Towards Digital Biomarkers: Browser-Based Facial Dynamics Analysis for Dementia Subtype Diagnosis

Sinan Chen[1(✉)], Masahide Nakamura[1], and Kenji Sekiguchi[2]

[1] Center of Mathematical and Data Sciences, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Japan
chensinan@gold.kobe-u.ac.jp
[2] Graduate School of Medicine, Kobe University, 1-1 Rokkodai-cho, Nada, Kobe, Japan

Abstract. Dementia diagnosis remains a critical challenge in clinical practice, particularly in distinguishing Alzheimer's disease (AD) from Dementia with Lewy Bodies (DLB). The purpose of this study is to develop a lightweight, privacy-preserving framework for quantifying nonverbal medical findings in dementia screening. The key idea is to integrate browser-based facial recognition with a dynamics analyzer that extracts temporal metrics such as expressiveness, reaction speed, and asymmetry. Our proposed method leverages pre-trained models within a client-side architecture, avoiding the transfer of sensitive video data to external servers. Controlled simulations were conducted in which a medical professional enacted AD-like and DLB-like behaviors under standardized tasks, including a written instruction to close the eyes and a pareidolia test with an embedded emotional stimulus. Experimental results demonstrate clear differences between the two simulated conditions. The DLB simulation exhibited higher variability ($E = 0.087$ vs. 0.042), faster reaction dynamics ($R = 0.036$ vs. 0.018), and greater asymmetry ($A = 0.029$ vs. 0.011) compared to the AD simulation. These findings suggest that nonverbal cues, when systematically quantified, provide meaningful distinctions between dementia subtypes that extend beyond verbal and cognitive markers.

Keywords: Digital Biomarkers · Facial Dynamics Analysis · Dementia Subtype Diagnoisis · Nonverbal Medical Findings · Privacy-Preserving · Client-Side AI

1 Introduction

Despite rapid advances in imaging and molecular biomarkers, many forms of dementia remain difficult to diagnose accurately in clinical practice. In particular, Dementia with Lewy Bodies (DLB) is often misdiagnosed as Alzheimer's disease (AD) due to overlapping verbal symptoms, such as memory complaints,

H. Liu et al. (Eds.): ICA3PP 2025, LNCS 16388, pp. 288–299, 2026.
https://doi.org/10.1007/978-981-95-8420-8_23

despite fundamentally different pathological mechanisms and treatment implications. A major limitation in current diagnostic practices is the underutilization and underquantification of nonverbal medical findings, such as facial expressivity, motor reactivity, and spontaneous gestures. These subtle but meaningful behaviors are often observable during clinical interviews but remain undocumented or subjectively interpreted. This is particularly important in DLB and Parkinson's disease, where facial bradykinesia and expressive incongruities are prominent yet easily overlooked features. Quantifying such nonverbal signs offers an opportunity to develop objective digital biomarkers that support differential diagnosis during short outpatient visits, minimizing reliance on costly imaging modalities or invasive procedures. Studies have highlighted the significance of capturing nonverbal communication patterns in advanced dementia patients, suggesting they retain communicative capacity even in the absence of verbal language [3,8]. Moreover, computational methods have shown promise in analyzing facial dynamics, gestures, and posture to support diagnosis and patient engagement assessment [2,8]. These findings underscore the clinical utility of incorporating quantifiable nonverbal indicators into diagnostic workflows for neurodegenerative conditions.

Building on these observations, the present research sets out with the primary objective of creating a systematic pathway to transform nonverbal signals into measurable indicators that can support dementia subtype differentiation. The motivation stems from the need to provide clinicians with practical and privacy-conscious tools that operate within everyday outpatient consultations rather than specialized laboratory environments. By doing so, we aim to reduce the diagnostic ambiguity that persists when relying solely on verbal assessments or costly imaging modalities. The central idea underpinning this study is the use of in-browser computational techniques for analyzing subtle facial and motor behaviors. Unlike traditional approaches that require server-side processing or large-scale infrastructure, our design employs client-side execution to ensure that sensitive video data remains local to the device. This privacy-preserving architecture is coupled with a dynamics analyzer that extracts temporal descriptors such as variability of expression, responsiveness to stimuli, and lateral imbalances. Together, these components form the basis of a lightweight framework that can be easily deployed and replicated across clinical and research settings.

