AFairDNet: Actively Empowering Fair Multisensor Emotion Recognition with Chain-of-Thought on Diffused Biosignals

Jatin Chhabria*, Ritik Verma*, Sreyasee Das Bhattacharjee, Wenyao Xu, Wei Bo
Department of Computer Science & Engineering,
State University of New York at Buffalo, NY, USA
{jatinjay, ritikver, sreyasee, wenyaoxu, weibo}@buffalo.edu

2025 IEEE 21st International Conference on Body Sensor Networks (BSN) | 979-8-3315-5454-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/BSN66969.2025.11337342

***Abstract*—The scarcity of reliable, extensive datasets hampers the training of effective models for wearable healthcare technology. This data gap frequently introduces biases into training sets, which then carry over into the models themselves. Such inherent biases pose substantial fairness challenges, particularly in sensitive healthcare scenarios. To this end, we propose *AFairDNet*, an effective active learning framework that utilizes a small collection of annotated data to create an initial classifier, and then continually refines it by incorporating synthesized 'hard' signals, representing areas where the model's training is currently insufficient. To ensure both creativity and ethical responsibility in these generated signals, we enhance the signal generation process using Chain of Thought (CoT) reasoning. The model employs real-time iterative CoT refinement of the model's text prompts to condition the multisensor signal diffuser, ensuring that the synthesized multisensor biosignals are not only of high quality but also semantically faithful. Extensive evaluations using two large publicly available multisensor emotion recognition datasets demonstrate that by leveraging a small yet comprehensive collection of synthesized samples (i.e., around 1.4% of the total training set), *AFairDNet* may boost a baseline classifier's performance, outperforming the state-of-the-art methods. More precisely, in addition to achieving $1.5 - 3\%$ higher accuracy than current supervised and self-supervised baselines, *AFairDNet* also boasts an impressive *Total Fairness Score*, signaling its potential for more responsible and transparent AI-driven synthesized signal generation.**

***Index Terms*—Fairness, Chain-of-Thought, Wearables, Biosignals, Conditional Diffusion, Emotion Recognition**

I. Introduction

Recognizing and understanding human emotions [1] is a critical first step in facilitating effective intra- and inter-human, as well as human-computer interactions to ensure personal objectives. A variety of physiological data (such as electrocardiogram (ECG), photoplethysmography (PPG), electrodermal activity (EDA), and skin temperature) generated from a range of user-friendly wearable devices have sparked significant research interest, due to their ability to uninterruptedly measure and track bodily states without interruption that reflect emotional conditions [2].

While significant progress has been made [3], [4], continuously monitoring multiple physiological signals remains a complex task fraught with pragmatic challenges. Despite the widespread presence of practical difficulties (e.g., scarcity, noise, bias), the vast majority of existing literature, aside from a few recent works [5], [6], fails to address these practical difficulties in algorithm building. This oversight means that consistently achieving strong performance of the majority of these state-of-the-art models may not translate into various application settings due to several hurdles: *First*, Emotional perception is subjective and may vary from person to person depending on their personal social and demographic backgrounds. Given this, the data-intensive nature of the emotion recognition models risks propagating such societal or demographic biases, potentially generating fundamentally discriminatory results. To this end, large language models' reasoning abilities offer a promising solution to reduce discriminatory outputs. For instance, the WESAD dataset exhibits a significant gender imbalance, with approximately 79.7% of participants being male while only 20.3% female. This imbalance can introduce bias and can lead to under-representation of female physiological patterns. *Secondly*, imbalances in class-specific data distributions in the training set present another significant challenge, particularly for intricate tasks like ours, i.e., recognizing and tracking human emotions. For instance, the CASE dataset exemplifies this, exhibiting

