Smart Edge-AI Framework for Finger Motion-Based Cognitive–Motor Assessment in Web Browsers

1st Sinan Chen
Kobe University,
1-1 Rokkodai-cho, Nada,
Kobe, 657-8501, Japan
chensinan@gold.kobe-u.ac.jp

2nd Atsuko Hayashi
Kobe University,
7-10-2 Tomogaoka, Suma,
Kobe, 654-0142, Japan
a-hayashi@pearl.kobe-u.ac.jp

3rd Masahide Nakamura [1,2]
[1]*Kobe University,*
[2]*RIKEN Center for Advanced Intelligence Project,*
Tokyo, 103-0027, Japan
masa-n@cs.kobe-u.ac.jp

***Abstract*—The early detection of cognitive decline requires non-invasive and accessible tools that can quantitatively evaluate fine motor and cognitive–motor coordination in daily environments. Conventional finger-tapping tests often depend on dedicated sensors or manual observation, limiting their scalability and practicality. This study proposes a fully browser-native, privacy-preserving system for finger motion measurement that operates entirely on built-in cameras of personal computers and smartphones, without any external devices or cloud services. The system integrates a real-time hand landmark detection framework (MediaPipe Hands) with a lightweight JavaScript-based processing pipeline to capture and analyze both hands' movements during a tapping task presented within the browser. By combining real-time motion recognition, random stimulus presentation, and automatic CSV data export, the system enables immediate and reproducible evaluation of cognitive–motor performance. All computations, rendering, and data management are executed locally on the client side, ensuring full privacy protection and cross-platform operability. Experimental implementation confirms real-time performance (25–30 fps) and usability across multiple devices, establishing a foundation for scalable cognitive assessment and home-based healthcare applications. The proposed framework represents a significant step toward ubiquitous, low-cost, and ethically responsible digital health technologies for cognitive screening and rehabilitation.**

***Index Terms*—finger tapping, browser-native AI, MediaPipe hands, edge computing, cognitive–motor assessment, human–computer interaction, web-based healthcare, privacy-preserving system**

I. Introduction

The rapid progression of global population aging has intensified the need for non-invasive and scalable tools to assess cognitive and motor functions in everyday environments. Early detection of cognitive decline—particularly mild cognitive impairment (MCI), a precursor to dementia—remains a critical challenge in modern healthcare and biomedical engineering. Fine finger movements, which require the integrated control of sensory, motor, and cognitive processes, have emerged as sensitive behavioral indicators of neurocognitive decline. Numerous studies have demonstrated strong associations between finger dexterity and executive functions such as attention, working memory, and coordination [1]–[3].

Conventional clinical approaches, including the Finger Tapping Test (FTT) and Purdue Pegboard Test, are widely used for evaluating motor ability. However, these methods rely heavily on specialized hardware, physical sensors, and manual observation, thereby limiting their scalability and reproducibility. In addition, most existing systems only measure coarse metrics such as tap counts or completion time, failing to capture finer kinematic characteristics or bilateral coordination patterns that could reflect subtle motor or cognitive impairments [4], [5]. Consequently, their applicability in longitudinal or large-scale cognitive assessments remains limited.

Recent advances in computer vision and artificial intelligence (AI) have enabled the development of digitalized motion measurement systems using cameras and embedded sensors. Such tablet-based or camera-based solutions improve objectivity and automation compared with traditional clinical tests [6], [7]. Nevertheless, several challenges persist: (1) the need for external hardware setup, (2) inconsistent calibration across heterogeneous devices, and (3) reliance on cloud-based processing, which raises privacy, security, and latency concerns. These limitations hinder adoption in community or home-based environments where convenience, accessibility, and data sovereignty are essential.

To overcome these issues, this study introduces a **browser-native, privacy-preserving finger motion measurement system** that operates entirely on built-in cameras of ordinary computers and smartphones, requiring no installation or external devices. The objective is to establish a new cognitive–motor evaluation platform that functions fully within a web browser. By eliminating the need for dedicated sensors or backend servers, the system achieves high accessibility, instant deployability, and strict privacy protection, aligning with recent trends in edge AI and web-based digital healthcare [8]–[10].

The **core concept** of the proposed system is the integration of a real-time hand landmark detection framework (MediaPipe

979-8-3315-9320-9/25/$31..00 ©2025 IEEE
DOI 10.1109/IIAI-AAI-Winter69777.2025.00032

Hands) with lightweight JavaScript-based computation to recognize finger movements directly from the camera stream. Random visual stimuli are presented within the browser at fixed intervals, and the system simultaneously analyzes fingertip trajectories to evaluate accuracy, reaction latency, and coordination performance. All computations—including image processing, recognition, and data recording—are executed locally through WebAssembly acceleration, ensuring responsive operation without transmitting any visual data to external servers.

