Interdisciplinary Journal of Information, Knowledge, and Management

An Official Publication of the Informing Science Institute InformingScience.org

IJIKM.org

Volume 19, 2024

IMPROVING THE ACCURACY OF FACIAL MICRO-EXPRESSION RECOGNITION: SPATIO-TEMPORAL DEEP LEARNING WITH ENHANCED DATA AUGMENTATION AND CLASS BALANCING

Budhi Irawan*	Telkom University, Bandung, Indonesia and Bandung Institute of Technology, Bandung, Indonesia	budhiirawan@telkomuniversity.ac.id
Rinaldi Munir	Bandung Institute of Technology, Bandung, Indonesia	rinaldi@informatika.org
Nugraha Priya Utama	Bandung Institute of Technology, Bandung, Indonesia	utama@informatika.org
Ayu Purwarianti	Bandung Institute of Technology, Bandung, Indonesia	ayu@itb.ac.id

* Corresponding author

ABSTRACT

Aim/Purpose	This study presents a novel deep learning-based framework designed to enhance spontaneous micro-expression recognition by effectively increasing the amount and variety of data and balancing the class distribution to improve recognition accuracy.
Background	Micro-expression recognition using deep learning requires large amounts of data. Micro-expression datasets are relatively small, and their class distribution is not balanced.
Methodology	This study developed a framework using a deep learning-based model to recognize spontaneous micro-expressions on a person's face. The framework also includes several technical stages, including image and data preprocessing. In data preprocessing, data augmentation is carried out to increase the amount and variety of data and class balancing to balance the distribution of sample classes in the dataset.

Accepting Editor Geoffrey Z. Liu │ Received: June 17, 2024 │ Revised: October 15, 2024 │ Accepted: October 16, 2024.

Cite as: Irawan, B., Munir, R., Utama, N. P., & Purwarianti, A. (2024). Improving the accuracy of facial micro-expression recognition: Spatio-temporal deep learning with enhanced data augmentation and class balancing. *Interdisciplinary Journal of Information, Knowledge, and Management, 19,* Article 31. https://doi.org/10.28945/5386

Contribution	This study's essential contribution lies in enhancing the accuracy of micro-expression recognition and overcoming the limited amount of data and imbalanced class distribution that typically leads to overfitting.
Findings	The results indicate that the proposed framework, with its data preprocessing stages and deep learning model, significantly increases the accuracy of micro-expression recognition by overcoming dataset limitations and producing a balanced class distribution. This leads to improved micro-expression recognition accuracy using deep learning techniques.
Recommendations for Practitioners	Practitioners can utilize the model produced by the proposed framework, which was developed to recognize spontaneous micro-expressions on a person's face, by implementing it as an emotional analysis application based on facial micro-expressions.
Recommendations for Researchers	Researchers involved in the development of a spontaneous micro-expression recognition framework for analyzing hidden emotions from a person's face are playing an essential role in advancing this field and continue to search for more innovative deep learning-based solutions that continue to explore techniques to increase the amount and variety of data and find solutions to balancing the number of sample classes in various micro-expression datasets. They can further improvise to develop deep learning model architectures that are more suitable and relevant according to the needs of recognition tasks and the various characteristics of different datasets.
Impact on Society	The proposed framework could significantly impact society by providing a reliable model for recognizing spontaneous micro-expressions in real-world applications, ranging from security systems and criminal investigations to healthcare and emotional analysis.
Future Research	Developing a spontaneous micro-expression recognition framework based on spatial and temporal flow requires the learning model to classify optimal features. Our future work will focus more on exploring micro-expression features by developing various alternative learning models and increasing the weights of spatial and temporal features.
Keywords	micro-expression, spatio-temporal, data augmentation, class balancing, 3DCNN, vision, deep learning

INTRODUCTION

Micro-expressions are brief, involuntary facial expressions that reveal genuine emotions, often lasting less than one second. They are challenging to recognize due to their rapid occurrence and subtle nature. An expert trained using micro-expression training tools (M. Wei et al., 2022) will still need help recognizing micro-expressions that arise in someone. Micro-expressions are facial movements that reveal a person's genuine emotions but try to hide them (Sun et al., 2022; Yang et al., 2022). Micro-expressions are uncontrolled facial expressions that only last for a period of 1/15 second to 1/25 second (N. Liu et al., 2020). Genuine emotions expressed through micro-expressions are usually ignored because they are invisible and disguised. Micro-expressions occur when someone deliberately tries to hide their feelings from other people.

Despite advancements in micro-expression recognition, significant challenges still need to be addressed, particularly in achieving high accuracy and performance due to limited datasets and imbalanced class distribution. This condition directly impacts the model's ability to recognize and understand various micro-expressions. The presence of these problems highlights the urgency of carrying out updates in related work to overcome these challenges. The scientific literature still needs to offer adequate solutions to these difficulties in micro-expression recognition. Therefore, further work is fundamental to developing new strategies to increase the accuracy of micro-expression recognition and reduce the risk of overfitting in the developed models.

In contrast to micro-expressions, macro-expressions are facial expressions that last longer, from half a second to several seconds, so they are easily recognized and analyzed by others. These expressions occur consciously and can be controlled or faked, such as smiling to show friendliness or frowning to indicate concern. Macro-expressions express emotions such as happiness, sadness, anger, fear, surprise, and disgust in a clear and long-lasting way, helping in communication and building relationships between individuals. They play an essential role in social interactions by conveying emotions and intentions.

RELATED WORK

Several traditional machine-learning approaches have been proposed in micro-expression recognition to address the challenges of recognizing subtle facial expressions. Some research studies propose different traditional machine-learning approaches for micro-expression recognition. The first study (Pan et al., 2020) uses a Hierarchical Support Vector Machine (H-SVM) to handle sample category distribution imbalances, emphasizing a multi-level fusion of features to enhance recognition accuracy. The second study (K.-H. Liu et al., 2021) employs a genetic algorithm (GA) to optimize feature extraction, combining it with a convolutional neural network (CNN) to achieve better recognition performance. The third study (J. Wei et al., 2022) introduces a Local Binary Pattern from the Five Intersecting Planes (LBP-FIP) approach, which captures dynamic texture features in horizontal, vertical, and oblique directions, improving traditional LBP-TOP methods.

