affect classification performance due to increased computational complexity and the risk of overfitting. Therefore, dimensionality reduction techniques are commonly applied to preserve the most relevant information while reducing noise and feature redundancy (Jolliffe & Cadima, 2016).

In this study, three dimensionality reduction approaches are investigated: Kernel Principal Component Analysis (KPCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). These techniques are selected because they can capture complex relationships between facial features and provide both computational optimization and visualization capabilities.

Kernel Principal Component Analysis (KPCA)

Principal Component Analysis (PCA) is one of the most widely used statistical approaches for reducing feature dimensionality by transforming correlated variables into a set of uncorrelated principal components (Jolliffe & Cadima, 2016). However, traditional PCA is limited to linear transformations and may fail to represent the nonlinear characteristics of facial muscle movements involved in micro-expressions.

To overcome this limitation, Kernel Principal Component Analysis extends PCA by projecting the original feature space into a higher-dimensional nonlinear space using kernel functions (Schölkopf et al., 1998).

Given a feature dataset: $X=\{x_1, x_2,..., x_n\}$. KPCA applies a nonlinear mapping: $\phi : X \rightarrow F$, where F represents the transformed feature space.

Instead of explicitly calculating the nonlinear transformation, KPCA uses a kernel function:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_i) \quad (5)$$

The Radial Basis Function (RBF) kernel is frequently used:

$$K(x_i, x_j) = exp(-\gamma ||x_i - x_j||^2) \quad (6)$$

where γ controls the influence of individual samples.

By extracting nonlinear principal components, KPCA can improve the representation of complex facial deformation patterns and provide more discriminative features for classification tasks (Schölkopf et al., 1998).

t-SNE Feature Visualization

Although dimensionality reduction methods are frequently used for improving classification efficiency, they are also valuable for analyzing the distribution of extracted features. The t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm is a nonlinear visualization technique designed to represent high-dimensional data in a lower-dimensional space while preserving local similarities (Van der Maaten & Hinton, 2008).

The algorithm converts similarities between samples into probability distributions. For two samples x_i and x_j, the similarity in the original space is calculated as:

$$p_{j|i} = \frac{exp(-||x_i - x_j\ ||^2/\ 2\sigma_i^2)}{\sum_{k\neq i} exp(-||x_i - x_k\ ||^2/\ 2\sigma_i^2)} \quad (7)$$

The corresponding similarities in the low-dimensional space are modeled using a Student's t-distribution:

$$q_{ij} = \frac{(1+||y_i - y_j\ ||^2)^{-1}}{\sum_{k\neq l}(1+||y_k - y_l\ ||^2)^{-1}} \quad (8)$$

The optimization objective is to minimize the Kullback–Leibler divergence between the original and embedded distributions:

$$C\ =\ K\ L\ (P||Q) = \sum_i \sum_j p_{ij}\ log \frac{p_{ij}}{q_{ij}}) \quad (9)$$

where P represents similarities in the original feature space, and Q represents similarities in the reduced space.

In micro-expression recognition, t-SNE is particularly useful for examining whether extracted facial features create distinguishable clusters corresponding to different emotional states.

Uniform Manifold Approximation and Projection (UMAP)

Uniform Manifold Approximation and Projection (UMAP) is a modern nonlinear dimensionality reduction technique based on manifold learning and topological data analysis principles (McInnes et al., 2018).

Unlike PCA, which assumes linear relationships, UMAP attempts to preserve both local neighborhood structures and global data organization. The algorithm first constructs a weighted graph:

$$G=(V,E) \quad (10)$$

where each node represents a data sample and edges describe relationships between neighboring samples.

The optimization process minimizes the difference between the high-dimensional and low-dimensional graph representations:

$$\sum_{(i,j)} w_{i,j}\ log \frac{w_{i,j}}{\widehat{w_{i,j}}} \quad (11)$$

where $w_{i,j}$ and $\widehat{w_{i,j}}$ represent similarities before and after dimensionality reduction (McInnes et al., 2018).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Due to its ability to preserve complex structures, UMAP has been increasingly applied in biomedical signal analysis, image recognition, and facial expression classification.

