Research and Application of Wifi CSI for Diagnosing Emotions in Mental Healthcare

Minh Quan Dang   
*Faculty of Advanced Science and Technology*   
*The University of Danang – University of Science and Technology*  
Danang, Vietnam  
[123210100@sv1.dut.udn.vn](mailto:123210100@sv1.dut.udn.vn)

Van Duc Le  
*Faculty of Advanced Science and Technology*   
*The University of Danang – University of Science and Technology*Danang, Vietnam  
[123210094@sv1.dut.udn.vn](mailto:123210094@sv1.dut.udn.vn)

Quang Nhu Quynh Nguyen  
*Faculty of Advanced Science and Technology*   
*The University of Danang – University of Science and Technology*Danang, Vietnam  
nqnquynh@dut.udn.vn

Le Hoa Nguyen\*  
*Faculty of Advanced Science and Technology*   
*The University of Danang – University of Science and Technology*Danang, Vietnam  
ORCID: 0000-0002-3316-2394

*Abstract*—Mental healthcare remains a critical challenge, especially in Vietnam, where 3.2 million people (3.1% of the population) suffer from depression, with the highest prevalence (5.4%) among young adults aged 18-29, particularly women. Traditional emotion monitoring methods, such as self-reports and wearable sensors, are often invasive, unreliable, and unsuitable for real-time assessment. We propose a novel, non-invasive approach using WiFi Channel State Information (CSI) to analyze subtle respiratory patterns for emotion recognition. Our system leverages a Vision Transformer (ViT) model to classify emotions based on spectrogram representations of CSI signals. To address the data-hungry nature of ViTs and improve performance, we apply Knowledge Distillation, using a vision-based facial expression model as a teacher for our CSI-based model. This approach enhances training efficiency and boosts classification accuracy, achieving 72.97% in recognizing key emotions (happy, anger, neutral, sadness).

Keywords—WiFi Channel State Information (CSI), Vision Transformer (ViT), emotion recognition, knowledge distillation, mental healthcare, spectrogram representation, non-invasive monitoring, real-time emotion analysis, deep learning, facial expression recognition

# Introduction

Emotion plays a vital role in human social communication, influencing both physiological and psychological states. Positive emotions enhance health and productivity, while negative emotions can harm well-being and, if prolonged, may lead to depression or even suicide. Unlike mood, a conscious and sustained mental state, emotion arises spontaneously and is often accompanied by physiological responses in the brain, heart, skin, muscles, and other organs.

In Vietnam, 3.2 million people (3.1% of the population) suffer from depression, with the highest rate (5.4%) among people aged 18-29, especially women [1]. Due to the complexity of the mutual interaction between physiology and psychology in emotion, the accurate and timely recognition of human emotions is still limited in our understanding and remains the goal of relevant scientific research and industry.

Emotion recognition has been applied in many fields such as safe driving [2], health care especially in mental health monitoring [3], human-computer interaction [4], etc. Emotion recognition methods such as using human physical signals [7-8], facial expressions [5], vocal patterns [6], etc., have the advantage of being easy to collect and have been studied for many years. However, reliability cannot be guaranteed, because it is relatively easy for people to control physical signals such as facial expressions or speech to conceal their true emotions, especially in social communication.

Several studies have explored alternative modalities for emotion detection. Traditional approaches commonly use image-based techniques, such as facial expression recognition through cameras, to capture visual cues like smiles or frowns for emotion classification. Deep learning methods employing Convolutional Neural Networks (CNNs) on facial images have achieved high accuracy but raise privacy concerns due to continuous visual monitoring. Other studies have examined physiological signals such as electroencephalogram (EEG), where power spectral density or principal component analysis combined with machine learning models like SVM have been effectively applied for emotion classification [7–8]. More recent research has used WiFi signals to infer facial expressions indirectly [9], detecting facial movements rather than underlying physiological signals. In contrast, our approach leverages WiFi CSI to recognize emotions through breathing patterns and body movements, providing a non-invasive and privacy-preserving alternative that eliminates reliance on facial data or visual input. The contributions of our study are as follows:

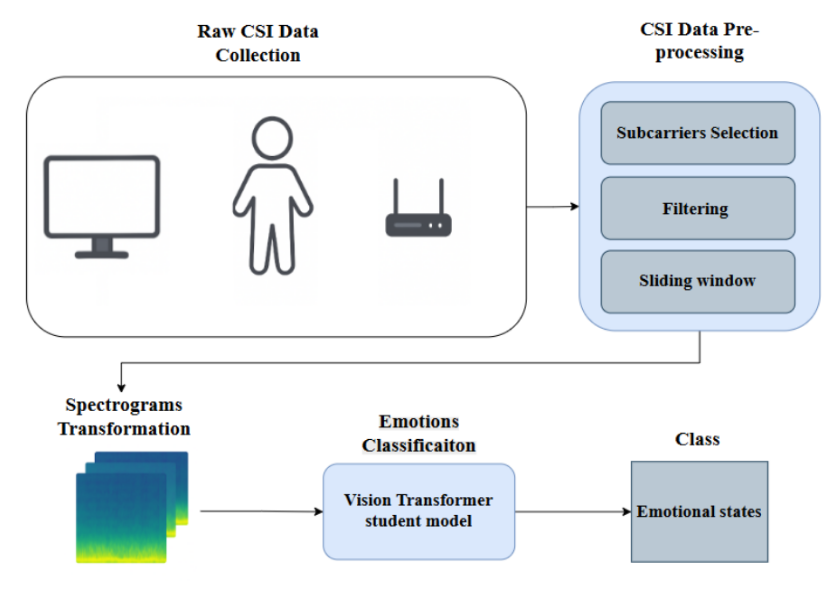
* Constructed a novel, privacy-preserving CSI–emotion dataset by recording breathing patterns of WiFi CSI signals collected via an ESP32 setup.
* Pioneered cross-modal knowledge distillation between facial-expression and CSI-based emotion recognition by transferring knowledge from a ViT trained on facial data to a ViT for CSI data, improving performance by ≈7–8%.
* Developed a secure, real-time, non-wearable emotion detection system (no cameras or wearables) that achieves ≈72% accuracy, making it practical for deployment in education, office, and healthcare settings.

# methodology

Our proposed system illustrates in Fig. 1. is designed to monitor and recognize emotional states through breathing patterns by leveraging WiFi CSI. The overall process consists of four main stages: raw CSI data collection, data preprocessing, Spectrograms transformation and emotions classification

## Wifi CSI

### How Wifi CSI works: This study uses WiFi CSI to detect breathing rhythms, as illustrated in Fig. 2. CSI provides

Fig. 1. System Design for Emotions Monitoring Using WiFi CSI

### detailed information about the wireless channel between devices, helping assess channel quality and optimize network performance [10]. In this research, CSI data from transmitted packets are collected, processed, and analyzed to identify channel variations caused by human motion [11].

CSI measures the physical wireless channel at the subcarrier level, revealing how transmitted signals interact with the environment. These interactions capture subtle signal variations caused by movements such as breathing, body movements. By analyzing frequency and time domain features, small body motions can be detected and used to extract movement patterns [12]. Examining these variations enables mapping CSI data to spatial geometry for real-time monitoring of respiratory rhythms [13].

### CSI Data Collection: CSI data is gathered through antennas on Wi-Fi devices. In this study, two ESP32 devices (one as a station and one as an access point) are used to collect CSI data. The receiver ESP32 records information about the transmission channel. CSI reflects various channel

A diagram of a person's path

AI-generated content may be incorrect.

Fig. 2. How Wi-Fi CSI interacts.

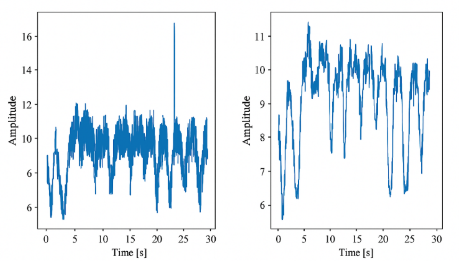
### characteristics, such as signal attenuation, delay, and signal dispersion. These parameters are extracted from feedback of transmitted Wi-Fi packets in the complex domain. CSI measures the amplitude and phase shift of each subcarrier, providing fine-grained channel information that reflects environmental or human-induced variations, enabling applications such as activity or emotion recognition.