Traditional machine learning methods, such as SVM, Genetic Algorithm, and LBP-FIP, offer advantages in interpretability and computational speed but often need help handling large-scale and high-dimensional data. The complexity of micro-expressions and the need to process subtle variations exceed these methods' capabilities, leading to limitations in accuracy and adaptability. These challenges have ultimately driven a shift toward deep learning approaches, which are more adept at handling large datasets and extracting complex patterns in micro-expression data despite requiring extensive computational resources and data for training.

Several deep learning models have since been developed to tackle the challenge of micro-expression recognition, primarily due to the scarcity of video-based training samples and the imbalanced distribution of emotion classes. These deep learning methods often require improvements as they generally need to learn numerous features and parameters to achieve higher accuracy. For example, Teja Reddy et al. (2019) focused on methods to detect video frames containing micro-expressions, explicitly addressing the detection of frames from the onset, apex, to the offset stages of expressions. He et al. (2020) proposed a CNN model combining face detection and recognition with Eulerian Video Magnification (EVM) to enhance micro-expression detection. Moreover, CNN has been effectively used to classify seven universal micro-expressions (Ayyalasomayajula et al., 2021). Jiao et al. (2021) suggested employing 3DCNN architectures to improve micro-expression recognition across various frameworks, with subsequent studies indicating that multiple enhancements to the 3DCNN model could lead to higher recognition accuracy (Bayu & Setyanto, 2022).

Recent advancements in micro-expression recognition have been characterized by further integrating CNN with other approaches to boost recognition accuracy. Guowen and Xi (2023) explored the use of CNN in conjunction with transformer-based models, such as the multiple branch neural networks STCN (Swin Transformer and ConvNeXt), to address localization issues in micro-expression actions and preserve the spatial facial structure. Their experiments on the CASME and SMIC datasets demonstrated that the STCN network significantly improved recognition accuracy. Zhou et al. (2023) introduced the Divided Block Multiscale Convolution Network (BDMCNet), which utilizes optical flow feature images between the onset and apex frames of micro-expression sequences to extract more detailed and multiscale features. Furthermore, Shang et al. (2023) proposed a novel approach with a spatio-temporal capsule network (STCP-Net) that reduces recognition time while maintaining high accuracy. The STCP-Net framework consists of a joint prediction module, differential feature extraction, a spatio-temporal capsule module, and a fully connected layer, enhancing temporal analysis and feature integration.

One of the significant challenges in micro-expression recognition is creating a practical framework that consistently achieves satisfactory recognition accuracy, primarily due to dataset limitations. Despite advancements, the field still needs to work on dataset size, diversity, and the imbalance of class distributions. The initial stages of micro-expression recognition often involve extensive preprocessing to ensure data quality. However, Wang et al. (2023) and Gupta (2023) highlighted that preprocessing can sometimes lead to problems such as missing or damaged features in the datasets, exacerbating recognition difficulties. The current micro-expression datasets are limited and often suffer from unbalanced class distributions, which cause deep learning models to overfit, favouring the majority class while neglecting the minority class. This imbalance reduces the model's effectiveness in recognizing and accurately classifying underrepresented expressions, a problem that remains a critical area of focus in this field.

Efforts to enhance micro-expression recognition have increasingly turned to hybrid approaches that combine deep learning with classical techniques to address internal challenges like limited datasets and class imbalances. The primary goal of these approaches is to develop a practical framework that substantially improves recognition accuracy. This goal involves refining both model architectures and preprocessing techniques. The hybrid approach leverages classical methods alongside deep learning architectures to better handle diverse data patterns, balance class distributions, and optimize model training. Key contributions in this area include advanced preprocessing techniques such as data augmentation, class weight adjustments, and Synthetic Minority Over-sampling Technique (SMOTE) to adjust class imbalances effectively. These methods ensure a more robust recognition system with higher accuracy and performance in practical applications.

METHODOLOGY

This work develops a framework consisting of three main stages: multi-level preprocessing, which includes image and data preprocessing; the second stage is classification; and the final stage involves measuring the level of accuracy, as shown in Figure 1.

STAGE 1: IMAGE PREPROCESSING

In the first stage, image frames from video clips contained in the dataset are processed. This process involves converting each video clip into a series of image frames. Facial landmark detection is then performed using 68 facial landmark features, explicitly focusing on the regions around the eyes and mouth. Black masking is applied to the eyes and mouth regions to reduce distractions. After that, a cropping process is performed on the face area, and the image size is adjusted to 128x128 pixels, followed by conversion to grayscale to standardize the data format.

The emotion classes from each dataset are grouped into three primary categories for better classification: negative, positive, and surprise. For example, in the CASME II dataset, classes such as 'repression,' 'angry,' 'disgust,' 'fear,' and 'sadness' are categorized under the 'negative' class. At the same time, 'happiness' falls under the 'positive' class, and 'surprise' is its class. This simplification helps reduce the complexity of the classification problem, making the model more focused on recognizing significant micro-expressions.

