

Home-based Dry Eye Assessment via Blink Kinematics Using mmWave and Clinical Knowledge Distillation

Meng Xue^{†*}, Wentao Xie^{†*}, Zuohuizi Yi[‡], Zhilong Zhang[#], Shumao Wu[#],
Yinan Zhu[†], Qian Zhang^{†◊}, Changzheng Chen^{‡◊}

[†]The Hong Kong University of Science and Technology

[‡]Eye Center, Wuhan University Renmin Hospital, [#]Wuhan University

ABSTRACT

Tear Film Break-Up Time (TBUT) is a critical clinical parameter in the management of dry eye disease (DED). However, traditional TBUT assessments rely on costly and time-consuming clinical procedures, while existing home-based solutions fail to provide precise TBUT values. In this work, we present *Blinic*, a contactless system leveraging commercial millimeter-wave (mmWave) radar to predict precise TBUT values and assess DED severity grades at home. *Blinic* incorporates detailed blink kinematics that are closely linked to TBUT. To address the challenge of predicting TBUT directly from radar data, we propose a teacher-student learning framework. The teacher model, trained on electronic health records (EHRs) including image-based diagnostic tests, transfers medical insights to the student model, which uses radar-captured blink dynamics. This knowledge transfer is further enhanced by a fine-tuned large language model, DryEye-LLM, which is based on clinical diagnostic reports and employs unsupervised domain adaptation to align EHRs with radar data. To ensure accurate blink motion capture, *Blinic* employs an antenna-coded MIMO mmWave radar design. Additionally, a query-based multitask learning module simultaneously predicts TBUT and DED severity grades, addressing potential conflicts in feature representation. Evaluated on 192 participants in collaboration with an eye clinic, *Blinic* demonstrates achieving a mean absolute error of 2.73 seconds for TBUT with an average accuracy of 90.54% for DED grading in real-world settings, providing a practical solution for home-based DED management.

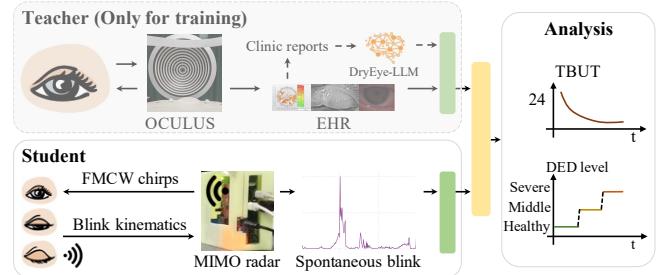


Figure 1: Overview of *Blinic*. *Blinic* uses a mmWave FMCW radar to passively and contactlessly assess dry eye progression via TBUT quantification.

CCS CONCEPTS

- Human-centered computing → Ubiquitous and mobile computing.

KEYWORDS

Dry eye disease assessment, blink kinematics using mmWave, electronic health records, knowledge distillation

ACM Reference Format:

M. Xue, W. Xie, Z. Yi, Z. Zhang, S. Wu, Y. Zhu, Q. Zhang, C. Chen. 2025. Home-based Dry Eye Assessment via Blink Kinematics Using mmWave and Clinical Knowledge Distillation. In *The 31st Annual International Conference on Mobile Computing and Networking (ACM Mobicom '25)*, November 4–8, 2025, Hong Kong, China. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3680207.3765268>

1 INTRODUCTION

Dry Eye Disease (DED), one of the most prevalent ocular conditions and a leading cause of ophthalmological consultations, affects an estimated 8.7% to 64% of the global population [32]. Its prevalence has significantly increased due to factors such as aging and prolonged digital screen exposure [32]. As a chronic condition, DED cannot be cured, and patients must focus on managing symptoms to minimize its impact on daily life. Tear Film Break-Up Time (TBUT) [7] is an objective clinical parameter to evaluate the severity of DED symptoms. It serves as a critical tool for adjusting treatment plans to meet individual patient needs [42] in managing DED [10, 24, 30, 43]. Given the modern lifestyle, which

* Co-first authors. [◊] Co-corresponding authors.



This work is licensed under a Creative Commons Attribution International 4.0 License.

often involves prolonged digital screen use, DED symptoms can fluctuate rapidly. This symptom variability calls for frequent TBUT assessments to ensure timely adjustments to DED treatment [31].

Despite its importance in DED management, the traditional TBUT test relies on expensive clinical devices [6], which is operated by ophthalmologists. These tests are only available in clinical settings, are time-intensive, and incur high costs. Additionally, they require significant patient cooperation during the procedure. These limitations make frequent testing and long-term TBUT monitoring impractical [37], which fails to align with the needs of DED patients.

In response, several research works try to design home-based dry eye assessment systems to support easy and prompt DED management. SDE [45] proposes a millimeter-wave (mmWave) radar-based system to classify subjects into different DED categories by analyzing the blink pattern. While this work successfully demonstrates the feasibility of assessing DED at home, it only achieves two-class classification to differentiate DED patients from healthy individuals, without the capability to measure TBUT. To measure TBUT in a home setup, DEDector [20] employs an external optical attachment on a smartphone and uses optical image-processing methods to screen for abnormal TBUT. However, DEDector also achieves only a two-class classification, distinguishing between normal and abnormal TBUT subjects, and it cannot provide precise TBUT values, which are crucial for DED management and medication adjustments.

In this work, we explore the feasibility of designing a home-based solution to assess the exact TBUT values for better DED management. Inspired by the previous work, SDE [45], we opt to leverage mmWave radars to predict TBUT from the captured eye blink dynamics, due to mmWave's contactless and privacy-preserving nature. Different from SDE where coarse-grained blinking patterns were explored, we observe from biomedical research that the detailed blinking kinematics, such as incomplete blink frequency and eye-closure speed, are closely related to the formation of abnormal TBUT [26, 32, 39] (see Section 2.3). These findings suggest that incorporating blink kinematic features captured from radars could potentially help to predict the precise TBUT values.

Despite this biomedically-grounded approach, realizing *Blinic* faces fundamental challenges.

- First, analyzing TBUT from blink kinematics is extremely challenging since it involves a subtle biomedical process - the break-up of an ultra-thin tear film, and only medical equipment can precisely capture such a process. In fact, even in clinics, assessing DED severity (characterized by TBUT) is not straightforward. It requires multiple medical tests (such as meibomian gland dysfunction (MGD) test [9, 36], tear meniscus height (TMH) test [29] and etc.) to be administered

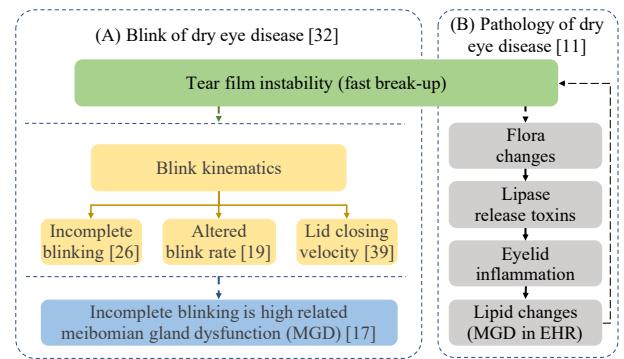


Figure 2: Motivation for inferring TBUT by aligning blink kinematic features with EHR data.

before the doctor can conclude the assessment. Therefore, there is no simple solution to model the relationship between blink kinematics and TBUT.

- Hence, in this work, we do not use an end-to-end solution to predict TBUT from radar-captured blink kinematics. Instead, we gain insights from the clinical practice that uses several intermediate tests to support the assessment. Specifically, we first design a teacher model that learns to predict TBUT based on patients' electronic health records (EHRs; e.g., image-based TMH and MGD data), thereby mimicking clinical DED assessment practices (see Sec. 2). Then, we design a student model with radar-captured blink kinematics as the input, while the objective of this teacher-student scheme is to force the blink features extracted from the student model to align with the EHR features extracted from the teacher model. In this way, we implicitly insert medical insights into the student model to boost its TBUT prediction ability.

- Second, however, aligning blink and EHR features is non-trivial. This is because of the large domain discrepancy between EHR data and mmWave data in which EHR is structured and high-dimensional, while blink data is temporal and low-dimensional.

- To narrow the gap between these two, we propose to include text-based diagnostic reports in the training of the above teacher model. The benefit of this design is that these diagnostic reports associated with the MGD and TMH tests are much easier to interpret than medical images due to the natural language format. Therefore, including these diagnostic reports should ease the alignment between the teacher and student models. In this work, we develop a dry eye large language model (DryEye-LLM) through prompt tuning with the diagnostic reports and serve the embedding encoded by DryEye-LLM to the teacher model. In addition, we leverage an unsupervised domain adaptation approach to align the teacher model trained on mmWave and EHR data with the student model, which is trained on mmWave data only.