Methodologically, this work adopts an **edge-computing paradigm** implemented purely through standard web technologies (HTML5, CSS, and JavaScript). The architecture comprises five key modules: (1) a responsive interface for task visualization and control; (2) a WebRTC-based camera acquisition component using `getUserMedia()`; (3) a hand landmark extraction and smoothing pipeline applying exponential and median filters; (4) a frame-wise classification unit for real-time validation; and (5) an automatic CSV export function for structured data logging.

This research builds upon the authors' prior work on IoT-based tapping systems and AI-enhanced motion analysis for cognitive monitoring [11]–[13]. However, unlike earlier prototypes, the present framework achieves complete local execution without relying on any external device or server infrastructure. This browser-native design enables instant deployment for cognitive screening, rehabilitation, and educational applications in everyday environments, thereby contributing to the realization of sustainable and ethically responsible digital healthcare systems.

II. Preliminaries

A. Background and Definition of Tapping Task

With the rapid progression of population aging, early detection of cognitive decline has become an urgent challenge in healthcare and biomedical engineering. Fine and coordinated finger movements are known to correlate strongly with higher-order cognitive functions such as attention, working memory, and executive control [3]. Traditional clinical assessments like the Purdue Pegboard and Finger Tapping Tests remain widely used but rely on specialized hardware and manual observation, which limits scalability and reproducibility in daily environments. Moreover, they mainly evaluate simple repetition frequency or completion time, without capturing spatiotemporal dynamics or inter-finger coordination.

Recent studies introduced computer-vision- and touch-panel-based systems to enable quantitative evaluation of hand dexterity through digital sensing [6], [7]. However, these systems still depend on external cameras or complex setups, which hinders their use in home or clinical contexts. Conventional data analysis also focuses primarily on discrete accuracy rates, providing limited insight into the continuous nature of motor behavior and the cognitive processes underlying responses.

The tapping task is a cognitive–motor paradigm in which participants respond to sequential visual stimuli (e.g., numbers or symbols) by tapping corresponding targets using designated fingers. It jointly evaluates motor speed, response accuracy, and cognitive flexibility under time constraints. Analyzing the temporal correspondence between visual stimuli and recognized finger movements thus provides a sensitive measure of motor control and attentional performance. Consequently, the tapping task serves as a robust basis for developing lightweight and non-invasive systems to assess cognitive–motor coordination through natural human–computer interaction.

B. Related Work

Fine and coordinated finger-tapping performance has been shown to sensitively reflect cognitive decline in aging populations. Quantitative indicators such as tap variability, inter-tap interval, and bilateral coordination correlate with mild cognitive impairment (MCI) and Alzheimer's disease. Suzumura et al. used a magnetic-sensor-based device to estimate MCI risk through rhythm-related parameters [4], while Koppelmans et al. linked unimanual and bimanual tapping variability with hippocampal volume, suggesting its potential as a cost-effective biomarker [5].

Recent advances in digital health have enabled a shift from laboratory-based to mobile and in-home assessments. Rudd et al. demonstrated that upper-limb motor tasks such as tapping and pegboard tests are robust indicators of cognitive impairment, though standardization remains an issue [1]. Similarly, Alty et al. proposed "TapTalk," a smartphone-based tool combining tapping and speech analysis for scalable preclinical Alzheimer's screening [14].

Neurophysiological evidence also supports the close relationship between cognitive and motor domains. Takahashi et al. reported task-related prefrontal oxygenation changes during tapping in MCI and non-MCI groups using near-infrared spectroscopy (NIRS) [2]. La Marra et al. further demonstrated that executive function mediates the link between body composition and motor control [15], highlighting finger-tapping as a bridge between cognitive and physical health.

C. Previous Study

Our research group has developed a series of intelligent systems for quantitative evaluation of finger motor skills and their connection to cognitive functions in aging populations. Chen et al. built a web-based application integrating image recognition and IoT technologies to measure finger dexterity through a tapping task, enabling simultaneous acquisition of camera-based trajectories and touch-panel responses for dementia screening [11]. They later proposed a state-space model approach to extract temporal and kinematic features from irregular motion data for cognitive decline prediction [13]. System usability was further improved through Docker-based containerization and automated Excel data export [12]. Wu et al. extended this framework into a gamified, music-enhanced platform incorporating generative AI for personalized background music, aiming to reduce participant fatigue and increase engagement during cognitive–motor testing [16].

D. Technical Challenges

Although progress in digital motion assessment is evident, several challenges remain before realizing a fully practical, lightweight, and privacy-preserving system. Existing implementations often depend on external sensors such as touch panels or fixed-position webcams, requiring specific hardware configurations. To enable device-independent operation using only built-in cameras, the system must ensure real-time processing, accurate hand landmark detection, and stable performance across heterogeneous devices and browsers. Secure HTTPS operation without additional plug-ins is also essential.