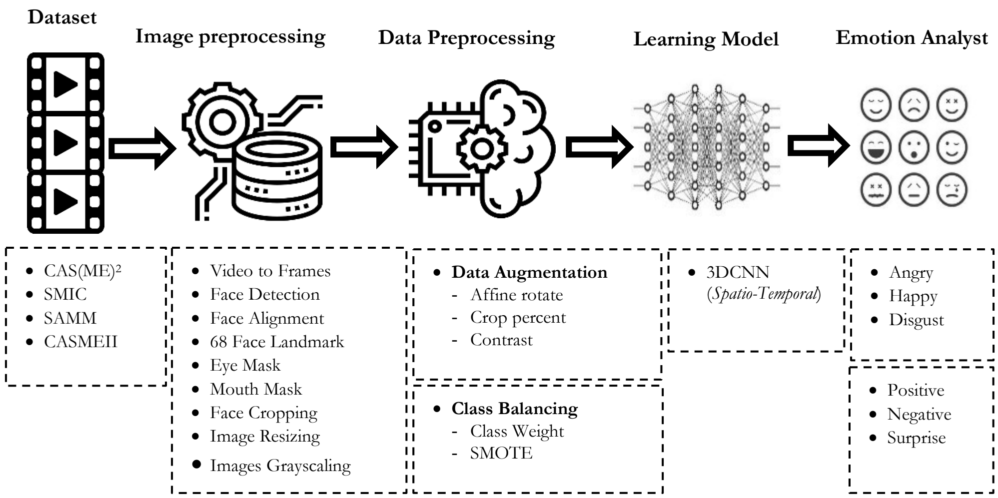

Figure 1. Design of a proposed framework for micro-expression recognition

In the image preprocessing stage, different resolution settings are applied to the dataset, converting video clips into sequential image frames with initial resolutions of 640x480 pixels for the CAS(ME)² and SMIC datasets, 960x560 pixels for the SAMM dataset, and 280x340 pixels for the CASME II dataset, as illustrated in Figure 2.

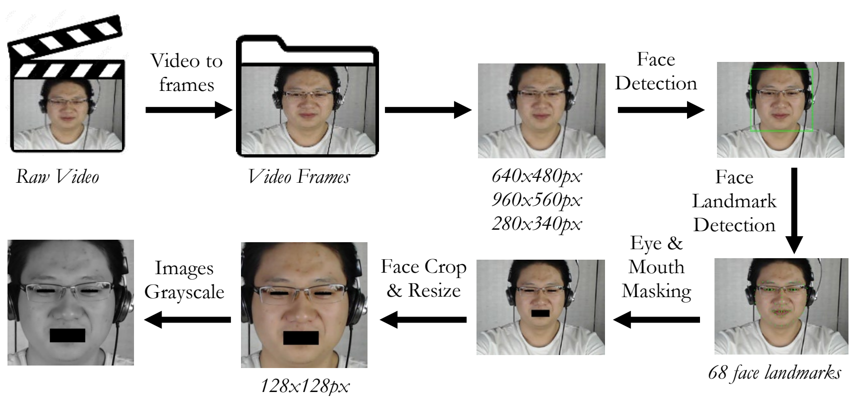

Figure 2. Image preprocessing stages

STAGE 2: DATA PREPROCESSING

The data preprocessing stage uses data augmentation techniques to increase the number and variety of datasets. These techniques include affine rotation, cropping percentages, and contrast and brightness level adjustments. These transformations help to artificially expand the dataset, thereby improving the model's ability to generalize across different variations in micro-expressions.

In addition, class balancing techniques are applied to address the issue of imbalanced class distributions. Two main approaches are used: Class Weight and SMOTE (Synthetic Minority Over-Sampling Technique). The Class Weight method assigns a higher weight to underrepresented classes, guiding the model to focus more on these minority expressions. SMOTE generates synthetic samples for the minority classes, ensuring that these classes are not underrepresented during the training process, as shown in Figure 3.

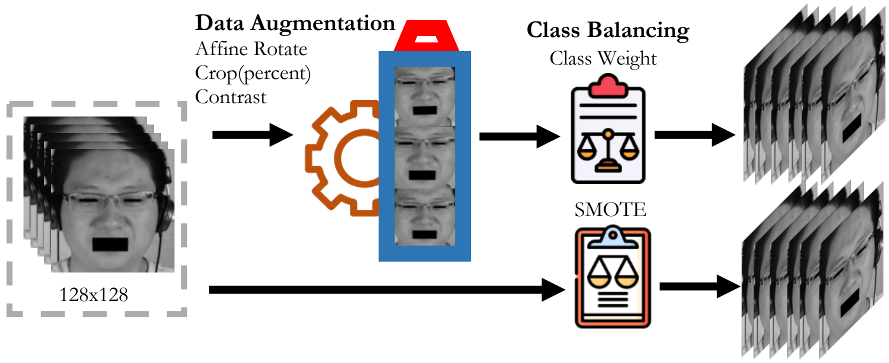

Figure 3. Data preprocessing stages

Stage 3: Classification Using 3DCNN Architecture

The 3DCNN (spatio-temporal convolutional neural network) model is used for micro-expression classification. This model is designed to capture both spatial and temporal features of micro-expressions. The architecture begins with a convolutional layer consisting of 64 filters and a 3x3 kernel size, followed by ReLU activation functions to introduce non-linearity. Pooling layers with a size of 2x2 are applied to reduce the dimensionality of the feature maps, and dropout layers are included to prevent overfitting. The training process utilizes the Adam optimizer with a learning rate of 0.001, and the model is trained over 50 epochs. The design of the 3DCNN classification model in this work is shown in Figure 4.

The framework has four variations, each tailored to emphasize different aspects of data handling and model performance:

1. MER-3DCNNST: Includes only image preprocessing and the classification model.
2. MER-DA3DCNNST: Incorporates image preprocessing, data augmentation, and the classification model.
3. MER-SCB3DCNNST: Combines image preprocessing, SMOTE class balancing, and the classification model.
4. MER-DACWB3DCNNST: Integrates image preprocessing, data augmentation, class weight balancing, and the classification model to provide a comprehensive solution for class imbalance.

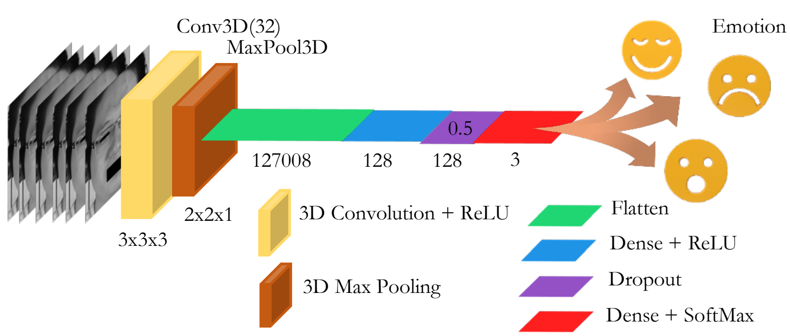

Figure 4. 3DCNN classification model

Data Augmentation

Data augmentation is a crucial technique for increasing the amount of training data and improving the diversity of the dataset. Transformations applied include affine rotations (up to 15 degrees), cropping (10% of the image area), and contrast adjustments (up to 30%). This process creates new variations of existing images, enhancing the model's ability to learn from more diversified data, thus leading to better recognition of subtle micro-expression changes, as illustrated in Figure 5.