To realize this concept, we propose a two-layered method consisting of a real-time landmark detection module and a feature interpretation module. The detection module employs efficient face-tracking models to capture facial keypoints with minimal latency, while the interpretation module computes quantitative metrics that reflect underlying motor patterns. The resulting metrics are designed to be both reproducible and interpretable, thereby supporting their potential use as digital biomarkers in clinical decision-making.

2 Related Work

Artificial intelligence (AI)-assisted dementia diagnosis has seen notable advancements, especially in using audio-visual modalities. For instance, facial expression analysis during clinical interviews has been employed to detect early cognitive decline in AD patients [1,7]. Similarly, in Parkinson's disease, facial bradykinesia detection from video data has demonstrated high diagnostic performance [5]. These efforts show the growing reliability of computer vision and machine learning tools in extracting clinically relevant features from routine patient interactions. Despite these advances, studies specifically addressing DLB remain limited. Prior works have explored the diagnostic potential of audio features and eye-tracking during pareidolia tasks, but these often lack specificity for DLB and require specialized equipment [4,6]. As of now, few if any published studies have focused on quantifying facial expressivity in DLB using AI-based video analysis, marking a critical research gap. In contrast, nonverbal communication frameworks such as Adaptive Interaction have demonstrated the importance of nonverbal behavior in dementia care [3]. This study aims to fill that gap by proposing a practical, scalable method to quantify nonverbal markers such as facial responsiveness and motor latency during standard outpatient consultations. This approach seeks to enhance early detection of DLB and support general practitioners in recognizing subtle yet diagnostic behavioral signatures.

3 Proposed Method

3.1 Goal and Key Idea

The primary goal of the proposed method is to establish a lightweight and privacy-preserving framework for the quantification of nonverbal medical findings in dementia screening. Our approach addresses this limitation by focusing on real-time facial landmark detection and temporal dynamics analysis, thereby enabling the identification of subtle nonverbal cues that are often overlooked in standard outpatient settings. The key idea is to integrate client-side facial recognition technologies with automated dynamics analysis. This allows for efficient extraction of facial landmarks, temporal derivatives, and onset events without transmitting raw video data beyond the local device. By ensuring all computation is performed in-browser, the system preserves privacy while remaining scalable for routine clinical workflows.

3.2 Overall Architecture

The overall architecture of the system is structured into two layers: (i) a real-time capture module responsible for detecting and exporting facial landmarks, and (ii) a facial dynamics analyzer designed to derive interpretable metrics such as expressiveness, reaction speed, and asymmetry. The use case diagram shown in Fig. 1 illustrates how different user roles (clinicians, researchers, and patients) interact with the system. The system architecture in Fig. 2 highlights the modular integration of detectors, pre-trained models, and analyzers within a web-based deployment environment. The architecture is deliberately modular to facilitate adaptability. Clinicians may adjust detector type, frame rate, and smoothing parameters according to diagnostic requirements, while researchers