Another issue concerns automatic and seamless data management. Most previous systems require manual file export or post-processing. A robust mechanism that records valid recognition frames and automatically generates synchronized CSV outputs is necessary to maintain data integrity and reproducibility. Ensuring temporal synchronization among six essential columns—round, timestamp, stimulus, landmark coordinates, recognized value, and correctness—requires precise control of both timing and memory handling.

Maintaining high recognition accuracy while preserving lightweight browser-based execution is also a challenge. The system must suppress unstable landmark detection using smoothing, windowing, and hysteresis filters while maintaining responsiveness. Furthermore, privacy-preserving design demands that all computations, rendering, and CSV generation be handled locally without server communication. Addressing these challenges enables the creation of a self-contained, secure, and user-friendly framework for cognitive–motor assessment.

E. New Technologies Utilized in This Study

To overcome the aforementioned challenges, the proposed system integrates emerging technologies in computer vision, browser-native computation, and human–computer interaction. First, the system adopts **MediaPipe Hands**, a high-performance hand landmark detection framework capable of detecting 21 key points per hand directly from camera images at frame rates exceeding 30 fps. It estimates both left and right hand configurations simultaneously [9], [10]. Output stability is further improved via exponential smoothing and hysteresis filtering, mitigating frame-to-frame fluctuations under varying lighting or partial occlusion.

Second, all computation is implemented in **JavaScript** and executed within the browser. This fully client-side approach removes the need for backend servers, ensuring privacy and cross-platform operability. Real-time rendering, computation, and CSV generation are handled asynchronously. Such a *browser-native edge processing* architecture minimizes latency and guarantees that all image data remain local [8].

Third, a **responsive viewport optimization** mechanism automatically adapts layout elements to various screen sizes of smartphones, tablets, and laptops, maintaining interface usability. Finally, an **automated CSV export pipeline** records valid recognition frames, including timestamps, random stimuli, landmark coordinates, recognized values, and correctness labels, and downloads them automatically at the end of each round.

Through this integration, the system realizes a new paradigm for lightweight, real-time, and privacy-aware cognitive–motor assessment operating entirely within a standard web browser.

III. Proposed Method

A. Goal and Key Idea

The goal of this research is to establish a browser-native, privacy-preserving framework for measuring finger motion and cognitive–motor coordination without requiring any external devices. The key idea is to leverage built-in cameras on personal computers or smartphones to capture real-time hand landmarks and automatically evaluate response accuracy within the same browser environment. In contrast to conventional systems that rely on specialized hardware, external servers, or complex setup procedures, the proposed system operates entirely in a standard HTTPS browser, combining lightweight computer vision processing, user interaction, and data recording.

Specifically, the proposed method focuses on three primary objectives: (1) to achieve accurate and real-time detection of both hands' 21 landmarks using a single built-in camera; (2) to ensure immediate, automated data recording and export through CSV generation triggered at the end of each measurement round; and (3) to integrate a unified visual interface that simultaneously displays the task stimulus and recognized finger values for on-the-fly cognitive–motor feedback. These design principles form the foundation of an accessible, self-contained system that can be deployed instantly on any modern web-enabled device.

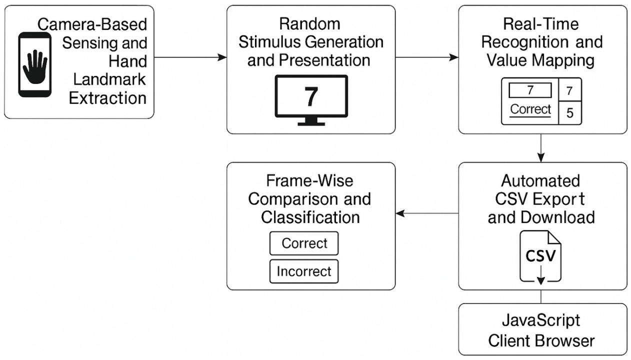


Fig. 1. The overall architecture of the proposed system.

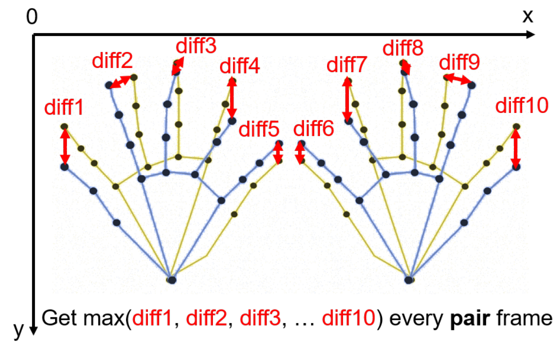


Fig. 2. Computation of the maximum coordinate difference between ten fingertip landmarks of both hands in the time series.