### CSI Data Pre-processing: The raw CSI dataset contains 128 numerical values for 64 subcarriers, each with real and imaginary components. The amplitude for each subcarrier is computed as the square root of the sum of the squares of these components.

### Not all subcarriers contain meaningful information. Some are null subcarriers with negligible energy, and a signal-to-noise ratio (SNR) close to zero, contributing mostly noise; thus, they are excluded during preprocessing. After the removal, 52 subcarriers remain for analysis [14], including pilot subcarriers used for channel estimation and equalization, and data subcarriers that carry the actual transmitted information [15].

The raw CSI values from the ESP32 are complex numbers with amplitude and phase components. Due to hardware imperfections and synchronization errors, amplitude data contain noise, while phase information is often too distorted for reliable use [16]. Therefore, this study focuses on CSI amplitude for breathing-based emotion detection. As shown in Fig. 3(a), the amplitude streams exhibit bursty noise and sudden spikes that distort the respiratory waveform. To address this, a Hampel filter [17] removes outliers, followed by a Butterworth low-pass filter to smooth the signal. The Hampel filter identifies outliers based on each point’s deviation from the local median, scaled by the median absolute deviation.

Samples exceeding a set of multiples of this deviation are flagged as outliers and replaced with the median, producing a smooth and more robust signal. After outlier suppression, a Butterworth low-pass filter removes residual high-frequency noise from Wi-Fi hardware and environmental interference. Its maximally flat passband preserves the breathing waveform while attenuating irelevant components. The effectiveness of this process is illutrated in Fig. 3(b).



a) b)

Fig. 3. CSI amplitude signal a) before filtering and b) after Hampel and Butterworth filtering.

### Spectrogram Changes Across Emotion: Different emotional states alter respiratory patterns [31–33], which can be captured through WiFi CSI. In this work, spectrogram analysis provides a time–frequency view of breathing dynamics, revealing distinctive features for each emotion. As shown in Fig. 4, spectrograms for sadness, neutral, happiness, and anger exhibit distinct spectral patterns, highlighting CSI spectrograms’ ability to differentiate emotions—supporting non-invasive emotion recognition for mental healthcare and human–computer interaction.

### Fig. 4 compares the amplitude signals and their corresponding spectrograms for the four emotional states: sadness (Fig. 4(a)), neutrality (Fig. 4(b)), happiness (Fig.4(c))**,** and anger (Fig. 4(d))**.**

### Distinct differences can be observed in both the time-domain waveforms and the frequency-domain representations, reflecting the variations in breathing patterns across emotions.

In the sad state, the breathing signal is slow, deep, and highly stable, appearing in the spectrogram as a narrow, steady low-frequency band with consistent warm color intensity. The neutral state also shows stable breathing but at a slightly faster rate, reflected by a sharper low-frequency band with moderate, evenly distributed energy.

For more expressive emotions, breathing become irregular and scattered.

In the happy state, laughter or giggles cause amplitude fluctuations, shown as bursts of higher-frequency energy in the spectrogram. The angry state exhibits sharp peaks and bursts from forceful breaths, forming multiple frequency bands with intense warm colors across low and high frequencies.

A few different colored waves

AI-generated content may be incorrect.a) Sad b) Neutral

A group of different colored lines

AI-generated content may be incorrect. c) Happy d) Angry

Fig. 4. Amplitude and Spectrogram Variations Across Emotions.

Taken together, these observations highlight a clear distinction between calmer emotions such as sadness and neutrality, which produce stable low-frequency spectrograms and more expressive emotions such as happiness and anger, which generate irregular, high-energy, and wide-frequency patterns. This contrast underlines the effectiveness of spectrogram analysis in capturing discriminative features of emotional states from CSI-derived breathing signals.

