2024 IEEE 10th International Conference on Collaboration and Internet Computing (CIC) | 979-8-3503-8670-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/CIC62241.2024.00013

Towards Contactless Human Concentration Monitoring Using mmWave Signals

Yuan Ge*, Yi Wei*, Xiaonan Guo*, Yucheng Xie†, Yan Wang‡, Jerry Cheng§, Yingying Chen¶
*George Mason University, USA
†Yeshiva University, USA
‡Temple University, USA
§New York Institute of Technology, USA
¶Rutgers University, USA
Email: *{yge3, ywei8, xguo8}@gmu.edu, †yucheng.xie@yu.edu, ‡y.wang@temple.edu,
§jcheng18@nyit.edu, ¶yingche@scarletmail.rutgers.edu

***Abstract*—Maintaining concentration in today's complex and distracting environments is increasingly challenging, with significant impacts on productivity, learning outcomes, and safety. Traditional methods like self-reporting and observational studies are subjective and labor-intensive. Current approaches, including camera-based systems and wearable sensors, raise privacy concerns and require continuous physical interaction. To address these limitations, we propose a novel, contactless concentration monitoring system using mmWave technology. Our system leverages Commercial-Off-The-Shelf (COTS) mmWave devices to detect concentration-related activities, such as eye blinking, nodding, yawning and leg shaking. In particular, we enhance the activity detection and overcome the limited field of view (FOV) of mmWave devices through spatial decomposition based on Delay-and-Sum (DAS) beamforming technologies. Moreover, we mitigate interference in concurrent activities by exploiting the distinct frequency ranges associated with each concentration-related activity based on Short-Time Fourier Transform (STFT). A CNN model, integrated with domain adaptation techniques, ensures robust performance in diverse environments. Experiments involving 10 volunteers demonstrated an overall accuracy of 95.3% in detecting human activities. The system maintained robust performance at distances up to 150 cm and across different office environments. Our method offers a contactless and privacy-preserving alternative to current approaches, making it suitable for applications such as classroom monitoring, workplace productivity observance, and cognitive health monitoring.**

***Index Terms*—mmWave sensing, Concentration monitoring, Activity recognition, Machine Learning**

I. Introduction

Concentration is a fundamental cognitive function that plays a crucial role in determining a person's productivity, learning outcomes, and personal safety across a wide range of environments, from classrooms to workplaces and safety-critical settings [7]. Despite its importance, maintaining focus has become increasingly challenging due to the demands of modern life. The growing complexity of daily tasks and the constant distractions reveal a significant gap in our ability to effectively monitor and sustain concentration. This gap is particularly concerning given the rising prevalence of attention-related disorders, such as Adult Attention Deficit Hyperactivity Disorder (ADHD), with recent studies reporting a substantial increase in diagnoses among adults [2]. Therefore, effective concentration monitoring is highly desired, as it is crucial for identifying factors that contribute to distractions, and for enabling the design of more focused and productive environments. To meet this need, in this paper, we develop a system that leverages millimeter wave (mmWave) technology to perform concentration monitoring. As illustrated in Fig. 1, such a system can be particularly useful in educational environments by helping teachers identify periods of high student engagement, allowing for optimized lesson pacing and content delivery. It could also assist in early detection of attention difficulties, enabling timely interventions for students who may need additional support. In the field of psychology, the technology could be used to count the unconcentrated behaviors and help the psychologists to analyze the correlation between concentration levels and those distracted behaviors.

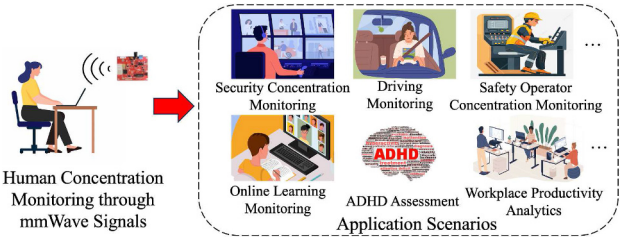


Fig. 1: Application scenarios for human concentration monitoring using mmWave signals.

Traditional approaches, such as self-reporting [5] and observational studies [22], have been widely used in concentration assessment. Self-reporting provides a direct insight into an individual's perceived concentration levels and is straightforward to implement. However, self-reporting is subjective and can be influenced by self-awareness or a desire to give favorable impressions [20]. Observational studies, conducted by trained professionals, offer detailed qualitative data on concentration behaviors but are labor-intensive and may unintentionally alter the subject's natural behavior due to the observer's presence [23]. Researchers have explored wearable sensors to measure physiological indicators of concentration such as heart rate variability (HRV) [4], skin conductance [24], and even brain

979-8-3503-8670-7/24/$31.00 ©2024 IEEE
DOI 10.1109/CIC62241.2024.00013

activity through electroencephalogram (EEG) [27]. These sensors provide accurate, real-time data and can be used in various environments. However, they require direct contact with the user, which can be uncomfortable or impractical for long-term use.

Recent research has identified several activities which are strongly correlated with concentration levels. These activities include leg shaking, eye blinking, nodding, and yawning. Leg shaking can indicate restlessness or a decline of focus [26], and is also associated with ADHD [1]. Changes in eye blinking rate can indicate varying levels of cognitive engagement [17]. Similarly, frequent nodding or yawning might suggest fatigue or waning attention levels [12] [33], providing crucial insights into an individual's concentration patterns over time. These behaviors serve as valuable indicators for assessing an individual's cognitive state, making them ideal for contactless concentration monitoring. Building on these findings, researchers have developed camera-based systems to capture subtle facial expressions [30], eye movements [6], and body language [21] to infer concentration levels. While these systems provide the advantage of passive monitoring without requiring active participation, they often raise significant privacy concerns.

Recently, millimeter wave (mmWave) technology has been integrated into current and next-generation wireless protocols, such as WiGig (IEEE 802.11ad and 802.11ay) [11] and 5G [9], expanding its potential for widespread adoption and application. Leveraging this technology, researchers have successfully applied mmWave sensing to a wide range of activity recognition tasks. For instance, Wang *et al.* [29] demonstrate the use of WiFi signals for human activity recognition. Liu *et al.* [16] develop a mmWave-based system that accurately recognizes arm gestures. These successes motivate us to leverage mmWave technology for concentration monitoring based on recognizing concentration-related activities. However, existing mmWave-based approaches for activity recognition cannot be directly applied to the task of concentration-related activity monitoring given the following challenges: (1) The 120-degree field of view (FOV) of Commercial-off-the-Shelf (COTS) mmWave devices limits their ability to capture full-body movements, especially when the mmWave device is positioned on a desk to monitor concentration while someone is working at a computer. In such a setup, the device typically focuses on upper body movements, but may miss crucial lower body indicators like leg shaking, which can also signal changes in concentration; (2) Separating concurrent-related activities is challenging, as subtle movements such as eye blinking often occur simultaneously with other activity like yawning or nodding, complicating accurate detection and classification of concentrated versus unconcentrated states; (3) Deploying the system in new environments can be challenging, as even small changes in the layout of desks or nearby objects may alter the signal propagation path, potentially affecting system performance.

