# Investigation And Comparative Analysis Of Algorithms About Recognition Of Micro Mimics For Analysis Of Person Using Emotional AI

Oleksandr Yaremchenko, Petro Pukach

*Lviv Polytechnic National University, Bandera str. 12, Lviv, 79013, Ukraine*

**Abstract**

This paper suggests utilizing micro mimics, subtle facial muscle movements that are challenging to detect with the naked eye, for evaluating psychological states through artificial intelligence. The research aims to develop and enhance methods for analyzing micro-mimics to precisely identify emotions and individuals' psychological states. In this study, we carried out an experimental examination of the proposed method using video recordings of people experiencing various emotional states. Our findings indicate that the proposed method effectively recognizes emotions and psychological states with high accuracy. This study contributes to the field of emotional AI and presents new possibilities for assessing psychological states using micro-mimics. The results of this study could be valuable in numerous applications, such as mental health, human-computer interaction, and social robotics.

**Keywords**

Micro mimics, emotional AI, psychological state, artificial intelligence, facial expression recognition, machine learning, video analysis, emotion recognition, human-computer interaction, mental health.

## 1. Introduction

The field of emotional AI has seen remarkable progress in recent years, with researchers and practitioners leveraging machine learning algorithms to identify subtle patterns in facial expressions that reveal a person's emotional and psychological state. One of the most promising approaches in this field is the analysis of micro mimics, brief facial expressions that can reveal an individual's true emotions even if they are attempting to conceal them. The study of micro mimics is particularly relevant in high-stakes environments, such as negotiations, interviews, and other situations where individuals may feel pressure to conceal their true emotions. Why it matters: As artificial intelligence learns to interpret and respond to human emotion, senior leaders should consider how it could change their industries and play a critical role in their firms [1].

The relevance of this research is clear: accurate and reliable methods for assessing the psychological state of individuals can have a significant impact on mental health, healthcare, education, marketing, and other fields. By improving and developing methods for assessing the psychological state using micro mimics and artificial intelligence, researchers can provide more accurate and nuanced insights into the emotional and psychological state of individuals. These insights can be used to inform diagnosis, treatment, and support for individuals with mental health conditions, as well as to develop more effective marketing and educational campaigns that resonate with their target audience.

The goal of this research is to advance our understanding of micro mimics and their potential applications in assessing the psychological state of individuals. Specifically, we aim to develop machine learning algorithms that can analyze micro mimics in real-time, providing accurate and reliable insights

MoMLeT+DS 2023: 5th International Workshop on Modern Machine Learning Technologies and Data Science, June, 3, 2023, Lviv, Ukraine
EMAIL: Oleksandr.D.Yaremchenko@lpnu.ua (O. Yaremchenko); Petro.Y.Pukach@lpnu.ua (P. Pukach)
ORCID: 0009-0001-2002-2704 (O. Yaremchenko); 0000-0002-0359-5025 (P. Pukach)

into a person's emotional and psychological state. To achieve this goal, we will undertake several tasks, including a review of the literature on micro mimics analysis and emotional AI, the development of machine learning algorithms for real-time analysis of micro mimics, and the evaluation of these algorithms in a variety of settings, including mental health diagnosis, healthcare, education, marketing, and human-computer interaction.

One of the key challenges in this field is the limited availability of high-quality micro mimics datasets, as well as the need for robust and reliable algorithms that can analyze these subtle and fleeting facial expressions in real-time. However, recent advances in deep learning and computer vision hold promise for addressing these challenges and developing more accurate and reliable methods for micro mimics analysis.

Overall, this research has the potential to make a significant contribution to the field of emotional AI, providing new insights into the emotional and psychological state of individuals and opening up new avenues for diagnosis, treatment, and support. As AI continues to advance in its ability to interpret and respond to human emotion, this research could play a critical role in shaping the future of mental health, healthcare, education, marketing, and other fields.

## 2. Related Works

Talking about micro mimics we can not forget about Paul Ekman. Paul Ekman is a renowned psychologist and a pioneer in the study of emotions and facial expressions. His research has significantly contributed to our understanding of micro-expressions and their role in emotion recognition. Ekman's work has laid the foundation for the development of technologies that recognize and analyze micro-expressions [2]. In the 1960s and 1970s, Ekman and his colleagues conducted groundbreaking research on the universality of facial expressions. They demonstrated that certain facial expressions of emotion are universally recognized, regardless of cultural background. This finding suggested that these expressions have a biological basis and are not merely culturally learned behaviors [3]. In the course of his research, Ekman discovered micro-expressions, which are very brief, involuntary facial expressions that occur when a person tries to conceal or suppress their emotions. These expressions can last as little as 1/25th of a second and are difficult to recognize with the naked eye. Ekman also co-developed the Facial Action Coding System (FACS) [4], a comprehensive tool for objectively measuring facial movements. FACS is widely used in psychology and computer science research, as well as in the development of emotion recognition technologies.

Ekman's work has inspired researchers and technologists to develop algorithms and systems that can recognize and analyze micro expressions, leading to the emergence of the emotion recognition technology field. His research has had a profound impact on various industries, including security, marketing, mental health, and human-computer interaction [5, 6].

In recent years, scholarly interest in micro expressions has grown considerably. As shown in Figure 1, after the release of two open-source micro expression databases in 2013, the number of articles related to micro expressions has risen annually. Since 2018, the Micro-Expression Grand Challenge (MEGC) workshop, which is part of the IEEE International Conference on Automatic Face and Gesture Recognition, has helped popularize the subject within the computer vision and machine learning communities.