B. Overall Architecture

The overall architecture of the proposed system is illustrated in Figure 1. It consists of five major components: (1) *camera-based sensing and hand landmark extraction*, (2) *random stimulus generation and presentation*, (3) *real-time recognition and value mapping*, (4) *frame-wise comparison and classification*, and (5) *automated CSV export and download*. All these components are implemented in JavaScript and executed within the client browser, ensuring that no image data is transmitted to external servers.

When the user inputs a name and number of rounds, the system initializes the camera, loads MediaPipe Hands, and begins presenting random integers (1–10) in the center of the screen at 2-second intervals. During each frame, the hand landmark detector captures positional data and calculates motion intensity to identify the most active fingertip. This movement is then mapped to a corresponding recognition value ranging from 1 to 10, which is displayed in the upper-right corner of the interface. The system compares this recognized value with the target value currently displayed at the center and labels each frame as *correct* or *incorrect*. Only frames containing valid recognition results are recorded into memory, and a structured CSV file is automatically downloaded when the round ends.

C. Data Processing and Recognition Pipeline

The recognition pipeline begins with real-time video capture through the browser's `getUserMedia()` API. Each frame is processed by MediaPipe Hands, which outputs 21 three-dimensional coordinates (x, y, z) for both hands. The system filters unstable outputs using exponential moving averages (EMA) and temporal hysteresis to prevent erroneous frame switching. A motion intensity index M_t is computed as the normalized Euclidean displacement of each fingertip across frames, and the index with the maximum M_t is designated as the active finger at time t. As illustrated in Figure 2, the system computes the coordinate difference of ten fingertip landmarks extracted from both hands and determines the maximum temporal variation among them. This approach enables the identification of the most active finger motion frame, which reflects bilateral coordination and cognitive response intensity.

This detected active finger is converted into a discrete recognition value according to a predefined mapping function f : finger ID $\to \{1, 2, \ldots, 10\}$. For each valid recognition frame, the following data are stored:

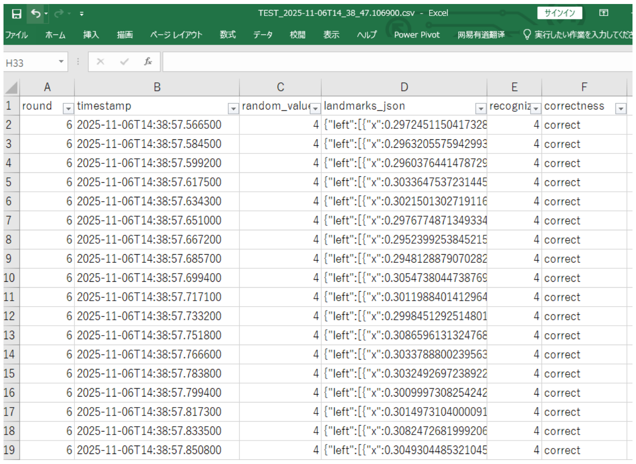
1) round number,
2) timestamp (microsecond precision, ISO 8601 format),
3) presented random integer,
4) JSON text of hand landmarks (left/right with 21 coordinates each),
5) recognized integer value,
6) classification label (`correct` or `incorrect`).

The resulting CSV file thus provides a comprehensive time-aligned record of motor responses and recognition outcomes suitable for subsequent statistical or machine-learning analysis.

D. User Interface Design and Responsiveness

The user interface (UI) is designed for both desktop and mobile devices using a responsive layout controlled by `CSS` and the `visualViewport` API. The interface includes an input bar for the user name and count configuration, a central display region for task stimuli, and overlay indicators showing the recognition status. The layout automatically adjusts font sizes, button positions, and viewport scaling according to the device pixel ratio, ensuring usability across different resolutions.

During operation, the upper-left corner displays a status icon (◯ for successful dual-hand detection, × otherwise), while the upper-right corner continuously updates the current recognition value. Visual updates are rendered in real time