To overcome the aforementioned challenges, we develop a system that contactlessly monitors concentration-related activities using mmWave technology. Our approach employs a multi-stage signal processing pipeline that includes noise reduction and spatial-temporal feature extraction to isolate subtle concentration-related movements. To overcome the limited FOV of COTS mmWave devices, we develop a spatial decomposition approach using the Delay-and-Sum (DAS) beamforming technique to enhance signals reflected from specific body parts involved in different activities. For detecting lower body movements, particularly leg shaking, we resort to an innovative indirect approach. By monitoring subtle movements induced in the belly area, we can infer leg activity even when the legs are obscured (e.g., under a table). This approach exploits the fact that leg shaking induces subtle yet detectable vibrations that propagate through the body. Moreover, to tackle the challenge of concurrent activity monitoring, we leverage frequency domain analysis to detect dominant frequencies of different activities. By decomposing the mmWave signals into their frequency components, we can simultaneously monitor multiple activities while mitigating interference between them. This process, combined with the spatial information extracted earlier, ensures accurate differentiation of concurrent activities. To enhance adaptability across different environments, we first employ a Convolutional Neural Network (CNN)-based model, which is highly effective at extracting spatial features from activity data. Further, to handle variations in the environment, we integrate domain adaptation techniques, making the feature extraction process domain-independent. This enables the system to reliably detect concentration-related activities with the presence of environmental factors. The main contributions of our work are as follows:

1) We develop a novel mmWave-based system for contactless and privacy-preserving concentration monitoring. This system addresses the growing need for unobtrusive methods to assess cognitive states in various settings, including workplaces, educational environments, and healthcare facilities.
2) We implement the beamforming technologies to address the limited FOV of mmWave radar, utilize an indirect monitoring for belly movements, allowing for the detection of subtle shakes associated with concentration-related activity.
3) We design a novel peak frequency identification algorithm that accurately captures and separates concurrent concentration-related activities through frequency analysis, overcoming the challenge of concurrent activities.
4) We integrate domain adaptation techniques into the system, ensuring the feature extraction process is robust and environment-independent, allowing reliable detection of concentration-related activities across diverse settings with varying layouts and surrounding objects.
5) We conduct experiments with 10 volunteers performing multiple concentration-related activities under different environmental conditions to evaluate the system's performance across various real-world scenarios. Results show that our system can achieve an overall accuracy of 95.3% for concentration-related activities identification.

II. Related Work

Sensor-based. Traditional methods for concentration monitoring primarily rely on wearable sensors [10, 27]. Han *et al.* [10] present a stress monitoring system based on three physiological signals: electrocardiogram (ECG), photoplethysmogram (PPG), and galvanic skin response (GSR) using Shimmer3 ECG, Shimmer3 GSR+, and Empatica E4 wearable sensors. Similarly, Velnath *et al.* [27] propose to extract different features from the collected EEG signals. The level of concentration is determined by comparing the features extracted from individuals of different age groups. However, wearable sensors, particularly those based on brain wave measurements such as EEG electrode caps or patches, can be cumbersome and intrusive. Furthermore, the need to remove and put back these devices during temporary breaks in monitoring disrupts the user experience and may lead to inconsistent data collection.

Camera-based. Camera-based technologies have emerged as a popular method for detecting user concentration levels due to their relative convenience and non-invasive nature. These systems typically analyze limb movements, eye behavior, pupil dilation, or facial expressions to infer a user's level of concentration [14, 19, 25]. Meriem *et al.* [19] find that students' emotions, inferred through facial expressions, are related to their attention levels. They develop a computer vision-based method to classify attention into three levels by correlating these emotions with students' concentration during class. Moreover, Lee *et al.* [14] propose a personal attention level monitoring system that focuses on users' pupil responses and blinking patterns while they perform online tasks on a computer. Tanaka *et al.* [25] utilize a camera based eye-tracker to explore a pipeline for constructing machine learning models to recognize the state of concentration using eye-gaze data during reading. However, camera-based solutions, while effective, are vulnerable to environmental variables, especially lighting conditions, which can compromise data accuracy and reliability. Furthermore, the persistent capture of visual information raises significant privacy concerns, potentially deterring widespread adoption.

RF-Based. To address the mentioned weaknesses, researchers have explored WiFi-based solutions for their convenience and sensing capabilities [8, 28]. Guo *et al.* [8] propose a device-free exercise recognition and assessment scheme using existing WiFi infrastructures. Wang *et al.* [28] study the domain variation problem and design a robust WiFi sensing framework. While WiFi technology can recognize user actions, it has not yet been used to infer concentration levels. Meanwhile, mmWave sensing is gaining attention with the development of IoT, 5G, and autonomous driving technologies [31]. It offers contactless, fine-grained sensing of humans and objects [32]. Due to its low cost and non-intrusive nature, mmWave-based human activity sensing has become a significant research area. Cardillo *et al.* [3] use 120 GHz radar to detect head movements and eye blinking, aiding communication for individuals with neurodegenerative disorders. Juncen *et al.* [12] develop techniques to filter noise from driving-related activities, accurately detecting driver fatigue. Thus, we can leverage RF-based action recognition to infer concentration. Our mmWave radar-based approach provides a non-intrusive, privacy-preserving alternative to camera, EEG, and eye-tracker systems, allowing continuous monitoring without the discomfort of wearables. By integrating signal processing and machine learning, our method isolates and analyzes concentration-related movements, offering reliable performance in various settings like education and cognitive health.

III. Preliminaries

A. mmWave Radar Fundamentals

This work utilizes an FMCW mmWave radar to detect user macro and micro-actions. The radar continuously transmits chirp signals that linearly sweep through a frequency bandwidth B over a chirp duration of T_c. The sweep slope is therefore $S = \frac{B}{T_c}$. The received signal is a delayed version of the transmitted signal due to the time it takes to travel through space. By calculating the frequency difference between the received and transmitted signals (i.e., beat frequency), we can directly determine the propagation time of the electromagnetic wave. Using the speed of the electromagnetic wave c, the propagation distance of the FMCW signal in space can be accurately calculated.