Here are a few notable research works related to micro expression recognition using different algorithms: Li, Xiaobai, Xiaopeng Hong, Antti Moilanen, Xiaohua Huang, Tomas Pfister, Guoying Zhao, and Matti Pietikäinen. "Towards reading hidden emotions: A comparative study of spontaneous micro expression spotting and recognition methods." IEEE Transactions on Affective Computing 9, no. 4 (2017): 563-577.

This work presents a comparative study of various micro expression spotting and recognition methods, including LBP-TOP, CNN, and CNN-LSTM. The study evaluates these methods on three benchmark datasets: CASME II, SMIC, and SAMM. The results indicate that the combination of CNN and LSTM (CNN-LSTM) offers the best performance in recognizing spontaneous micro expressions.

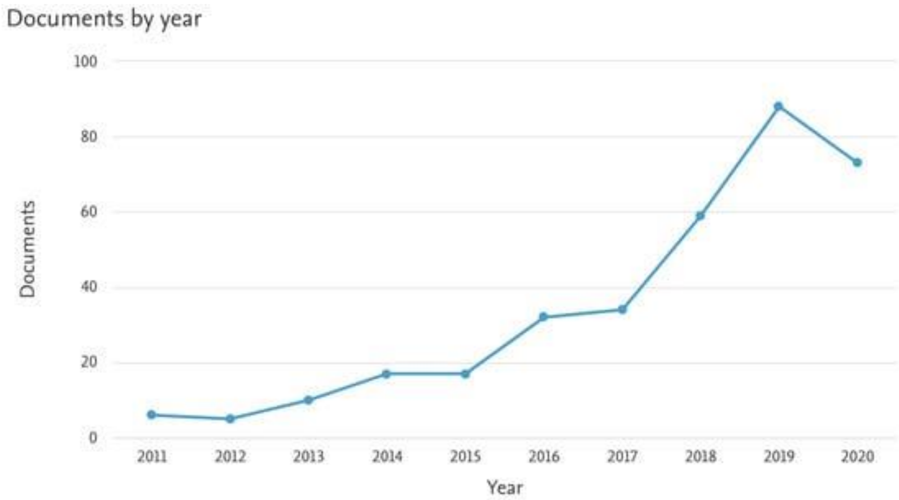


**Figure 1:** The number of micro expression recognition publications from 2011 to 2020 (Data Source: Scopus).

Huang, Xiaohua, Guoying Zhao, Xiaopeng Hong, Wenming Zheng, and Matti Pietikäinen. "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns and SVM." In Asian conference on computer vision, pp. 162-177. Springer, Cham, 2014.

This paper proposes a novel method called Spatiotemporal Completed Local Quantized Patterns (STCLQP) for spontaneous facial micro-expression analysis. The authors use SVM for classification and evaluate the proposed method on the CASME dataset. The results show that the STCLQP method outperforms other local pattern-based methods, such as LBP-TOP and LBP-SIP.

Liu, Yan, Jun Li, and Wenjing Zheng. "Micro-expression recognition based on 3D convolutional neural networks." In 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 1578-1582. IEEE, 2018.

In this work, the authors propose a micro-expression recognition method based on 3D Convolutional Neural Networks (3D-CNNs). The method captures both spatial and temporal information in micro-expression video clips. The study evaluates the proposed method on the CASME II dataset, and the results demonstrate that the 3D-CNN model outperforms traditional LBP-TOP-based methods.

Le Ngo, Anh Cat, John See, and Raphael C.-W. Phan. "Learning Deep Spatiotemporal Features for Micro-expression Recognition Using 3D CNN and Optical Flow." In 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1486-1495. IEEE, 2021.

This paper presents a deep learning-based method for micro-expression recognition that combines 3D CNN and optical flow. The proposed method extracts spatiotemporal features from micro-expression video clips and uses optical flow to improve temporal feature extraction. The method is evaluated on the CASME II, SMIC, and SAMM datasets, showing promising results compared to other state-of-the-art methods. These studies demonstrate the application of different algorithms and techniques in micro-expression recognition tasks. The results suggest that hybrid models and 3D-CNNs tend to perform better than traditional methods. However, the optimal approach will depend on the specific dataset and task. Various companies and organizations are using micro-expression recognition technology for applications such as security, marketing, and mental health. Some notable companies and their applications include:

Affectiva: Affectiva, an emotion recognition technology company spun off from the MIT Media Lab, develops software that can detect and analyze facial expressions in real-time. Their technology is used in various applications, including market research, automotive safety, and mental health.

Emotient (acquired by Apple): Emotient was a company specializing in emotion recognition through facial expression analysis. Apple acquired Emotient in 2016, and it is speculated that their technology has been integrated into various Apple products and services, such as Animoji, Memoji, and potentially, their rumored augmented reality (AR) glasses.

nViso: nViso is a company that offers emotion recognition solutions based on 3D facial imaging technology. Their applications include customer experience enhancement, market research, and gaming.

Eyeris: Eyeris develops emotion recognition technology for various industries, such as automotive, robotics, and smart home devices. Their EmoVu technology can recognize micro-expressions in real-time and has been used to improve driver safety and user experiences in connected cars.

Kairos: Kairos is a company focused on facial recognition and emotion analysis. They provide solutions for various industries, including marketing, security, and entertainment. Their emotion analysis technology can be used to gauge audience reactions to advertisements, movies, or other content.

Cognitec Systems: Cognitec Systems is a company specializing in facial recognition technology for various applications, such as access control, surveillance, and marketing. Their technology includes emotion recognition capabilities that can detect and analyze micro-expressions.