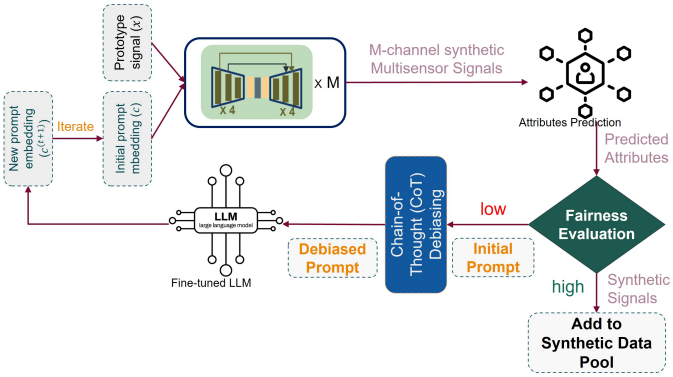

Fig. 1. Overview of *Fairness-Aware Synthetic Multisensor Signal Generation* process, where the 'initial prompt' is converted into a 'debiased prompt' to provide an improved context to the next iteration of the signal generation process. For example, an 'initial prompt' described as "Step by Step generate 60-second length BVP, EDA, and TEMP signal, ensuring fair demographic attributes may be considered. Avoid dominant attribute values and categories in the dataset." is automatically updated via fairness-aware Chain-of-Thought (CoT) reasoning to a 'debiased prompt' as"Generate 60-second length BVP, EDA, and TEMP signal, labeled with Arousal and Valence. Include only the details of mean and standard deviation in the prompts."

*These authors contributed equally to this work

a severely skewed distribution with only 4.44% negative, 81.2% neutral, and 14.36% positive samples across its three categories. *Finally*, due to limited annotated samples, ensuring model generalization is another persistent concern; To address these, we propose *AFairDNet*, an effective active learning-based framework that can upgrade a baseline classifier (both in terms of decision accuracy and fairness) on the fly via generating a fair set of synthesized samples. Its primary contributions include:

1) *A fairness-aware Chain-of-Thought (CoT) reasoning within multiple mode-specific conditional signal diffusion models* to refine and guide the generation of signals that closely mimic the identified rare patterns of shortlisted 'hard' signals (i.e., signals which were not correctly classified by the existing version of the classifier) on the fly in a more equitable manner.
2) *An iterative training approach* highlights how continuous model refinement excels in evolving problem environments and captures instance-level "hardness" from both modal and multimodal data.
3) *Extensive evaluation* proves that the proposed *AFairDNet* improves the baseline classifier's performance to supersede existing state-of-the-art supervised and self-supervised methodologies in the identification of diverse human mental health states, including stress and arousal.

II. Proposed Method

From a dataset $\mathcal{D}$ of multisensor signals (comprising M sensor-specific 1D time sequences), our objective is to evaluate the emotional state y_i of a subject based on their biosignal x_i and its associated demographic description c_i. In other words, we have $x_i = \{s_i^1, s_i^2, \ldots, s_i^M\}$, where $s_i^m \in \mathbb{R}^{N\times 1}$ and N is the signal length and $m \in \{1, \cdot, M\}$. The proposed multisensor fusion network *AFairDNet* is comprised of four modules: *Signal Embedding*; *Multisensor Classifier*; *Active Learning-based Model Training*; and *Fairness-Aware Synthetic Multisensor Signal Generation* (an overview is shown in Figure 1).

A. Signal Embeddding

A Temporal Convolution Network (TCN) [7] encodes each normalized s_i into an embedding $\mathbf{e}_i \in \mathbb{R}^{N\times d_e}$, where $d_e >> 1$ is the embedding dimension set by final TCN layer's filter size.

B. Multisensor Classifier

A multi-sensor signal x_i, represented by M TCN-generated sensor-specific encoders $\{\mathbf{e}_i^1, \mathbf{e}_i^2, \ldots, \mathbf{e}_i^M\}$, is compactly represented as $\mathbf{t}_i = \mathbf{e}_i^1 \oplus \mathbf{e}_i^2 \oplus \ldots \mathbf{e}_i^M$. This $\mathbf{t}_i$, paired with ground truth label y_i, trains a classifier head (θ), which is a three-layer perceptron with GeLU activation and dropout. We optimize the model using the Cross Entropy (CE) loss between the predicted output $P(\mathbf{t}_i|\theta)$ and y_i.