1) Range Estimation: The range information reveals the user's location and the relative positions of different body parts, such as the arms, stomach, legs, and head. To determine the range of these body parts, we apply a Fast Fourier Transform (FFT), specifically a range-FFT, on the time-domain intermediate frequency (IF) signal. When the user is within the field of view, the strong frequency response from their body creates peaks at various IF frequencies, corresponding to different body parts. The distance between each reflected point and the radar can then be calculated as follows:

$$d = \frac{f_{IF} \cdot c \cdot T_c}{2 \cdot B} = \frac{f_{IF} \cdot c}{2 \cdot S}, \tag{1}$$

where f_{IF} is the frequency of the intermediate frequency (IF) signal, c is the speed of the light, T_c is the period of one chirp, B is the bandwidth of the FMCW radar, and S is the chirp slope. The centimeter-level distance resolution of FMCW millimeter-wave radar enables it to precisely differentiate between the positions of an individual's head, limbs, and torso. Using range-FFT, we can divide the received signal into different range bins based on the distance from the reflection point to the radar's receiving antenna.

2) Angle Estimation: There are also situations where multiple reflection points fall into the same range bin but originate from different angles. For instance, when a user faces the mmWave radar, the distances from both arms to the radar's receiving antenna are nearly identical. For the same signal source, the distance of its reflected signal to different receiving antennas varies slightly, leading to small phase differences. The distance d between the reflected signal and the receiving

antenna is related to the distance l between the receiving antennas and the incident angle θ of the signal source. Knowing the arrangement and spacing of the receiving antennas, as well as the phase difference ω of the received signal, allows us to accurately calculate the angle of the signal source relative to the receiving antenna within the FOV. This phase difference across multiple TX antennas can then be used to estimate the angle of arrival (AOA) as follows:

$$\theta = \sin^{-1}(\frac{\lambda \cdot \omega}{2\pi l}). \tag{2}$$

3) Micro Displacement Estimation: Concentration-related activities involve subtle movements that require high-precision detection. To accurately capture these micro-movements, we need a detection granularity on the order of millimeters. The standard FMCW radar range resolution, typically around 4 cm, is not precise enough for detecting such subtle motions. By unwrapping the IF signal phase, we can extract micro displacements of the signal and achieve the required millimeter-level granularity because phase measurements are inherently more sensitive than amplitude measurements and the wavelength of mmWave signals is on the same order as the movements we aim to detect. The phase difference $\Delta\Phi(t, t - T_c)$ of the IF signal in a single range bin between two consecutive chirps at time t allows us to calculate the micro-distance change $\Delta d(t, t - T_c)$ between time t and $t - T_c$:

$$\Delta d(t, t - T_c) \approx \frac{c \cdot \Delta\Phi(t, t - T_c)}{4\pi f_c}. \tag{3}$$

In this paper, we utilize the extracted phase information from the IF signal to monitor concentration-related movements.

B. Feasibility Study

We conduct experiments with a volunteer to test the capability of millimeter-wave radar in detecting concentration-related activities. Specifically, we used the AWR1642 FMCW millimeter-wave radar, positioning it at a fixed location. The distance between the radar and the volunteer was set to 0.5 meters. To ensure consistency in data collection, the volunteer performed each activity with a 5-second interval, which helps distinguish between small and large movements related to concentration. The volunteer is asked to perform specific activities, such as eye blinking, leg shaking, nodding, and yawning, to assess the radar's ability to detect these behaviors.

We extract phase difference information from mmWave signals to detect concentration-related activities and compare the phase patterns of different activities. The phase data is derived from raw mmWave radar signals and processed to capture subtle movements associated with each activity. Fig. 2 shows these phase plots, where each graph illustrates the phase changes over time for specific activities like blinking, shaking, nodding, and yawning. By analyzing this phase data, we can observe and differentiate between the distinct phase change patterns associated with each activity. In this figure, the x-axis represents the duration of the activities, while the y-axis shows the phase changes. The ground truth, captured through camera recordings during the experiments, aligns with the phase

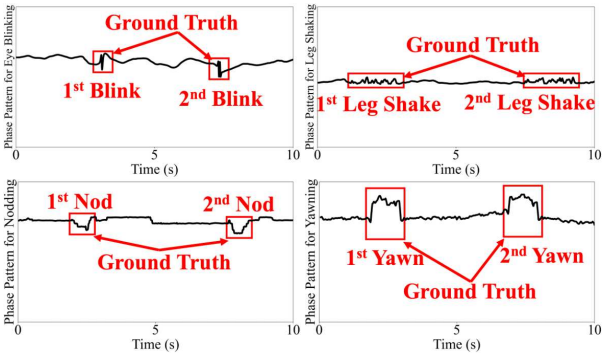


Fig. 2: Phase patterns corresponding to four concentration-related activities: eye blinking, leg shaking, nodding, and yawning. Each activity is performed twice with a 5-second interval between occurrences. The ground truth for each activity is marked with a red rectangle.

changes, confirming the occurrence of specific activities. For example, the "Blink" plot reveals subtle, rapid phase changes corresponding to each blink, reflecting the small, quick nature of this action. The "Shake" plot displays more pronounced, periodic phase changes with higher amplitude, consistent with the vigorous, repetitive motion of leg shaking. The "Nod" plot shows smoother, more gradual phase variations, indicative of the moderate, rhythmic motion of nodding. Finally, the "Yawn" plot exhibits significant amplitude in its phase changes, reflecting the larger, sustained motion of yawning. These observations demonstrate that each activity has unique phase characteristics, with different frequency components and the amplitude of phase changes. This allows us to distinguish between the activities based on their specific phase information.