## Machine learning model

The ViT model has become a leading approach in image processing, particularly for facial expression recognition (FER), due to its ability to capture long-range dependencies via self-attention [18]. Unlike CNNs, which extract local features through convolutions and pooling, ViTs split images into fixed-size patches and embed them as a sequence, allowing self-attention to model global contextual relationships [19]. This architecture has been validated on benchmark datasets like FER-2013, AffectNet, and RAF-DB, where ViT-based models consistently achieve state-of-the-art results [20]. Transformer-based approaches outperform CNNs under occlusion and pose variation, thanks to dynamic attention across facial regions [21]. Moreover, ViTs scale well under constrained settings: when trained on ImageNet without external data, with distillation, and fine-tuned at higher resolutions, they achieve competitive or top-tier performance [22]. These strengths make ViTs both practical and effective for FER, even in challenging real-world conditions [23].

## Features knowledge distillation

Although ViTs excel in facial expression recognition, they require large-scale training data, limiting use in scenarios with few labeled samples [24]. To overcome this, we apply knowledge distillation: a compact student ViT model learns from a pre-trained teacher ViT model with the identical architecture trained on a large FER dataset. This enables the student to acquire meaningful features and robust representations while reducing reliance on extensive labeled data [25].

Recent studies show that distilling knowledge from a Transformer teacher to a CNN student, by aligning student features with the teacher’s attention space, helps the student capture global relational features and improve performance even across different architectures [26]. The teacher, trained on high-resolution face data, provides attention-guided feature maps that the student mimics via a similarity loss, enhancing generalization in challenging or data-scarce scenarios [27]. This is especially important in FER, where subtle expression changes demand robust feature representations. Empirical results indicate that such distillation can yield lightweight student models with much lower computational cost and recognition accuracy only ~1–2% below that of larger teacher models [28]

Our approach aligns with [27], where attention map matching from a high-resolution teacher helps a student generalize with limited inputs. Similarly, [29] show that attention-based distillation guides the student to focus on key semantic regions. By applying cross-modal knowledge distillation from FER to CSI spectrograms, our method reduces ViTs’ data dependency while preserving strong representational performance, enabling effective FER in limited-data scenarios.

# experiment setup

## ESP32 configuration

The hardware setup consists of two ESP32 modules, as illustrated in Fig. 5(a)), and Fig. 5(b)) The ESP32-S runs in Station (STA) mode, transmitting Wi-Fi packets, while the ESP32-S3 acts as an Access Point (AP) to receive signals and collect CSI data, capturing the wireless channel characteristics. Both modules use compatible firmware to maintain stable communication in a controlled indoor environment.

## Environment for data collection

Data were collected in a controlled indoor environment with seated participants to minimize external influences. Each followed emotion-specific breathing patterns: happiness (giggling, mouth exhalation) [30], anger (fast, irregular deep nasal breaths), sadness (slow, deep breaths with sighs) [31], and neutral (steady rhythm) [32]. Participants watched emotion-specific videos to elicit the target emotional states, ensuring consistency, while facial expressions were recorded to provide visual context for emotion analysis.

Eight university students participated, maintaining an upright posture while breathing through the ESP32 setup (Fig. 5(b)). WiFi CSI was captured at 100 packets/s, converted into 100 Hz spectrograms, and resized to

224×224 pixel for ViT input. Each spectrogram was paired with a simultaneously captured, resized, and normalized facial image.

During each session, ESP32 modules continuously transmit and receive Wi-Fi packets, capturing CSI fluctuations from breathing-induced channel changes. Each 5–10-minute session yields data processed into breathing features, totaling 1.5 hours per emotion and 6 hours overall for training and evaluation.

A computer with a computer open and a computer with a computer open

AI-generated content may be incorrect.

a) b)

Fig. 5. ESP32 Hardware Setup a) for CSI Data Collection and b) set up and participant right posture to collect data

## Model training setup

### Student and teacher models setup: A knowledge distillation framework is implemented using two ViT models with identical architectures.