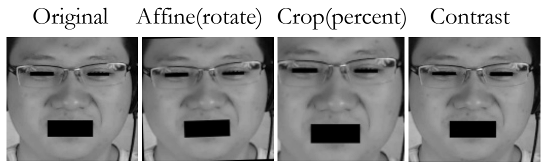

Figure 5. Data augmentation

Class Balancing

Class balancing is crucial for handling disparities in the number of samples across classes in a dataset, especially in classification tasks where imbalanced data can lead to biased models that favor the majority class and overlook minority groups. This issue is particularly problematic in micro-expression recognition, where some classes have significantly fewer samples. Two techniques were applied to address this: Class Weight and SMOTE. The Class Weight method assigns higher weights to minority classes, increasing the model's focus on learning from these underrepresented classes and enhancing its ability to classify them (Sagoolmuang, 2021) correctly. This approach helps mitigate the lower accuracy often associated with infrequent facial expressions, ensuring better recognition and improved performance in identifying these minority classes.

Giving class weights is also expected to help reduce overfitting in the majority class. Giving weight to class will prevent the model from focusing too much on the majority class and help the model be more balanced in learning all classes. Equation formulas for calculating class weights in class imbalance often follow simple approaches involving proportions or comparing sample sizes between classes. The general equation for calculating class weights can be stated by using Eq. (1):

$$W_k = \frac{N}{n_k} \tag{1}$$

where W_k is the class weight for class k, N is the total number of samples in the training data, and n_k is the number of samples in class *k*

From this equation, the class weight for the minority class will be higher than the majority class. This equation creates the inverse of the sample proportions within the class. A superior class weight for a minority class gives that class more influence in the machine learning process, thereby helping to overcome class imbalance.

Class weights allow the model to focus more on minority classes by assigning them higher importance, encouraging the model to make better predictions and reduce errors for these underrepresented classes, ultimately enhancing its accuracy in identifying infrequent classes. This technique helps address class distribution imbalances and promotes a fairer classification process. Additionally, the Synthetic Minority Over-Sampling Technique (SMOTE) is another effective method for class balancing, particularly in micro-expression recognition, where some expressions are less common (Balakrishnan et al., 2022). SMOTE generates synthetic samples for minority classes, helping to balance the dataset and prevent the model from becoming biased toward the majority class.

Experiment Setup

The experiment focuses on optimizing model hyperparameters and ensuring generalizability to real-world conditions. It addresses diverse micro-expression variations to enhance the model's accuracy and robustness.

Hyperparameter Tuning

The ADAM optimizer is utilized for its adaptive learning rate and effective handling of gradient fluctuations, which enhances the training stability of the 3DCNN model. Categorical Cross-Entropy is applied as the loss function due to its accuracy in handling multi-category classification tasks, making it suitable for micro-expression recognition.

A small batch size (16, 24, or 32) improves data handling and reduces overfitting. At the same time, a low number of epochs (5 or 10) is selected to assess model performance quickly without extensive training time. The data is divided into 80% for training, 10% for validation, and 10% for testing, ensuring that the model learns efficiently and generalizes well to new, unseen data.

Evaluation Metrics

In the context of classification evaluation, specific metrics are used to evaluate the model's performance in predicting the target class of a dataset. Some fundamental concepts to comprehend include True Positive, False Positive, True Negative, and False Negative. True Positive (TP) signifies instances where the model accurately predicts a positive sample, aligning with the ground truth. Essentially, TP denotes the count of positive instances correctly predicted by the model. On the other hand, a False Positive (FP) arises when the model erroneously predicts a sample as positive despite being negative. In the context of classification, FP represents the number of negative cases incorrectly predicted as positive by the model. True Negative (TN) occurs when the model correctly predicts a sample as negative, and the prediction matches reality. TN represents the number of negative cases correctly predicted by the model. False Negative (FN) occurs when the model incorrectly predicts a sample as negative, even though the sample is positive. In the context of classification, FN represents the number of positive cases incorrectly predicted as negative by the model.

Evaluation metrics are measures or parameters used to evaluate the performance of a model or algorithm in predicting or classifying data. Evaluation metrics show how well the model solves a given problem. In classification, evaluation metrics generally involve accuracy, precision, recall, and F1-Score.

Accuracy measures the extent to which the classification model correctly classifies the entire dataset. Accuracy is computed by dividing the number of correct predictions (TP + TN) by the total sample count. The formula for accuracy by using Eq. (2):

$$Accuracy = \frac{TP+TN}{Total\ Sample} \tag{2}$$

Precision is a measure of how many of the positive predictions are correct. This is the ratio of TP to the sum of TP and FP. Precision shows how reliable the model is in predicting the positive class. The formula for precision by using Eq. (3):

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall is a measure of how many of all positive values were predicted correctly. This is the ratio of TP to the sum of TP and FN. Recall shows how well the model can identify all positive class instances. The formula for the recall by using Eq. (4):

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

The F1-Score is an assessment metric amalgamating precision and recall into a singular value. It represents a harmonic mean of precision and recall, ensuring equilibrium between the two. F1-Score proves beneficial in scenarios with an imbalance between positive and negative classes within the dataset. The formula for F1-Score by using Eq. (5):

$$F1 - Score = 2 \text{ x } \frac{PrecissionxRecall}{Precission+Recall} \quad (5)$$

Result and Discussion

The results of experiments carried out on four micro-expression datasets by adding data augmentation methods and class balancing methods, as well as scenarios for setting the parameters needed during the learning and classification process, can be presented in graphical form. Each stage of the experiment is carried out by calculating evaluation metrics in accuracy and F1-Score. The graphs presented include information from the entire proposed framework, dataset, image processing, data preprocessing implementation, accuracy, and F1-Score. The type of graph presented is a line graph with additional information on accuracy and F1 score for each experiment based on the dataset used. Apart from knowing the accuracy calculation results and the highest F1 score for each proposed framework, the graph presented also provides a complete table of the results of all experiments that have been carried out.