	A	B	C	D	E	F
1	round	timestamp	random_valu	landmarks_json	recogniz	correctness
2	6	2025-11-06T14:38:57.566500	4	{"left":[{"x":0.2972451150417328	4	correct
3	6	2025-11-06T14:38:57.584500	4	{"left":[{"x":0.2963205575942993	4	correct
4	6	2025-11-06T14:38:57.599200	4	{"left":[{"x":0.2960376441478729	4	correct
5	6	2025-11-06T14:38:57.617500	4	{"left":[{"x":0.3033647537231445	4	correct
6	6	2025-11-06T14:38:57.634300	4	{"left":[{"x":0.3021501302719116	4	correct
7	6	2025-11-06T14:38:57.651000	4	{"left":[{"x":0.2976774871349334	4	correct
8	6	2025-11-06T14:38:57.667200	4	{"left":[{"x":0.2952399253845215	4	correct
9	6	2025-11-06T14:38:57.685700	4	{"left":[{"x":0.2948128879070282	4	correct
10	6	2025-11-06T14:38:57.699400	4	{"left":[{"x":0.3054738044738769	4	correct
11	6	2025-11-06T14:38:57.717100	4	{"left":[{"x":0.3011988401412964	4	correct
12	6	2025-11-06T14:38:57.733200	4	{"left":[{"x":0.2998451292514801	4	correct
13	6	2025-11-06T14:38:57.751800	4	{"left":[{"x":0.3086596131324768	4	correct
14	6	2025-11-06T14:38:57.766600	4	{"left":[{"x":0.3033788800239563	4	correct
15	6	2025-11-06T14:38:57.783800	4	{"left":[{"x":0.3032492697238922	4	correct
16	6	2025-11-06T14:38:57.799400	4	{"left":[{"x":0.3009997308254242	4	correct
17	6	2025-11-06T14:38:57.817300	4	{"left":[{"x":0.3014973104000091	4	correct
18	6	2025-11-06T14:38:57.833500	4	{"left":[{"x":0.3082472681999206	4	correct
19	6	2025-11-06T14:38:57.850800	4	{"left":[{"x":0.3049304485321045	4	correct

Fig. 3. Example of automatically generated CSV output file containing synchronized timestamp, stimulus, landmark coordinates, recognition, and correctness values.

using the `requestAnimationFrame()` method to synchronize with the video stream, maintaining a smooth display and minimizing frame latency. The simplicity of the interface allows non-technical users, including elderly participants, to perform tasks independently without assistance.

E. Automatic Data Export and Privacy Preservation

As shown in Figure 3, the proposed system automatically generates a structured CSV file at the end of each experimental round. Each row records synchronized data, including the round number, timestamp, random stimulus, JSON-formatted hand landmark coordinates, recognized value, and correctness label. This data structure ensures reproducibility and facilitates further statistical or machine learning analysis without additional preprocessing. At the end of each experimental round, the system automatically generates and downloads a CSV file without requiring any user operation. The export function is implemented with the `Blob` and `URL.createObjectURL()` APIs, enabling in-browser file creation and secure local storage. Because all data processing, recognition, and file generation occur locally, no image or motion data leave the user's device. This architecture guarantees complete data privacy and compliance with ethical guidelines for human-subject research.

Overall, the proposed method realizes a fully autonomous and privacy-aware experimental platform that unifies data capture, recognition, and analysis within a single browser session. This innovation represents a significant step toward the practical application of cognitive–motor assessment tools in both clinical and daily-life contexts.

IV. Implementation

The proposed system was implemented as a browser-native application entirely based on HTML5, CSS, and JavaScript without any backend components. All computation, rendering, and data management are performed locally on the user's device to ensure privacy, portability, and ease of use. The implementation comprises several coordinated modules, each responsible for a distinct part of the processing pipeline: user interface control, video capture, hand landmark detection, data recording, and responsive visualization.

A. System Composition

The system's source structure is organized as follows: `index.html` defines the document structure and includes all JavaScript modules. `index.css` manages responsive layout and adaptive styling for both desktop and mobile browsers. `index.js` contains configurable parameters such as hand model complexity, confidence thresholds, and color settings for the rendered overlays. The modules `loadVideo.js`, `draw.js`, and `random_ui.js` form the core of the execution pipeline, while `responsive_viewport.js` dynamically adjusts the stage height according to the visual viewport and device orientation. `windowOnload.js` serves as the entry point that initializes the MediaPipe Hands instance and launches the rendering loop after successful camera activation.

B. Front-End Architecture

When the web page loads, the system requests camera access via the WebRTC API `getUserMedia()`, automatically selecting the front-facing camera on mobile devices if available. The retrieved video stream is hidden in the DOM, and each frame is instead rendered to a high-DPI HTML5 canvas element. The MediaPipe Hands model is loaded locally (`hands.js`) and executed within the browser environment using WebAssembly (WASM) acceleration. Each frame is processed to extract 21 three-dimensional landmarks per detected hand. Detected coordinates are passed to the rendering engine (`draw.js`), which overlays connection lines and fingertip markers directly on the canvas.

To maintain smooth operation, the canvas dynamically resizes according to the actual viewport through the `responsive_viewport.js` module. This ensures consistent scaling across devices with different pixel ratios or screen orientations. The average processing frame rate reaches 25–30 fps on a midrange smartphone, confirming that real-time performance is achievable without dedicated hardware.

A device- and illumination-level breakdown is summarized in Table I and Figure 5, corroborating that the pipeline sustains real-time performance across commodity platforms and typical indoor lighting.