### The student model is initialized with the teacher’s pretrained weights, with the first 11 encoder layers (0–10)

A graph with numbers and symbols

AI-generated content may be incorrect.

Fig. 6. Confusion matrix of Teacher model on test set

### frozen and only the final layers fine-tuned. The teacher model, pretrained on a large-scale facial expression dataset of 48×48 grayscale images across seven emotions (angry, disgust, fear, happy, sad, surprise, neutral), achieves 99% accuracy on our four-class subset (happy, sad, angry, neutral) as shown in Fig. 6.

### Training loop setup: The proposed dual-model architecture processes paired spectrogram–facial image data, as illustrated in Fig. 7. Spectrograms derived from WiFi CSI signals are input to the student model, while temporally aligned facial images are fed to the frozen teacher model , which provides high-level feature guidance. During training, only the student’s parameters are updated. We employ feature-level knowledge distillation for the training loop, where the pretrained FER teacher transfers feature representations from its encoder to guide the student. The student model is optimized using two loss components: (1) the Mean Squared Error (MSE) [33] between the final encoder outputs of both models (scaled by 0.001) for feature-level alignment, and (2) cross-entropy loss [34]between the student’s predictions and the ground-truth emotion labels.

### Training is conducted end-to-end using the Adam optimizer [35], which combines adaptive momentum estimation with RMSprop for efficient updates.

This dual-loss approach, combining both high-level feature distillation (via MSE) and label information (via cross-entropy), allows the student model to effectively learn the mapping between Wi-Fi spectrogram patterns and corresponding facial expressions, successfully bridging the modality gap between these two distinct input domains. After training, the student model operates independently, requiring only CSI-derived spectrograms for emotion recognition.

A diagram of a student model

AI-generated content may be incorrect.

Fig. 7. Feature-Based Knowledge Distillation Diagram

The dataset includes two modalities: facial emotion images for teacher model pretraining and CSI-derived

A green and orange rectangular bars with numbers

AI-generated content may be incorrect.

Fig. 8. Comparison of Model predictions on test data

spectrograms for student model training. It was split into 70% training, 15% validation, and 15% testing to enable effective learning, hyperparameter tuning, and fair evaluation on unseen data.

# Results

## Model performance comparison

Our experimental results demonstrate the effectiveness of incorporating knowledge distillation in training the student ViT model with spectrogram inputs derived from CSI data. The Distilled student achieves an accuracy of 72.97%, marking a substantial improvement over the Baseline ViT train without distillation, which attains 65.18% as shown in Fig. 8.

This 7.8% improvement highlights the effectiveness of feature-based knowledge distillation in enabling a spectrogram-only model to achieve higher accuracy while remaining efficient and fully deployable without facial images. The significant performance gain confirms the effectiveness of transferring knowledge from the visual to the wireless signal domain.

## Model prediction accuracy analysis

The confusion matrix in Fig. 9 shows that the model often confuses sad with neutral and happy with angry, likely due to overlapping breathing patterns and similar spectrogram features. Sadness, however, is detected strongly with 70.6% precision and 84% recall, suggesting potential for mental health monitoring, particularly in early depression detection.

A diagram of different emotions

AI-generated content may be incorrect.

Fig. 9. Confusion matrix of Student model on test set

A graph of different colored rectangular bars

AI-generated content may be incorrect.

Fig. 10. Accuracy of different segment length of 1200, 600, 300

## Impact of different segment length

We further investigated how different segment lengths affect model accuracy by testing three configurations are 300, 600 and 1200 packages corresponding to 3, 6 and 12 second. And we gain the accuracy result of each segment length as shown in Fig. 10.

The shortest segment length of 300 packets (3 seconds) yields the lowest classification accuracy of 56.57%, as this brief duration proves insufficient for capturing discrimative patterns in spectrogram generation

In contrast, the 600-packet (6 s) and 1200-packet (12 s) segments both achieve strong and comparable results, with accuracies of 72.97% and 70.81%, respectively. These findings suggest that capturing around 6 seconds of data provides sufficient temporal context for reliable emotion recognition. Extending the observation period to 12 seconds offers minimal additional benefit, indicating diminishing returns in accuracy improvement beyond this duration.