Figure 6 shows a graph of the accuracy of micro-expression recognition using the 3DCNN model using four datasets and applying the four proposed frameworks MER-3DCNNST, MER-DA3DCNNST, MER-SCB3DCNNST, and DACWB3DCNNST.

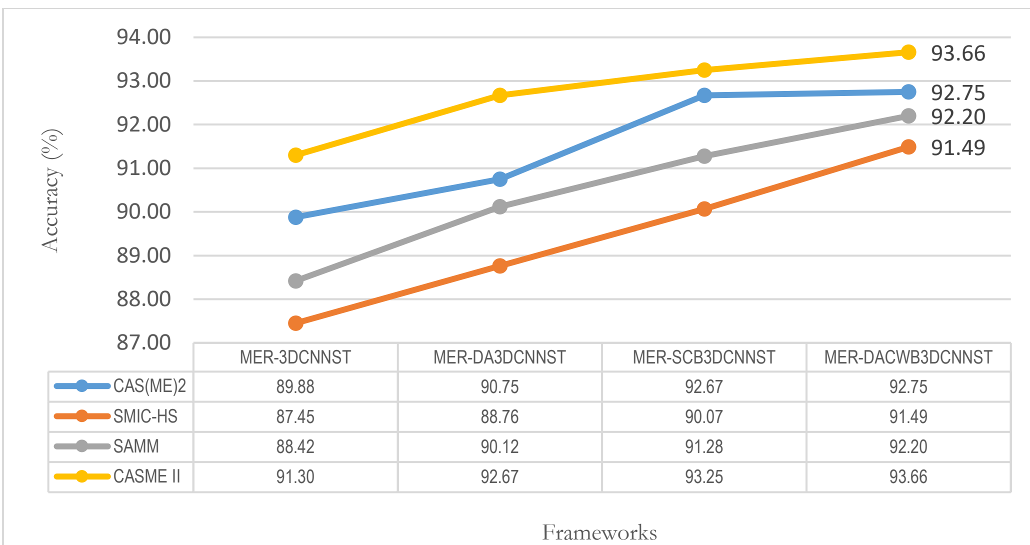

Figure 6. Accuracy graph for micro-expression recognition with the 3DCNN model using four datasets with frameworks: MER-3DCNNST, MER-DA3DCNNST, MER-SCB3DCNNST and MER-DACWB3DCNNST

Figure 7 shows the F1-Score graph for the same four proposed frameworks. This graph shows the highest accuracy and F1-Score values in experiments that apply the DACWB3DCNNST framework, namely scenarios that apply data augmentation and class balancing techniques using the class weight dataset method.

The experimental results highlight the effectiveness of combining advanced preprocessing techniques with the 3DCNN architecture in enhancing micro-expression recognition. The model efficiently identifies subtle expressions by focusing on specific facial regions like the eyes and mouth during image preprocessing, aligning with previous studies that emphasize the importance of localized facial features in improving accuracy.

Data augmentation and class balancing methods were crucial in addressing data limitations and class imbalances, both common challenges in micro-expression analysis. Techniques like rotation, cropping, and contrast adjustments expanded the training data variety, improving the model's generalization ability. Using class weights and SMOTE further helped balance class distributions, reducing the bias towards majority classes and leading to more stable and reliable recognition results.

Integrating these preprocessing techniques with the 3DCNN's ability to analyze spatial and temporal features provided a deeper understanding of micro-expression dynamics. This comprehensive approach fills gaps in existing methodologies and sets a new standard for precision in micro-expression analysis, offering a solid foundation for future research in this field.

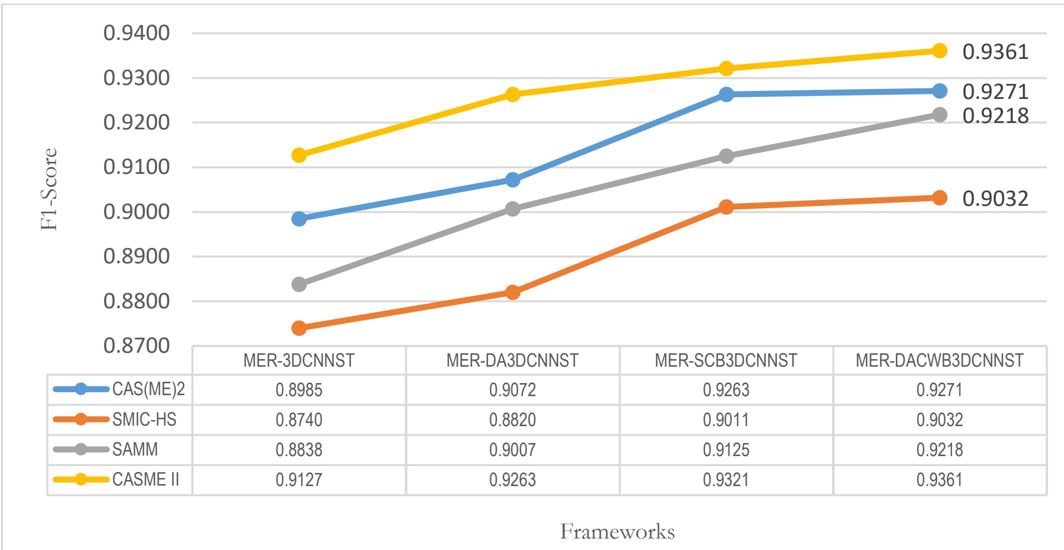

Figure 7. F1-score graph for micro-expression recognition with the 3DCNN model using four datasets with frameworks: MER-3DCNNST, MER-DA3DCNNST, MER-SCB3DCNNST and MER-DACWB3DCNNST

The findings from the proposed framework reveal significant advancements in recognizing spontaneous micro-expressions, primarily due to the combined impact of data augmentation and class balancing techniques. The data augmentation process expanded the diversity of training samples by employing methods like affine rotations, cropping, and contrast adjustments, allowing the model to learn more robust features. These techniques increased the model's accuracy by up to 5% on the CASME II dataset compared to unaugmented data. Additionally, the Class Weight method and SMOTE were crucial in addressing the class imbalance, improving the recall of minority classes by generating synthetic samples, and assigning higher importance to underrepresented expressions. These strategies ensured the model developed a balanced approach in identifying common and less frequent micro-expressions, reducing bias towards majority classes.