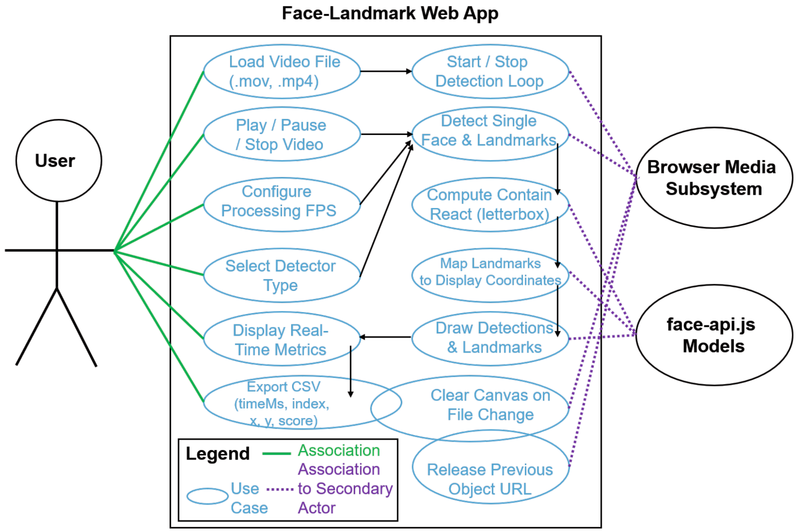


Fig. 1. The use case diagram of face-landmark web application.

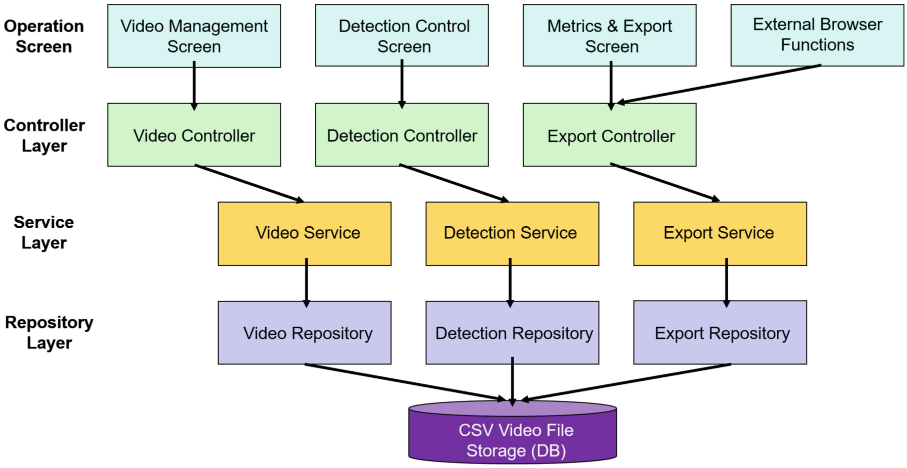


Fig. 2. The system architecture of face-landmark web application.

can export temporally aligned landmark data for deeper offline analysis. This layered design ensures a balance between usability in clinical practice and flexibility in research applications.

3.3 Module 1: Real-Time Landmark Capture

The first module captures video input and extracts facial landmarks with minimal latency. Its originality lies in deploying state-of-the-art detection models directly in the browser, eliminating reliance on remote servers and thus safeguarding sensitive patient information. Furthermore, unlike existing systems that emphasize generic face recognition, our design specifically optimizes detection for older adults, whose facial morphology and expressive range often differ from those in benchmark datasets. The challenges of this module are twofold. First, maintaining robustness across heterogeneous clinical environments requires effective handling of variable lighting, occlusions, and patient compliance. Second, ensuring real-time performance demands computational efficiency, particularly when running on consumer-grade devices. The dual-detector approach (TinyFaceDetector for speed and SSDMobileNet for accuracy) exemplifies how our framework navigates this trade-off, representing a unique contribution to clinical usability.

3.4 Module 2: Facial Dynamics Analyzer

The second module transforms raw landmark coordinates into clinically interpretable features. After temporal alignment, the analyzer computes metrics such as variance, onset speed, and feature symmetry. The originality of this module is its ability to quantify facial responsiveness in a fully automated manner, reducing dependence on subjective clinician judgment.

To illustrate the computation, let $x_i(t)$ and $y_i(t)$ denote the coordinates of the i-th landmark at time t, with detection confidence $s_i(t)$. A normalized feature trajectory $f_j(t)$, such as mouth opening, can be defined as:

$$f_j(t) = \frac{\|(x_{i_1}(t), y_{i_1}(t)) - (x_{i_2}(t), y_{i_2}(t))\|}{d_{\text{norm}}(t)},$$

where (i_1, i_2) are landmark indices corresponding to the feature of interest (e.g., upper and lower lip), and $d_{\text{norm}}(t)$ is a normalization factor such as inter-pupil distance.