# conclusion

This research developed a privacy-preserving emotion detection system using WiFi CSI, achieving 72.97% accuracy in recognizing emotions from breathing patterns and body movements without images. Pioneering in using knowledge distillation from a facial ViT model to a CSI-based ViT improved accuracy by 7.8%. The proposed CSI-emotion dataset and optimal segment length (600 packets, 6 s) highlight the system’s potential for real-time emotion monitoring in healthcare, education, and workplace settings.

##### Acknowledgment

The authors are deeply grateful to Mr. Huu Tai Phung for his insightful discussions, valuable support in conducting the study, and constructive feedback on the research findings.

##### References

1. H. Sơn, “Báo động trẻ hóa bệnh nhân trầm cảm,” Báo Nhân Dân Điện Tử, Apr. 13, 2024. [Online]. Available: <https://nhandan.vn/bao-dong-tre-hoa-benh-nhan-tram-cam-post804351.html>
2. S. De Nadai, M. D’Incà, F. Parodi, M. Benza, A. Trotta, E. Zero, L. Zero, and R. Sacile, “Enhancing safety of transport by road by on-line monitoring of driver emotions,” in Proc. 11th Syst. Syst. Eng. Conf. (SoSE), Kongsberg, Norway, Jun. 2016, pp. 1–4.
3. R. Guo, S. Li, L. He, W. Gao, H. Qi, and G. Owens, “Pervasive and unobtrusive emotion sensing for human mental health,” in Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthcare, Venice, Italy, May. 2013, pp. 436–439.
4. R. Cowie et al., "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, Jan 2001, doi: 10.1109/79.911197.
5. Y. D. Zhang, Z. J. Yang, H. M. Lu, X. X. Zhou, P. Phillips, Q. M. Liu, and S. H. Wang, “Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation,” IEEE Access, vol. 4, pp. 8375–8385, 2016.
6. Q. Mao, M. Dong, Z. Huang, and Y. Zhan, “Learning salient features for speech emotion recognition using convolutional neural networks,” IEEE Trans. Multimedia, vol. 16, no. 8, pp. 2203–2213, 2014.
7. M. A. Tran and L. H. Nguyen, “EEG-based emotion recognition using principal component analysis and support vector machine,” in Proc. 2024 IEEE Int. Conf. Control & Autom., Electron., Robot., Internet Things, Artif. Intell. (CERIA), pp. 1–5, 2024.
8. M. A. Tran, L. H. Nguyen, A. Turnip, C. T. K. Le, N. Q. Luong, and V. B. Vo, “EEG-based emotion feature extraction using power spectral density,” in Proc. 2023 15th Int. Conf. Knowl. Syst. Eng. (KSE), pp. 1–4, 2023.
9. Y. Chen, R. Ou, Z. Li, and K. Wu, “WiFace: Facial Expression Recognition Using Wi-Fi Signals,” IEEE Trans. Mobile Comput., vol. 21, no. 1, pp. 378–391, Jan. 2022, doi: 10.1109/TMC.2020.3001989.
10. Z. Yang, Y. Zhang, G. Chi, and G. Zhang, “Hands-on Wireless Sensing with Wi-Fi: A Tutorial,” MobiSense Group, Tsinghua Univ., Beijing, China. [Online]. Available: <https://tns.thss.tsinghua.edu.cn/>
11. H. Xu, Y. Chen, J. Liu, et al., “CSI-based localization and tracking in Wi-Fi networks,” IEEE Trans. Mobile Comput., vol. 13, no. 11, pp. 2451–2463, Nov. 2014.
12. C. Wu, F. Zhang, Y. Hu, and K. J. R. Liu, “GaitWay: Monitoring and recognizing gait speed through the walls,” IEEE Trans. Mobile Comput., vol. 20, no. 6, pp. 2186–2199, Jun. 2021.
13. F. Zhang, C. Chen, B. Wang, and K. J. R. Liu, “WiSpeed: A statistical electromagnetic approach for device-free indoor speed estimation,” IEEE Internet Things J., vol. 5, no. 3, pp. 2163–2177, Jun. 2018.