These companies represent a small sample of the organizations working with micro-expression recognition technology. The technology is continuously evolving, and its applications are expanding into new fields and industries as more research is conducted in this area.

## 3. Methods and Materials

When working on micro-expression recognition for emotional AI, you can consider several algorithms and techniques. Some popular ones are:

Convolutional Neural Networks (CNNs) [7]: CNNs are widely used in image recognition tasks and have shown good performance in facial expression recognition. They can automatically learn features from input images, making them ideal for micro-expression recognition. Some popular CNN architectures include VGG, ResNet, and Inception.

Recurrent Neural Networks (RNNs) [8]: Since micro-expressions are temporal in nature, RNNs can be useful in modeling the sequential information. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are popular variants of RNNs used in this context.

3D Convolutional Neural Networks (3D-CNNs): 3D-CNNs can be used to capture both spatial and temporal information in video data. They are well-suited for micro-expression recognition tasks since they take into account the temporal aspect of the expressions.

Temporal Convolutional Networks (TCNs): TCNs are a recent development in the field of deep learning that combine the strengths of both CNNs and RNNs. They offer a powerful way to model temporal dependencies and can be used for micro-expression recognition as well.

Hybrid models: Combining different types of neural networks can improve performance on certain tasks. For example, you could use a CNN for spatial feature extraction and an RNN or TCN for temporal feature extraction. This would create a hybrid model that capitalizes on the strengths of each network.

Comparing these algorithms is subjective and will depend on your specific requirements and constraints. Some factors to consider when choosing an algorithm include:

Performance: The ability to accurately recognize micro-expressions in a variety of contexts and lighting conditions is crucial. You may need to experiment with different algorithms and architectures to find the one that performs best on your dataset.

Computational efficiency: Some models may be more computationally expensive than others, which can impact training and inference times. Consider your available hardware and whether real-time processing is necessary when choosing an algorithm.

Ease of implementation: Depending on your expertise and the available resources, some algorithms may be easier to implement than others. Open-source implementations and pre-trained models can save time and effort.

Robustness: The ability to generalize to new, unseen data is crucial in micro-expression recognition. Make sure the chosen algorithm can handle variations in lighting, head pose, and facial occlusion.

To make an informed decision, you should test different algorithms and architectures on your dataset, compare their performance, and consider the factors mentioned above

It's important to note that the performance of these algorithms varies across datasets and implementations. However, here's a general comparison of the algorithms based on their performance and other parameters:

Convolutional Neural Networks (CNNs):

Performance: Good performance on spatial features and facial expression recognition. However, they lack the ability to capture temporal information, which is crucial for micro-expression recognition.

Computational Efficiency: Moderate. Training and inference times depend on the depth and complexity of the network. GPUs can be used to speed up computations.

Ease of Implementation: Relatively easy to implement with many open-source libraries and pre-trained models available.

Robustness: Good generalization capabilities, but limited in handling temporal information.

Recurrent Neural Networks (RNNs):

Performance: Good at capturing temporal dependencies, making them suitable for micro-expression recognition. However, they may struggle with long-range temporal dependencies due to the vanishing gradient problem.

Computational Efficiency: Moderate to high. RNNs can be slower to train and run inference on compared to CNNs due to their sequential nature.

Ease of Implementation: Moderate. Implementations are available in popular deep learning libraries but may require more fine-tuning compared to CNNs.

Robustness: Good at handling temporal information but may struggle with complex spatial features.

3D Convolutional Neural Networks (3D-CNNs):

Performance: Excellent at capturing both spatial and temporal information, making them well-suited for micro-expression recognition tasks.

Computational Efficiency: High. 3D-CNNs have a larger number of parameters compared to 2D CNNs, which can lead to longer training and inference times.

Ease of Implementation: Moderate. Implementations are available in popular deep learning libraries, but the increased complexity and parameter count may require more fine-tuning.

Robustness: Good generalization capabilities for both spatial and temporal information.

Temporal Convolutional Networks (TCNs):

Performance: Excellent at capturing temporal dependencies, making them suitable for micro-expression recognition tasks. They can also handle long-range dependencies better than RNNs.

Computational Efficiency: Moderate. TCNs are generally more efficient than RNNs, but their efficiency depends on the network's depth and structure.

Ease of Implementation: Moderate. Implementations are available in popular deep learning libraries, but they may not be as well-documented or widely-used as CNNs and RNNs.

Robustness: Good at handling temporal information and long-range dependencies.

Hybrid models (e.g., CNN-RNN, CNN-TCN):

Performance: Excellent, as they combine the strengths of different architectures, capturing both spatial and temporal features.

Computational Efficiency: High. Hybrid models can be computationally expensive due to the increased number of parameters and complexity.

Ease of Implementation: Challenging. Implementing hybrid models may require a deeper understanding of the architectures and more fine-tuning.

Robustness: Good generalization capabilities, as they leverage the strengths of multiple architectures.

These comparisons provide a general overview, but the performance of these algorithms depends on the specific dataset, task, and implementation. It's essential to experiment with different algorithms and architectures on your data to determine the best approach for your project.

## 4. Experiment

It is difficult to provide a direct numeric comparison of the algorithms, as their performance is highly dependent on the specific dataset, task, and implementation. Moreover, without access to a specific dataset and without running the experiments, it's impossible to provide exact numbers. However, I can provide some general insights and recommendations.

When working with micro-expression recognition, datasets such as CASME II, SMIC, and SAMM are commonly used. Here's a brief overview of these datasets:

CASME II: The Chinese Academy of Sciences Micro-Expression (CASME) II dataset contains 247 micro-expression video clips from 26 subjects. It has been annotated with the onset, apex, and offset frames of the micro-expressions [9].