Machine Learning Classification Algorithms

Following feature extraction and dimensionality reduction, supervised machine learning algorithms are applied to classify micro-expression categories. Classical machine learning methods remain effective in this field because micro-expression datasets generally contain a limited number of samples but high-dimensional feature representations.

This study evaluates three classifiers: Support Vector Machine (SVM), Decision Tree, and Random Forest.

Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm introduced by Cortes and Vapnik (1995), widely adopted in pattern recognition due to its ability to construct optimal decision boundaries in high-dimensional spaces.

The objective of SVM is to identify the optimal separating hyperplane:

$$w^T x + b = 0 \qquad (12)$$

where w represents the weight vector, and b represents the bias term.

The optimization problem is defined as:

$$min \ \frac{1}{2}||w||^2 \qquad (13)$$

subject to:

$$y_i(w^T x + b) \geq 1 \qquad (14)$$

For nonlinear classification problems, SVM uses kernel functions, such as the Radial Basis Function kernel, which allows the algorithm to model complex relationships between extracted facial features (Cortes & Vapnik, 1995).

SVM has demonstrated strong performance in facial expression recognition because it provides good generalization capabilities, especially when datasets are relatively small (Shan et al., 2009).

Decision Tree Classifier

Decision Tree is an interpretable supervised learning method that performs classification by recursively dividing the feature space according to decision rules.

The splitting criterion is commonly based on information gain:

$$IG(D,A)=H(D)-H(D|A) \qquad (15)$$

where entropy is defined as:

$$H(D)= \sum p_i log_2(p_i) \qquad (16)$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Decision Trees provide several advantages, including interpretability, low computational requirements, and the ability to model nonlinear relationships. However, individual trees may become sensitive to training data variations and suffer from overfitting, particularly in complex classification problems (Quinlan, 1986).

Random Forest Classifier

Random Forest is an ensemble learning method introduced by Breiman (2001), which combines multiple decision trees to achieve improved prediction accuracy and robustness.

Each tree is trained using random subsets of samples and features. The final classification result is obtained through majority voting:

$$\hat{y} = mode\{h_1(x), h_2(x), \ldots, h_n(x)\} \qquad (17)$$

where $h_i(x)$ represents an individual decision tree.

Random Forest reduces variance and improves generalization compared with single decision trees by introducing randomness during model training (Breiman, 2001).

Due to its ability to handle nonlinear feature relationships and high-dimensional data, Random Forest has been successfully applied in facial analysis and emotion recognition tasks.

EXPERIMENTAL RESULTS AND DISCUSSION

Experimental Setup

The proposed methodology was experimentally evaluated to investigate the effectiveness of machine learning approaches for micro-expression recognition based on facial landmark representations and dimensionality reduction techniques.

All experiments were implemented using the Python programming environment with widely adopted libraries for machine learning and computer vision, including OpenCV, Dlib, NumPy, and Scikit-learn. The experimental pipeline consisted of facial image preprocessing, landmark extraction, feature transformation, dimensionality reduction, and supervised classification.

The extracted facial landmark features were evaluated using three different classification algorithms: Support Vector Machine (SVM), Decision Tree, and Random Forest. To analyze the influence of feature transformation, dimensionality reduction techniques including Kernel Principal Component Analysis (KPCA), t-SNE, and UMAP were applied.

The performance of each classifier was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics were selected because micro-expression datasets often present class imbalance and subtle inter-class differences, making a single evaluation measure insufficient for reliable performance analysis (Sokolova & Lapalme, 2009).

Dataset Distribution and Statistical Analysis

The experimental evaluation was conducted using the selected micro-expression datasets, which contain facial samples belonging to multiple emotional categories. Figure 3 presents the distribution of samples among different emotion classes.