The 3D Convolutional Neural Network (3DCNN) architecture demonstrated exceptional performance in capturing both spatial and temporal features of micro-expressions, leading to notable increases in accuracy and F1-scores, with the MER-DACWB3DCNNST framework achieving up to 92.75% accuracy on the CAS(ME)^2 dataset. Compared to existing state-of-the-art methods, this framework consistently outperformed them by an average of 3-4% accuracy and F1-score metrics across multiple datasets. These results validate the framework's ability to effectively handle challenges associated with limited data and class distribution imbalances, establishing its potential as a robust solution for micro-expression recognition in practical applications.

COMPARISON WITH PREVIOUS WORKS

Tables 1-4 present the results of the comparison of accuracy and F1-Score between the current latest approach and the proposed approach using the CAS(ME)², SMIC, SAMM, and CASME II datasets, as well as the implementation of the proposed framework, namely MER-3DCNNST, MER - DA3DCNNST, MER-SCB3DCNNST, and DACWB3DCNNST. The table shows that the proposed approach shows significant overall performance improvement compared to state-of-the-art approaches.

Table 1. Comparison of accuracy and F1-score with state-of-the-art for CAS(ME)² dataset

Year	Method	Accuracy (%)	F1-score
2020	LEARNET (Verma et al., 2020)	76.33	-
2021	MERASTC (Gupta, 2023)	91.20	0.9070
2021	MSFME-IR (Sharma et al., 2021)	-	0.8103
2024	MER-3DCNNST (Ours)	89.98	0.8985
2024	MER-DA3DCNNST (Ours)	90.75	0.9072
2024	MER-SCB3DCNNST (Ours)	92.67	0.9263
2024	**MER-DACWB3DCNNST (Ours)**	**92.75**	**0.9271**

Table 2. Comparison of accuracy and F1-score with state-of-the-art for SMIC dataset

Year	Method	Accuracy (%)	F1-score
2022	ODCNN-MER (Min et al., 2022)	74.80	-
2022	DMER-KD (Sun et al., 2022)	76.06	0.7100
2022	SMER-3DCNNLV (Irawan et al., 2023)	87.20	0.8679
2023	RENGAN-MER (Rakesh Kumar & Bhanu, 2023)	85.52	0.8053
2023	MER-STCN (Shang et al., 2023)	85.41	0.8429
2023	MERASTC (Gupta, 2023)	79.30	0.7900
2023	AGAN-TFMER (Zhang et al., 2023)	82.20	0.8073
2023	MBERT-FMER (Nguyen et al., 2023)	85.50	0.8384
2024	MER-3DCNNST (Ours)	87.45	0.8740
2024	MER-DA3DCNNST (Ours)	88.76	0.8820
2024	MER-SCB3DCNNST (Ours)	90.07	0.9011
2024	**MER-DACWB3DCNNST (Ours)**	**91.49**	**0.9032**

Table 3. Comparison of accuracy and F1-score with state-of-the-art for SAMM dataset

Year	Method	Accuracy (%)	F1-score
2022	GFAUG-MEC (Leong et al., 2022)	61.90	0.4690
2022	DMER-KD (Sun et al., 2022)	86.74	0.8300
2023	DBM-CNET (Zhou et al., 2023)	-	0.6243
2023	MER-LRMIF (Huang et al., 2023)	87.25	0.8621
2023	ADMT-MER (Wang et al., 2023)	81.28	0.8168
2023	MERASTC (Gupta, 2023)	83.80	0.8440
2023	AGAN-TFMER (Zhang et al., 2023)	79.28	0.7643
2023	MBERT-FMER (Nguyen et al., 2023)	83.36	0.8475
2024	MER-3DCNNST (Ours)	88.42	0.8838
2024	MER-DA3DCNNST (Ours)	90.12	0.9007
2024	MER-SCB3DCNNST (Ours)	91.28	0.9125
2024	**MER-DACWB3DCNNST (Ours)**	**92.20**	**0.9218**

Table 4. Comparison of accuracy and F1-Score with state-of-the-art for CASME II dataset

Year	Method	Accuracy (%)	F1-score
2022	ODCNN-MER (Min et al., 2022)	89.89	-
2022	GFAUG-MEC (Leong et al., 2022)	71.10	0.5910
2022	VIT-BiLSTM (Chen et al., 2022)	86.70	0.8640
2022	DMER-KD (Sun et al., 2022)	72.61	0.6700
2023	RENGAN-MER (Rakesh Kumar & Bhanu, 2023)	76.83	0.7543
2023	DBM-CNET (Zhou et al., 2023)	-	0.6653
2023	MER-LRMIF (Huang et al., 2023)	90.82	0.8820
2023	MER-STCN (Shang et al., 2023)	91.46	0.8977
2023	ADMT-MER (Wang et al., 2023)	92.37	0.9210
2023	MERASTC (Gupta, 2023)	85.40	0.8620
2023	FRL-AGDTF (Zhai et al., 2023)	78.00	0.7500
2023	AGAN-TFMER (Zhang et al., 2023)	90.24	0.9221
2024	MER-3DCNNST (Ours)	91.30	0.9127
2024	MER-DA3DCNNST (Ours)	92.67	0.9263
2024	MER-SCB3DCNNST (Ours)	93.25	0.9321
2024	**MER-DACWB3DCNNST (Ours)**	**93.66**	**0.9361**

This work has the best accuracy value obtained from the DACWB3DCNNST framework. In this framework, data augmentation techniques are carried out, namely rotation, cropping, changes in brightness and contrast, and class balancing using the class weight method. Better accuracy results can occur when using class weights than SMOTE because class weights provide a direct approach by giving additional weight to minority classes in model training.