- Finally, realizing DED management at home still faces a few practical issues that require special designs. (i) Using

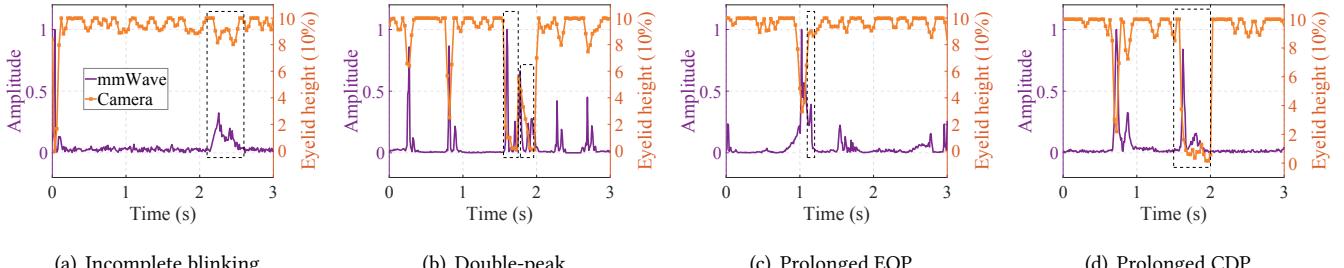


Figure 3: Synchronized DED-associated blinking patterns via mmWave radar and camera. Definitions: Eyelid height (EH) 100% = fully open; early opening phase (EOP, EH: 0%→97%); closed phase (CDP, EH: 100%→0%).

mmWave radars to measure detailed blink kinematics is challenging due to their coarser granularity compared with the tiny blink motions. (ii) It is a common clinical practice to assess the TBUT value together with an assessment of the DED severity grade, where these two can be sometimes paradoxical, say the same TBUT but with different DED severity grades.

▷ To solve the challenge (i), we design an antenna-coded, multiple-input multiple-output (MIMO) mmWave radar system to accurately measure blinking motions while eliminating environmental interference. To solve the challenge (ii), we design a query-based multitask learning module to simultaneously predict these two metrics. This query-based module can address the feature representation paradox in co-learning of TBUT regression and DED severity grading, where partial and total feature representations conflict.

In this work, we present *Blinic*, a fully passive and contactless DED management system that can assess TBUT and the DED severity grade. As illustrated in Figure 1, *Blinic* is developed with a commercial off-the-shelf (COTS) mmWave radar, together with the above algorithm designs. We collaborate with an eye clinic and evaluate *Blinic* on 192 participants, including 44 healthy subjects, 83 mild DED patients, and 65 severe DED patients. Additionally, the training of our deep-learning models involves the EHRs of an extra 1119 DED patients. To our knowledge, this represents the largest dataset scale in home-based DED research to date. The results confirm that *Blinic* can accurately assess DED in real-world environments, achieving a mean absolute error of 2.73 seconds for TBUT with an average accuracy of 90.54% for DED grading.

We summarize our key contributions as follows.

- To the best of our knowledge, this work is the first to infer TBUT from blink kinematics. We achieve TBUT assessment with only a COTS mmWave radar, which supports convenient and home-based DED management.
- We design a teacher-student learning scheme to distill the pathological knowledge from EHR data to blink kinematics captured by mmWave radar and develop a DryEye-LLM

fine-tuned by clinical reports to ease the knowledge transfer.

- We also design algorithms to solve a few practical challenges, including an antenna-coded MIMO mmWave radar scheme to acquire fine-grained spontaneous blink kinematics, and a query-based module for multi-task learning module to co-learn the potential paradox feature representations of TBUT and DED grading.
- In collaboration with a hospital-based eye center, we evaluate *Blinic* on the largest DED dataset to date. The results demonstrate that *Blinic* can accurately assess dry eye disease.

2 BACKGROUND AND MOTIVATION

2.1 TBUT Measurements

A key pathophysiological concept underlying dry eye disease is the disruption of tear film homeostasis, characterized by its accelerated break-up and instability [16]. The tear film, formed during blinking through secretions from meibomian glands (lipid layer), lacrimal glands (aqueous layer), and conjunctival goblet cells (mucin layer), consists of three distinct layers [32, 37]. Therefore, in clinical practice, measuring TBUT serves as a crucial diagnostic examination, with techniques evolving from Fluorescein Break-Up Time (FBUT) to Non-Invasive Breakup Time (NIBUT) [7, 35]. The FBUT uses an intrusive manner to stain the tear film and NIBUT requires expensive medical equipment, such as the OCULUS Keratograph® 5M [6].

When DED occurs, the structural integrity and functional properties of the tear film undergo significant alterations [44]. As illustrated in Figure 2, this process is driven by a self-perpetuating cycle [11]: initial tear film instability alters the ocular surface flora, leading to the release of lipase and toxins. These substances induce eyelid inflammation, which in turn modifies the lipid composition of the tear film. This altered lipid composition further destabilizes the tear film, thereby perpetuating the cycle and underscoring the progressive nature of the disease.

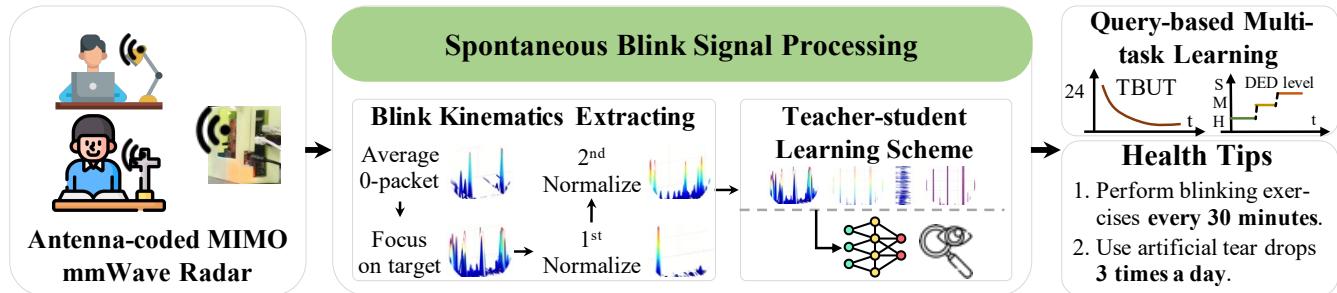


Figure 4: System framework of *Blinic*.

2.2 Blink Kinematics

Recent biomechanical studies [17, 19, 26, 32, 39, 44] have revealed significant alterations in the blink patterns of patients with DED. Spontaneous blinking is a natural, unconscious action primarily used to maintain eye health, while reflexive and voluntary blinking are responses to external stimuli or actions controlled by the individual.

Based on the DED diagnostic consensus [44], blink closure analysis is a recommendation of appropriate tests for diagnosis and assessment of dry eye. Morphological changes in meibomian glands (e.g., atrophy, obstruction) directly indicate the pathological process of DED. Additionally, TMH serves as a critical clinical indicator for assessing tear volume and distribution dynamics in dry eye examinations, where its abnormality directly correlates with compromised tear film functionality and signals underlying tear film disruption. As shown in Figure 2, blink kinematic features [32] are related to the vicious cycle of DED, such as incomplete blinking, altered blink rate, and lid closing velocity. Given that meibomian gland-secreted lipids maintain tear film integrity by reducing evaporation, compromised lipid secretion leads to accelerated tear film breakup. This consequently manifests as increased incomplete blink frequency [26], thus establishing incomplete blinks as a clinically quantifiable biomarker for DED severity. An altered blink rate [19] has been validated to be used to support dry eye disease diagnosis. A reduced eye closure velocity [39] has been confirmed related to clinical examinations, such as FBUT using a high-speed camera. Additionally, incomplete blinking [17] is highly related to MGD. Therefore, blink kinematic features could potentially surpass the performance limitations of traditional TBUT measurement techniques.

2.3 Motivation

Key idea. The frequency of hospital visits is primarily determined by disease progression and treatment response following initial diagnosis [16, 37, 44]. However, TBUT measurements are still a daunting task in hospital visits due to the large number of patients and the extensive medical resources

required. The recent surge in mobile device proliferation has the potential to disrupt this stalemate. For example, prior work [20] has shown the possibility of using portable devices to measure TBUT values. However, a significant obstacle lies in the absence of precise TBUT data essential for continuous and systematic monitoring of disease progression. *To solve this problem, we harness the strong correlation between blink kinematics and DED pathology in EHR data, which enables accurate TBUT measurement solely through a COTS radar device.*

DED blink pattern captured by Blinic. To confirm Blinic's use of blink kinematics-DED pathology correlations in EHR data, we analyze mmWave-derived features [39] under the same setup as Section 6.1. Ground truth comes from video-based eyelid height measurements (100% = fully open) [39]. Figure 3 shows mmWave-captured DED patterns clinically validated in studies: incomplete blinks, rapid blinking, prolonged eyelid opening phase (EOP, from initial movement to 97% open), and closed phase (CDP, from closure start to reopening). As shown in Figure 3(a), incomplete blinking refers to a condition where the eyelids do not fully close during a blink, leaving part of the eye surface (usually the lower portion of the cornea) exposed. As shown in Figure 3(b), the double-peak occurs in DED patients who have difficulty forming a tear film, resulting in severe partial blinks. As shown in Figures 3(c) and 3(d), prolonged EOP and CDP occur in cases of DED. This demonstrates that Blinic can effectively capture these DED-related blinking patterns.

3 DESIGN OVERVIEW

As illustrated in Figure 4, Blinic comprises four major components, i.e., COTS radar, blink denoising, blink feature, and analysis results.

- **Antenna-coded MIMO mmWave radar.** To capture spontaneous blinking, we design an antenna-coded MIMO mmWave radar (Section 4.1).

- **Blink kinematics extracting.** To clarify blink kinematic features, we design a series of noise removal algorithms for blink motions signals (Section 4.2).

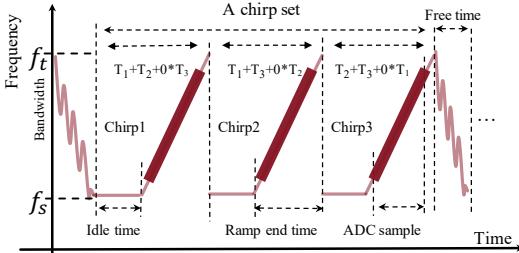


Figure 5: Schematic of *Blinic*'s three transmitters emitting signals with binary phase modulation coding.