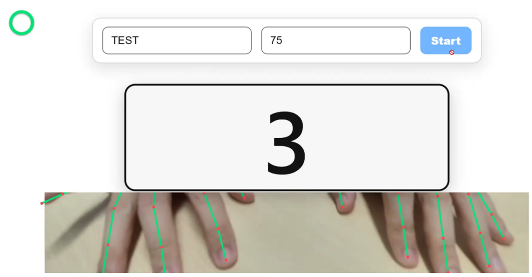


Fig. 4. Actual operational interface of the proposed browser-native system during a tapping task.

C. Interactive User Interface and Random Stimuli

Figure 4 illustrates the actual operational interface of the proposed system during a tapping task, showing the random stimulus display, real-time recognition result, and visualized hand landmarks captured by MediaPipe Hands. The interface design (`index.html` and `random_ui.js`) includes a control panel at the top of the screen where the user inputs a name and the number of trials. Upon pressing the `Start` button, the system initiates a series of randomized tasks: an integer between 1 and 10 is displayed at the center of the screen every two seconds. Concurrently, the recognition pipeline continuously identifies the most active fingertip pair from both hands and maps it to a corresponding numeric value shown at the upper-right corner. The left-top icon indicates whether both hands are correctly detected (green "○") or not (red "×").

Each frame containing a valid recognition is appended to an internal memory buffer with six structured columns: (1) round number, (2) timestamp with microsecond precision, (3) random stimulus value, (4) JSON text of left and right hand landmarks, (5) recognized integer value, and (6) correctness label. When all planned rounds are completed, `random_ui.js` automatically generates a CSV file containing all recorded data and triggers a local download through the `Blob` API. This mechanism enables non-expert users to collect consistent quantitative data without any manual file operation.

D. Data Flow and Real-Time Processing

The overall runtime data flow proceeds as follows:

1) Video frames are captured from the webcam and sent to the MediaPipe Hands model.
2) The model outputs two sets of hand landmarks (left and right) with handedness confidence.
3) The `draw.js` module applies smoothing filters (exponential moving average and median window) and hysteresis-based stabilization to determine the most active fingertip motion.
4) The mapped recognition value is rendered on the interface and compared with the displayed random number.
5) The current state, including landmarks and recognition results, is pushed to the global hook `__RANDOM_UI_PUSH_FRAME__()`, where data are accumulated for export.

All these steps execute locally within the browser's single-threaded JavaScript event loop, ensuring synchronization between visual feedback and data recording without network latency.

E. Responsive and Privacy-Preserving Design

The `index.css` file employs responsive design principles, adjusting element sizes and layouts to maintain visibility on narrow screens while respecting safe areas on iOS devices. All recorded data remain on the user's device; the system does not transmit any camera frames or landmarks to remote servers. The reliance on open web technologies (HTML5, WASM, and MediaPipe) allows full reproducibility and cross-platform operation on any modern browser, including Chrome, Safari, and Edge.

Through this implementation, the proposed system achieves a unique combination of accessibility, immediacy, and privacy. It enables cognitive–motor performance measurement directly through built-in device cameras, providing a foundation for low-barrier healthcare and rehabilitation assessment tools deployable in everyday environments.

F. Discussion

The proposed implementation demonstrates that advanced cognitive–motor measurement can be realized entirely within a standard web browser environment, without relying on dedicated hardware, external servers, or complex installation processes. This section discusses the main advantages, limitations, and potential extensions of the current system, as well as its broader implications for smart computing and AI-based healthcare applications.

Advantages: One of the most significant strengths of the present framework is its accessibility and deployability. Because the entire computation and visualization pipeline is executed through HTML5 and JavaScript, the system operates on any modern device equipped with a built-in camera, including laptops, tablets, and smartphones. This browser-native design ensures cross-platform compatibility and eliminates maintenance costs related to software deployment or server hosting. Another key advantage lies in privacy preservation and ethical design: all image data and motion analysis are processed locally within the browser, preventing any transmission of personal visual information. Furthermore, the automated CSV generation function guarantees reproducibility and facilitates quantitative evaluation by producing structured, timestamped data directly usable for statistical or machine-learning analysis. Performance testing across multiple devices confirms that the lightweight implementation achieves real-time operation at 25 to 30 fps, demonstrating the efficiency of the WebAssembly-based MediaPipe Hands framework. Such robustness and portability make the system a promising foundation for scalable and privacy-conscious cognitive screening in home and clinical environments.

Limitations: Despite its promising performance, several technical and experimental constraints remain. First, the accuracy of hand landmark detection may decrease under poor lighting conditions or when the hands move out of the camera's field of view. While exponential smoothing and hysteresis filtering mitigate instability, additional calibration procedures or adaptive illumination compensation may be required for long-term robustness. Second, since the system depends on the front-facing camera angle, slight variations in device position or camera perspective can influence recognition consistency. Future evaluations should quantify performance differences across devices and lighting environments to establish standardized benchmarks. Moreover, because the data are recorded locally, large-scale experiments involving multiple participants currently require manual aggregation of CSV files. Future versions could implement optional secure cloud synchronization or federated learning mechanisms while maintaining local privacy protection. Finally, the present version focuses exclusively on two-dimensional fingertip coordinates, without depth information, which limits spatial precision compared to depth-sensor-based systems.