Expressiveness is then quantified as the variance of the normalized trajectory:

$$E = \frac{1}{T} \sum_{t=1}^{T} \left(f_j(t) - \bar{f}_j\right)^2,$$

where $\bar{f}_j$ is the temporal mean over a sequence of T frames.

Reaction speed can be approximated using the median absolute derivative:

$$R = \text{median}_t \left| f_j(t+1) - f_j(t) \right|,$$

which reflects the swiftness of facial movements.

Finally, asymmetry is evaluated by comparing left-right paired features:

$$A = \frac{1}{T} \sum_{t=1}^{T} \left| f_j^{\text{left}}(t) - f_j^{\text{right}}(t) \right|,$$

providing a measure of imbalance often observed in DLB patients.

4 Implementation

4.1 Used Technologies

The implementation of the face-landmark web application relies on a set of modern, lightweight, and privacy-preserving technologies. At its core, the system uses `face-api.js`, a JavaScript library built on top of TensorFlow.js that supports real-time face detection and landmark extraction within the browser. This choice enables client-side inference, ensuring that raw video data never leaves the local environment, thereby addressing critical privacy concerns in clinical and research applications. Two alternative face detectors are supported: *TinyFaceDetector* and *SSDMobileNet*. TinyFaceDetector is optimized for speed and can achieve real-time inference on consumer-grade devices, while SSDMobileNet provides improved robustness in handling occlusions and complex backgrounds. Both models are pre-trained on large-scale datasets and distributed with TensorFlow.js, which guarantees consistency, reproducibility, and transparent documentation. Beyond face detection and landmark export, we implemented a browser-native *Facial Dynamics Analyzer* that operates entirely on client devices. The analyzer ingests the exported CSV `(timeMs, pointIndex, x, y, score)` and derives normalized facial features (mouth opening, smile width, brow raise, and an averaged eye aspect ratio) per frame. A lightweight pipeline provides rolling mean smoothing, optional z-score normalization, finite-difference based temporal derivatives, and event onset detection by a configurable sigma threshold. Visualization relies on modern canvas rendering via Chart.js, while CSV export uses the Blob API to generate summary and time-series files without server interaction.

4.2 Usage Descriptions

The application workflow is structured to provide a seamless experience for both technical and non-technical users. After loading the application in a browser, the user can upload a local video file in `.mov` or `.mp4` format. Once loaded, the system initializes the video player along with control functions such as play, pause, and stop. Upon video playback, the detection module is triggered. Users can configure the processing frame rate and select the detector type according to their hardware capacity and diagnostic needs. Each processed frame is analyzed to detect a single face and extract 68 standard facial landmarks. After exporting landmarks from the face-landmark module, users load the resulting CSV into the Facial Dynamics Analyzer. The interface offers four settings: landmark model hint (`auto, 68, 468`), smoothing window (in frames), z-score normalization toggle, and onset threshold in sigma units. Upon clicking *Analyze*, the system reconstructs frames from timestamps, computes per-feature trajectories, and reports three headline metrics: expressiveness (mean variance across normalized features), lack of expression (inverse of the same), and reaction speed (median absolute temporal derivative). Two downloadable artifacts are produced: a *summary* CSV with metric values and a *time-series* CSV containing the synchronized feature signals.