- *Teacher-student learning scheme*. To statistically correlate eyelid motion patterns with DED pathological biomarkers, we propose a teacher-student structure that transfers clinical EHR insights to mmWave-derived blink features (Section 5.1, Section 5.2, Section 5.3).

- *Query-based multi-task learning*. To measure TBUT, we design a query-based module using task-related attention to extract specific outputs from blink features (Section 5.4).

4 CAPTURING BLINK KINEMATICS

Since conventional mmWave radar's spatial resolution is insufficient to detect submillimeter-scale motions induced by blinks, we devise an antenna-coded MIMO mmWave radar system for measuring spontaneous blinks coupled with interference elimination.

4.1 Antenna-coded MIMO mmWave Radar

We use two steps to achieve antenna-coded MIMO mmWave radar, *i.e.*, *antenna setup* and *MIMO radar*.

- **Antenna setup.** We use a commercial mmWave radar [1] to capture spontaneous blinks. Our choice of mmWave is driven by: i) Cost-effectiveness and high resolution: the radar's core chip costs \$40, yet achieves 0.5 ns temporal resolution (at 1 GHz bandwidth), critical for detecting incomplete blinks as brief as 5 ms [39]. ii) Privacy-safe contactless sensing: widely adopted in industry [2–5] and academia [12, 13, 28, 34, 40, 47] due to non-intrusive design. iii) mmWave can avoid occlusions caused by hair or glasses, whereas cameras rely heavily on good lighting conditions. Moreover, mmWave is more sensitive to the micro-motion of the eyelid (in the range of centimeters to millimeters) compared to smartphone cameras, which are limited by pixel-based resolution. The radar's mmWave signal operates in the frequency range of 77 GHz to 81 GHz using frequency-modulated continuous wave (FMCW) technology. The field of view (FOV) is 30° vertically and 120° horizontally. As shown in Figure 5, each chirp has a ramp end time of 60 μs, an idle time of 100 μs, 512 sampling points, and a slope of 29.982 MHz/μs.

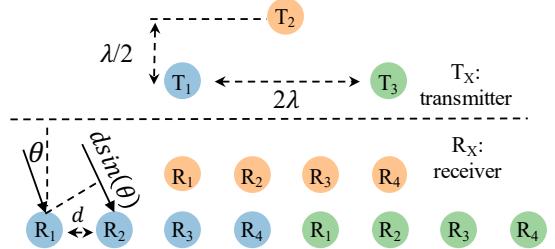


Figure 6: *Blinic*'s multiple-input-multiple-output (MIMO) mmWave radar configuration.

Although mmWave radar achieves high temporal resolution, its spatial resolution remains limited. A widely adopted strategy for enhancing spatial resolution involves expanding the transceiver antenna array through MIMO configurations, which effectively increases the virtual aperture size. However, adding extra antennas significantly increases the complexity and cost of mmWave front-end devices. Therefore, we design an antenna-coded MIMO system to generate virtual receivers by using a chirp-component time-division transmission method. As shown in Figure 5, we use two antennas to send a same chirp-component one time, which can improve transmission gain of 3 dB compared to single-antenna. By leveraging double antennas, the sending signals can be denoted as $S_a = S_{T_1} + S_{T_2}$, $S_b = S_{T_1} + S_{T_3}$, and $S_c = S_{T_2} + S_{T_3}$ in a time-division manner. Moreover, the sending chirp S_{T_x} can be denoted as:

$$S_{T_x}(t) = \exp(-j2\pi(f_s + \frac{B}{2T}t)t), \quad (1)$$

where f_s denotes initial frequency, B denotes sweep bandwidth, and T denotes sweep time.

- **MIMO radar.** After configuring the multiple transmitters, we can utilize the 4-receivers to achieve MIMO mechanism. As shown in Figure 6, we can turn the 4-receivers into 12-receivers by MIMO mechanism. Let R_r^s denote the received signal from the s -th transmitter to the r -th receiver, where $s \in 1, 2, 3$ and $r \in 1, 2, 3, 4$. For instance, the signal R_4^3 corresponds to transmission from transmitter T_1 to receiver R_4 . The complete received signal model can be expressed as:

$$R_r^s(t) = \sum_{i=1}^N \alpha_i \exp(-j2\pi(f_s + \frac{B}{2T}(t - \tau_i)(t - \tau_i))), \quad (2)$$

where N , α_i , τ_i denote the number of multipath, attenuation coefficient, i -th path of N . For transmitted signals, the received signals across the four receivers exhibit phase differences ranging from 0 to 3ω , where ω represents the angular frequency modulation induced by the chirp waveform. According to our MIMO setup, the phase information can be represented as:

$$\varphi(n) = (n - 1) \times d \times \sin(\theta), \quad (3)$$

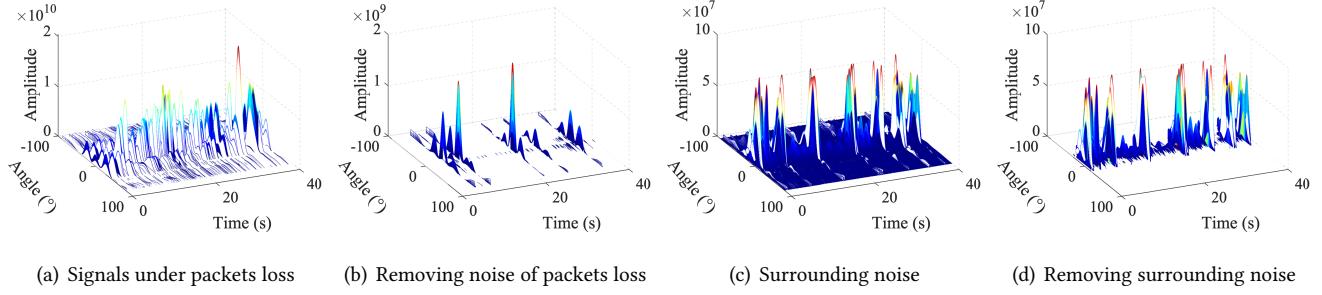


Figure 7: The removal of noise caused by packet loss and surrounding surfaces.

where n denotes the virtual receiver numbers. Based on the design of S_a , S_b , and S_c , we can decode the signals received at the receiver from different timestamps to separate the transmitted signals from T_1 , T_2 , and T_3 as $\frac{S_a+S_b-S_c}{2}$, $\frac{(S_a-S_b+S_c)}{2}$, and $\frac{-S_a+S_b+S_c}{2}$. In this way, without increasing the number of physical transmitting and receiving antennas, we achieve virtual antenna pairs through the MIMO method, thereby enhancing the spatial resolution of mmWave.

4.2 Blink Kinematics Extracting

We utilize two steps to extract blink kinematics, *i.e.*, *capturing blinking* and *eliminating noise in blinking*.

- **Capturing blinking.** Inspired by SDE [45], the frame chirps variance (FCV) has a larger value on speed dimension when there is blinking. Therefore, we can leverage their findings to denote the distance between the user's head and mmWave radar. Specifically, based on the $S_{T_x}(t)$ and $R_x(t)$, we can get the intermediate frequency (IF) of each virtual receiver antenna as:

$$IF(t) = \sum_{i=1}^N \alpha_i \exp \left(j2\pi \left(\frac{B(\tau_i^2 - 2t\tau_i)}{2T} - f_s \tau_i \right) \right). \quad (4)$$

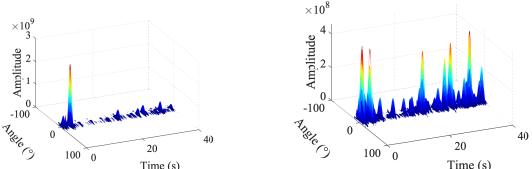
Based on the $IF(t)$, we can get the position of user's eyes based on the FCV. Specifically, we first calculate the range-Doppler information by applying a two-dimensional FFT, *i.e.*, first performing a range FFT on $IF(t)$ with size $R_f = 1024$, followed by a Doppler FFT with size $D_f = 249$. Then, we calculate the variance on the Doppler dimension, which denote speeds. Finally, we sum the variance on the Doppler axis and get the largest value, which denotes the position of user's eye. By applying an inverse transform to the velocity dimension of the angular FFT (range-Doppler, 180) and calculating the variance along this dimension, a blink image with angular information can be obtained. Since the velocity dimension is not easily perceptible, the velocity data is first converted back to the original domain before calculating the variance. As shown in Figure 7(b), we can observe continuous spontaneous blinking motion, where a high variance indicates a blink and a low variance suggests no blink.

- **Eliminating noise in blinking.** After capturing continuous spontaneous blinking data, the remaining challenge is how to remove noise in the blinking signals, *i.e.*, noise from packet loss, surrounding surfaces, and outliers.

- ▷ **Noise caused by packet loss.** Interference in the multipath environment can lead to inaccurate signal estimation, causing errors between transmitted and received signals, and resulting in packet loss. As shown in Figure 7(a), the packet loss can bring much noise in blinking. To address this issue, we first detect whether the received mmWave signals contain packets with data equal to zero. Then, we replace the zero packets' mmWave data with the average of the surrounding packets. As shown in Figure 7(a) and 7(b), this approach helps to minimize the impact of zero packets on the mmWave signals.