C. Active Learning-based Model Training

To address the challenge of imbalanced datasets, where certain minority classes or rare patterns are under-represented, this paper proposes a classifier-agnostic active learning method, which automatically identifies and shortlists a limited number of "hard" samples for the model to revisit during subsequent training phases. An effective uncertainty sampling strategy [8] is adopted to pinpoint ambiguous data patterns. These shortlisted prototypes act as the inputs to the following *fairness-Aware Synthetic Multisensor Signal Generation* module. This module kicks off a fairness-aware Chain-of-Thought (CoT) reasoning process, which can then fine-tune the multisensor signal diffusor composed of M sensor-specific, pretrained diffusers. The selection criteria for these "hard" samples include: (1) instances the current classifier misidentifies; (2) samples located near the classifier's decision boundary; and (3) data points where the confidence scores for different possible labels are very similar. The uncertainty scores for a sample $x_i \in \mathcal{D}$ is computed as:$U_1(x_i|\theta) = max(P(y_i^p = y_i|x), P(y_i^p \neq y_i|x))$ and $U_2(x_i|\theta) = |P(y_i^p = y_i|x) - P(y_i^p \neq y_i|x))|$. For a given sample x_i, y_i^p is the class label predicted by the underlying classifier (θ). In our experiments, we labeled a sample x_i as 'hard' if either $U_1(x_i|\theta) < \beta$ or $U_2(x_i|\theta) < \eta$, and we chose $\beta = 0.7$ and $\eta = 0.5$.

D. Fairness-Aware Synthetic Multisensor Signal Generation

1) Initial Signal Generation:: The proposed *Multisensor signal diffusor* conditioned with the BERT embedding ($\mathbf{C}_i$) of a comprehensive metadata description (c_i) of the input multisensor signal (x_i) as a prototype, is used to generate the initial set of synthetic signals $\mathcal{S}_i := \{x_i^{syn_j}\}_{syn_j}$, where $x_i^{syn_j} = \{s_i^{syn_j,1}, s_i^{syn_j,2}, \ldots, s_i^{syn_j,M}\}$ and $s_i^{syn_j,m} \in \mathbb{R}^{N\times 1}$. This initial synthetic signal generation process provides the input to the subsequent bias assessments and makes refinements. In our experiments, the architecture of each of the M sensor-specific conditional diffusion models is designed using a mode-specific, finetuned version of the pretrained BioDiffusion model [9].

Within each mode-specific U-Net architecture, the forward phase involves augmenting each residual block with both the text embedding vector $\mathbf{C}_i$ and the current diffusion timestep. Conversely, in the backward phase, the mode-specific diffusion model processes noise sampled from a normal distribution. This noise is enhanced by two additional inputs: a textual metadata description covering a variety of statistical features (e.g., range, kurtosis, interquartile range) and an example signal whose pattern the model should replicate during synthesis.

2) Attribute Prediction:: For each key attribute (e.g., age, gender), we design a two-layer perceptron that uses the concatenated TCN-generated multisensor encoder $\mathbf{t}_i^{syn}$ as the network input, *GeLU* activation (followed by dropout), and the last Softmax layer. We apply the *Cross Entropy classification loss* on the model's prediction. In our experiments, we have used two genders (M/F) and four age baskets $[18, 25], [25, 35], [35, 50], [50, 100]$, which determine the number of units in the last Softmax layer in the attribute-specific perceptron model.

3) Bias Evaluation:: To evaluate the existing bias within the initial synthetic signal collection, we use normalized entropy $E_{a,i} = -\frac{1}{log(v)} \sum_{v=1}^{V} p(a_v)(log(p(a_v))$ and $L_i = \frac{1}{|\mathcal{S}_i|} \sum_{x^{syn}\in\mathcal{S}_i} cos(\phi(c_i), \phi(c_i^{syn}))$, where $p(a_v)$ is the probability of a_v approximated from $\mathcal{D} \cup \mathcal{S}_i$ and V is the number the possible values (or range of values) for the attribute a_v.