14. S. M. Hernandez and E. Bulut, “Wi-Fi sensing on the edge: Signal processing techniques and challenges for real-world systems,” IEEE Commun. Surveys Tuts., vol. 25, no. 1, pp. 46–76, 2022.
15. M. Roopak, Y. Ran, X. Chen, G. Y. Tian, and S. Parkinson, “Channel state information based physical layer authentication for Wi-Fi sensing systems using deep learning in Internet of Things networks,” IET Wireless Sensor Syst., vol. 14, no. 16, Sep. 2024, doi: 10.1049/wss2.12093.
16. H. Zhu, E. Dong, M. Xu, H. Lv, and F. Wu, “Commodity Wi-Fi-based wireless sensing advancements over the past five years,” Sensors, vol. 24, no. 22, p. 7195, Nov. 2024, doi: 10.3390/s24227195.
17. L. Davies and U. Gather, “The identification of multiple outliers,” J. Amer. Stat. Assoc., vol. 88, no. 423, pp. 782–792, 1993.
18. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, pp. 5998–6008, 2017.
19. S. Khan et al., “Transformers in vision: A survey,” ACM Comput. Surveys, vol. 54, no. 10s, 2022.
20. S. Bobojanov, B. M. Kim, M. Arabboev, and S. Begmatov, “Comparative analysis of vision transformer models for facial emotion recognition using augmented balanced datasets,” Appl. Sci., vol. 13, no. 22, art. no. 12271, 2023, doi: 10.3390/app132212271.
21. F. Ma, B. Sun, and S. Li, “Facial Expression Recognition with Visual Transformers and Attentional Selective Fusion,” arXiv preprint arXiv:2103.16854, v3, Feb. 2022.
22. H. Touvron et al., “Training data-efficient image transformers and distillation through attention,” in Proc. 38th Int. Conf. Mach. Learn. (ICML), 2021, pp. 10347–10357.
23. H. Li, M. Sui, F. Zhao, Z. Zha, and F. Wu, “MViT: Mask Vision Transformer for Facial Expression Recognition in the wild,” arXiv preprint arXiv:2106.04520, Jun. 2021.
24. A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.
25. G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” arXiv preprint arXiv:1503.02531, 2015.
26. Y. Liu et al., “Cross-architecture knowledge distillation,” in Proc. Asian Conf. Comput. Vis. (ACCV), 2022.
27. S. Shin, J. Lee, J. Lee, Y. Yu, and K. Lee, “Teaching where to look: Attention similarity knowledge distillation for low-resolution face recognition,” in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. 13672, pp. 631–647, 2022.
28. K. Lee, S. Kim, and E. C. Lee, “Fast and accurate facial expression image classification and regression method based on knowledge distillation,” Appl. Sci., vol. 13, no. 11, art. no. 6409, May 2023, doi: 10.3390/app13116409.
29. N. A. Talemi, H. Kashiani, and N. M. Nasrabadi, “CATFace: Cross-attribute-guided transformer with self-attention distillation for low-quality face recognition,” IEEE Trans. Biom. Behav. Identity Sci., vol. 6, no. 1, pp. 132–146, Jan. 2024.
30. F. A. Boiten, N. H. Frijda, and C. J. Wientjes, “Emotions and respiratory patterns: Review and critical analysis,” Int. J. Psychophysiol., vol. 17, no. 2, pp. 103–128, 1994.
31. P. Philippot, G. Chapelle, and S. Blairy, “Respiratory feedback in the generation of emotion,” Cogn. Emotion, vol. 16, no. 5, pp. 605–627, 2002.
32. R. Jerath and C. Beveridge, “Respiratory rhythm, autonomic modulation, and the spectrum of emotions: The future of emotion recognition and modulation,” Front. Psychol., vol. 11, art. no. 1980, 2020.
33. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
34. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
35. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2014.