- ▷ **Noise caused by surrounding surfaces.** Except for the target user's reflected signals, the virtual receivers can get other object's reflection. As shown in Figure 7(c), at the position of target user's eye, different angles in the radar's FOV also have reflections. Therefore, we first calculate the average value over a window length of 40 seconds after removing zero packets, and then we count the number of values that exceed this average. Finally, based on the observation that the blinking segments exhibit greater fluctuations and contain more values above the average, we calculate the center of spontaneous blinking. As shown in Figure 7(d), to signify the continuous spontaneous blinking, we reserve 70° around the center of blinking.

- ▷ **Noise caused by outliers.** To undermine abnormal large values at 3 s as shown in Figure 8(a), we design a two-level normalization. At the first-level normalization, we first compute the signal peaks in Figure 8(a) and the median of these peaks. Then, we use four hundred times the median value as the threshold to reduce the effect of abnormal values. As shown in Figure 8(a) and 8(b), *Blinic* can effectively undermine some large range of values. At the second-level normalization, we improve the robustness of blinking by normalizing signals to the range of 0 to 1.



(a) Signals under outliers noise (b) After the first level normalization
Figure 8: The removal of noise caused by outliers.

5 FROM BLINK KINEMATICS TO TBUT

After extracting blink kinematics from the mmWave data, *Blinic* employs a deep-learning model to predict TBUT and DED severity grading. However, as discussed in Section 1 and Section 2, achieving this prediction is extremely challenging due to the complexity of the assessment. Therefore, drawing insights from the clinical practice and biomechanism of blink kinematics, we first use various types of medical data to train a teacher model. Then, we use this teacher model to supervise the training of a student model which involves only the mmWave data with an unsupervised domain adaptation (UDA) scheme.

5.1 Teacher-student Learning Scheme

Blinking kinematics reflect DED symptoms associated with pathological mechanisms. However, a substantial information disparity exists between clinical biomarkers (*i.e.*, TBUT, TMH, MGD) and blinking kinematic measurements. Therefore, to bridge this gap, we design a teacher-student knowledge distillation framework that transfers DED diagnostic features from EHR to blink kinematic patterns.

As shown in Figure 9, the teacher model has the full-modality data with labels regarding DED and the student has only the blink kinematics from mmWave signals. The full modal data of teacher's contain X_mmWave, X_EHR, and X_ClinicReport. The X_mmWave denotes blink kinematics from mmWave radar, X_EHR denotes clinical examinations of both TMH, and MGD, and X_ClinicReport denotes clinical reports of clinical examinations, *i.e.*, TBUT, TMH, and MGD.

5.2 Teacher Learning with Full Modal

Since the teacher model incorporates multiple modalities, such as blink kinematics, EHR, and clinical reports, extracting DED features from data with diverse structures and formats presents a significant challenge. Inspired by the success of large language models (LLMs) [15] in generalization, we utilize LLM to promote teacher extracting DED features from multimodal data. Specifically, we split the teacher learning into two parts: *two-phase LLM optimization: clinical report fine-tuning and feature caching, multimodal learning for full modal*.

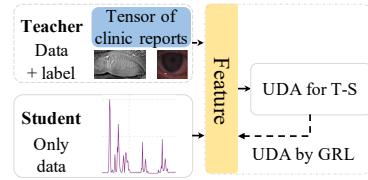


Figure 9: The teacher-student structure.

- **Two-phase LLM optimization: clinical report fine-tuning and feature caching.** Recently, LLMs [15], as high-capacity transformer-based architectures trained on massive and diverse text corpora, have demonstrated exceptional generalization abilities across linguistic and reasoning tasks, particularly in contextual medical knowledge extraction. These models contain a wealth of medical information, such as that from college medicine. However, there are two issues regarding the extraction of medical knowledge from the model. First, general LLMs are not better than a specific disease model according to a recent study from Harvard [27]. Second, performing inference with LLMs incurs significant latency and exhibits high memory consumption.

To address the first issue, we develop DryEye-LLM, a domain-specific LLM for DED diagnosis, through prompt-based fine-tuning on de-identified clinical reports. Specifically, we leverage the latest LLM model Flan-t5 [15] with 3 billion parameters as backbones to conduct prompt-tuning. A patients' diagnostic reports are fixed as "*Prompt: [The first TBUT time is 2.17s, and the average time is 3.50s, and the tear meniscus is 0.25mm, Meibomian glands have openings that are lost by more than 1/3.]*" to minimize the variation of LLM's output. Then, we fine-tune DryEye-LLM using the format fixed prompt by adding both an embedding layer after the pretrained tokenizer of Flan-t5 and a linear layer after the pre-trained encoder of Flan-t5. Thus, the prompt tuning process can be supervised by a loss of cross entropy. Additionally, to optimize the fine-tuning process, we employ quantization, converting the model's parameters from high precision (*e.g.*, 32-bit floating-point) to lower precision (*e.g.*, 8-bit integers). The quantization method significantly enhances calculate speed while simultaneously reducing memory footprint.

To address the second issue, we cached the feature tensor from the fine-tuned DryEye-LLM instead of performing feature inference for each participant. Specifically, we first extracted feature tensors from the fine-tuned DryEye-LLM using activation hooks. A forward hook was then attached to the final_layer_norm layer of the transformer architecture. The captured feature tensors with dimensions [1, 2048] were subsequently cached. This approach enables efficient teacher-student knowledge distillation while circumventing the computational constraints of direct LLM deployment.

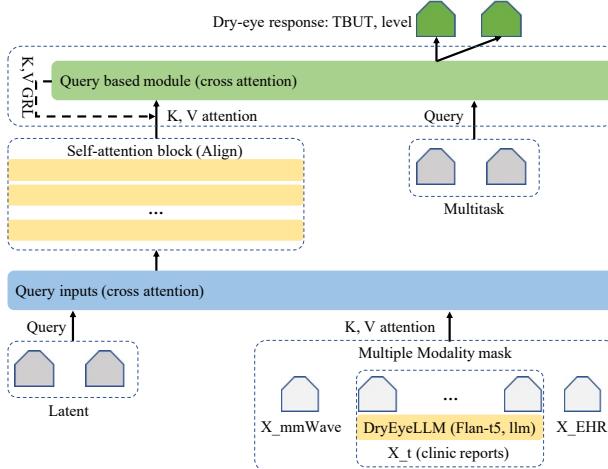


Figure 10: Foundation model-based multimodal learning framework for DED assessment.

- **Multimodal learning for full modal.** Different clinical examinations reflect the pathology of DED from different aspects, such as the Placido rings in the tear break-up time test, the lower eyelid bands in the tear meniscus height measurement, and the gland ducts in the meibomian gland function analysis, resulting high dimension data. Inspired by Perceiver [25], we use cross-attention to direct the model towards DED features in multi-modal data. This mechanism also decouples model depth from input size, enhancing compatibility with the teacher model. After pre-fusing multimodal data, we process features through 26 self-attention layers [18]. Depth-wise weight sharing [25] further cuts trainable parameters.

As shown in Figure 10, the cross-attention mechanism has two inputs, *i.e.*, "Query" and "K, V attention". The "Query" denotes the latent vector, which is the DED features. And, the "K, V attention" denotes the multimodal input data, *i.e.*, X_{mmWave} , feature tensors from DryEye-LLM, and X_{EHR} . X_{mmWave} denotes blink data from mmWave signals as shown in Figure 11 and X_{EHR} denotes the clinical examinations data. Additionally, *Blinic* serializes the input data and then add a Fourier transform feature to each modality's information to maintain the structured characteristics lost during the serialization of the EHR data, such as the curved bands in the lower eyelid area of the tear and height measurement.

5.3 Student Learning with mmWave

As illustrated in Figure 10, we implement a teacher-student structure with unsupervised domain adaptation [46] to train the student model exclusively on X_{mmWave} data. To enforce modality isolation, we apply a binary mask that nullifies feature contributions from DryEye-LLM and X_{EHR} in the student's input. The mask activates (1) for blink kinematics and deactivates (0) for other modalities. We integrate a gradient

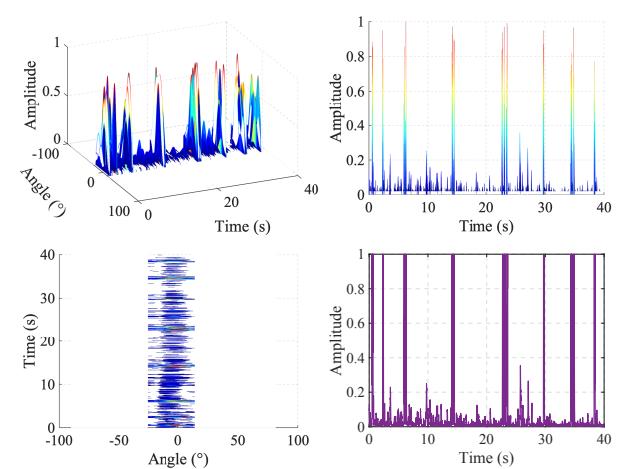


Figure 11: The mmWave modality from spontaneous blink.

reverse layer (GRL) [21] into the key-value (K, V) projections of the final self-attention block, as annotated in Figure 10. The DED features extracted from full modal data (teacher domain) can be transferred into blink kinematics (student domain):

$$\mathcal{L}_{ts} = -[1/(N_t + N_s)] \sum_{i=1}^{N_t+N_s} [D_i \log(p(D_i)) + (1 - D_i) \log(1 - p(D_i))], \quad (5)$$

where N_t and N_s are the numbers of samples in the teacher and student domains, respectively. D_i is the domain label for sample i , where $D_i = 1$ if the sample is from the teacher domain, and $D_i = 0$ if it is from the student domain. $p(D_i)$ is the predicted probability that sample i belongs to the teacher domain.