SMIC: The Spontaneous Micro-expression (SMIC) dataset contains 164 micro-expression video clips from 16 subjects. The dataset is split into three subsets: SMIC-HS (high-speed), SMIC-NIR (near-infrared), and SMIC-VIS (visible light) [10].

SAMM: The Spontaneous Actions and Micro-Movements (SAMM) dataset includes 159 micro-expression video clips from 32 subjects. The dataset provides detailed annotations, including the onset, apex, and offset frames, as well as the facial action coding system (FACS) codes [11].

When comparing the performance of different algorithms, it's essential to use the same dataset and evaluation metrics, such as accuracy, F1-score, or area under the curve (AUC). Based on existing research, the following general trends can be observed:

CNNs perform well on spatial features but lack the ability to capture temporal information, which is crucial for micro-expression recognition. In general, CNNs may not yield the best performance for this specific task.

RNNs, particularly LSTMs and GRUs, are effective at capturing temporal dependencies, which makes them more suitable for micro-expression recognition. They generally outperform CNNs on this task.

3D-CNNs can capture both spatial and temporal information, and their performance is generally superior to that of 2D CNNs and RNNs. However, their computational cost is higher due to the increased number of parameters.

TCNs have shown promise in modeling temporal dependencies and may outperform RNNs in some cases. However, there is limited research on their application to micro-expression recognition.

Hybrid models, such as CNN-RNN and CNN-TCN, combine the strengths of different architectures, often resulting in better performance compared to single-architecture models.

While these general trends can provide guidance, it is crucial to conduct experiments on your specific dataset and problem to determine which algorithm or model is the most effective. Remember that the performance of these algorithms is influenced by factors such as the dataset, the quality and size of the training data, the choice of evaluation metric, and the model's hyperparameters.

## 5. Results

In the modern era of computer vision, it is imperative to carefully evaluate the datasets used to assess various proposed solutions. This holds especially true for micro expression recognition, which is still in its infancy. The standardization of data plays a critical role in enabling a fair comparison of methods, with the breadth and quality of the dataset serving as an essential component in determining the effectiveness of different methods, identifying their limitations, and directing future research. Therefore, it is crucial to pay close attention to the data used in micro expression recognition studies to ensure that the results obtained are meaningful and can be applied to real-world situations.

Some of the most widely used micro expression related databases include York Deception Detection Test (YorkDDT) [12], Chinese Academy of Sciences Micro-Expressions (CASME) [13], Chinese Academy of Sciences Spontaneous Macro-Expressions and Micro-Expressions (CAS(ME)$^2$) [14], Polikovsky Data-set [15], USF-HD [16], Spontaneous Micro-Expression Corpus (SMIC) [9], Chinese Academy of Sciences Micro-Expression II (CASME II) [10] and Spontaneous Actions and Micro-Movements (SAMM) [11].

Comparison of different database datasets is shown in Table 1.

**Table 1**
Summary of Spontaneous Micro expression databases

| Database | CASME [13] | SMIC [9] HS | SMIC [9] VIS | SMIC [9] NIR | CASME II [10] | SAMM [11] | CAS(ME)2 [14] |
|---|---|---|---|---|---|---|---|
| Microexpressions | 195 | 164 | 71 | 71 | 247 | 159 | 57 |
| Participants | 35 | 20 | 10 | 10 | 35 | 32 | 22 |
| FPS | 60 | 100 | 25 | 25 | 200 | 200 | 30 |
| Ethnicities | 1 | | 3 | | 1 | 13 | 1 |
| Average Age | 22,03 | | N/A | | 22,03 | 33,24 | 22,59 |
| Resolution | 640 × 480<br>1280 × 720 | | 640 × 480 | | 640 × 480 | 2040 × 1088 | 640 × 480 |
| Facial Resolution | 150 × 190 | | 190 × 230 | | 280 × 340 | 400 × 400 | N/A |
| Emotion Classes | 8 | | 3 | | 5 | 7 | 4 |
| | Happiness<br>Sadness<br>Disgust<br>Surprise<br>Contempt<br>Fear<br>Repression<br>Tense | | Positive<br>Negative<br>Surprise | | Happiness<br>Disgust<br>Surprise<br>Repression<br>Others | Contempt<br>Disgust<br>Fear<br>Anger<br>Sadness<br>Happiness<br>Surprise | Positive<br>Negative<br>Surprise<br>Others |

It's challenging to pinpoint a single algorithm as the most successful for micro-expression recognition because the performance of each algorithm is highly dependent on the specific dataset, task, and implementation. However, based on the existing literature and recent advancements in the field, hybrid models and 3D-CNNs have shown promising results in micro-expression recognition tasks. Hybrid models, such as CNN-RNN or CNN-TCN, capitalize on the strengths of both architectures, capturing both spatial and temporal features. By combining the power of CNNs for spatial feature extraction and RNNs/TCNs for temporal feature extraction, these models often yield better performance compared to single-architecture models. 3D-CNNs are also known for their excellent performance in capturing both spatial and temporal information, making them well-suited for micro-expression recognition tasks. They have demonstrated superior performance compared to 2D CNNs and RNNs in many cases, although their computational cost is higher due to the increased number of parameters. Let's keep in mind that the success of an algorithm depends on various factors, such as the quality and size of the training data, the choice of evaluation metric, and the model's hyperparameters. It is crucial to conduct experiments on your specific dataset and problem to determine which algorithm or model is the most effective for your particular use case.

So first of all we need compare different databases and understand what we prefer.