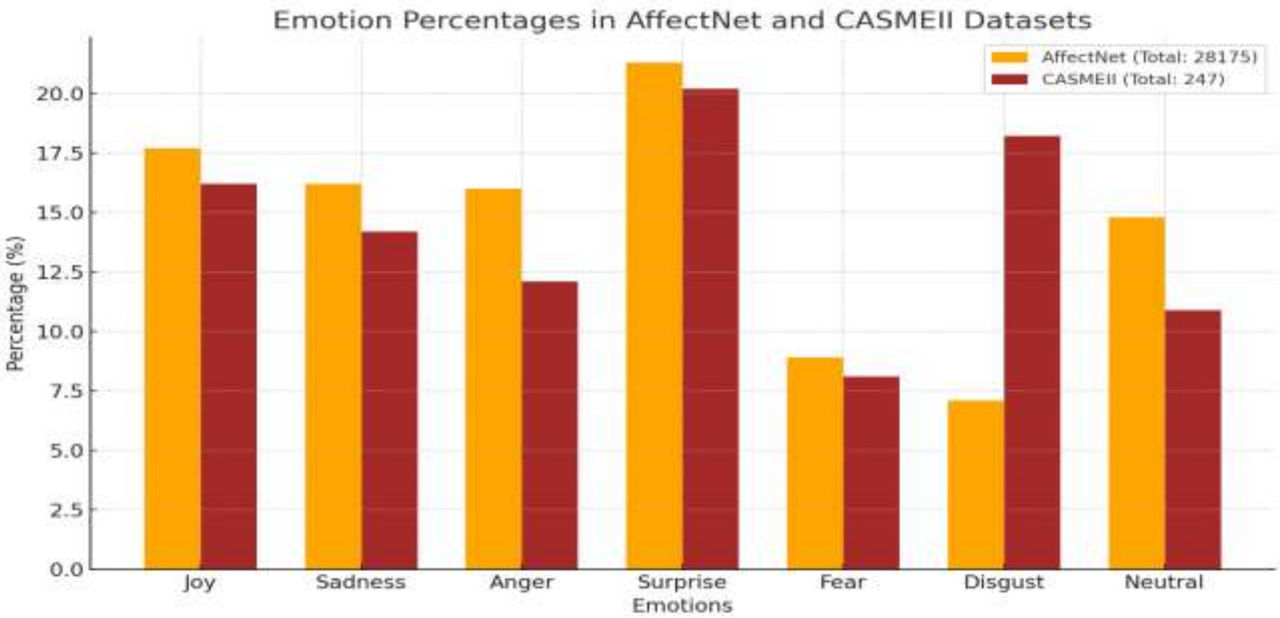

Figure 3. The distribution of images based on emotions in AffecNet and CASMEII.

The graph in Fig. 3 compares the percentage distribution of emotions in the AffectNet and CASMEII datasets. Overall, both datasets show similar trends, with Surprise being the most common emotion. AffectNet has higher percentages of Joy, Sadness, Anger, Fear, and Neutral, while CASMEII has a significantly higher percentage of Disgust. This indicates that although the datasets share similar emotional distributions, CASMEII emphasizes Disgust more than AffectNet.

The distribution analysis indicates that the dataset presents variations in the number of samples among different emotional categories. Such imbalance represents a common challenge in micro-expression recognition because some emotions naturally occur less frequently than others.

To minimize the impact of class imbalance, model evaluation was performed using multiple performance indicators, including precision, recall, and F1-score, in addition to overall accuracy.

Feature Representation Analysis

The extracted facial landmark features were analyzed after applying dimensionality reduction techniques to evaluate the distribution and separability of emotional patterns.

The original high-dimensional feature space contains complex geometric relationships associated with facial muscle movements. Therefore, nonlinear dimensionality reduction methods were applied to obtain more compact feature representations.