5.4 Query-based Multi-task Learning

Although dry eye grading is primarily determined by the TBUT metric, it also incorporates other clinical tests such as MGD and TMH. To address this complexity, we propose a query-based multi-task learning module designed to jointly model the shared feature representations between TBUT and dry eye grading. As illustrated in Figure 10, this module leverages integrated cross-attention mechanisms to decode task-specific information from distilled DED features through learnable queries. While sharing a common feature space and query tensor across tasks, we employ distinct loss functions: cross-entropy loss \mathcal{L}_{ce} for ordinal DED grading and mean squared error \mathcal{L}_{mse} for TBUT regression.

The composite objective function is formulated as: $\mathcal{L}_{DED} = \mathcal{L}_{ce}(D_t, Y_t) + \mathcal{L}_{mse}(D_t, Y_t) + \delta \cdot \mathcal{L}_{ts}(D_t, D_s)$, where $\mathcal{L}_{ce}(D_t, Y_t)$ is the classification loss in the teacher domain, $\mathcal{L}_{mse}(D_t, Y_t)$ is the regression loss for TBUT prediction, $\mathcal{L}_{ts}(D_t, D_s)$ is the transferring loss that aligns the teacher D_t and student D_s ,

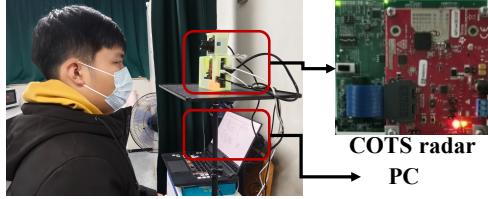


Figure 12: Typical deployments.

and $\delta = 0.3$ controls the trade-off between transferring and task-specific objectives.

6 EVALUATION

We conduct extensive experiments to evaluate *Blinic*'s performance in real-world environments, aiming to address three key research questions:

- **(Q1)** Does *Blinic* accurately measure TBUT?
- **(Q2)** Does *Blinic*'s architectural design enhance its effectiveness in DED assessment?
- **(Q3)** How robust is *Blinic* under varying experimental configurations?

6.1 Experiment Setup

As illustrated in Figure 12, we conducted experiments in a clinically realistic environment¹, *i.e.*, an ophthalmic examination room. Under standardized conditions, the distance between mmWave radar and users is set to around 30cm, with the angles at 0°. Participants were seated facing the device without performing specific tasks, each lasting five minutes. Participants were not given specific posture instructions and sat in self-selected postures. First, we implement a prototype of *Blinic* using an unmodified commercial mmWave radar [1]. The algorithms are developed using Matlab 2023b, Python 3, and PyTorch 2.3. As illustrated in Figure 13, the recorded spontaneous blink is shown on the right alongside the left image captured by a camera mounted above the mmWave radar. Then, the data collection procedure is integrated into patients' examination workflows, where doctors performed three clinical tests: TBUT, TMR, and MGD.

Participants demographics. Our study cohort comprised 192 participants (101 female, 91 male; age range: 20–88 years; height range: 150–180 cm) with complete mmWave sensor data and matched electronic health records (EHR). This cohort included 44 healthy controls, 83 mild DED patients, 65 severe DED patients. In addition, we also collected 1119 participants with only clinic tests.

Diagnostic protocol. Participants were clinically diagnosed by ophthalmologists, with DED patients requiring at least one symptom (*e.g.*, dryness, blurred vision), abnormal tear stability (TBUT \leq 14s or Schirmer I \leq 10mm/5min), and



Figure 13: Examples of recorded blink data.

without comorbid conditions (*i.e.*, impaired blinking, keratitis, trichiasis). Healthy controls showed no ocular/systemic diseases and normal vision (BCVA \geq 1.0). All participants retained withdrawal rights.

Model training. To build an effective and robust model, our training dataset consists of two components: teacher labeled data $D_t = (X_i, Y_i)$ and 30% student unlabeled data $D_s = (X_j)$.

Evaluation metrics. Following the leave-one-participant-out cross-validation protocol from [14], we assess *Blinic*'s performance through two quantitative measures: i) Accuracy: The proportion of correctly classified test samples. ii) Mean Absolute Error (MAE): The average absolute deviation between predicted and clinician-measured TBUT values across participants.

6.2 Overall Performance

To answer the question **(Q1)**, we evaluate the performance of *Blinic* with only blink kinematics from mmWave radar. Since *Blinic* utilizes a teacher-student learning scheme to measure TBUT from blink kinematics, we also evaluate the performance of teacher model using only clinical reports or EHR raw data.

Performance of *Blinic*. Figures 14(a) and 14(b) present the system's performance in TBUT quantification and DED severity classification, respectively. The experimental cohort comprising 192 participants was prospectively stratified into four distinct groups: Ug1 (n=40), Ug2 (n=55), Ug3 (n=42), and Ug4 (n=55).

As delineated in Figure 14(a), group-level MAE for TBUT measurements were 2.41s (Ug1), 2.32s (Ug2), 3.57s (Ug3), and 2.61s (Ug4), with the integrated system (*Blinic*) achieving an overall MAE of 2.73s across all participants. Complementary results in Figure 14(b) demonstrate DED classification accuracies of 89.34% (Ug1), 92.68% (Ug2), 86.53% (Ug3), and 93.04% (Ug4), culminating in a system-wide classification accuracy of 90.40%.

Figure 14(c) further reveals strong concordance between *Blinic*-derived TBUT measurements and clinical gold-standard assessments, particularly among high-performing users within each cohort (mean absolute difference \leq 0.8s). These findings collectively demonstrate that *Blinic* enables precise non-invasive TBUT quantification through blink kinematic analysis.

¹The experiments are approved by the Ethics Committee of both our institutions and the partnering hospital.

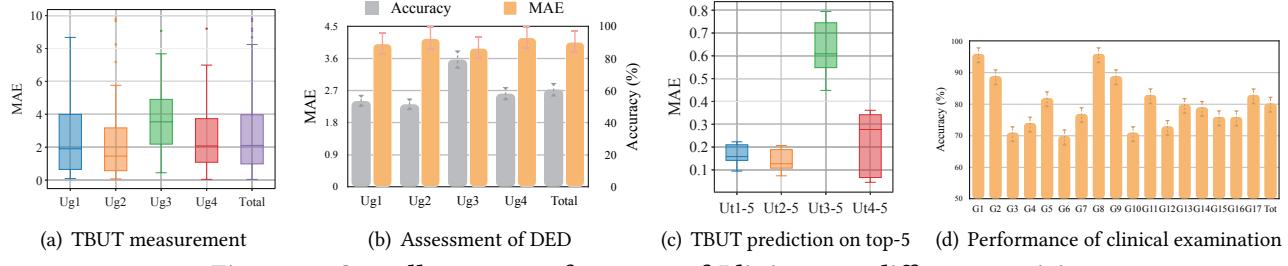
Figure 14: Overall system performance of *Blinic* across different participants.

Table 1: Performance of various DED detection.

Metrics	DEDetector [20]		SDE [45]	<i>Blinic</i>	
	Binary	Binary	Binary	Grading	
Accuracy	80.43%		94.22%	98.01%	87.0%
F1-score	75.67%		>91.43%*	93.17%	84.35%
TBUT	-	-	-	-	2.41s

*: SDE [45] does not report a concrete F1-score.

Table 1 compares various DED detection methods. Since state-of-the-art (SOTA) methods only perform binary classification, to ensure fair comparison, we defined two classes: (1) a single 'DED' class encompassing all patients across severity stages; and (2) a 'Normal' class. As shown in Table 1, *Blinic* achieves better performance than the SOTA methods. *Blinic*'s grading accuracy (i.e., multi-class severity classification) is lower than its binary screening performance, primarily due to misclassification between moderate and severe cases.

Performance of teacher models trained exclusively on clinical reports. Figure 18 presents the diagnostic accuracy of DED severity grading using clinical reports. The experimental protocol consisted of two phases: First, the clinical report corpus was stratified into training and testing subsets with a 7:3 partitioning ratio. Three foundational architectures were subsequently fine-tuned on the training data to develop DryEye-LLM variants: Flan-T5 large language model (Flan-T5-LLM) [15] with 3B/5B parameters and BERT-Large (0.34B). The 3B model achieved the highest accuracy (98.44%), followed by the 0.34B model (96.87%), while the 11B variant unexpectedly underperformed (92.9%). Though larger models typically show better performance, this inverse pattern demonstrates that architectural optimization surpasses mere parameter scaling for medical tasks. The results align with clinical AI specialization research, indicating targeted mid-sized models outperform generic large models in diagnostics [27].

Performance of teacher models trained exclusively on EHR raw data. Figure 14(d) demonstrates the diagnostic accuracy for DED severity grading using multi-modal EHR, comprising MGD imaging, TMR measurements, and TBUT video recordings. We implemented the model architecture

shown in Figure 10, replacing UDA components and attention query mechanisms with a conventional classification head. The experimental design comprised two sequential phases: First, partitioning of raw EHR data from 1,119 participants into 17 balanced cohorts, with a 7:3 training-testing ratio. Second, cross-cohort validation of feature extraction capability. As shown in Figure 14(d), we can see that *Blinic* can achieve accuracies ranging from 70% to 96%. The variation results demonstrate that the effective extraction of DED features from raw EHR data is a non-trivial task. Therefore, we design a teacher-student learning scheme to transfer DED knowledge, enabling the possibility of measuring TBUT from blink kinematics.