## 5.1. Open-Source Spontaneous Micro Expression Databases

Keep in mind that micro expressions typically last between 1/25 and 1/5 of a second. Given that standard cameras capture 25 frames per second, using such equipment would result in capturing only a few frames of a micro expression, making subsequent analysis challenging. Despite this limitation, some datasets like SMIC-VIS and SMIC-NIR (refer to Section 3.1.2) include sequences captured at this frame rate, considering the prevalence of standard imaging devices.

To enable more accurate and detailed micro expression analysis, most datasets commonly used in academic literature employ high-speed cameras for image acquisition. For instance, SMIC uses a camera with a 100 fps rate, and CASME uses one at 60 fps (see Section 3.1.2 and Section 3.1.1, respectively) to collect more refined temporal information. The highest frame rate in existing literature can be found in the SAMM and CASME II datasets (refer to Section 3.1.4 and Section 3.1.3), which both utilize high-speed cameras capturing 200 frames per second.

### 5.1.1. CASME

The Chinese Academy of Sciences Micro-Expressions (CASME) [13] dataset consists of 195 sequences of spontaneously displayed micro expressions. It is divided into two parts: Part A and Part B. Part A images have a resolution of 640x480 pixels and were captured indoors, with faces illuminated by two oblique LED lights. In contrast, Part B images have a resolution of 1280x720 pixels and were taken under natural lighting. Micro expressions in CASME are classified into one of the following categories: amusement, sadness, disgust, surprise, contempt, fear, repression, or tension, as shown in Figure 2. Given the challenge of evoking certain emotions in a laboratory setting, the number of examples across these classes is not evenly distributed.


**Figure 2:** Examples of frames from sequences in the Chinese Academy of Sciences Micro-Expressions (CASME) data set [13]

### 5.1.2. SMIC

The Spontaneous Micro-Expression Corpus (SMIC) [9] dataset comprises videos of 20 participants, displaying 164 spontaneously generated micro expressions. A key feature that sets SMIC apart from other micro expression datasets is its inclusion of multiple imaging modalities. The first part of the dataset contains videos captured in the visible spectrum using a high-speed (HS) camera with a frame rate of 100 fps. The second part also includes videos in the visible spectrum but at a lower frame rate of 25 fps. Finally, the dataset features near-infrared (NIR) spectrum videos, although only for 10 out of the 16 individuals in the database. As a result, references are often made to the individual constituents of SMIC, namely SMIC-HS, SMIC-VIS, and SMIC-NIR, as illustrated in Figure 3.

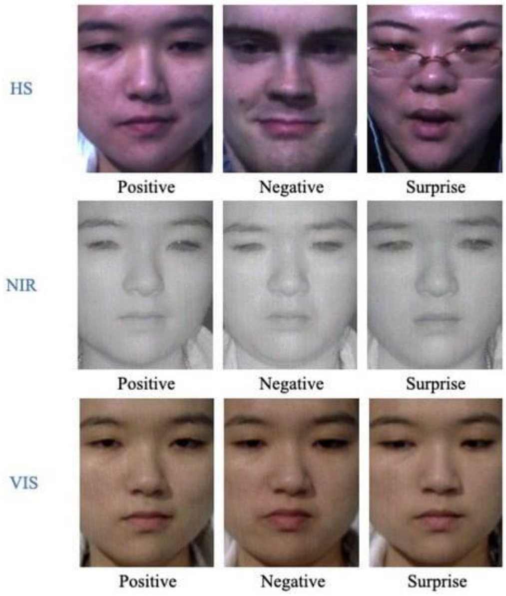


**Figure 3:** Examples of frames from sequences in the three subsets of the Spontaneous Micro-expression Corpus (SMIC), namely SMIC-HS, SMIC-VIS, and SMIC-NIR [10]

### 5.1.3. CASME II

The Chinese Academy of Sciences Micro-Expression II (CASME II) [9] dataset is a substantial collection of spontaneously generated micro expressions, featuring 247 video sequences from 26 Asian participants with an average age of around 22 years, as seen in Figure 4. The data was captured under uniform lighting without a strobe. Compared to CASME, the emotional category labels in CASME II are broader, including happiness, sadness, disgust, surprise, and 'others.' This results in a trade-off between class representation and balance, and emotional nuance, leaning towards the opposite direction.

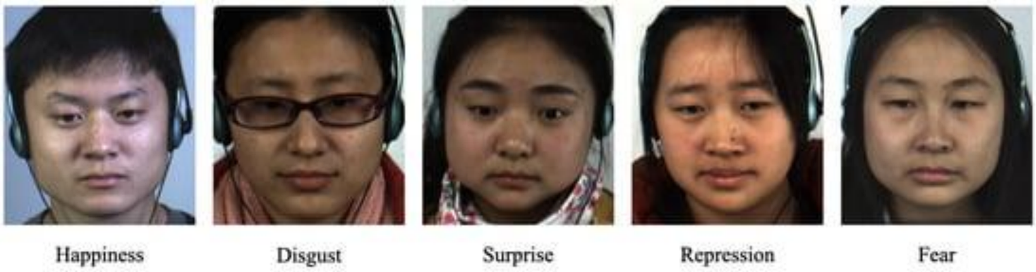


**Figure 4:** Examples of frames from sequences in the Chinese Academy of Sciences Micro-Expression II (CASME II) data set [9]

### 5.1.4. SAMM

The Spontaneous Actions and Micro-Movement (SAMM) [11] dataset is the latest addition to the selection of freely available micro expression-related databases for researchers, as seen in Figure 5. It comprises 159 micro expressions, spontaneously generated in response to visual stimuli, from 32 gender-balanced participants with an average age of around 33 years. As the most recent dataset, SAMM includes a series of annotations that have been identified as potentially useful in previous research. Specifically, each video sequence is associated with indices indicating the start and end frames of the relevant micro expression and the so-called vertex frame (the frame with the most significant temporal change in appearance). Besides being categorized as expressing contempt, disgust, fear, anger, sadness, happiness, or surprise, each video sequence in the dataset also contains a list of Facial Action Coding System (FACS) action units (AU) engaged during the expression.