The term c_i^{syn} is the metadata description of $x^{syn} \in \mathcal{S}_i$ and ϕ represents the CLIP's text encoder [10]. Finally, we use a comprehensive *fairness score* $F_{\mathcal{S}_i} = \alpha_1\big(\sum_{a\in\{age,gender\}} E_{a,i}\big) + \alpha_2(\sigma(L_i))$ to decide on the requirement of subsequent debiasing process, described below. The sigmoid function is denoted by $\sigma()$. In all our experiments, we use $\alpha_1 = \alpha_2 = 0.5$.

4) Chain-of-Thought Debiasing and Finalizing the Signal Collection: : The zeroshot debiasing approach adopted from [11] is used for the initial iteration of Chain-of-Thought with prompts like "Step by Step generate 60-second length BVP, EDA, and TEMP signal, ensuring fair demographic attributes may be considered. Avoid dominant attribute values and categories in the dataset." However, if there is a consistently low *fairness score* in the subsequent iterations, the prompt is appropriately modified. for instance, if $E_{age,i}$ is consistently low, the prompt is modified as "think again, focusing specifically on improving fairness on age attribute." This iterative improvement process is terminated using a predefined stopping criterion (e.g., maximum number of iterations is performed, consecutive iterations of improvement do not change the *fairness score*).

TABLE I
DISTRIBUTION OF SAMPLES

Dataset	Task	Category (no. of samples)
WESAD	Emotion-3	baseline (58692), stress (33221), amusement (18584)
CASE	Valence-3	negative (3958), neutral (72283), positive (12785)
CASE	Arousal-3	low (2228), medium (75738), high (11060)

III. EXPERIMENTS

The proposed *AFairDNet* is evaluated using the two largest publicly available biosignal-based affective datasets - CASE [12] and WESAD [13], which exhibit significant class imbalances, as reported in table I. The WESAD dataset categorizes physiological responses as amusement, stress, or baseline. Complementing this, the CASE dataset supports two distinct classification tasks: sorting physiological signals by three valence levels (negative, neutral, positive) (aka. Valence-3) or three arousal levels: low, medium, high (aka. Arousal-3). Together, these resources establish a comprehensive framework for studying the intricate links between physiological signals and emotional states, pushing forward both affective computing and physiological research.

A. Data Processing

To ensure a consistent sampling frequency, we downsampled all physiological signals in both datasets to 4Hz. Following this, we segmented the data into 60-second windows. These windows featured a high degree of overlap: 99.5% for the WESAD dataset and 99% for the CASE dataset. When a segment contained multiple labels, we assigned the majority label to that segment, a method consistent with prior research [14]. To mitigate inter-subject variability in physiological responses, we applied Z-score normalization to each subject's recorded data, as detailed by [15]. This normalization step helps ensure that differences in physiological signals are primarily due to emotional states rather than individual biological variations.

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED *AFairDNet* NETWORK MODEL WITH MULTIPLE STATE-OF-THE-ART METHODS USING THE *Accuracy* AND *F1Score* METRICS. THE *Multisensor Classifier* METHOD USES THE CLASSIFIER DESCRIBED IN SECTION II(B). IT IS TRAINED USING THE ACTIVE LEARNING BASED METHOD EXPLAINED IN SECTION II(B), WHEREIN THE SYNTHESIZED MULTISENSOR SIGNALS USING THE PRETRAINED BIODIFFUSION MODEL [9] ARE DIRECTLY USED IN THE NEXT ITERATION OF ACTIVE LEARNING, BUT NO FAIRNESS CHECK WAS PERFORMED

Dataset	Task	Method	*Accuracy*	*F1Score*
WESAD	Emotion-3	WESAD-Wrist [13]	75.21	64.12
		SimpDCNN [16]	78.3	74.59
		RF [13]	76.17	66.33
		LDA [13]	68.85	58.18
		SigRep [14]	78.13	77.35
		SSL [3]	78.7	75.98
		S&T [17]	69.84	73.86
		Multisensor Classifier	77.41	74.50
		AFairDNet (ours)	81.78	79.47
CASE	Valence-3	SSL [3]	78.99	76.66
		SimpDCNN [16]	59.2	51.95
		SigRep [14]	64.83	60.25
		MULT [18]	63.14	62.5
		CorrNet [19]	65.14	53.00
		S&T [17]	70.28	59.87
		Multisensor Classifier	75.62	74.48
		AFairDNet (ours)	80.12	78.36
CASE	Arousal-3	SSL [3]	85.38	82.63
		SimpDCNN [16]	56.8	53.85
		SigRep [14]	65.07	61.08
		MULT [18]	62.15	58.48
		CorrNet [19]	58.22	55.00
		S&T [17]	68.36	58.22
		Multisensor Classifier	83.97	81.78
		AFairDNet (ours)	86.97	84.84