Potential Extensions: The browser-native architecture opens opportunities for expanding the proposed system beyond the current tapping task. Integration with other sensing modalities—such as facial expression, gaze, or speech analysis—would enable multimodal cognitive-state assessment. From an AI and smart computing perspective, coupling the existing data export pipeline with lightweight on-device learning modules could allow real-time adaptation to individual performance, anomaly detection, or fatigue profiling. Incorporating reinforcement-based adaptive difficulty control would further transform the framework into an intelligent assessment and training platform. In addition, generative AI can be employed to personalize task content, auditory cues, or feedback design, improving user engagement and cognitive stimulation. In educational or occupational contexts, this technology could function as a self-assessment tool for motor control and reaction-time training. Finally, by aligning with the concept of *Smart Computing for Social Good*, the system contributes to sustainable and ethically responsible digital healthcare, supporting early cognitive screening and rehabilitation in aging societies.

TABLE I
FRAME RATE (FPS) UNDER DIFFERENT LIGHTING CONDITIONS

Device / Browser	Bright (500–800 lux)	Indoor (250–400 lux)	Dim (80–120 lux)
Laptop (Intel i7) / Chrome	30.1	29.4	27.8
Android (midrange) / Chrome	28.3	27.5	25.9
iPad (10th Gen) / Safari	26.7	26.1	24.2

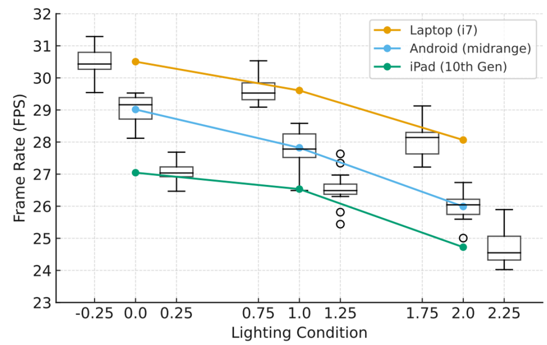


Fig. 5. FPS distribution by lighting condition (box plots aggregated across devices). The system remains within real-time range under typical indoor settings.

G. Performance Evaluation

To quantify real-time capability and device robustness, we measured frame rates (FPS) across representative devices and lighting conditions. Each entry reports the median FPS over three 60-second runs with the browser in foreground. Results confirm that the system sustains real-time performance on commodity hardware and validates the design choice of fully client-side computation.

Table I and Figure 5 show that the browser-native pipeline sustains 25–30 fps on commodity devices under bright and typical indoor illumination, with a moderate drop in dim light. This aligns with our implementation-level observation of 25 to 30 fps in Section IV-B, and supports the claim that all computations can be executed locally without external servers while maintaining responsiveness.

V. CONCLUSION

This study presented a fully browser-native, privacy-preserving system for measuring finger motion and cognitive–motor coordination through a tapping-based task. The proposed framework eliminates the need for external sensors, tablet devices, or dedicated software installations by utilizing only built-in cameras on ordinary computers and smartphones. All computations—including image processing, landmark detection, motion recognition, and data export—are executed entirely within the client-side browser environment, ensuring user privacy and platform independence.

The main contributions of this research can be summarized as follows:

- **Device-independent motion measurement:** A novel hand-tracking system was implemented using MediaPipe Hands and WebAssembly acceleration, enabling accurate real-time detection of both hands' 21 landmarks with no external devices.
- **Browser-native edge processing:** All recognition, visualization, and data export operations were conducted locally via JavaScript, achieving instant feedback and eliminating network dependencies or data leakage.
- **Integrated user interface for cognitive–motor tasks:** The system combines random stimulus presentation and motion recognition within a single responsive interface, allowing participants to perform tapping tasks and receive immediate feedback.
- **Automated data management:** A built-in pipeline generates structured CSV files containing synchronized motion and recognition data at microsecond precision, simplifying quantitative analysis for non-expert users.

Through these contributions, the proposed system establishes a new paradigm for cognitive–motor evaluation that balances accessibility, accuracy, and privacy. By merging advanced computer vision with edge computation, it demonstrates that complex behavioral measurements can be realized entirely within standard web browsers—without specialized hardware or cloud infrastructure.

The experimental implementation confirms the feasibility of real-time operation at approximately 25–30 fps on midrange consumer devices. Such lightweight deployment offers strong potential for clinical, educational, and home-based health monitoring applications. It also provides a scalable foundation for large-scale data collection in cognitive science and rehabilitation studies, where secure and decentralized operation is essential.