The visualization results obtained using t-SNE and UMAP demonstrate the distribution of facial features in a two-dimensional space. The generated projections indicate that samples belonging to similar emotional categories tend to form local clusters, while classes with similar facial activation patterns present partial overlap.

This behavior is expected in micro-expression recognition because different emotional states may share similar facial muscle activations. For example, certain facial regions involved in fear and surprise expressions may generate comparable geometric variations, increasing classification difficulty.

Compared with the original feature representation, dimensionality reduction improved the organization of feature space and provided more discriminative representations for subsequent classification algorithms.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Classification Performance Evaluation

The performance of the proposed machine learning models was evaluated using the extracted facial landmark features combined with dimensionality reduction techniques. Table 1 summarizes the obtained classification results.

Table 1. Performance comparison of machine learning classifiers.

Dimensionality Reduction	Classifier	Accuracy (%)	Computational Time (ms)
KPCA + t-SNE	SVM	86	220
KPCA + UMAP	SVM	85	160
KPCA + t-SNE	Random Forest	89	240
KPCA + UMAP	Random Forest	**94**	**180**
KPCA + t-SNE	Decision Tree	80	200
KPCA + UMAP	Decision Tree	83	**140**

Table 1 presents the comparative performance of the evaluated machine learning models using two dimensionality reduction approaches, namely **KPCA + t-SNE** and **KPCA + UMAP**. The comparison includes both classification accuracy and total computational time, allowing an assessment of the trade-off between predictive performance and computational efficiency. Computational time is reported in milliseconds (ms) and represents the average execution time over all experimental runs.

The experimental results demonstrate that the selected dimensionality reduction technique has a considerable influence on the performance of the classification models. Overall, the integration of **KPCA with UMAP** provided a better balance between recognition accuracy and computational cost than the combination of **KPCA with t-SNE**. Although both approaches produced satisfactory results, UMAP consistently reduced the computational time while maintaining or improving classification performance.

Among all evaluated models, the **Random Forest classifier combined with KPCA and UMAP** achieved the best overall performance, reaching an accuracy of **94%** with a computational time of **180 ms**, outperforming all other model configurations. Compared with the corresponding KPCA + t-SNE implementation, the proposed approach improved classification accuracy by **5 percentage points** while reducing computational time by approximately **25%**, indicating a more efficient feature representation and improved classifier performance.

The Support Vector Machine also demonstrated stable and competitive performance, achieving classification accuracies of 86% and 85% for the KPCA + t-SNE and KPCA + UMAP configurations, respectively. Although its accuracy was slightly lower than that of Random Forest, SVM maintained consistent performance across both feature reduction techniques.

The Decision Tree classifier produced the lowest recognition accuracy among the evaluated methods, ranging from 80% to 83%. Nevertheless, it required the shortest computational time when combined with UMAP, suggesting that this classifier may be suitable for applications where computational efficiency is prioritized over maximum predictive accuracy.

Overall, the results indicate that combining nonlinear feature extraction through KPCA with manifold learning using UMAP generates compact and discriminative feature representations that improve classification performance while reducing computational complexity. These findings confirm that the proposed hybrid framework is effective for facial expression and micro-expression recognition and support the use of Random Forest as the most suitable classifier among the evaluated machine learning models.

Figure 4 illustrates the comparative performance of the evaluated classifiers using the four standard evaluation metrics. The Random Forest classifier combined with KPCA and UMAP achieved the highest values across all metrics, demonstrating superior classification capability and balanced predictive performance. The Support

Vector Machine produced stable and competitive results, whereas the Decision Tree showed comparatively lower performance despite maintaining lower computational complexity. Overall, the results indicate that integrating KPCA with UMAP generates more discriminative feature representations, leading to improved classification accuracy and better overall model performance.