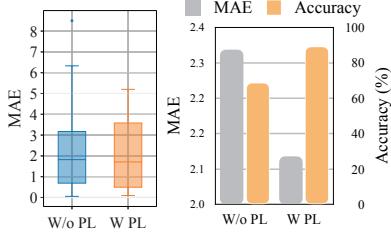
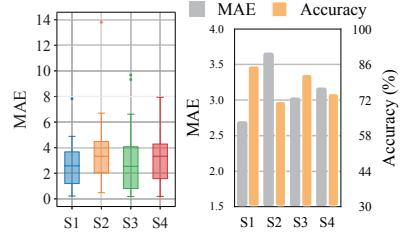
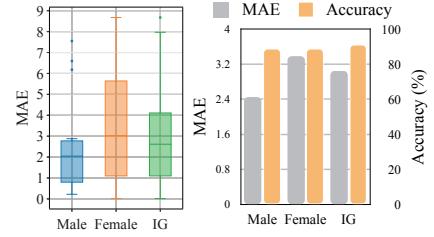
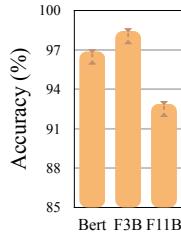
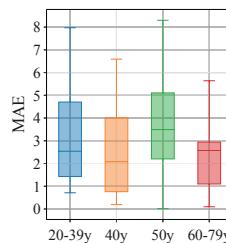
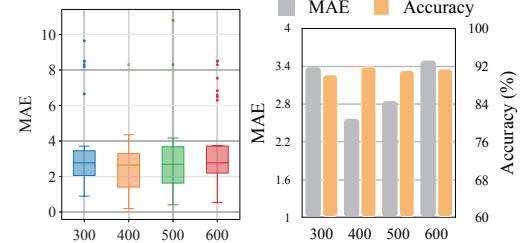
The *Blinic* model contains 0.26 billion trainable parameters with a total parameter count of 1.05 billion. This depth-wise weight sharing architecture enhances inference efficiency through memory optimization, achieving an inference latency of 125 ms on NVIDIA GeForce RTX 4090. For users' end devices, we evaluated inference latency on a 2019 MacBook Pro laptop (2.4 GHz Quad-Core Intel Core i5 CPU, 8 GB RAM), measuring 1.373 s.

6.3 Ablation Study

To investigate (Q2), we conduct empirical evaluations of *Blinic*'s performance across two key factors: teacher model absence and packet loss of mmWave signal.

6.3.1 Evaluation of Knowledge Distillation Framework. In this section, we empirically evaluate the effectiveness of *Blinic* to clarify how the teacher-student learning scheme benefits the measurement of TBUT from blink kinematics. Specifically, we validate knowledge distillation from a teacher model trained on: Scenario 1 (S1): a combination of MGD, TMR, and diagnostic reports, Scenario 2 (S2): blink kinematics from mmWave signals alone (without distillation), Scenario 3 (S3): MGD with diagnostic reports, and Scenario 4 (S4): TMR with diagnostic reports.

As shown in Figure 16, knowledge transfer via teacher-student distillation critically enhances TBUT measurement. S1 achieves optimal performance: lowest TBUT MAE (2.71s) and highest DED classification accuracy (85.71%). Conversely,

**Figure 15: Impact of packet loss.****Figure 16: Impact of distillation.****Figure 17: Impact of gender.****Figure 18: LLM size.****Figure 19: Impact of ages.****Figure 20: Impact of outliers.**

S2 without distillation yields the highest TBUT MAE (3.67s) and lowest accuracy (71.43%). Notably, S3 (TBUT MAE: 3.18s; accuracy: 82.25%) outperforms S4 (TBUT MAE: 3.04s; accuracy: 74.65%), indicating that MGD provides more clinically significant information than TMH. This aligns with physiological mechanisms: meibomian glands offer multifaceted data on dry eye etiology, classification, and severity, whereas tear meniscus height primarily reflects tear volume and offers relatively limited information. Consequently, distilled DED knowledge enables student models using only blink kinematics to achieve effective TBUT measurement.

6.3.2 Evaluation of Packet Loss. We designed an antenna-coded phase MIMO system to enhance spatial resolution, though this implementation introduced packet loss-induced noise. To mitigate this limitation, we implemented a packet loss detection algorithm for blink data outlier removal, with implementation details provided in Section 4.2 (*Noise caused by packet loss*). Through empirical evaluation with 35 participants, we compared *Blinic*'s performance with and without the detection algorithm.

Figure 15 demonstrates that *Blinic* achieves an average accuracy of 89.34% diagnostic with the packet loss detection algorithm, compared to 68.91% without it. Similarly, the TBUT measurement MAE improves from 3.35s to 2.11s when employing the algorithm. These results confirm the algorithm's critical role in performance optimization. This improvement stems from the algorithm's capacity to enhance spontaneous blink data quality, enabling more effective extraction of DED biomarkers. Furthermore, the baseline MAE

(≤2.35s) indicates *Blinic*'s inherent capability to detect partial abnormal blink patterns even without noise mitigation.

6.4 Robustness

To answer the question (Q3), we empirically evaluate the effectiveness of *Blinic* under different settings.

6.4.1 Impact of Angle Range. The receivers capture reflections not only from the target user's head but also from surrounding surfaces. To address this multipath interference, we developed an angular filtering mechanism that selectively processes signals within natural head movement ranges (detailed in Section 4.2: *Noise caused by surrounding surfaces*). Through empirical evaluation with 37 participants, we compared *Blinic*'s performance under different ranges, *i.e.*, five incremental ranges from 50° to 90° (10° increments) and one extended 120° range.

Figure 21 reveals an inverse relationship between angular range and system performance. The MAE for TBUT measurement demonstrates progressive improvement, decreasing from 3.62s at 50° to an optimal 2.57s at 70°, then increasing to 3.64s at 120°. This pattern is mirrored in diagnostic accuracy, which rises from 86.49% (50°) to peak at 92.11%, before declining to 62.16% (120°). Concurrent TBUT prediction errors follow this pattern, with MAE values measuring 3.62s (50°), reaching optimal precision at 2.57s (70°), and deteriorating to 3.64s (120°). These results establish 70° as the optimal angular detection range, achieving minimal TBUT estimation error (MAE=2.57s) with an average accuracy of 91.89%. This configuration is therefore implemented as the system default.

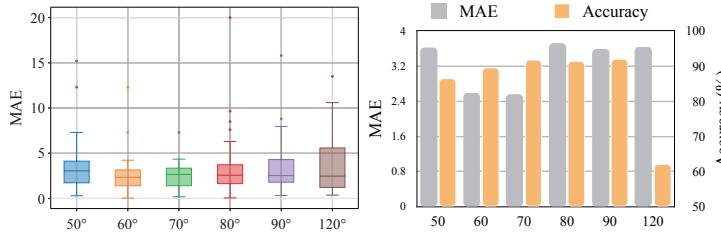


Figure 21: Impact of angle range.

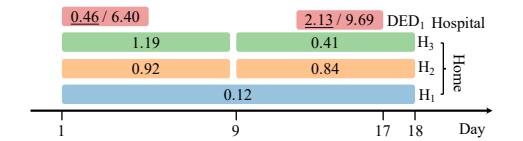


Figure 23: Long-term consistency of Blinic.

6.4.2 Impact of Distance. Given the uncontrolled nature of home environments where device-user distance may vary significantly, we empirically evaluated *Blinic*'s performance to validate its effectiveness at different user-device distances: 30, 40, 50, and 60 cm. Data were collected from three randomly selected DED patients (two severe cases and one moderate case). Participants first underwent clinical dry eye examinations at a hospital, followed by at-home testing at varying distances.

The average TBUT MAE values were 1.42s (30 cm), 1.94s (40 cm), 1.96s (50 cm), and 2.05s (60 cm). Although *Blinic*'s performance decreases with increasing distance, the MAE remains ≤ 2.05 s even at the maximum tested distance of 60 cm. This demonstrates *Blinic*'s viability for practical at-home applications.

6.4.3 Impact of Outliers. Since spontaneous blinks may be interfered by abnormal large values, we utilize an abnormal value threshold to remove its impact in Section 4.2: *Noise caused by outliers*). Through empirical evaluation with 37 participants, we compared *Blinic*'s performance under different thresholds, i.e., 300, 400, 500, and 800.

Figure 20 reveals *Blinic*'s performance degradation with increasing anomaly detection thresholds. At thresholds 300 / 400 / 500 / 800, TBUT MAE measures 3.38 / 2.57 / 2.85 / 3.58s while DED grading accuracy reaches 90.30% / 91.89% / 91.14% / 91.41%. The 400 threshold achieves optimal balance: lower thresholds (< 400) introduce noise through oversensitive blink detection, whereas higher thresholds (> 400) filter diagnostically useful features. This trade-off establishes 400 as the operational default.

6.4.4 Impact of Ages. We conducted stratified evaluations across four demographic cohorts: young adults (20-39 years, n=10), middle-aged (40-49 years, n=19), senior (50-59 years, n=12), and elderly (60-79 years, n=21) participants.

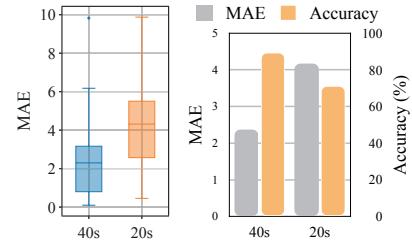


Figure 22: Impact of data length.