B. Results

1) Comparative Study: We evaluated our model's performance using *F1Score* and *Accuracy*, aligning with standard evaluation practices, wherein the objective is to recognize diverse emotion and affective state categories demonstrated by the subjects in different datasets. Following established baseline methodologies [14], our primary evaluation employed Leave-One-Subject-Out (LOSO) cross-validation. For the ablation study, we adopted a more stringent evaluation protocol. We randomly divided the data subject-wise into a $3:1:1$ ratio. Average categorical prediction *Accuracy* and the average *F1Score* calculated across all P iterations are reported in the tables. The proposed *AFairDNet* is evaluated against multiple state-of-the-art baseline models. Table 1 reports the comparative performance. Across multiple experimental settings, *AFairDNet* consistently outperforms existing baseline methods, including the top-performing SSL model [3]. Specifically, *AFairDNet* shows an approximate 1.5% F1-score improvement for the Valence-3 task and a 2% F1-score improvement for the Arousal-3 task on the CASE dataset. Furthermore, on the WESAD dataset, ActDiffNet achieves roughly 3% higher *Accuracy* and 4% higher *F1Scores*.

2) Ablation Study: : To evaluate the model's effectiveness in generating a useful and diverse set of samples that are generated equitably, we use varying-sized synthetic sample collections generated by the *Fairness-Aware Synthetic Multisensor Signal Generation* module to fine-tune the multisensor classifier head (please see Section II(B)), as described in

TABLE III
ABLATION STUDY THAT REPORTS THE PERFORMANCE OF *AFairDNet* IN A VARIETY OF EXPERIMENTAL SETTINGS, WHEREIN A VARIED NUMBER OF SYNTHETIC SAMPLES ARE GENERATED BY THE *Fairness-Aware Synthetic Multi- sensor Signal Generation* MODULE TO FINE-TUNE THE MULTISENSOR CLASSIFIER HEAD

Dataset(Task)	#Syn. Samples	*Avg. Fairness Score*	*Accuracy*	*F1Score*
WESAD (Emotion-3)	0	0.78	77.41	74.5
	600	1.13	78.27	75.96
	1200	1.17	79.18	77.08
	2400	1.21	81.78	79.47
CASE (Valence-3)	0	0.91	75.62	74.48
	600	1.14	77.58	77.60
	1200	1.13	78.17	77.04
	2400	1.28	80.12	78.36
CASE (Arousal-3)	0	0.92	83.97	81.78
	600	1.17	84.34	83.92
	1200	1.21	85.12	84.03
	2400	1.33	86.97	84.84

Section II(D). Table 2 reports both correctness (*Accuracy* and *F1Score*) and *Avg. Fairness Score* that is computed as the average of the fairness scores for all identified prototypes used for generating the synthesized signals in a given experimental setting. As observed, by leveraging a small yet comprehensive collection of synthesized samples, *AFairDNet* may boost a baseline classifier's performance significantly. More precisely, via targeted fine-tuning with a set of only $1,200$ synthesized signals mimicking the prototypes (i.e. around 1.4% of the total training size), which were initially found as 'hard' by the baseline classifier, the model achieves comparable performance as reported by the best performing baseline SSL [3] in all three experiments and finally with $2,400$ synthesized samples, the proposed *AFairDNet* outperforms. As reported in the table, the model not only demonstrates a dominating precision performance but also shows a consistently robust *Avg. Fairness Score*.