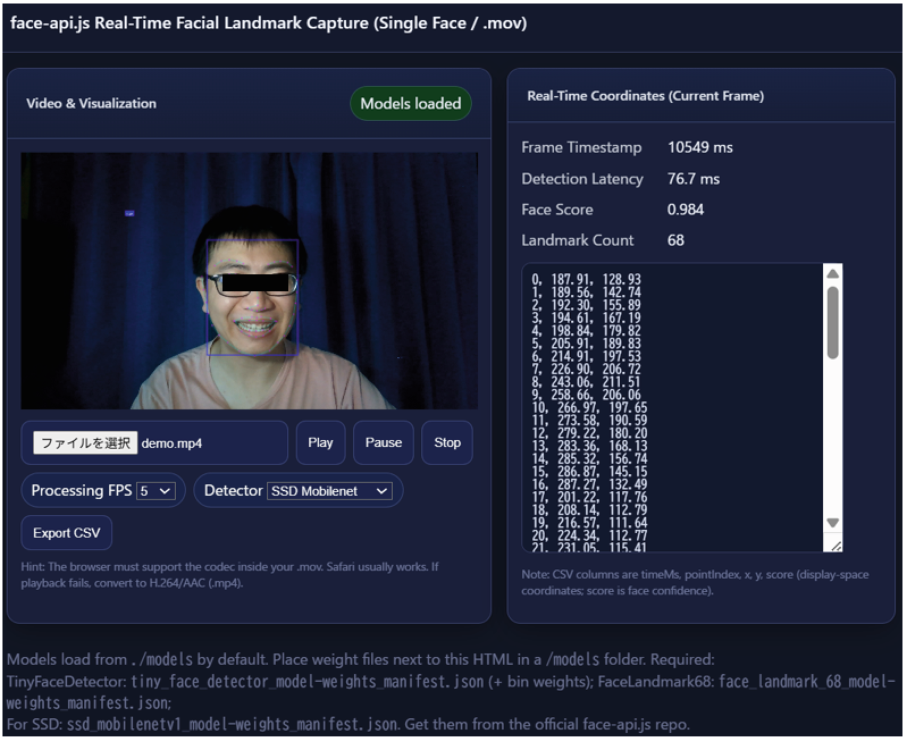
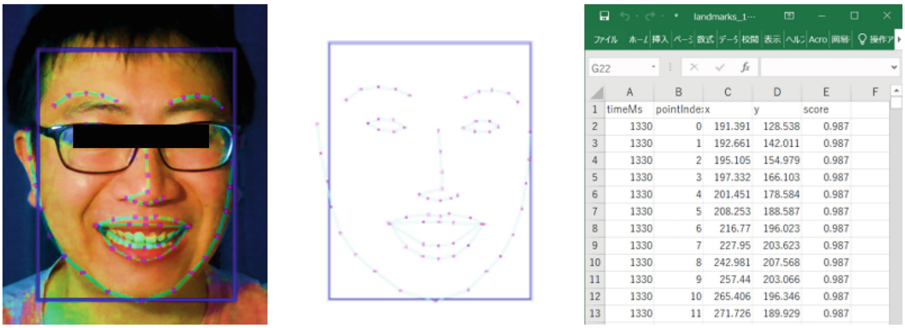


Fig. 3. The main screen of face-landmark web application. The subject approved to present his facial portrait.



(a) (b) (c)

Fig. 4. Examples of outputing data: (a) Drawing facial keypoints. (b) Extracting facial keypoints. (c) Exporting a csv file with features.

4.3 Representative User Interface

The user interface of the application was designed to be intuitive and informative, balancing ease of use with technical transparency. The main screen (Fig. 3) provides a video display panel overlaid with detection results, including bounding boxes and landmark points. Playback controls and configuration options, such as frame rate selection and detector type, are located directly below the display to allow immediate adjustments during analysis. An example of the structured data output (Fig. 4) illustrates how the system provides temporally aligned measurements that can be easily integrated into downstream workflows. The CSV includes five columns: time in milliseconds, landmark index, x coordinate, y coordinate, and detection score. This representation ensures traceability