IV. System Design

The proposed system is designed to continuously monitor concentration-related activities using mmWave technology by extracting phase features from radar signals. The system is capable of detecting both concentration-related activities—such as eye blinks, leg shaking, nodding, yawning, and non-concentration-related activities. As illustrated in Fig. 3, the *Signal Preprocessing Module* processes the collected mmWave signals to mitigate environmental impacts and reducing noise using the proposed two-stage filtering approach. Next, the *Enhansing Activity Detection Through Spatial Decomposition Module* employs a delay-and-sum (DAS) beamforming technique to enhance signals reflected from specific body parts associated with different activities (e.g., eye blinking in the upper body, leg shaking in the lower body). This module also determines the distances and angles for extracting each activity using range-angle heatmap. Furthermore, the *Distinguish Concurrent Activities Module* mitigate the interference in concurrent activities by employing dominant frequency detection through Short-Time Fourier Transform (STFT) on the extracted phase information. For continuous monitoring, the *Concentration-Related Activities Recognition Module* further segments the extracted phase data, isolating individual activity

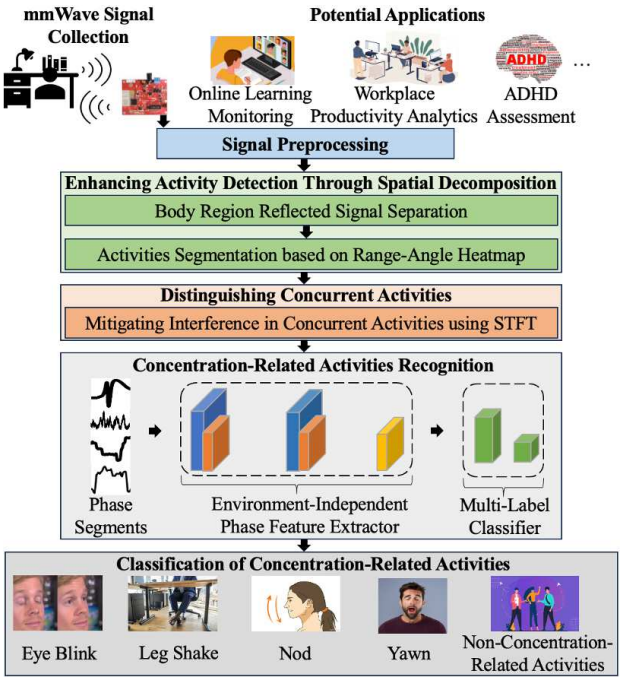


Fig. 3: System overview of the proposed system.

instances. Each segment is then fed into a CNN model for multi-label classification. In addition, the system incorporates Domain Adaptation to ensure the CNN model's feature extractor remains domain-independent, capable of extracting reliable activity features even the domain (e.g., environment) has been changed.

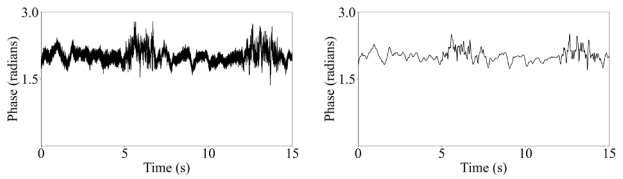
V. Methodology

A. Signal Preprocessing

Raw mmWave signals are inherently noisy and subject to interference from static objects and non-target movements, which can reduce the accuracy of target detection and range estimation. To enhance signal quality and isolate the relevant information, we employ a two-stage filtering approach. First, we apply a Finite Impulse Response (FIR) low-pass band filter to attenuate high-frequency noise while preserving the frequency content of concentration-related activities. Specifically, the output $y[n]$ of the FIR filter is given by:

$$y[n] = \sum_{k=0}^{M} b_k x[n-k], \tag{4}$$

where b_k are filter coefficients, $x[n]$ is the raw IF data, and M is the filter order. For our application, we set the cutoff frequency to 10 Hz to preserve movements in the 0.1-8 Hz range. By applying the FIR filter on the IF signal, the system attenuates unwanted high-frequency noise while preserving the beat frequency (i.e., difference in frequency between the transmitted chirp signal and the received reflected signal.) that contains the range information for the target. This step helps to improve the signal-to-noise ratio (SNR), leading to more accurate target detection and range estimation. After



(a) Original Phase vs. Time (b) Noise Mitigated Phase vs. Time

Fig. 4: Human body reflected mmWave signal phase pattern before and after the proposed two-stage noise mitigation.

filtering, we perform a range-FFT on the filtered IF signal, and extract phase information from the range-FFT output. To further reduce noise while preserving essential signal features, we apply a Savitzky-Golay smoothing filter to the extracted phase data. The smoothed output y_i^* is given by:

$$y_i^* = \sum_{n=-m}^{m} c_n y_{i+n}, \tag{5}$$

where c_n are convolution coefficients, $2m+1$ is the window size, and y_i is the phase data extracted from the range-FFT output. We optimize this filter using a 3rd-degree polynomial and a window size of 50 samples, to smooth rapid fluctuations while maintaining the distinct peaks of eye blinks and the periodic patterns of leg shaking. Fig. 4 shows the extracted phase data before and after applying the proposed two-stage filter, highlighting the effectiveness of the filtering method in reducing noise in the mmWave signals.

B. Enhancing Activity Detection Through Spatial Decomposition

1) Body Part-Specific Signal Extraction via Beamforming: In typical scenarios where a user is seated at a desk, working in front of a computer, the mmWave radar is fixed in place, capturing all movements within its FOV, which is generally limited to 120 degrees. In this configuration, concentration-related activities may occur simultaneously and such concurrent nature makes it challenging to distinguish and isolate the signals associated with each body part, particularly when lower body movements are obscured (e.g., legs are under a table). To address this challenge, we develop a body part-specific signal extraction based on Delay-and-Sum (DAS) beamforming technique [12]. In DAS beamforming, the signals received by an array of antennas are combined by applying appropriate time delays to each signal, such that signals from a desired direction are aligned and summed constructively. By enhancing radar signals at specific angles, beamforming allows us to selectively focus on spatial regions corresponding to different body parts, improving our ability to isolate and analyze concentration-related activities. In particular, the DAS beamforming output $y(t)$ for a given direction θ is expressed as:

$$y(t,\theta) = \sum_{n=1}^{N} w_n x_n(t - \tau_n(\theta)), \tag{6}$$

where $x_n(t)$ is the signal received by the n-th antenna, w_n is the weighting factor, and $\tau_n(\theta)$ is the time delay applied to

steer the beam in the direction θ. For upper body detection, the system focuses on angles within the radar's FOV that correspond to the head and torso, typically in the range $\theta = 30^o - 90^o$. In this range, the system can analyze reflected signals to detect concentration-related activities such as eye blinking, nodding, and yawning. To detect lower body activities, particularly leg movements like shaking or tapping, the system indirectly monitors subtle movements in the belly area, which typically corresponds $\theta = 0^o - 30^o$. Leg movements generate small but detectable shifts in the body's posture and motion, which propagate upwards and cause minor vibrations or movements in the torso, especially the belly area. This approach allows the system to infer leg movement even when the legs are obscured (e.g., under a table) and are not within the radar's direct line of sight.