CONCLUSION

The proposed framework has successfully improved micro-expression recognition by integrating data augmentation, class balancing techniques, and a 3D Convolutional Neural Network (3DCNN) architecture. These strategies effectively addressed data limitations and class imbalances, leading to higher accuracy and F1-scores across various datasets than existing methods. The model's ability to capture both spatial and temporal features of micro-expressions sets a new benchmark in the field, highlighting its robustness and precision in handling subtle facial expressions.

For future research, a key focus should be enhancing the model by incorporating attention mechanisms to prioritize crucial facial regions, enabling more precise extraction of micro-expression features. Combining these attention-based approaches with advanced techniques like transformer architectures could further boost feature representation and model interpretability. Developing hybrid models that integrate traditional machine learning and deep learning methods and expanding training datasets to include diverse facial expressions will help create more generalized and globally applicable recognition systems. These advancements will pave the way for more accurate and efficient micro-expression analysis in real-world applications.

REFERENCES

Ayyalasomayajula, S. C., Ionescu, B., & Ionescu, D. (2021, May). A CNN approach to micro-expressions detection. *Proceedings of the IEEE 15th International Symposium on Applied Computational Intelligence and Informatics, Timisoara, Romania*, 345–350. https://doi.org/10.1109/SACI51354.2021.9465542

Balakrishnan, A., Medikonda, J., Namboothiri, P. K., & Natarajan, M. (2022, October). Parkinson's disease stage classification with gait analysis using machine learning techniques and SMOTE-based approach for class imbalance problem. *Proceedings of the IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, Shivamogga, India*, 277–281. https://doi.org/10.1109/DISCOVER55800.2022.9974754

Bayu, W. S. P., & Setyanto, A. (2022, August). 3D CNN for micro expression detection. *Proceedings of the 5th International Conference on Information and Communications Technology*, *Yogyakarta, Indonesia,* 397–401. https://doi.org/10.1109/ICOIACT55506.2022.9972194

Chen, H., Cui, J., Zhang, Y., & Zhang, Y. (2022, September). VIT and Bi-LSTM for micro-expressions recognition. *Proceedings of the IEEE 5th International Conference on Information Systems and Computer Aided Education, Dalian, China,* 946–951. https://doi.org/10.1109/ICISCAE55891.2022.9927522

Guowen, F., & Xi, L. (2023, September). Micro-expression recognition based on dual branch neural network. *Proceedings of the 2nd International Conference on Artificial Intelligence and Computer Information Technology, Yichang, China*, 1–4. https://doi.org/10.1109/AICIT59054.2023.10278020

Gupta, P. (2023). MERASTC: Micro-expression recognition using effective feature encodings and 2D convolutional neural network. *IEEE Transactions on Affective Computing*, *14*(2), 1431–1441. https://doi.org/10.1109/TAFFC.2021.3061967

He, Y., Wang, S.-J., Li, J., & Yap, M. H. (2020, November). Spotting macro-and micro-expressions intervals in long video sequences. *Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition, Buenos Aires, Argentina*, 742–748. https://doi.org/10.1109/FG47880.2020.00036

Huang, P., Wang, L., Cai, T., & Guo, K. (2023, October). Micro-expression recognition with layered relations and more input frames. *Proceedings of the IEEE International Conference on Image Processing*, *Kuala Lumpur, Malaysia,* 2210–2214. https://doi.org/10.1109/icip49359.2023.10222395

Irawan, B., Utama, N. P., Munir, R., & Purwarianti, A. (2023, October). Spontaneous micro-expression recognition using 3DCNN on long videos for emotion analysis. *Proceedings of the 10th International Conference on Advanced Informatics: Concept, Theory and Application, Lombok, Indonesia*, 1–6. https://doi.org/10.1109/ICAICTA59291.2023.10390247

Jiao, Y., Jing, M., Hu, Y., & Sun, K. (2021, July). Research on a micro-expression recognition algorithm based on 3D-CNN. *Proceedings of the 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field*, *Xi'an, China*, 221–225. https://doi.org/10.1109/ICMSP53480.2021.9513351

Leong, S.-M., Noman, F., Phan, R. C.-W., Baskaran, V. M., & Ting, C.-M. (2022, October). Graphex: Facial action unit graph for micro-expression classification. *Proceedings of the IEEE International Conference on Image Processing, Bordeaux, France*, 3296–3300. https://doi.org/10.1109/ICIP46576.2022.9897873

Liu, K.-H., Jin, Q.-S., Xu, H.-C., Gan, Y.-S., & Liong, S.-T. (2021). Micro-expression recognition using advanced genetic algorithm. *Signal Processing: Image Communication*, *93*, 116153. https://doi.org/10.1016/j.image.2021.116153

Liu, N., Liu, X., Zhang, Z., Xu, X., & Chen, T. (2020, November). Offset or onset frame: A multi-stream convolutional neural network with CapsuleNet module for micro-expression recognition. *Proceedings of the 5th International Conference on Intelligent Informatics and Biomedical Sciences*, *Okinawa, Japan,* 236–240. https://doi.org/10.1109/ICIIBMS50712.2020.9336412

Min, K. S., Asyraf Zulkifley, M., & Mohamed Kamari, N. A. (2022, May). Optimized dense convolutional neural networks for micro-expression recognition. *Proceedings of the IEEE 12th Symposium on Computer Applications and Industrial Electronics, Penang, Malaysia,*, 288–293. https://doi.org/10.1109/ISCAIE54458.2022.9794470

Nguyen, X.-B., Duong, C. N., Li, X., Gauch, S., Seo, H.-S., & Luu, K. (2023, June). Micron-BERT: BERT-based facial micro-expression recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *Vancouver, BC, Canada*, 1482–1492. https://doi.org/10.1109/CVPR52729.2023.00149

Pan, H., Xie, L., Lv, Z., Li, J., & Wang, Z. (2020). Hierarchical support vector machine for facial micro-expression recognition. *Multimedia Tools and Applications*, *79*, 31451–31465. https://doi.org/10.1007/s11042-020-09475-4

Rakesh Kumar, A. J., & Bhanu, B. (2023, June). Relational edge-node graph attention network for classification of micro-expressions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, *Vancouver, BC, Canada,* 5819–5828. https://doi.org/10.1109/CVPRW59228.2023.00618