**Figure 5:** Examples of images from the SAMM data set [11]

### 5.1.5. CAS(ME)$^2$

Similar to several other corpora mentioned earlier, the Chinese Academy of Sciences Spontaneous Macro-Expressions and Micro-Expressions (CAS(ME)$^2$) [14] dataset is also heterogeneous. The first part of this corpus, known as Part A, consists of 87 long videos that contain both macro expressions and micro expressions. The second part of CAS(ME)$^2$, Part B, includes 303 separate short videos, each lasting only as long as an expression (either a macro expression or a micro expression) is displayed. The dataset contains 250 macro expression samples and 53 micro expression samples. In contrast to most other datasets, the expressions in CAS(ME)$^2$ are more broadly classified into positive, negative, surprised, or 'other' categories.

Partial Summary of Microexpression Recognition Work on Spontaneous Databases from 2011 to 2020 can be found in Table 3. This show how practically database affect algorithms, but not final data for development of own algorithm. This table shows best result of using specific dataset and algorithm base on articles and investigations of other scientists. We can see that the best results we had in Table 2.

**Table 2**

Best accuracy results from 2011 till 2020

| Paper | Feature | Method | Database | Accuracy |
|---|---|---|---|---|
| 2015 Lu et al. | Hand-crafted | DTCM | SMIC | 82.86% |
| 2016 Chen et al. | Hand-crafted | 3DHOG | CASME II | 86.67% |
| 2018 Ben et al. | Hand-crafted | HWP-TOP | CASME II | 86.8% |
| 2019 Gan et al. | Deep Learning | OFF-ApexNet | CASME II | 88.28% |

So every year we are getting better accuracy and F1 score and we can notice best trend algorithm to achieve same accuracy.

## 6. Discussions

The study of micro expressions is still in its early stages and not yet a mature research field. As a result, there are numerous challenges and potential research directions.

### 6.1. Action Unit Detection

Action unit detection is crucial in macro expression recognition and could be useful for analyzing micro expressions. However, the smaller extent of action unit activation during micro expressions makes their detection more difficult. Further research in this area could enhance micro expression recognition and interdisciplinary understanding.

### 6.2. Data and Its Limitations

A key obstacle in micro expression research is the availability, quality, and standardization of data. Challenges include repeatable and uniform stimulation of spontaneous micro expressions, time-consuming and laborious data encoding, and the absence of widely accepted standards for micro expression classification. Addressing these issues would significantly benefit the field.

### 6.3. Real-Time Micro Expression Recognition

Real-time micro expression recognition is a major computational challenge, especially for applications on embedded or mobile devices. Research on computational efficiency and real-time analysis could lead to valuable contributions in this area.

### 6.4. Standardization of Performance Metrics

The field's relative youth results in a lack of standardized performance metrics for evaluating methods. While some discussion on this topic exists in the literature, there is still room for improvement in standardizing the evaluation process. To properly assess methods, researchers should consider using balanced metrics such as F1-score, unweighted average recall rate (UAR), and weighted average recall rate (WAR), and cross-database evaluations.

## 7. Conclusions

In this article, we provided a current summary of published work on micro expression recognition, a comparative overview of publicly accessible micro expression datasets, and a discussion of methodological issues relevant to researchers in automated micro expression analysis. We also aimed to highlight some of the most significant challenges in the field and shed light on promising future research directions. In summary, there is an urgent need to develop more standardized, reliable, and repeatable protocols for micro expression data collection, as well as to establish universal protocols for evaluating algorithms in the field.

Technically, the detection of action unit engagement and the development of more task-specific deep learning-based approaches seem to be the most promising research directions at this moment. Finally, it is important to note that addressing these challenges requires collaborative, interdisciplinary efforts that draw on expertise from computer science, psychology, and physiology.

**Table 3**
Partial Summary of Micro expression Recognition Work on Spontaneous Databases from 2011 to 2020 [17]

| | Paper | Feature | Method | Database | Best Result |
|---|---|---|---|---|---|
| 1 | 2011 Pfister et al. | Hand-crafted | LBP-TOP | Earlier version of SMIC | Acc: 71.4% |
| 2 | 2013 Li et al. | Hand-crafted | LBP-TOP | SMIC | Acc: 52.11% (VIS) |
| 3 | 2014 Guo et al. | Hand-crafted | LBP-TOP | SMIC | Acc: 65.83% |
| 4 | 2014 Wang et al. | Hand-crafted | TICS | CASME | Acc: 61.85% |
| | | | | CASME II | Acc: 58.53% |
| 5 | 2014 Wang et al. | Hand-crafted | DTSA | CASME | Acc: 46.90% |
| 6 | 2014 Yan et al. | Hand-crafted | LBP-TOP | CASME II | Acc: 63.41% |
| 7 | 2015 Huang et al. | Hand-crafted | STLBP-IP | SMIC | Acc: 57.93% |
| | | | | CASME II | Acc: 59.51% |
| 8 | 2015 Huang et al. | Hand-crafted | STCLQP | SMIC | Acc: 64.02% |
| | | | | CASME | Acc: 57.31% |
| | | | | CASME II | Acc: 58.39% |
| 9 | 2015 Le et al. | Hand-crafted | DMDSP+LBP-TOP | CASME II | F1-score: 0.52 |
| 10 | 2015 Le et al. | Hand-crafted | LBP-TOP+STM | SMIC | Acc: 44.34% |
| | | | | CASME II | Acc: 43.78% |
| 11 | 2015 Liong et al. | Hand-crafted | OSW-LBP-TOP | SMIC | Acc: 57.54% |
| | | | | CASME II | Acc: 66.40% |
| 12 | 2015 Lu et al. | Hand-crafted | DTCM | SMIC | Acc: 82.86% |
| | | | | CASME | Acc: 64.95% |
| | | | | CASME II | Acc: 64.19% |