2) Activities Localization based on Range-Angle Heatmap: Building on the beamforming technique, we enhance the system's ability to detect and localize concentration-related activities by leveraging range-angle heatmaps within each beamformed region. This approach improves the detection of subtle movements, such as eye blinking, yawning, and nodding. To accurately identify and track these activities, we develop an activity localization algorithm that effectively localizes these movements. For a given activity a in region k, we calculate the signal amplitude $A_{a,k}(t, \theta, R)$:

$$A_{a,k}(t, \theta, R) = |y_k(t, \theta, R)|, \tag{7}$$

where $|y_k(t, \theta, R)|$ represents the magnitude of the filtered signal at time t, angle θ, and range R for region k. This amplitude information is then used to construct a range-angle heatmap that visualizes the spatial distribution of signal intensities across different regions of the body. On the obtained heatmap, we apply clustering and temporal tracking of high-amplitude regions. By grouping areas of consistent signal intensity over time, we can accurately localize and differentiate simultaneous activities occurring in both the upper and lower body. This range-angle heatmap approach allows the system to focus on localized areas of interest, ensuring precise detection even when multiple activities occur simultaneously. We identify the key parameters, such as angles and distances, that correspond to the highest and most consistent signal amplitudes for each movement, allowing us to effectively localize each activity. Fig. 5 shows the range-angle heatmap generated following the application of the Delay-and-Sum (DAS) beamforming technique, with a participant seated 50 cm in front of the mmWave radar. This visualization enables precise localization and tracking of multiple activities across the body, enhancing the system's ability to detect and differentiate between subtle movements in both the upper and lower body regions.

C. Distinguishing Concurrent Activities

1) Mitigating Interference in Concurrent Activities Using STFT: Building on the spatial separation of concurrent movements via beamforming and range-angle heatmaps, this section addresses the challenge of mitigating interference between simultaneous activities. In seated scenarios, leg movements often propagate through the body, generating vibrations that influence the radar's phase readings. These larger motions complicate the detection of subtle movements, as the radar captures a composite signal that reflects both leg shaking and smaller activities such as eye blinking. As a result, distinguishing between these overlapping signals is non-trivial.

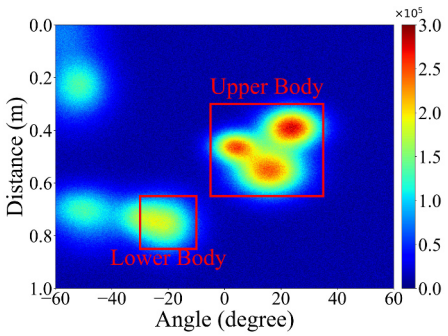


Fig. 5: Range-angle heatmap after applying the proposed DAS Beamforming technique, captured while a volunteer sits 50 cm in front of the mmWave radar.

To address challenge, we exploit the distinct frequency ranges associated with each concentration-related activity. Each activity has a characteristic frequency range: eye blinking (0.5 - 2 Hz) [13], leg shaking (4 - 8 Hz) [18], nodding (0.5 - 2 Hz) [12], and yawning (0.1 - 0.5 Hz) [33]. However, since some activities like eye blinking and nodding share the same frequency range, we combine the frequency analysis with the spatial decomposition results from the beamforming step. By leveraging both the spatial and frequency-domain characteristics, we can accurately differentiate activities based on their location and their corresponding frequency signatures. We apply a Short-Time Fourier Transform (STFT) to the filtered radar signal, enabling time-frequency analysis that captures the spectral content of the signal at different time intervals. This method helps isolate the frequency components corresponding to each activity and mitigate interference from concurrent movements. After that, we use a peak detection algorithm to identify the dominant frequencies. The process is as follows: **(1) Magnitude Spectrum Calculation:** We compute the magnitude spectrum: $|X[m, k]|$ from the STFT, yielding a time-frequency representation of the mmWave signal $x[n]$. This representation reveals the signal's spectral content at specific time intervals m. **(2) Local Maximum Identification:** We detect local maxima in $|X[m, k]|$ that exceed a predetermined threshold δ, allowing us to identify the dominant frequencies:

$$\begin{aligned} P[m] = \{k : |X[m, k]| > |X[m, k-1]| \text{ and} \\ |X[m, k]| > |X[m, k+1]| \text{ and } |X[m, k]| > \delta\}. \end{aligned} \tag{8}$$

(3) **Peak Selection and Sorting:** The identified peaks in $P[m]$ are sorted by magnitude, and the top N_{peaks} are selected. (4) **Filtering Peaks by Frequency Range:** The peaks are then filtered based on the expected frequency ranges for each activity. For instance, frequencies between 0.5 - 2 Hz are retained for detecting eye blinking and nodding, while higher

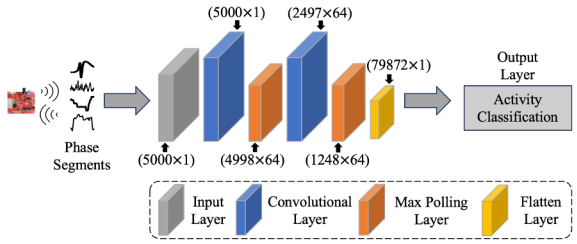


Fig. 6: CNN model achitecture for the proposed system.

frequencies (4 - 8 Hz) are used for leg shaking. This filtering step ensures that only the relevant frequencies for each target activity are considered.

D. Concentration-Related Activities Recognition

1) Activity-Related Signal Extraction and Segmentation: After mitigating interference between concurrent activities using frequency-domain analysis and spatial decomposition, we proceed to extract and segment phase information for each activity identified within the corresponding frequency peaks. To ensure consistency across different measurements, we first normalize the unwrapped phase facilitating the segmentation process. This process divides the continuous phase signal into discrete segments, with each segment potentially corresponding to an occurrence of a distinct activity or movement. The segmentation process is described as follows:

$$S_i(t) = \begin{cases} 1, & \text{if } \frac{1}{W}\sum_{j=t-W/2}^{t+W/2} |\phi_{\text{norm}}(j) - \mu_i| > \tau \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

where $\phi_{\text{norm}}(t)$ is the normalized phase signal, $S_i(t)$ is the segmentation result for activity i at time t, W is the sliding window size, μ_i is the mean phase value for activity i, and τ is the threshold for activity i. The mean phase is derived empirically for each activity, based on its typical phase behavior observed during training. The system detects an activity by identifying when the normalized phase deviates significantly from this mean phase value. The thresholds are set for each activity according to its characteristic phase patterns, allowing the segmentation to accurately capture the start and end of each activity.