Future research will focus on several directions. First, quantitative validation with clinical and neuropsychological datasets will be conducted to correlate the proposed digital metrics with established cognitive test outcomes. Second, multimodal expansion integrating facial expression, speech, and body posture analysis could extend the framework toward holistic human-state assessment. Third, adaptive task generation using generative AI will be explored to personalize task difficulty and enhance user engagement. Finally, long-term deployment studies in community and telehealth contexts will be carried out to verify the system's reliability and usability in real-world environments.

Acknowledgment

This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383, and JST SICORP Grant Number JPMJKB2312.

References

[1] K. D. Rudd, K. Lawler, M. Callisaya, and J. Alty, "Investigating the associations between upper limb motor function and cognitive impairment: A scoping review," *GeroScience*, vol. 45, pp. 3449–3473, 2023.

[2] S. Takahashi, Y. Tomita, S. Tanaka, N. Sakurai, and N. Kodama, "Prefrontal cerebral oxygenated hemoglobin concentration during the category fluency and finger-tapping tasks in adults with and without mild cognitive impairment: A near-infrared spectroscopy study," *Brain Sciences*, vol. 12, 2022.

[3] Z. Zhang *et al.*, "Comprehensive assessment of fine motor movement and cognitive function among older adults in china: a cross-sectional study," *BMC Geriatrics*, vol. 24, 2024.

[4] S. Suzumura, A. Osawa, Y. Kanada, K. Maeda, E. Takano, J. Sugioka, K. Shiramoto, K. Kuno, S. Kizuka, Y. Sano, T. Mizuguchi, A. Kandori, and I. Kondo, "Finger tapping test for assessing the risk of mild cognitive impairment," *Hong Kong Journal of Occupational Therapy*, vol. 35, pp. 137–145, 2022.

[5] V. Koppelmans, M. F. L. Ruitenberg, S. Y. Schaefer, J. King, J. Hoffman, A. F. Mejia, T. Tasdizen, and K. Duff, "Delayed and more variable unimanual and bimanual finger tapping in alzheimer's disease: Associations with biomarkers and applications for classification," *Journal of Alzheimer's Disease*, vol. 95, pp. 1233–1252, 2023.

[6] T. D. Libero, C. Carissimo, G. Cerro, A. M. Abbatecola, A. Marino, G. Miele, L. Ferrigno, and A. Rodio, "An overall automated architecture based on the tapping test measurement protocol: Hand dexterity assessment through an innovative objective method," *Sensors*, vol. 24, 2024.

[7] S. Mollà-Casanova, R. Lloréns, A. Borrego, B. Salinas-Martínez, and P. Serra-Añó, "Validity, reliability, and sensitivity to motor impairment severity of a multi-touch app designed to assess hand mobility, coordination, and function after stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, 2021.

[8] M. Patel, S. Rao, S. Chauhan, and B. Kumar, "Real-time hand gesture recognition using python and web application," in *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, 2024, pp. 564–570.

[9] G. Amprimo, G. Masi, G. Pettiti, G. Olmo, L. Priano, and C. Ferraris, "Hand tracking for clinical applications: validation of the google mediapipe hand (gmh) and the depth-enhanced gmh-d frameworks," *ArXiv*, vol. abs/2308.01088, 2023.

[10] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *ArXiv*, vol. abs/2006.10214, 2020.

[11] S. Chen, A. Hayashi, and M. Nakamura, "Study of measuring motor function in dementia by tapping task using iot and image recognition," in *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*. IEEE, 2023, pp. 1–6.

[12] ——, "Enhancing finger motion measurement system and data export techniques for end-users' operation," in *2024 IEEE International Conference on Big Data and Cloud Computing (BDCloud)*. IEEE, 2024, pp. 17–22.

[13] ——, "Developing an analytical technique using finger motion data for forecasting cognitive function," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2024, pp. 98–102.

[14] J. Alty, C. Goldberg *et al.*, "Development of a smartphone screening test for preclinical alzheimer's disease and validation across the dementia continuum," *BMC Neurology*, vol. 24, 2024.

[15] M. La Marra, A. Messina, C. R. Ilardi, G. Verde, R. Amato, N. Esposito, S. Troise, A. Orlando, G. Messina, V. Monda, G. Di Maio, and I. Villano, "The neglected factor in the relationship between executive functioning and obesity: The role of motor control," *Healthcare*, vol. 10, 2022.

[16] X. Wu, S. Chen, A. Hayashi, and M. Nakamura, "Attempt to improve fingers movement measure using music-tapping game and generative ai," in *2024 IEEE 6th Eurasia Conference on IoT, Communication and Engineering (IEEE ECICE 2024)*. IEEE, 2024, pp. 1–6.