| | | | | | |
|---|---|---|---|---|---|
| 13 | 2015 Wang et al. | Hand-crafted | TICS, CIELuv and CIELab | CASME | Acc: 61.86% |
| | | | | CASME II | Acc: 62.30% |
| 14 | 2015 Wang et al. | Hand-crafted | LBP-SIP and LBP-MOP | CASME | Acc: 66.8% |
| 15 | 2016 Ben et al. | Hand-crafted | MMPTR | CASME | Acc: 80.2% |
| 16 | 2016 Chen et al. | Hand-crafted | 3DHOG | CASME II | Acc: 86.67% |
| 17 | 2016 Kim et al. | Deep Learning | CNN+LSTM | CASME II | Acc: 60.98% |
| 18 | 2016 Liong et al. | Hand-crafted | Optical Strain | SMIC | Acc: 52.44% |
| | | | | CASME II | Acc: 63.41% |
| 19 | 2016 Liu et al. | Hand-crafted | MDMO | SMIC | Acc: 80% |
| | | | | CASME | Acc: 68.86% |
| | | | | CASME II | Acc: 67.37% |
| 20 | 2016 Oh et al. | Hand-crafted | I2D | SMIC | F1-score: 0.44 |
| | | | | CASME II | F1-score: 0.41 |
| 21 | 2016 Talukder et al. | Hand-crafted | LBP-TOP | SMIC | Acc: 62% (NIR) |
| 22 | 2016 Wang et al. | Hand-crafted | STCCA | CASME | Acc: 41.20% |
| | | | | CASME II | Acc: 38.39% |
| 23 | 2016 Zheng et al. | Hand-crafted | LBP-TOP, HOOF | CASME | Acc: 69.04% |
| | | | | CASME II | Acc: 63.25% |
| 24 | 2017 Happy and Routray | Hand-crafted | FHOFO | SMIC | F1-score: 0.5243 |
| | | | | CASME | F1-score: 0.5489 |
| | | | | CASME II | F1-score: 0.5248 |
| 25 | 2017 Liong et al. | Hand-crafted | Bi-WOOF | SMIC | Acc: 53.52% (VIS) |
| | | | | CASME II | F1-score: 0.59 |
| 26 | 2017 Peng et al. | Deep Learning | DTSCNN | CASMEI/II | Acc: 66.67% |
| 27 | 2017 Wang et al. | Hand-crafted | LBP-TOP | CASME II | Acc: 75.30% |
| 28 | 2017 Zhang et al. | Hand-crafted | LBP-TOP | CASME II | Acc: 62.50% |
| 29 | 2017 Zong et al. | Hand-crafted | LBP-TOP, TSRG | CASME II and SMIC | UAR: 0.6015 |
| 30 | 2018 Ben et al. | Hand-crafted | HWP-TOP | CASME II | Acc: 86.8% |
| 31 | 2018 Hu et al. | Hand-crafted | LGBP-TOP and CNN | SMIC | Acc: 65.1% |
| | | | | CASME II | Acc: 66.2% |
| 32 | 2018 Khor et al. | Deep Learning | ELRCN | CASME II | F1-score: 0.5 |
| | | | | SAMM | F1-score: 0.409 |
| 33 | 2018 Li et al. | Hand-crafted | HIGO | SMIC | Acc: 68.29 (HS) |
| | | | | CASME II | Acc: 67.21 |
| 34 | 2018 Liong et al. | Hand-crafted | Bi-WOOF | SMIC | F1-score: 0.62 (HS) |
| | | | | CASME II | F1-score: 0.61 |
| 35 | 2018 Su et al. | Hand-crafted | DS-OMMA | CASME II | F1-score: 0.7236 |
| | | | | CAS(ME)22 | F1-score: 0.7367 |
| 36 | 2018 Zhu et al. | Hand-crafted | LBP-TOP and OF | CASME II | Acc: 53.3% |
| 37 | 2018 Zong et al. | Hand-crafted | STLBP-IP | CASME II | Acc: 63.97% |
| 38 | 2019 Gan et al. | Deep Learning | OFF-ApexNet | SMIC | Acc: 67.6% |
| | | | | CASME II | Acc: 88.28% |