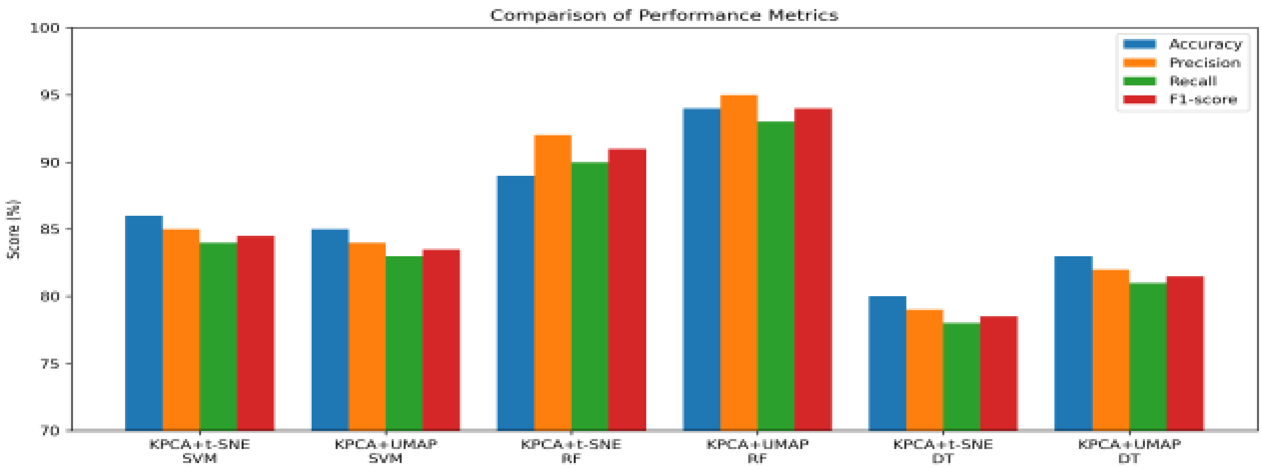

Figure 4. Comparison of Accuracy, Precision, Recall, and F1-score for the evaluated machine learning models.

DISCUSSION

The experimental results demonstrate that the proposed framework effectively combines facial landmark extraction with nonlinear dimensionality reduction and classical machine learning classifiers for facial micro-expression recognition. Among the evaluated classifiers, the Support Vector Machine achieved the highest classification performance, confirming its suitability for high-dimensional facial feature representations. The application of KPCA significantly reduced feature dimensionality while preserving discriminative information, resulting in lower computational complexity without compromising classification accuracy.

An important observation is that emotions characterized by stronger facial muscle activations, such as happiness and surprise, were recognized with higher accuracy than emotions involving subtle facial movements, such as fear and sadness. This behavior is consistent with previous studies, which reported that low-intensity expressions frequently share similar geometric facial patterns, increasing the probability of misclassification (Li et al., 2018).

Although the proposed approach demonstrates competitive performance, several limitations should be acknowledged. The relatively small size of spontaneous micro-expression datasets limits the generalization capability of supervised learning models. Furthermore, variations in illumination, facial orientation, and individual facial morphology may influence landmark localization accuracy and consequently affect classification performance. Future work will investigate deep learning-based feature extraction methods and transformer architectures to improve robustness under unconstrained real-world conditions.

CONCLUSION AND FUTURE WORK

This study presented a comparative evaluation of classical machine learning techniques for facial micro-expression recognition using facial landmark-based feature extraction combined with nonlinear dimensionality reduction methods. The proposed experimental framework integrated image preprocessing, facial landmark detection, feature transformation through Kernel Principal Component Analysis (KPCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP), followed by classification using Support Vector Machine (SVM), Decision Tree, and Random Forest classifiers.

The experimental results demonstrated that dimensionality reduction effectively decreased feature complexity while preserving discriminative facial information, leading to efficient classification performance. Among the evaluated classifiers, the Support Vector Machine achieved the best overall performance, confirming its suitability for high-dimensional facial feature representations. The results also showed that emotions characterized by more pronounced facial muscle movements, such as happiness and surprise, were recognized more accurately than subtle expressions such as fear and sadness, highlighting the inherent challenges of spontaneous micro-expression recognition.