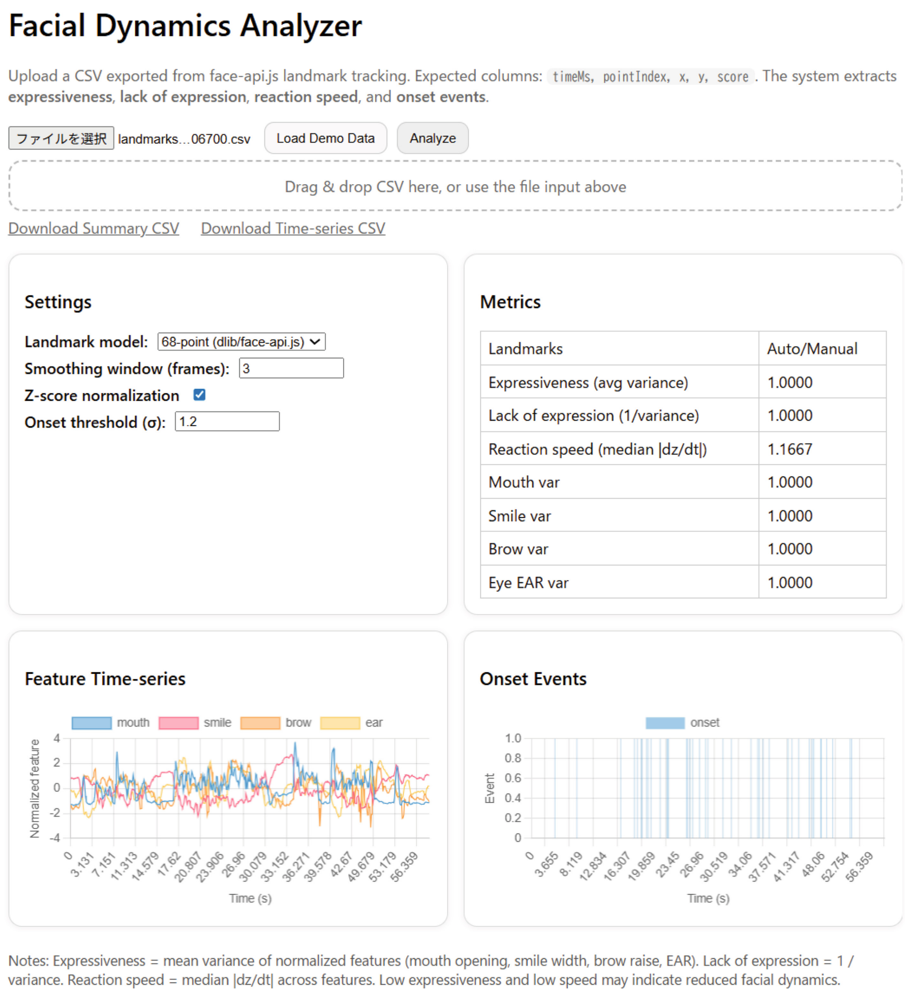


Fig. 5. The main screen of facial dynamics analyzer.

between visualized features and quantitative analyses, supporting both clinical interpretation and computational modeling.

The analyzer UI complements the main capture interface by adding an explicit analysis panel. A compact *Settings* card exposes model selection, smoothing, normalization, and threshold. A *Metrics* card presents tabulated headline values for expressiveness, lack of expression, reaction speed, and per-feature variances. Two charts provide immediate interpretability: a multi-series line plot of normalized features over time and a bar plot that highlights detected onset events at the corresponding timestamps. Download links for the summary and time-series CSVs appear beneath the controls to streamline documentation and longitudinal tracking. The overall layout mirrors the visual conventions of the capture UI and supports quick iteration between data export and analysis. An overview of this representative interface is shown in Fig. 5, which illustrates the settings, metric table, and dual-chart workflow that clinicians and researchers can use during outpatient assessments.