Figure 19 reveals significant age-dependent performance variations in *Blinic*. The system demonstrated TBUT prediction MAEs of 3.40s (20-39), 2.44s (40-49), 3.76s (50-59), and 2.17s (60-79), with corresponding diagnostic accuracies of 90.00%, 83.33%, 89.47%, and 95.24% respectively. Minimum MAE (2.17s) occurred in the elderly cohort (60-79 years), contrasting with the suboptimal performance in middle-aged participants. Notably, the system maintained robust performance (< 3.76) across all cohorts, demonstrating strong age generalizability.

6.4.5 Impact of Gender. We empirically evaluated the effectiveness of *Blinic* across different gender groups: male, female, and intergender. For each group, we selected 30 individuals for data collection.

As shown in Figure 17, the average MAE of TBUT prediction was 2.46s, 3.39s, and 3.05s for males, females, and intergender individuals, respectively. Similarly, the average DED accuracy of *Blinic* was 88.89%, 88.89%, and 91.30% for these three groups. These results indicate that *Blinic* performs consistently well across all gender categories. This consistency arises because *Blinic* analyzes spontaneous blinking patterns, which are indicative of ocular abnormalities and independent of gender.

6.4.6 Impact of Data Length. We empirically evaluated the effectiveness of *Blinic* across different data durations (20-second and 40-second intervals). In clinical practice [39], spontaneous blink analysis traditionally relies on 20-second recordings captured by high-speed cameras (≥ 1000 fps). For each duration group, we recruited 37 participants for data collection.

As illustrated in Figure 22, the average MAE of TBUT measurements is 2.37s and 4.17s for 40-second and 20-second mmWave data, respectively. Correspondingly, the DED grading accuracy is 89.47% and 71.05% for these two durations. Notably, *Blinic* demonstrates superior performance with 40-second data compared to 20-second recordings. This improvement likely stems from the extended 40-second window capturing more DED symptomatic patterns in spontaneous

blinks. Furthermore, even with 20-second data, *Blinic* maintains an accuracy exceeding 71.05%, suggesting its capability to detect partial DED manifestations.

6.4.7 Long-Term Consistency of Blinic. We conducted an 18-day longitudinal evaluation of *Blinic*'s measurement consistency with three healthy participants in home environments and one DED patient in clinical settings.

As shown in Figure 23, the system demonstrates sustained TBUT measurement accuracy through continuous tracking. Using initial clinical readings as baseline references: healthy participant H_1 maintained a MAE of 0.12 s throughout the 18-day observation window. Participant H_2 exhibited MAE fluctuations from 0.92 s at the 9-day mark to 0.84 s in subsequent measurements, while Participant H_3 showed more pronounced improvement from 1.19 s to 0.41 s over the same 9-day interval. Clinical validation with a DED patient: follow-up measurements at 17-day intervals yielded MAE values of 0.46 s (clinical TBUT 6.40 s) and 2.13 s (clinical TBUT 9.69 s), confirming *Blinic*'s capability to maintain diagnostic relevance beyond short-term usage.

7 RELATED WORK

As highlighted in the TFOS Dry Eye Workshop diagnostic guidelines [44], blink parameter analysis has emerged as a clinically significant biomarker, earning recognition as a recommended diagnostic tool for DED evaluation [32]. Thus, one leading concept is to use blink analysis for screening DED [33, 45, 48]. SDE [45] uses RF signals to capture fine-grained spontaneous blinking and extract DED biomarker representations, realizing a ubiquitous DED screening system. EyeScore [48] employs an iPhone to record eyelid movements for one minute within the app, analyzing blink rates and patterns as early clinical biomarkers for DED. DryEyeRhythm [33] utilizes the cameras of smart devices to assess users' blink characteristics for screening DED, achieving a positive predictive value of 91.3% and a negative predictive value of 69.1%. Other studies [8, 22, 23, 38, 41] leverage specific devices or additional tests for DED screening. While these works successfully demonstrate the feasibility of assessing DED at home, they do not facilitate the measurement of TBUT. To measure TBUT in a home setup, DEDector [20] employs an external optical attachment on smartphone and uses optical image-processing methods to screen for abnormal TBUT. However, DEDector can still only achieve a two-class classification that differentiates normal or abnormal TBUT subjects, while measuring precise TBUT values for management is not feasible.

In response to these limitations, *Blinic* explores the feasibility of designing a passive, contactless, home-based solution to accurately assess TBUT values for improved DED management. This approach aims to indicate disease progression

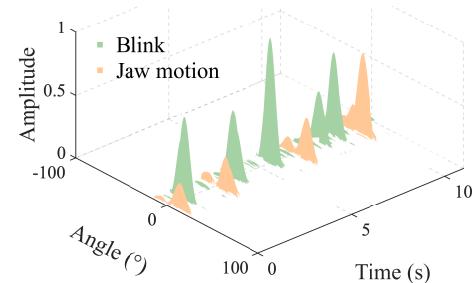


Figure 24: Angular separation of jaw motion.

and facilitate timely adjustments to treatment strategies and interventions.

8 DISCUSSION

Facial Movement Considerations. While *Blinic* focuses on detecting subtle eyelid movements during blinks, other motions like facial expressions or jaw movements can also occur. As shown in Figure 8(a), speech-induced movements generate interference comparable to head motion in terms of amplitude and duration characteristics. Since *Blinic* requires only ~40 seconds per assessment, we can reasonably instruct patients to remain silent and still during measurement to mitigate these artifacts. Notably, both speech and head movements constitute significant noise sources relative to small eyelid motions.

Furthermore, *Blinic*'s antenna-coded MIMO mmWave radar enhances angular resolution, enabling separation between eyelid and facial movements in the angle-Doppler domain. As demonstrated in Figure 24, rotating the radar by 90 degrees allows clear separation of blinks and minor jaw movements along the angle dimension.

9 CONCLUSION

In this paper, we propose *Blinic*, a fully passive and contactless DED assessment system based on mmWave signals. The system utilizes a teacher-student architecture to distill pathological knowledge from EHR data into blink kinematics captured via mmWave radar. *Blinic* then quantifies TBUT through analysis of blink signals. Experimental validation demonstrates the system's effectiveness, with results confirming its operational reliability. This study provides critical insights for developing DED assessment systems to enhance home-based management of the condition.

ACKNOWLEDGMENTS

We thank all reviewers and our shepherd for their insightful comments. This research is supported in part by RGC under Contract CERG 16206122, 16204523, 16205824, AoE/E-601/22-R, R6021-20, SRFS2425-6S05 and Contract R8015.