2) Feature Extractor: After segmenting the phase signal, the system uses a CNN model to classify concentration-related and non-concentration-related activities. The CNN's feature extractor captures both local features, such as a single blink or nod, and global features, like sustained leg shaking, enabling accurate classification of various activities. As shown in Fig. 6, the CNN processes a 5000-point phase segment through two 1D convolutional layers (64 filters) and max pooling layers, progressively extracting more complex features. The final output is flattened into a 1D vector for classification. This hierarchical structure allows the model to effectively differentiate between similar activities and adapt to individual movement variations.

3) Multi-Label Classifier: To distinguish between concentration-related and non-concentration-related activities, we implement a Multi-Label Classifier. This classifier interprets the features extracted by the Feature Extractor and translates them into activity predictions. The classifier architecture consists of a dense layer with 64 units, using ReLU activation to capture non-linear relationships. This is followed by an output layer with five units—four representing concentration-related activities and one for non-concentration-related activities—using sigmoid activation. The use of sigmoid activation allows the system to detect multiple activities simultaneously. By leveraging the CNN's feature extraction capabilities, the classifier can differentiate between subtle and similar movements, ensuring precise recognition of both concentration-related and non-concentration-related activities.

4) Domain Adaptation via Transfer Learningn: A key challenge is that new environments can impact system performance, as changes in surroundings affect the reflected radar signal. To address this, we integrate domain adaptation techniques into the feature extraction process, implementing an Adversarial Autoencoder (AAE) architecture with Maximum Mean Discrepancy (MMD) regularization [15]. The feature extractor acts as the encoder, producing latent representations optimized for both activity classification and environmental invariance. A decoder is introduced to reconstruct the original input from these latent features, while a discriminator aligns the feature distribution with a Laplace prior [32]. The system is optimized using a multi-component loss function:

$$L_{total} = \lambda_r L_r + \lambda_m L_m + \lambda_a L_a, \quad (10)$$

where L_r is the reconstruction loss defined by:

$$L_r = \frac{1}{N}\sum_{i=1}^{N} MSE(p_i, \hat{p}_i), \quad (11)$$

where Mean Squared Error (MSE) measures the difference between the original input p_i and its reconstruction $\hat{p}_i$, ensuring essential information is retained in the latent features. To calculate this, we introduce a decoder alongside our encoder (feature extractor). The encoder compresses the input into a latent representation, while the decoder attempts to reconstruct the original input from this representation. This process ensures that the extracted features retain essential information about the input. The Maximum Mean Discrepancy (MMD)-based Environment Alignment Loss (L_m) encourages the encoder to produce similar feature distributions across different environments:

$$L_m = \max\left(\left\|\frac{1}{N_u}\sum_{i=1}^{N_u} E(p_{u,i}) - \frac{1}{N_v}\sum_{i=1}^{N_v} E(p_{v,i})\right\|, 0\right), \quad (12)$$

where $E(\cdot)$ is our encoder function, and $p_{u,i}$ and $p_{v,i}$ represent samples from two different environments. The adversarial loss L_a ensures the latent features follow a Laplace distribution, helping capture variability in human movement:

$$L_a = \frac{1}{N}\sum_{i=1}^{N} MSE(h_i, l_i), \quad (13)$$

(a) (b)

Fig. 7: (a) Experiment setup (device displacement). (b) Environment illustration (Office 1, Office 2, Office 3).

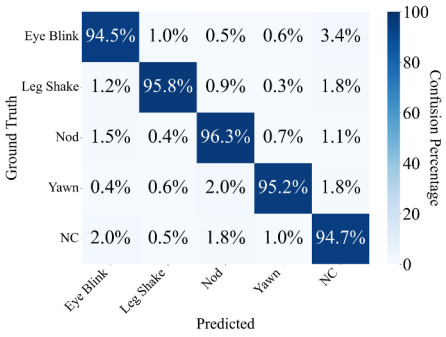


Fig. 8: Confusion matrix illustrating the classification performance for eye blinking, leg shaking, nodding, yawning, and non-concentration-related (NC) activities.

where h_i represents the latent features, and l_i are samples drawn from a Laplace distribution. By jointly optimizing these loss components, we ensure that the extracted features are discriminative for activity classification, invariant to environmental changes, and retain essential input information. This domain adaptation strategy enhances the system's generalization across diverse environments, ensuring reliable concentration-related activity detection.

VI. Performance and Evaluation

A. Evaluation Setup and Methodology

1) Device Configuration: We implement the proposed system using a single commercial COTS mmWave device: Texas Instruments AWR1642 mmWave radar with a DCA1000EVM data capture and streaming card. Our mmWave radar system operates at a starting frequency of $f_0 = 77$ GHz, utilizing 100 ADC (Analog-to-Digital Converter) samples corresponding to 100 range bins. The radar provides a range resolution of 3.85 cm and a FOV of 120° in elevation and 30° in azimuth, with an angular resolution of 14.32°. This configuration allows for high-precision detection of subtle movements associated with concentration-related activities.

2) Data Collection: Our study involved 10 volunteers (i.e., 8 males and 2 females), aged 24 to 31 years, who participated in experiments conducted within three different office environments. Each office varied in size and layout, allowing us to demonstrate the system's performance in different environmental conditions. Each participant was seated in a chair facing

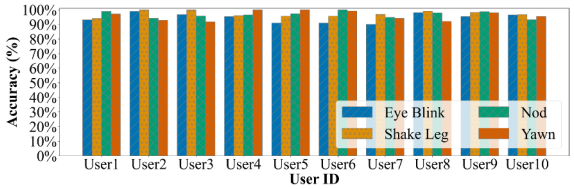


Fig. 9: Accuracy comparison for classifying eye blinking, leg shaking, nodding, and yawning across 10 different users.

a mmWave device positioned on a desk with its antennas directed toward their faces. They were asked to perform a series of concentration-related activities—such as natural eye blinking, shaking their leg at a comfortable pace, nodding, and yawning as if sleepy—for 60 seconds each at specific distances of 50 cm, 100 cm, and 150 cm from the device respectively, with short breaks between activities. Moreover, participants engaged in non-concentration-related behaviors like singing a song or having casual conversations to establish baseline data. As illustrated in Fig. 7(a) for device placement and Fig. 7(b) for the different office environments, was repeated across various settings and distances to assess the robustness of our system.