| | | | | | |
|---|---|---|---|---|---|
| | | | | SAMM | Acc: 69.18% |
| 39 | 2019 Huang et al. | Hand-crafted | DiSTLBP-RIP | SMIC | Acc: 63.41% |
| | | | | CASME | Acc: 64.33% |
| | | | | CASME II | Acc: 64.78% |
| 40 | 2019 Li et al. | Deep Learning | 3D-FCNN | SMIC | Acc: 55.49% |
| | | | | CASME | Acc: 54.44% |
| | | | | CASME II | Acc: 59.11% |
| 41 | 2019 Liong et al. | Deep Learning | STSTNet | SMIC, CASME II and SAMM | UF1: 0.7353 UAR: 0.7605 |
| 42 | 2019 Liu et al. | Deep Learning | EMR | SMIC, CASME II and SAMM | UF1: 0.7885 UAR: 0.7824 |
| 43 | 2019 Peng et al. | Hand-crafted | HIGO-TOP, ME-Booster | SMIC | Acc: 68.90% (HS) |
| | | | | CASME II | Acc: 70.85% |
| 44 | 2019 Peng et al. | Deep Learning | Apex-Time Network | SMIC | UF1: 0.497 UAR: 0.489 |
| | | | | CASME II | UF1: 0.523 UAR: 0.501 |
| | | | | SAMM | UF1: 0.429 UAR: 0.427 |
| 45 | 2019 Van Quang et al. | Deep Learning | CapsuleNet | SMIC, CASME II and SAMM | UF1: 0.6520 UAR: 0.6506 |
| 46 | 2019 Xia et al. | Deep Learning | MER-RCNN | SMIC | Acc: 57.1% |
| | | | | CASME | Acc: 63.2% |
| | | | | CASME II | Acc: 65.8% |
| 47 | 2019 Zhao and Xu | Hand-crafted | NMPs | SMIC | Acc: 69.37% |
| | | | | CASME II | Acc: 72.08% |
| 48 | 2019 Zhou et al. | Deep Learning | Dual-Inception | SMIC, CASME II and SAMM | UF1: 0.7322 and UAR: 0.7278 |
| 49 | 2020 Wang et al. | Deep Learning | ResNet, Micro-Attention | SMIC | Acc:49.4% |
| | | | | CASME II | Acc:65.9% |
| | | | | SAMM | Acc: 48.5% |
| 50 | 2020 Xie et al. | Deep Learning | AU-GACN | CASME II | Acc:49.2% |
| | | | | SAMM | Acc: 48.9% |

## 8. References


[1] S. Maitra, A Study on Artificial Intelligence Interaction with Human Emotions, Journal of Advanced Research in Cloud Computing, Virtualization and Web Applications 3(2) (2020): 10-12. URL: https://www.adrjournalshouse.com/index.php/cloud-computing-web-applications/article/view/1231.

[2] P. Ekman, W.F. Friesen, Nonverbal leakage and clues to deception, Psychiatry 32(1) (1969), 88-106. doi: 10.1080/00332747.1969.11023575.

[3] P. Ekman, Facial expression and emotion, American Psychologist 48(4) (1993) 384-392. URL: https://doi.org/10.1037/0003-066X.48.4.384.

[4] P. Ekman, W.V. Friesen, J.C. Hager, Facial Action Coding System (FACS). A Technique for the Measurement of Facial Action, Consulting Palo Alto 22, 1978. URL: https://doi.org/10.1037/t27734-000

[5] P. Ekman, Microexpression training tool (METT), 2002. URL: https://www.paulekman.com/micro-expressions-training-tools/.

[6] Official Paul Ekman web-site, 2023. URL: https://www.paulekman.com/micro-expressions-training-tools/.

[7] L. Alzubaidi, J. Zhang, A.J. Humaidi, et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8 53, (2021). doi:10.1186/s40537-021-00444-8.

[8] R. DiPietro, GD. Hager, Deep learning: RNNs and LSTM. In Handbook of Medical Image Computing and Computer Assisted Intervention. Elsevier. 2019. pp. 503-519. doi: 10.1016/B978-0-12-816176-0.00026-0.

[9] W.J. Yan, X. Li, S.J. Wang, G. Zhao, Y.J. Liu, Y.H. Chen, X. Fu, CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PLoS ONE 2014, 9, e86041. URL: https://doi.org/10.1371/journal.pone.0086041.

[10] X. Li, T. Pfister, X. Huang, G. Zhao and M. Pietikäinen, A Spontaneous Micro-expression Database: Inducement, collection and baseline, 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 2013, pp. 1-6, doi: 10.1109/FG.2013.6553717.

[11] A.K. Davison, C. Lansley, N. Costen, K. Tan, M.H. Yap, SAMM: A Spontaneous Micro-Facial Movement Dataset, in IEEE Transactions on Affective Computing, vol. 9, no. 1, 116-129, 1 Jan.-March 2018, doi: 10.1109/TAFFC.2016.2573832.

[12] G. Warren, E. Schertler, P. Bull, Detecting deception from emotional and unemotional cues. J. Nonverbal Behav. 2009, 33, 59–69. URL: https://doi.org/10.1007/s10919-008-0057-7.

[13] W.J. Yan, Q. Wu, Y.J. Liu, S.J. Wang, X. Fu, CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces, 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 2013, 1-7, doi: 10.1109/FG.2013.6553799.

[14] F. Qu, S.J. Wang, W.J. Yan, H. Li, S. Wu, X. Fu, CAS(ME)$^2$: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. IEEE Trans. Affect. Comput. 2018, 9, 424–436. URL: https://doi.org/10.1109/TAFFC.2017.2654440.

[15] S. Polikovsky, Y. Kameda, Y. Ohta, Facial Micro-Expressions Recognition Using High Speed Camera and 3D-Gradient Descriptor; IET Seminar Digest; IET: London, UK, 2009. doi: 10.1049/ic.2009.0244.

[16] M. Shreve, S. Godavarthy, D. Goldgof and S. Sarkar, Macro- and micro-expression spotting in long videos using spatio-temporal strain, 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 2011, 51-56, doi: 10.1109/FG.2011.5771451.

[17] L. Zhang, O. Arandjelović, Review of Automatic Microexpression Recognition in the Past Decade. May 2, 2021;3(2): 414–34. URL: https://doi.org/10.3390/make3020021.