REFERENCES

- [1] 2021. IWR1443BOOST. <https://www.ti.com/tool/IWR1443BOOST>.
- [2] 2022. Google's New Tech Can Read Your Body Language—Without Cameras. <https://www.wired.com/story/google-soli-atap-research-2022/>.
- [3] 2022. Nest Hub (2nd gen). https://store.google.com/us/product/nest_hub_2nd_gen?hl=en-US.
- [4] 2024. Nest Learning Thermostat (4th gen). https://store.google.com/us/product/nest_learning_thermostat_4th_gen?hl=en-US.
- [5] 2024. vayyarcare. <https://vayyar.com/care/>.
- [6] 2025. The OCULUS Keratograph® 5M. <https://www.oculus.de/us/products/keratograph-5m/#angebot>.
- [7] Mark B Abelson, George W Ousler, Lauren A Nally, Donna Welch, and Kathleen Krenzer. 2002. Alternative reference values for tear film break up time in normal and dry eye populations. *Lacrimal gland, tear film, and dry eye syndromes 3: basic science and clinical relevance part A and B* (2002), 1121–1125.
- [8] U Rajendra Acharya, Jen Hong Tan, Joel EW Koh, Vidya K Sudarshan, Sharon Yeo, Cheah Loon Too, Chua Kuang Chua, EYK Ng, and Louis Tong. 2015. Automated diagnosis of dry eye using infrared thermography images. *Infrared Physics & Technology* 71 (2015), 263–271.
- [9] Reiko Arita, Jun Suehiro, Tsuyoshi Haraguchi, Rika Shirakawa, Hideaki Tokoro, and Shiro Amano. 2014. Objective image analysis of the meibomian gland area. *British Journal of Ophthalmology* 98, 6 (2014), 746–755.
- [10] Penny Asbell, Algis J Vingrys, Jacqueline Tan, Abayomi Ogundele, Laura E Downie, Gary Jenkins, and Lee Shettle. 2018. Clinical outcomes of fixed versus as-needed use of artificial tears in dry eye disease: a 6-week, observer-masked phase 4 clinical trial. *Investigative Ophthalmology & Visual Science* 59, 6 (2018), 2275–2280.
- [11] Christophe Baudouin, Elisabeth M Messmer, Pasquale Aragona, Gerd Geerling, Yonca A Akova, José Benítez-del Castillo, Kostas G Boboridis, Jesús Merayo-Lloves, Maurizio Rolando, and Marc Labetoulle. 2016. Revisiting the vicious circle of dry eye disease: a focus on the pathophysiology of meibomian gland dysfunction. *British Journal of Ophthalmology* 100, 3 (2016), 300–306.
- [12] Zhaoxin Chang, Fusang Zhang, Jie Xiong, Weiyuan Chen, and Daqing Zhang. 2024. MSense: Boosting wireless sensing capability under motion interference. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 108–123.
- [13] Baicheng Chen, John Nolan, and Xinyu Zhang. 2024. MetaBioLiq: A Wearable Passive Metasurface Aided mmWave Sensing Platform for BioFluids. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 1192–1206.
- [14] Tao Chen, Yongjie Yang, Xiaoran Fan, Xiuzhen Guo, Jie Xiong, and Longfei Shangguan. 2024. Exploring the feasibility of remote cardiac auscultation using earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 357–372.
- [15] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2024. Scaling instruction-finetuned language models. *Journal of Machine Learning Research* 25, 70 (2024), 1–53.
- [16] Jennifer P Craig, Kelly K Nichols, Esen K Akpek, Barbara Caffery, Harminder S Dua, Choun-Ki Joo, Zuguo Liu, J Daniel Nelson, Jason J Nichols, Kazuo Tsubota, et al. 2017. TFOS DEWS II definition and classification report. *The ocular surface* 15, 3 (2017), 276–283.
- [17] Jennifer P Craig, Michael TM Wang, Dabin Kim, and Jung Min Lee. 2016. Exploring the predisposition of the Asian eye to development of dry eye. *The Ocular Surface* 14, 3 (2016), 385–392.
- [18] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
- [19] Kenta Fujio, Ken Nagino, Tianxiang Huang, Jaemyoung Sung, Yasutsugu Akasaki, Yuichi Okumura, Akie Midorikawa-Inomata, Keiichi Fujimoto, Atsuko Eguchi, Maria Miura, et al. 2023. Clinical utility of maximum blink interval measured by smartphone application DryEye-eRhythm to support dry eye disease diagnosis. *Scientific Reports* 13, 1 (2023), 13583.
- [20] Vaibhav Ganatra, Soumyasis Gun, Pallavi Joshi, Anand Balasubramanian, Kaushik Murali, Nipun Kwatra, and Mohit Jain. 2024. DEDector: Smartphone-Based Noninvasive Screening of Dry Eye Disease. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–26.
- [21] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Fran ois Laviolette, Mario March, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of machine learning research* 17, 59 (2016), 1–35.
- [22] Yuting Hong and Makoto Hasegawa. 2021. Proposal of Tear Meniscus Measurement for Minor Dry-eye Detection Using Smart-phone Camera and Ring-light. In *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. IEEE, 1–4.
- [23] Yuting Hong and Makoto Hasegawa. 2021. Study of minor dry-eye detection using smartphone camera based on deep learning. In *International Workshop on Advanced Imaging Technology (IWAIT) 2021*, Vol. 11766. SPIE, 621–626.
- [24] Sho Ishikawa, Takanori Sasaki, Takahumi Maruyama, Koichiro Murayama, and Kei Shinoda. 2023. Effectiveness and adherence of dry eye patients who switched from short-to long-acting diquafosol ophthalmic solution. *Journal of Clinical Medicine* 12, 13 (2023), 4495.
- [25] Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. 2021. Perceiver: General perception with iterative attention. In *International conference on machine learning*. PMLR, 4651–4664.
- [26] Ying Jie, Ruti Sella, Jun Feng, Maria L Gomez, and Natalie A Afshari. 2019. Evaluation of incomplete blinking as a measurement of dry eye disease. *The Ocular Surface* 17, 3 (2019), 440–446.
- [27] Eric Lehman, Evan Hernandez, Diwakar Mahajan, Jonas Wulff, Micah J Smith, Zachary Ziegler, Daniel Nadler, Peter Szolovits, Alistair Johnson, and Emily Alsentzer. 2023. Do we still need clinical language models?. In *Conference on health, inference, and learning*. PMLR, 578–597.
- [28] Yingcheng Liu, Guo Zhang, Christopher G Tarolli, Rumen Hristov, Stella Jensen-Roberts, Emma M Waddell, Taylor L Myers, Meghan E Pawlik, Julia M Soto, Renee M Wilson, et al. 2022. Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response. *Science Translational Medicine* 14, 663 (2022), eadc9669.
- [29] Julia C Mainstone, Adrian S Bruce, and Timothy R Golding. 1996. Tear meniscus measurement in the diagnosis of dry eye. *Current eye research* 15, 6 (1996), 653–661.
- [30] Cynthia Matossian, Micaela Crowley, Laura Periman, and Steven Sorkin. 2022. Personalized management of dry eye disease: beyond artificial tears. *Clinical Ophthalmology (Auckland, NZ)* 16 (2022), 3911.
- [31] Evan Michaelov, Caroline McKenna, Pierre Ibrahim, Manav Nayeni, Arpit Dang, and Rookaya Mather. 2022. Sj  gren's syndrome associated dry eye: impact on daily living and adherence to therapy. *Journal of Clinical Medicine* 11, 10 (2022), 2809.
- [32] Anthony Oganov, Ghasem Yazdanpanah, Sayena Jabbehdari, Aditya Belamkar, and Stephen Pflugfelder. 2023. Dry eye disease and blinking behaviors: A narrative review of methodologies for measuring blink dynamics and inducing blink response. *The Ocular Surface* 29 (2023),

- 166–174.
- [33] Yuichi Okumura, Takenori Inomata, Akie Midorikawa-Inomata, Jaemyoung Sung, Kenta Fujio, Yasutsugu Akasaki, Masahiro Nakamura, Masao Iwagami, Keiichi Fujimoto, Atsuko Eguchi, et al. 2022. DryEyeRhytm: A reliable and valid smartphone application for the diagnosis assistance of dry eye. *The Ocular Surface* 25 (2022), 19–25.
- [34] Xiaomin Ouyang, Xian Shuai, Yang Li, Li Pan, Xifan Zhang, Heming Fu, Sitong Cheng, Xinyan Wang, Shihua Cao, Jiang Xin, et al. 2024. ADMarker: A Multi-Modal Federated Learning System for Monitoring Digital Biomarkers of Alzheimer's Disease. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 404–419.
- [35] H Pult, Christine Purslow, and Paul James Murphy. 2011. The relationship between clinical signs and dry eye symptoms. *Eye* 25, 4 (2011), 502–510.
- [36] Sruthi Srinivasan, Kara Menzies, Luigina Sorbara, and Lyndon Jones. 2012. Infrared imaging of meibomian gland structure using a novel keratograph. *Optometry and Vision Science* 89, 5 (2012), 788–794.
- [37] Fiona Stapleton, Monica Alves, Vatinee Y Bunya, Isabelle Jalbert, Kaevalin Lekhanont, Florence Malet, Kyung-Sun Na, Debra Schaumberg, Miki Uchino, Jelle Vehof, et al. 2017. Tfos dews ii epidemiology report. *The ocular surface* 15, 3 (2017), 334–365.
- [38] Tai Yuan Su, Chen Kerh Hwa, Po Hsuan Liu, Ming Hong Wu, David O Chang, Po Fang Su, Shu Wen Chang, and Huihua Kenny Chiang. 2011. Noncontact detection of dry eye using a custom designed infrared thermal image system. *Journal of biomedical optics* 16, 4 (2011), 046009–046009.
- [39] Yuandong Su, Qingfeng Liang, Guanyu Su, Ningli Wang, Christophe Baudouin, and Antoine Labb  . 2018. Spontaneous eye blink patterns in dry eye: clinical correlations. *Investigative ophthalmology & visual science* 59, 12 (2018), 5149–5156.
- [40] Zehua Sun, Tao Ni, Yongliang Chen, Di Duan, Kai Liu, and Weitao Xu. 2024. Rf-egg: An rf solution for fine-grained multi-target and multi-task egg incubation sensing. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 528–542.
- [41] Maho Taniai, Kaito Okazaki, and Makoto Hasegawa. 2023. Improvement of Tear Meniscus Measurement for Dry Eye Detection Using Smartphone and Ring Light. In *2023 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. IEEE, 1–4.
- [42] Kazuo Tsubota. 2018. Short tear film breakup time-type dry eye. *Investigative ophthalmology & visual science* 59, 14 (2018), DES64–DES70.
- [43] Miki Uchino, Norihiko Yokoi, Jun Shimazaki, Yuichi Hori, Kazuo Tsubota, and Japan Dry Eye Society. 2022. Adherence to eye drops usage in dry eye patients and reasons for non-compliance: a web-based survey. *Journal of Clinical Medicine* 11, 2 (2022), 367.
- [44] James S Wolffsohn, Reiko Arita, Robin Chalmers, Ali Djalilian, Murat Dogru, Kathy Dumbleton, Preety K Gupta, Paul Karpecki, Sihem Lazreg, Heiko Pult, et al. 2017. TFOS DEWS II diagnostic methodology report. *The ocular surface* 15, 3 (2017), 539–574.
- [45] Meng Xue, Yuyang Zeng, Shengkang Gu, Qian Zhang, Bowei Tian, and Changzheng Chen. 2024. SDE: Early Screening for Dry Eye Disease with Wireless Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–23.
- [46] Jinyu Yang, Jingjing Liu, Ning Xu, and Junzhou Huang. 2023. Tvt: Transferable vision transformer for unsupervised domain adaptation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*. 520–530.
- [47] Yuzhe Yang, Yuan Yuan, Guo Zhang, Hao Wang, Ying-Cong Chen, Yingcheng Liu, Christopher G Tarolli, Daniel Crepeau, Jan Bukartyk, Mithri R Junna, et al. 2022. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nature medicine* 28, 10 (2022), 2207–2215.
- [48] Sydney Zhang and Julio Echegoyen. 2023. Design and Usability Study of a Point of Care mHealth App for Early Dry Eye Screening and Detection. *Journal of Clinical Medicine* 12, 20 (2023), 6479.