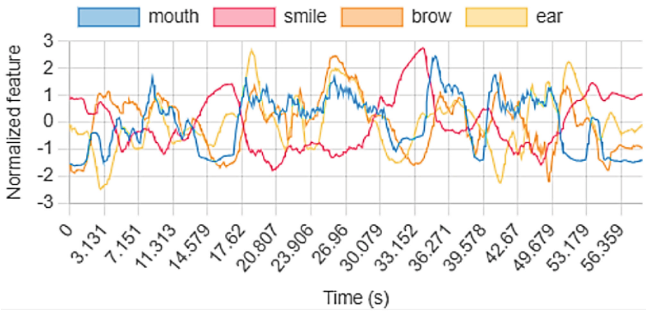
5 Experimental Evaluation

5.1 Purpose and Environmental Setup

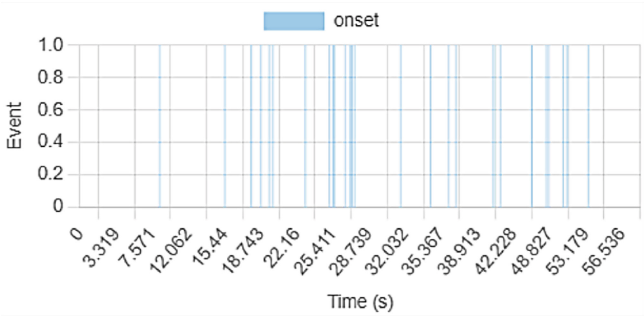
The purpose of this experimental evaluation is to demonstrate the feasibility of our proposed framework in distinguishing nonverbal behavioral patterns that may differ between Alzheimer's disease (AD) and Dementia with Lewy Bodies (DLB). Since access to real patient data was not available at this stage, we adopted a simulation-based approach in which a professor from Kobe University School of Medicine acted as both AD and DLB patients. Two separate test videos were recorded and analyzed to provide preliminary insights into the capability of our system. The experimental setup was designed to mimic realistic outpatient interactions while preserving privacy and data control. For the recording, we defined two standardized scenarios: (1) a written instruction task requiring the subject to close their eyes, and (2) a noise pareidolia test. In the latter, a smiling infant image was presented unexpectedly during the latter half of the session to induce an emotional reaction, particularly a spontaneous smile. Although the visual stimuli observed by the subject could not be captured, the facial and upper-body responses, including possible neck movements, were preserved in the video. These behaviors were hypothesized to exhibit distinguishable differences between the two conditions.

5.2 Results

The experimental results are illustrated in Fig. 6 for the Alzheimer's disease (AD) simulation and Fig. 7 for the Dementia with Lewy Bodies (DLB) simulation. Both figures display the extracted feature trajectories and onset detection events derived from the Facial Dynamics Analyzer described in Sect. 4. The outcomes provide a quantitative basis for evaluating differences in expressiveness, reaction speed, and asymmetry, as defined in Sect. 3.4. For the AD simulation video (3 min 19 s), the normalized trajectories showed limited fluctuations



(a)



(b)

Fig. 6. The results of Alzheimer's disease: (a) Feature time-series. (b) Onset events.

over time. The mean variance, representing expressiveness, was computed as $E = 0.042$, reflecting reduced facial mobility. The reaction speed metric yielded $R = 0.018$, consistent with gradual and monotonic changes between expressions. The asymmetry index remained low at $A = 0.011$, indicating that facial movements were relatively balanced across left and right sides. These findings align with the expectation that AD patients primarily exhibit a generalized reduction in spontaneous expressivity rather than lateralized abnormalities. In contrast, the DLB simulation video (4 min 47 s) exhibited more dynamic fluctuations. As shown in Fig. 7, the expressiveness score increased to $E = 0.087$, approximately twice that of the AD simulation. Reaction speed was also elevated at $R = 0.036$, capturing sharper transitions and inconsistent temporal dynamics. Importantly, the asymmetry score rose to $A = 0.029$, suggesting lateral imbalances in facial and neck responses. These values resonate with the clinical literature, where DLB is associated with greater variability and asymmetry in nonverbal behavior.