IV. CONCLUSION

In this work, we introduce *AFairDNet*, an innovative framework that blends active learning, chain-of-thought reasoning, and fairness-driven synthetic data to tackle multisensor emotion recognition. In particular, the model uses an internal chain-of-thought reasoning to guide its signal generation module that may effectively address critical demographic imbalances across attributes such as gender, age. This helps not only in ensuring superior precision performance on public datasets like CASE and WESAD but also promises impressive fairness scores, underscoring its model-agnostic design to ensure compatibility with both priority and open-source systems. An immediate future research would involve extending the chain-of-thought prompting to be adaptive, so that the model's internal reasoning can adjust automatically for each new user's unique personal contexts.

V. ACKNOWLEDGEMENT

The project was partially funded by the National Science Foundation, Award ID: 2347251

REFERENCES

[1] Vamsi Kumar Naidu Pallapothula, Sidharth Anand, Sreyasee Das Bhattacharjee, and Junsong Yuan, "Generalized multisensor wearable signal fusion for emotion recognition from noisy and incomplete data," *Smart Health*, p. 100571, 2025.

[2] Anubhav Bhatti, Behnam Behinaein, Paul Hungler, and Ali Etemad, "Attx: Attentive cross-connections for fusion of wearable signals in emotion recognition," *ACM Transactions on Computing for Healthcare*, vol. 5, no. 3, pp. 1–24, 2024.

[3] Yujin Wu, Mohamed Daoudi, and Ali Amad, "Transformer-based self-supervised multimodal representation learning for wearable emotion recognition," *IEEE Transactions on Affective Computing*, vol. 15, no. 1, pp. 157–172, 2023.

[4] Sirat Samyoun, Md Mofijul Islam, Tariq Iqbal, and John Stankovic, "M3sense: Affect-agnostic multitask representation learning using multimodal wearable sensors," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–32, 2022.

[5] Kleanthis Avramidis, Dominika Kunc, Bartosz Perz, Kranti Adsul, Tiantian Feng, Przemysław Kazienko, Stanisław Saganowski, and Shrikanth Narayanan, "Scaling representation learning from ubiquitous ecg with state-space models," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[6] Flavio Di Martino and Franca Delmastro, "Challenges and limitations in the synthetic generation of mhealth sensor data," *arXiv preprint arXiv:2505.14206*, 2025.

[7] Colin Lea, Rene Vidal, Austin Reiter, and Gregory D Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 47–54.

[8] Sreyasee Das Bhattacharjee, Ashit Talukder, and Bala Venkatram Balantrapu, "Active learning based news veracity detection with feature weighting and deep-shallow fusion," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 556–565.

[9] Xiaomin Li, Mykhailo Sakevych, Gentry Atkinson, and Vangelis Metsis, "Biodiffusion: A versatile diffusion model for biomedical signal synthesis," *arXiv e-prints*, pp. arXiv–2401, 2024.

[10] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al., "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.

[11] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa, "Large language models are zero-shot reasoners," *Advances in neural information processing systems*, vol. 35, pp. 22199–22213, 2022.

[12] Karan Sharma, Claudio Castellini, Egon L Van Den Broek, Alin Albu-Schaeffer, and Friedhelm Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific data*, vol. 6, no. 1, pp. 196, 2019.

[13] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[14] Vipula Dissanayake, Sachith Seneviratne, Rajib Rana, Elliott Wen, Tharindu Kaluarachchi, and Suranga Nanayakkara, "Sigrep: Toward robust wearable emotion recognition with contrastive representation learning," *IEEE Access*, vol. 10, pp. 18105–18120, 2022.

[15] Pritam Sarkar and Ali Etemad, "Self-supervised ecg representation learning for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1541–1554, 2020.

[16] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–30, 2019.

[17] Aaqib Saeed, Victor Ungureanu, and Beat Gfeller, "Sense and learn: Self-supervision for omnipresent sensors," *Machine Learning with Applications*, vol. 6, pp. 100152, 2021.

[18] Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in *Proceedings of the conference. Association for computational linguistics. Meeting*. NIH Public Access, 2019, vol. 2019, p. 6558.

[19] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar, "Cornet: Fine-grained emotion recognition for video watching using wearable physiological sensors," *Sensors*, vol. 21, no. 1, pp. 52, 2020.