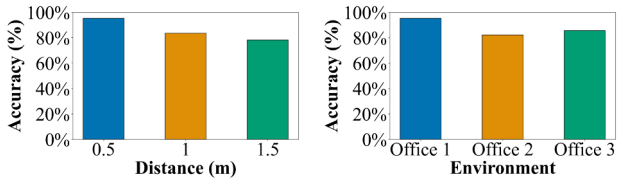
3) Evaluation Metrics: To assess our system's performance, we employ the following metrics: activities classification accuracy (i.e., the ratio of correctly classified instances to the total number of instances), confusion matrix (i.e, visual representation of predicted versus ground truth classes), precision (i.e, ratio of true positives to total predicted positives, indicating prediction accuracy), recall (i.e., ratio of true positives to all actual positives, measuring the model's ability to find all positive instances).

B. Performance of Activity Classification

We first examine the overall performance of our system for concentration-related activity detection. As demonstrated in Fig. 8, the confusion matrix shows the classification results for eye blinking, leg shaking, nodding, yawning, and non-concentration-related activities (NC) in percentages. The overall performance of the system achieves an accuracy of 95.3%. These results demonstrate the effectiveness of our system in recognizing and monitoring concentration-related activity with high accuracy. To assess the system's consistency across different users, we analyze individual performance data, as illustrated in Fig. 9. This result shows the concentration-related activities detection results by individual users, which achieves an average accuracy of 96.4% across all participants. This individual accuracy not only confirms the system's overall performance but also indicates its reliability and adaptability to different users.

C. Impact of Different Device-to-Participant Distances

To assess the impact of distance on system performance, we evaluate the system at three different device-to-participant distances (i.e., 50 cm, 100 cm, and 150 cm). As shown in Fig. 10(a), the system demonstrates promising performance across all tested distances, with variations in accuracy. At 50cm, the system achieves its highest accuracy of 95.3% due



(a) (b)

Distance	Precision	Recall	Environment	Precision	Recall
0.5m	0.9534	0.9530	Office 1	0.9534	0.9530
1.0m	0.8450	0.8250	Office 2	0.8350	0.8190
1.5m	0.7950	0.7710	Office 3	0.8620	0.8420
Average Precision: 0.8645			Average Precision: 0.8835		
Average Recall: 0.8497			Average Recall: 0.8713		

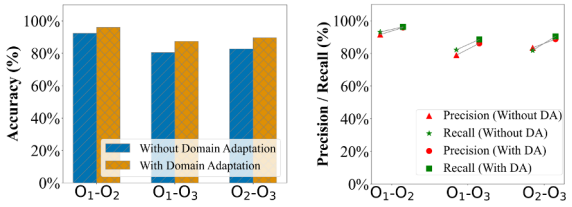
(c)

Fig. 10: (a) Impact of varying person-device distance (0.5m, 1m, and 1.5m). (b) Impact of different environments (Office 1, Office 2 and Office 3). (c) Precision and recall for activity classification across varying distances and environments.

to the strongest signal strength and highest resolution at this close range. As we extend the distance to 100cm, performance remains around 83.5%. Even at 150cm, the system continues to achieve a 78.3% accuracy rate, demonstrating its potential for longer-range applications. Besides, as shown in Fig. 10(c), the average precision for activity classification for different distances is 0.8645 and the average recall is 0.8497. These metrics confirm the system's performance in both accurately identifying activities and minimizing misclassifications.The observed decline in accuracy with increasing distance aligns with expectations due to reduced signal strength and resolution at greater distances. Note that the system maintains acceptable performance up to 150 cm, which covers typical usage scenarios in many applications. Future work could focus on improving long-range detection capabilities to extend the system's effective range.

D. Impact of Different Environments

To evaluate the system's robustness across diverse environmental conditions, we conducted experiments in three different office environments as shown in Fig. 7(b). We employed a cross-environment evaluation approach, where the system was trained using data from only one environment (e.g., office 1) and then tested across all three environments (offices 1, 2, and 3). This process was repeated for each environment, training in one office and evaluating performance in all three. This approach allows us to test the system's ability to generalize to unseen environments, simulating real-world deployment scenarios where retraining for each new location would be impractical or inconvenient. Fig. 10(b) illustrates the system's performance, demonstrating accuracies of 95.3%, 82.7%, and 85.2% for office 1, office 2 and office 3, respectively. These results yield a mean accuracy of 87.7% across all environments. The performance in Office 1 (95.3%) represents the system's capability in its training environment, while the accuracies in Office 2 (82.7%) and Office 3 (85.2%) reflect its generalization to unseen environments. Besides, as shown in Fig. 10(c),



(a) (b)

Fig. 11: (a) Activity classification accuracy without and with domain adaptation (DA) across different environment. (b) Activity classification precision and recall without and with domain adaptation (DA) across different environment. (e.g., $O_1 - O_2$ represents training data from Office 1 and testing data from Office 2).

the average precision for activity classification for different environments is 0.8835 and the average recall is 0.8713. These metrics confirm the system's performance in both accurately identifying activities and minimizing misclassifications. These results indicate that environmental factors do impact the system's performance. The performance variation under different environments can be further enhanced with domain adaptation techniques, which will be demonstrated in the next subsection.

E. Performance of Domain-independent Training across Different Environments

To further enhance our system's robustness and address the performance variations observed across different environments, we implement domain adaptation techniques. In this approach, we designate the data collected from one environment as the source domain, while using a few amount of data from another environment as the target domain. We evaluated three distinct source-target pairs (i.e., $O_1 - O_2, O_1 - O_3$, and $O_2 - O_3$ with O_1, O_2, O_3 represent office 1, office 2 and office 3, respectively) to assess the effectiveness of our domain adaptation strategy. Fig. 11(a) shows the results of comparing the system's performance with and without domain adaptation. The proposed system achieves a 91% activity classification accuracy in cross-environment scenarios when employing domain adaptation. This presents an average increase of 3.3 percentage points compared to the baseline performance without domain adaptation. Moreover, as it shown in Fig. 11(b), the precision has an average increase of 6.63% after applying the proposed domain adaptation. The recall has an average increase of 7% after the proposed domain adaptation. These results demonstrate the effectiveness of our approach in reducing the impact of environmental variations and improving the system's generalization capability across different settings.