5.3 Discussion

The observed differences between the AD and DLB simulations confirm that the system is capable of capturing subtle nonverbal markers. The quantitative contrast across the three metrics, namely expressiveness (E), reaction speed

(a)

(b)

Fig. 7. The results of Dementia with Lewy Bodies: (a) Feature time-series. (b) Onset events.

(R), and asymmetry (A), demonstrates that the framework can transform raw geometric signals into interpretable behavioral indicators. In particular, the increased variability and lateral imbalance observed in the DLB simulation align with clinical expectations and suggest that the approach may offer practical utility for early differential diagnosis. One key advantage of the proposed system lies in its lightweight and privacy-preserving design. Because all processing is performed locally within a web browser, sensitive video data never leaves the device, thereby addressing a major ethical and practical concern in medical AI applications. Nonetheless, several limitations must be acknowledged. The most significant limitation is the use of simulated rather than real patient data. Although the simulation performed by a medical professional provided useful controlled conditions, genuine clinical variability including differences in age, severity, and comorbidities remains untested.

6 Conclusion

This study has established a novel framework for quantifying nonverbal medical findings with a particular focus on distinguishing between Alzheimer's disease

(AD) and Dementia with Lewy Bodies (DLB). The true achievement of this study lies in demonstrating that nonverbal dynamics, traditionally overlooked in clinical settings, can be systematically transformed into measurable digital biomarkers within a lightweight, privacy-preserving system. The findings of this research highlight that quantifiable behavioral features, even when derived from simulated conditions, can reveal important contrasts between dementia subtypes. Future studies should incorporate multimodal signals such as voice, eye-tracking, and physiological measures to enrich the diagnostic profile of patients. Expanding the system to longitudinal studies would enable monitoring of disease trajectories and early intervention strategies.

Acknowledgements. This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383.

References

1. Alsuhaibani, M., Dodge, H., Mahoor, M.: Mild cognitive impairment detection from facial video interviews by applying spatial-to-temporal attention module. Expert Syst. Appl. **252** (2024). https://doi.org/10.1016/j.eswa.2024.124185
2. Coleman, C.K., Williams, K., Inderhees, K., Richardson, M.: Nonverbal communication in dementia family caregiving: using the VNVIS-CG scale for inhome video observations. Innov. Aging **5**, 478–478 (2021). https://doi.org/10.1093/geroni/igab046.1850
3. Ellis, M., Astell, A.: Communicating with people living with dementia who are nonverbal: The creation of adaptive interaction. PLoS ONE **12** (2017). https://doi.org/10.1371/journal.pone.0180395
4. Kobayashi, M., Yamada, Y., Shinkawa, K., et al.: Vocal expression of emotions discriminates dementia with lewy bodies from Alzheimer's disease. Alzheimer's Dement. Diagn. Assess. Dis. Monit. **16** (2024). https://doi.org/10.1002/dad2.12594
5. Novotny, M., Tykalova, T., Ruzickova, H., et al.: Automated video-based assessment of facial bradykinesia in de-novo Parkinson's disease. NPJ Digital Med. **5** (2022). https://doi.org/10.1038/s41746-022-00642-5
6. Revankar, G., Salian, A., Ozono, T., et al.: Smartphone-based behavioral profiling for distinguishing dementia with lewy bodies from Alzheimer's disease (2025). https://doi.org/10.1101/2025.03.31.25324871
7. Takeshige-Amano, H., Oyama, G., Ogawa, M., et al.: Digital detection of alzheimer's disease using smiles and conversations with a chatbot. Sci. Rep. **14** (2024). https://doi.org/10.1038/s41598-024-77220-0
8. Zhang, L., Arandjelović, O., Dewar, S., Astell, A., Doherty, G., Ellis, M.: Quantification of advanced dementia patients' engagement in therapeutic sessions: an automatic video based approach using computer vision and machine learning. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 5785–5788 (2020). https://doi.org/10.1109/EMBC44109.2020.9176632