Sagoolmuang, A. (2021, August). Power-weighted kNN classification for handling class imbalanced problem. *Proceedings of the 2nd International Conference on Big Data Analytics and Practices, Bangkok, Thailand*, 42–47. https://doi.org/10.1109/IBDAP52511.2021.9552164

Shang, Z., Liu, J., & Li, X. (2023). Micro-expression recognition based on spatio-temporal capsule network. *IEEE Access*, *11*, 13704–13713. https://doi.org/10.1109/ACCESS.2023.3242871

Sharma, P., Coleman, S., Yogarajah, P., Taggart, L., & Samarasinghe, P. (2021, January). Magnifying spontaneous facial micro expressions for improved recognition. *Proceedings of the 25th International Conference on Pattern Recognition*, *Milan, Italy*, 7930–7936. https://doi.org/10.1109/ICPR48806.2021.9412585

Sun, B., Cao, S., Li, D., He, J., & Yu, L. (2022). Dynamic micro-expression recognition using knowledge distillation. *IEEE Transactions on Affective Computing*, *13*(2), 1037–1043. https://doi.org/10.1109/TAFFC.2020.2986962

Teja Reddy, S. P., Teja Karri, S., Dubey, S. R., & Mukherjee, S. (2019, July). Spontaneous facial micro-expression recognition using 3D spatiotemporal convolutional neural networks. *International Joint Conference on Neural Networks*, *Budapest, Hungary*, 1–8. https://doi.org/10.1109/IJCNN.2019.8852419

Verma, M., Vipparthi, S. K., Singh, G., & Murala, S. (2020). LEARNet: Dynamic imaging network for micro expression recognition. *IEEE Transactions on Image Processing*, *29*, 1618–1627. https://doi.org/10.1109/TIP.2019.2912358

Wang, Y., Shi, H., & Wang, R. (2023). Action decouple multi-tasking for micro-expression recognition. *IEEE Access*, *11*, 82978–82988. https://doi.org/10.1109/ACCESS.2023.3301950

Wei, J., Lu, G., Yan, J., & Liu, H. (2022). Micro-expression recognition using local binary pattern from five intersecting planes. *Multimedia Tools and Applications*, *81*, 20643–20668. https://doi.org/10.1007/s11042-022-12360-x

Wei, M., Zheng, W., Zong, Y., & Jiang, X. (2022, May). A novel micro-expression recognition approach using attention-based magnification-adaptive networks. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Singapore*, 2420-2424. https://doi.org/10.1109/ICASSP43922.2022.9747232

Yang, P., Jin, H., & Li, Z. (2022, July). Combining attention mechanism and dual-stream 3D convolutional neural network for micro-expression recognition. *Proceedings of the 7th International Conference on Image, Vision and Computing, Xi'an, China*, 51–59. https://doi.org/10.1109/ICIVC55077.2022.9886046

Zhai, Z., Zhao, J., Long, C., Xu, W., He, S., & Zhao, H. (2023, June). Feature representation learning with adaptive displacement generation and transformer fusion for micro-expression recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, *Vancouver, BC, Canada,* 22086–22095. https://doi.org/10.1109/CVPR52729.2023.02115

Zhang, Y., Wang, H., Xu, Y., Mao, X., Xu, T., Zhao, S., & Chen, E. (2023, July). Adaptive graph attention network with temporal fusion for micro-expressions recognition. *Proceedings of the IEEE International Conference on Multimedia and Expo*, *Brisbane, Australia,* 1391–1396. https://doi.org/10.1109/ICME55011.2023.00241

Zhou, Q., Liu, S., Wang, Y., & Wang, J. (2023, January). Divided block multiscale convolutional network for micro-expression recognition. *Proceedings of the 1st International Conference on Cyber-Energy Systems and Intelligent Energy, Shenyang, China*, 1–5. https://doi.org/10.1109/ICCSIE55183.2023.10175242

AUTHORS

Budhi Irawan received a bachelor's degree in Computer Science from Padjadjaran University, Bandung – Indonesia, in 1999, an M.Sc. degree in computer engineering from the Bandung Institute of Technology, Bandung – Indonesia, in 2006, and is currently pursuing a Ph.D. degree from the School of Electrical Engineering and Informatics, Bandung Institute of Technology. He also works as a lecturer and researcher at Telkom University, Bandung - Indonesia. His research interests include Computer Vision and Artificial Intelligence.

Rinaldi Munir received a bachelor's degree in informatics engineering and an M.Sc. degree in digital image compression from the Bandung Institute of Technology (ITB), Bandung, Indonesia, in 1992 and 1999, respectively, and a Ph.D. degree in image watermarking from the School of Electrical Engineering and Informatics, ITB, in 2010. In 1993, he started his academic career as a Lecturer with the Department of Informatics, ITB. He is currently an Associate Professor with the School of Electrical Engineering and Informatics, ITB, and the Informatics Research Group. His research interests include cryptography and steganography-related topics, digital image processing, fuzzy logic, and numerical computation.

Nugraha Priya Utama (Member, IEEE) received a bachelor's degree in informatics from the Bandung Institute of Technology, Indonesia, in 2002 and master's and Ph.D. degrees from the Tokyo Institute of Technology in 2006 and 2009, respectively. His research interests include computer vision and neuroscience. He has been the Chair of the Artificial Intelligence Center, ITB, since 2023.

Ayu Purwarianti received her Ph.D. degree from the Toyohashi University of Technology in December 2007, with the dissertation title "Cross-Lingual Question Answering System (Indonesian Monolingual QA, Indonesian-English CLQA, and Indonesian-Japanese CLQA)." The dissertation was in the area of natural language processing, also known as computational linguistics, which is part of the artificial intelligence knowledge domain. Since then, she has been a Lecturer at the Bandung Institute of Technology (ITB). Other than teaching and doing research, her other activity is in the Indonesian Association for Computational Linguistics, where she was elected Chair from 2016 to 2018. She was also the Chair of the IEEE Education Chapter of the Indonesian Section from 2017 to 2019. Since 2015, she has been with IABEE. In 2018, she founded a start-up named Prosa.ai.