VII. Conclusion

In this paper, we propose a novel system for contactless human concentration monitoring using mmWave signals. The system can be deployed using a single COTS mmWave device while achieving high accuracy in detecting concentration-related activities. We design a multi-stage pipeline that addresses key challenges in concentration monitoring. Our ap-

proach employs beamforming techniques to enhance signals from specific body regions, enabling the detection of both upper and lower body movements. We utilize frequency domain analysis to differentiate multiple concurrent activities and indirect monitoring of belly movements to detect leg shaking outside the device's direct field of view. Moreover, a CNN model is implemented to classify concentration-related activities from the extracted features. With the integration of domain adaptation techniques, our system eliminates environment-specific characteristics from the extracted features, enabling robust activity recognition across different office settings. Experimental results demonstrate that our system can accurately detect and classify concentration-related activities, achieving an overall accuracy of 95.3%. We also demonstrate the system's robustness across varying distances and different environmental settings.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation Grants CNS2120396, CNS2329278, CCF2211163, IIS2311596, CNS2304766, CNS2329280, CNS2120350, IIS2311598, CNS2120276, CNS2329279, IIS2311597, CNS2145389.

REFERENCES

[1] Attention Deficit Disorder Association, "Adhd and stimming," *Attention Deficit Disorder Association*, 2023. [Online]. Available: https://add.org/stimming-adhd/

[2] BBC News, "Adult adhd: Huge rise in prescriptions in scotland," *BBC News*, July 2023. [Online]. Available: https://www.bbc.com/news/uk-scotland-66135145

[3] E. Cardillo, G. Sapienza, C. Li, and A. Caddemi, "Head motion and eyes blinking detection: A mm-wave radar for assisting people with neurodegenerative disorders," in *50th European Microwave Conference (EuMC)*. IEEE, 2021.

[4] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term hrv features as surrogates of short term hrv: A case study on mental stress detection in real life," *BMC medical informatics and decision making*, 2019.

[5] I. H. Chen, Y.-T. C. Yang, and S. W. Hsu, "Development and evaluation of a concentration questionnaire for students in classroom," in *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 2013.

[6] S. D'Mello, A. Olney, C. Williams, and P. Hays, "Gaze tutor: A gaze-reactive intelligent tutoring system," *International Journal of human-computer studies*, 2012.

[7] A. W. Gaillard, "Concentration, stress and performance," in *Performance under stress*. CRC Press, 2018.

[8] X. Guo, J. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "Device-free personalized fitness assistant using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018.

[9] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," 2015.

[10] H. J. Han, S. Labbaf, J. L. Borelli, N. Dutt, and A. M. Rahmani, "Objective stress monitoring based on wearable sensors in everyday settings," *Journal of Medical Engineering & Technology*, 2020.

[11] C. J. Hansen, "Wigig: Multi-gigabit wireless communications in the 60 ghz band," *IEEE Wireless Communications*, 2011.

[12] Z. Juncen, J. Cao, Y. Yang, W. Ren, and H. Han, "mmdrive: Fine-grained fatigue driving detection using mmwave radar," *ACM Transactions on Internet of Things*, 2023.

[13] K.-A. Kwon, R. J. Shipley, M. Edirisinghe, D. G. Ezra, G. Rose, S. M. Best, and R. E. Cameron, "High-speed camera characterization of voluntary eye blinking kinematics," *Journal of the Royal Society Interface*, 2013.

[14] G. Lee, A. Ojha, and M. Lee, "Concentration monitoring for intelligent tutoring system based on pupil and eye-blink," in *Proceedings of the 3rd International Conference on Human-Agent Interaction*, 2015.

[15] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[16] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2020.

[17] A. Magliacano, L. Catalano, L. Sagliano, A. Estraneo, and L. Trojano, "Spontaneous eye blinking during an auditory, an interoceptive and a visual task: The role of the sensory modality and the attentional focus," *Cortex*, 2023.

[18] D. Martino, C. Delorme, E. Pelosin, A. Hartmann, Y. Worbe, and L. Avanzino, "Abnormal lateralization of fine motor actions in tourette syndrome persists into adulthood," *PLoS One*, 2017.

[19] B. Meriem, H. Benlahmar, M. A. Naji, E. Sanaa, and K. Wijdane, "Determine the level of concentration of students in real time from their facial expressions," *International Journal of Advanced Computer Science and Applications*, 2022.

[20] P. M. Podsakoff, S. B. MacKenzie, and N. P. Podsakoff, "Sources of method bias in social science research and recommendations on how to control it," *Annual review of psychology*, 2012.

[21] M. Raca, R. Tormey, and P. Dillenbourg, "Sleepers' lag-study on motion and attention," in *Proceedings of the fourth international conference on learning analytics and knowledge*, 2014.

[22] H. A. Ruff and M. C. Capozzoli, "Development of attention and distractibility in the first 4 years of life." *Developmental psychology*, 2003.

[23] N. J. Salkind, *Encyclopedia of research design*. Sage, 2010.

[24] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, 2009.

[25] S. Tanaka, A. Tsuji, and K. Fujinami, "Eye-tracking for estimation of concentrating on reading texts," *International Journal of Activity and Behavior Computing*, 2024.

[26] M. Thompson, "Why do people shake their legs?" BetterHelp, May 2023. [Online]. Available: https://www.betterhelp.com/advice/general/why-do-people-shake-their-legs/

[27] R. Velnath, V. Prabhu, and S. Krishnakumar, "Analysis of eeg signal for the estimation of concentration level of humans," in *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2021.

[28] K. Wang, C. Shi, J. Cheng, Y. Wang, M. Xie, and Y. Chen, "Solving the wifi sensing dilemma in reality leveraging conformal prediction," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022.

[29] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015.

[30] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Transactions on Affective Computing*, 2014.

[31] Y. Xie, T. Zhang, X. Guo, Y. Wang, J. Cheng, Y. Chen, Y. Wei, and Y. Ge, "Palm-based user authentication through mmwave," in *2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2024.

[32] Y. X. Xie, X. Guo, Y. Wang, J. Q. Cheng, T. Zheng, Y. Chen, Y. Wei, and Y. Ge, "mmpalm: Unlocking ubiquitous user authentication through palm recognition with mmwave signals." IEEE Conference on Communications and Network Security (CNS),, 2024.

[33] H. Yang, L. Liu, W. Min, X. Yang, and X. Xiong, "Driver yawning detection based on subtle facial action recognition," *IEEE Transactions on Multimedia*, 2020.