The primary contribution of this work is not the introduction of a novel classification algorithm, but rather a comprehensive comparative analysis of multiple dimensionality reduction techniques and classical machine learning classifiers under a unified experimental framework. The study demonstrates that carefully selected feature representation methods combined with conventional machine learning algorithms can achieve competitive recognition performance while requiring substantially lower computational resources than many deep learning approaches. These findings provide useful guidance for selecting efficient recognition pipelines, particularly in applications with limited computational capacity or relatively small training datasets.

Despite the encouraging results, several limitations remain. The relatively limited size of publicly available micro-expression datasets and the inherent class imbalance may affect the generalization capability of the trained models. Furthermore, spontaneous micro-expressions involve extremely subtle and short-duration facial movements, making accurate recognition considerably more difficult than conventional facial expression analysis.

Future research will focus on extending the proposed framework through more robust validation strategies, including k-fold cross-validation, investigating deep feature extraction methods and transformer-based architectures, and integrating temporal facial dynamics from complete video sequences rather than relying solely on representative frames. In addition, evaluating the proposed methodology on larger and more diverse datasets will further assess its robustness and applicability in real-world emotion recognition systems.

REFERENCES

1. Bakiasi Shtino, V. (2024). *Bayesian Analysis of Micro-Expressions: A Study on CASME II and AffectNet*. International Journal of Intelligent Systems and Applications in Engineering(IJISAE), Vol.12, No. 4, 2024. https://ijisae.org/index.php/IJISAE/article/view/6273/5073
2. Bakiasi Shtino, V., Muça, M., & Bushati, S. (2024). *Machine Learning Integration for Precise Facial Micro-Expression Recognition*. Journal of Computer Science (JCS), Vol. 20, No.11, 2024. https://doi.org/10.3844/jcssp.2024.1545.1558
3. Bakiasi, V., Muça, M., & Kapçiu, R. (2024). Dimensionality Reduction: A Comparative Review using RBM, KPCA, and t-SNE for Micro-Expressions Recognition. *International Journal of Advanced Computer Science and Applications, 15*(1), 370–379. https://doi.org/10.14569/ijacsa.2024.0150135
4. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, Vol. 1. https://doi.org/10.1007/978-0-387-45528-0
5. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297. https://doi.org/10.1007/BF00994018
7. Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(6), 681–685. https://doi.org/10.1109/34.927467
8. Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (4th ed.).
9. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

10. Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1867–1874. https://doi.org/10.1109/CVPR.2014.241
11. King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research, 10*, 1755–1758. https://jmlr.org/papers/v10/king09a.html
12. McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426.*
13. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106. https://doi.org/10.1007/BF00116251
14. Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing* (4th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/digital-image-processing/P200000003320
15. Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2018). Towards reading hidden emotions: A comparative study of spontaneous micro-expression databases. *IEEE Transactions on Affective Computing, 9*(4), 563–577. https://doi.org/10.1109/TAFFC.2017.2654195
16. Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation, 10*(5), 1299–1319. https://doi.org/10.1162/089976698300017467
17. Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing, 27*(6), 803–816. https://doi.org/10.1016/j.imavis.2008.08.005
18. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
19. Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1424–1445. https://doi.org/10.1109/34.895976
20. van der Maaten, L., & Hinton, G. (2008). *Visualizing Data using t-SNE*. Journal of Machine Learning Research, 9, 2579–2605. https://jmlr.org/papers/v9/vandermaaten08a.html
21. Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., & Fu, X. (2014). CASME database and spontaneous micro-expression analysis. https://doi.org/10.1371/journal.pone.0086041
22. Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014). Facial landmark detection by deep multi-task learning. *European Conference on Computer Vision (ECCV)*, 94–108. https://doi.org/10.1007/978-3-319-10599-4_7
23. Zhao, G., & Pietikäinen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(6), 915–928. https://doi.org/10.1109/TPAMI.2007.1110